**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **57** |
| **Stop Coordinate** | **36780** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | **37127** |
| **Selected Function** | **Hypothetical 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? | *GeneMarkS andr GeneMark-M. folio show coding potential.* |
| Is this gene present in other annotated genomes? | *PhagesDB Blast hit: QuadZero\_56; e-value= 2e-60*  *Phamerator: phage: zepp; gene: 56; pham: 197878*(4-18-25) |
| Does the gene violate any major guiding principles? | *No* |
| **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 GeneMark start: 37127* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 37127*  *FInal score: -4,520*  *Z-value: 2,056*  *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)? | *It’s the longest ORF and doesn’t result in gene overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *does the start match the consensus start site predicted from Starterator? YES* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: QuadZero\_56, function unknown*  *E-value: 2e-60*    ***BlastP match from NCBI***  *Best Hit:* hypothetical protein SEA\_QUADZERO\_56 [Microbacterium phage QuadZero]  *E-value: 8e-79*    ***DNA Master***  *Best Hit: hypothetical protein SEA\_QUADZERO\_56 [Microbacterium phage QuadZero]*  *E-value: 0*  *Q#1: MNADRRVAGK*  *S#1: MNADRRIAGK* |
| **DECISION:** | *The gene should start at 37127* |

**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? | *List the most informative BlastP match from each source PhagesDB: QuadZero\_56, function unknown*  *NCBI:* hypothetical protein SEA\_QUADZERO\_56 [Microbacterium phage QuadZero]  *DNA Master: hypothetical protein SEA\_QUADZERO\_56 [Microbacterium phage 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? | *3ZDM\_E SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT- CONTAINING PROTEIN 2; CHAPERONE-SIGNALING PROTEIN COMPLEX; 1.803A {SACCHAROMYCES CEREVISIAE}*  Probability: 81.58%,  E-value: 14 |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *According to the July 30, 2025 Phamerator database version, Pham 197878 is flanked upstream by Pham 4886 and downstream by Pham 2778, forming a conserved syntenic arrangement among related phages.* |
| Is this gene a possible transmembrane protein? | *NO* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *NO* |
| **DECISION:** | *Hypothetical Protein* |