**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **25** |
| **Stop Coordinate** | **17796bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Overlap 1bp** |
| **Selected Start Coordinate** | **16927bp** |
| **Selected Function** | **endolysin** |

**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\_24 with e value= e-169*  *Phamerator phage: pham 229956 (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:16927*  *GeneMark Start Coordinate: 16927* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -5.092*  *Z-score:2.505*  *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 55 of the 104 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QIG58112 (Zepp)*  *E-value: 0*  *Q#1: MATYVQWGNA*  *S#1:MATYVQWGNA* |
| **DECISION:** | *Keep DNA Master pre start site 16927* |

**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: Zepp\_24, lysin A*  *e-value: e-169*  *NCBI: lysin A [Microbacterium phage Zepp]*  *Sequence ID: QIG58112.1*  *e-value: 0.0*  *DNA Master: lysin A [Microbacterium phage Zepp]* |
| 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: D,D-dipeptidase/D,D-carboxypeptidase; CENTER FOR STRUCTURAL GENOMICS OF INFECTIOUS DISEASES, CSGID, NATIONAL INSTITUTE O*  *Probability: 99.79*  *e-value: 8.6e-20* |
| 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: 235269 (103) ” and “235345 (80)” (May 16, 2025)* |
| 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, endolysin. According to the SEA-PHAGES approved function list, the term 'lysin A' is typically used when multiple lysins are present; however, in Wardwill, no other lysin or endolysin is encoded, so the function is annotated simply as 'endolysin'.* |
| **DECISION:** | *endolysin* |