**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **48** |
| **Stop Coordinate** | **32921** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | ***33895*** |
| **Selected Function** | ***glycosyltransferase*** |

**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 (different start site)* |
| 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: LilTerminator; e-value=0  Hasitha; e-value=0  QuadZero; e-value=0  Phamerator: phage: QuadZero, Hashita and LilTerminator Similarities found with gene 47; pham: 219441 ( (5-6-25) |
| 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 Start Coordinate: 33910*  *GeneMark Start Coordinate: 33895* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 33910*  *Final score: -6,976*  *Z-score: 1,382*  *Start site 33895*  *Final RBS\_start: -4,307*  *Z-score: 2,284*  ***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)? | *Start site 33910: It's not the longest ORF and does result in a gene overlap of 22 bp.*  *Start site 33895: It's not the longest ORF and does result in a gene overlap of 7 bp.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Start 20 corresponds to the auto-annotated start at 32924 (predicted by Glimmer), and it's not as well conserved in other phage genomes. It’s not the first start available.*  *Start 23 corresponds to the manual annotated start at 33895. It’s not conserved in other phage genomes.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: Zepp\_47, glycosyltransferase*  *E-value: 0*  *Q#1: VTDKMIELLHEMDSRRTEEWLYP*  *S#1: MTDKMIELLHEMDSRRTEEWLYP*  ***BlastP match from NCBI***  *Best Hit: glycosyltransferase [Microbacterium phage Zepp]*  *E-value: 0*  *Q#1: VTDKMIELLHEMDSRRTEEWLYP*  *S#1: MTDKMIELLHEMDSRRTEEWLYP*  ***DNA Master***  *Best Hit: QIG58135 glycosyltransferase [Microbacterium phage Zepp]*  *E-value: 0*  *Q#1: VTDKMIELLH*  *S#1: MTDKMIELLH* |
| **DECISION:** | *The gene should start in 33895.* |

**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: Zepp\_47, glycosyltransferase*  *NCBI: glycosyltransferase [Microbacterium phage Zepp]*  *DNA Master: QIG58135 glycosyltransferase [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? | *Best match: 3BCV\_A Putative glycosyltransferase protein*  *Probability: 98.91%, E-value: 9.1e-8* |
| 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 glycosyltransferase and downstream by a gene from Pham 243011, 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? | *YES* |
| **DECISION:** | *glycosyltransferase* |