**Student Gene Annotation Worksheet**

| **Basic Phage Information** | |
| --- | --- |
| **Phage Name** | **wardwill** |
| **Gene #** | **24** |
| **Stop Coordinate** | **16927bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap 16bp** |
| **Selected Start Coordinate** | **16670bp** |
| **Selected Function** | **Unknown** |

**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 Zepp\_23 with e value= 2e-44*  *Phamerator phage: pham 227359 (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: 16670*  *GeneMark Start Coordinate: 16670* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -2.505*  *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 40 of the 83 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QIG58111*  *E-value: 3.1E-44*  *Q#1: MSTIRAVTEQ*  *S#1: MSTIRAVTEQ* |
| **DECISION:** | *Keep DNA Master pre start site 16670* |

**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: LilTerminator\_Draft\_23, function unknown*  *e-value: 7e-45*  *NCBI: hypothetical protein SEA\_ZEPP\_23 [Microbacterium phage Zepp]*  *Sequence ID: QIG58111.1*  *e-value: 5e-53*  *DNA Master: hypothetical protein SEA\_ZEPP\_23 [Microbacterium phage Zapp]* |
| 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: Helix-turn-helix domain-containing protein; helix turn helix, HTH, DdrO, DNA BINDING PROTEIN; 1.26A {Escherichia coli}*  *Probability: 80.43*  *e-value: 3.7* |
| 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 “minor tail protein” and “lysin A”* |
| Is this gene a possible transmembrane protein? | *No, TMHMM result does not show posible transmembrane subunit.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes,*  *Hypothetical Protein* |
| **DECISION:** | *the function is unknown. “hypothetical protein”* |