**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **49** |
| **Stop Coordinate** | **33888** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **Yes, with gene 48 (22 pb)** |
| **Selected Start Coordinate** | ***34133*** |
| **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\_48; e-value=1e-26*  *Librie\_48; e-value=1e-26*  *Fulton\_48; e-value=1e-26*  *Phamerator: phage: similarities found with gene 48; pham: 231141* (5-6-25)  *No similarities found with gene 49* |
| 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: NA*  *GeneMark Start Coordinate: 34133* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 34133:*  *Final score: -2,523*  *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’s not the longest ORF and doesn’t result in gene overlap.*  *ORF length: 246 pb*  *Predicted start: 34133*  *Gap: 62 pb* |
| 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.*  *The start number called is the most often in the published annotations, it was called in 127 of the 150 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: GreenIvy\_47, function unknown*  *E-value: 6e-46*    ***BlastP match from NCBI***  *Best Hit:* hypothetical protein SEA\_GREENIVY\_47 [Microbacterium phage GreenIvy]  *E-value: 9e-54*    ***DNA Master***  *Best Hit: XEN16698 [hypothetical protein SEA\_GREENIVY]*  *E-value: 0*  *Q#1: MAQTIHLNGG*  *S#1: MAQTIHLNGG* |
| **DECISION:** | *The gene should start in 34133.* |

**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: XEN16698 [hypothetical protein SEA\_GREENIVY]* |
| 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? | [***4G2S\_B***](http://www.rcsb.org/pdb/explore/explore.do?structureId=4G2S) ***Protein prgH; FHA domain, CELL INVASION; 1.858A {Salmonella enterica subsp. enterica serovar Typhimurium}***  ***Probability: 61.82%, E-value: 25*** |
| 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 membrane protein belonging to Pham 4923, 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? | *NO* |
| **DECISION:** | *Hypothetical Protein* |