**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **47** |
| **Stop Coordinate** | **31986** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | ***32924*** |
| **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:  Hasitha\_46; e-value= 0  Fulton\_46; e-value= 0  CaptainRex\_46; e-value=0  Zayuliv\_46; e-value= 0  *Phamerator: phages:*  *Hasitha; Fulton; CaptainRex; Zayuliv. gene: 47; pham: 219.441 (05-05-25)* |
| 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: 32924*  *GeneMark Start Coordinate: 32777* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 32924*  *Final score: -2,845*  *Z-score: -2,843*  ***This is the best score***  *Start site 32777*  *Final RBS\_start: -6,143*  *Z-score: 1,255* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *Start site 32924: It’s the longest ORF and does not result in gene overlap.*  *Start site 32777: It’s not the longest ORF and does not result in gene overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Start 27 is the first start available which corresponds to the auto-annotated start at 32924 (predicted by Glimmer), and it's not as well conserved in other phage genomes.*  *The start site predicted by GeneMark was not found in Starterator.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *Start site 32924*  ***BlastP match from NCBI:***  *Best Hit: glycosyltransferase [Microbacterium phage Hasitha]*  *E-value: 0*  *Q#1: MNEHLALLRRCKVSWIYPIYVPSYSRAGGAP*  *S#1: MNEHLALLRRCKVSWIYPIYVPSYSRAGGAP*  ***PhagesDB***  *Best Hit: Hasitha\_46*  *E-value: 0*  *Q#1: MNEHLALLRRCKVSWIYPIYVP*  *S#1: MNEHLALLRRCKVSWIYPIYVP*  ***DNA Master***  *Best Hit: QZD99174 [Microbacterium phage Hasitha]*  *E-value: 0*  *Q#1: MNEHLALLRR*  *S#1: MNEHLALLRR* |
| **DECISION:** | *The gene should start in 32924* |

**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: Hasitha\_46 glycosyltransferase*  *NCBI: glycosyltransferase [Microbacterium phage Hasitha]*  *DNA Master: QZD99174 [Microbacterium phage Hasitha]* |
| 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: 99.81%, E-value: 5.5e-19* |
| 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 thymidylate kinase and downstream by a glycosyltransferase, 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? | *YES* |
| **DECISION:** | *glycosyltransferase* |