**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **45** |
| **Stop Coordinate** | **30623** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **NO** |
| **Selected Start Coordinate** | **31405** |
| **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:  Zepp\_44; e-value= e-147  Zayuliv\_44; e-value=e-147  Phamerator: pham: 228.311 (05-05-25)  No similatiries found |
| 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 Start Coordinate (type NA if not supported):: 31405*  *GeneMark Start Coordinate (type NA if not supported)::31405* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Final RBS\_start: -2.443*  *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)? | *It is the longest ORF*  *ORF lengTh: 783bp*  *Gap: 22bp*  *Predicted start: 31405* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *El codón de inicio predicho por Starterator es el mismo que el predicho por DNA Master. However, according to Starterator, this start was identified in only 22 out of the 199 genes within the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: Zepp\_44, function unknown*  *E-value: e-147*  *Q#1: MKKTARSNPQAHPHLEHLFKIGV*  *S#1: MKKTARSNPQAHPHLEHLFKIGV*  ***BlastP match from NCBI***  *Best Hit: hypothetical protein SEA\_ZEPP\_44 [Microbacterium phage Zepp]*  *E-value: 0*  *Q#1: MKKTARSNPQAHPHLEHLFKIGVLLTR*  *S#1: MKKTARSNPQAHPHLEHLFKIGVLL*  ***DNA Master***  *Best Hit: hypothetical protein SEA\_ZEPP\_44 [Microbacterium phage Zepp]*  *E-value: 0*  *Q#1: MKKTARSNPQ*  *S#1: MKKTARSNPQ* |
| **DECISION:** | *The gene should start at 31405* |

**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: Function unknown*  *NCBI:* Hypothetical protein  *DNA Master: hypothetical protein SEA\_ZEPP\_44* |
| 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? | *Best match: 2J8W\_B CYTOCHROME C'; C', HEME, IRON, TRANSPORT, CYTOCHROME, METAL-BINDING, ELECTRON TRANSFER, ELECTRON TRANSPORT, RUBRIVIVAX GELATINOSUS*  *Probability: 68.34%, E-value: 64* |
| 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, this gene is flanked upstream by a MazG-like nucleotide pyrophosphohydrolase and downstream by a thymidylate kinase, forming a conserved syntenic block 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* |