**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **50** |
| **Stop Coordinate** | **34193** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | **34333** |
| **Selected Function** | *membrane 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\_49, e-value= 2e-18*  *Librie\_49, e-value= 2e-18*  *Hasitha\_49, e-value= 2e-18*  *Phamerator: phage: zepp, librie, hashita*  *Similarities found with gene 49; pham: 4923 (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 and GeneMark Start Coordinate: 34333* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 34333*  *Final score: -5,547*  *Z-value: 1,784*  *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’s not the longest ORF and doesn’t result in gene overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start codon predicted by Starterator is the same as the one predicted by DNA Master (Glimmer)*  *The start number called is the most often in the published annotations.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: Zepp\_49, function unknown, 46*  *E-value: 2e-18*    ***BlastP match from NCBI***  *Best Hit: membrane protein [Microbacterium phage Zayuliv]*  *E-value: 3e-20*    ***DNA Master***  *Best Hit: QIG58137 [hypothetical protein ZEA\_ZEPP49]*  *E-value: 7,1e-21*  *Q#1: MNPFELLLLI*  *S#1: MNPFELLLLI* |
| **DECISION:** | *The gene should start at 34333* |

**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\_49, function unknown, 46*  *NCBI: membrane protein [Microbacterium phage Zayuliv]*  *DNA Master: QIG58137 [hypothetical protein ZEA\_ZEPP49]* |
| 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? | *Protocadherin-15; PCDH15, LHFPL5, protocadherin, tip link, hair cell, TMHS, hearing, MEMBRANE PROTEIN; HET: BMA, MAN, NAG;{Mus musculus}*  *Probability: 78.74%, E-value: 14*  *List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*  *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.* |
| 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 gene from Pham 243011 and downstream by a thymidylate synthase, forming a conserved syntenic block among related phages.* |
| Is this gene a possible transmembrane protein? | *YES* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *NO* |
| **DECISION:** | *membrane protein* |