**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **wardwill** |
| **Gene #** | **15** |
| **Stop Coordinate** | **8535bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Overlap 1bp** |
| **Selected Start Coordinate** | **8203** |
| **Selected Function** | **unknown** |

**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, Glimmer and GeneMark call it gene.* |
| Is there evidence for coding potential? | *GeneMarkS and GeneMark-host show coding potencial.* |
| Is this gene present in other annotated genomes? | *PhagesDB blast hit Zepp\_15 with e value=7e-58*  *Phamerator phage: pham 228360 (25/4/25)* |
| Does the gene violate any major guiding principles? | *The length exceeds 120 bp, there is no overlap greater than 30 bp, and the orientation of genes 14 and 16 is the same as that of gene 15.* |
| **DECISION:** | *Yes, it is a gene.* |

**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:8203*  *GeneMark Start Coordinate:8203* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -2.505*  *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)? | *Yes* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Yes, is the start number called the most often in the published annotations. It was called in 13 of the 13 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QIG58104 (SEA\_ZEPP\_15)*  *E-value: 0*  *Q#1: MTGIKPSSHG*  *S#1:* |
| **DECISION:** | *Keep DNA Master predicted start site 8203* |

**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: función unknown*  *NCBI: Microbacterium phage Zepp*  *Sequence ID: QIG58104.1*  *E-Value: 3e-72*  *Query 1 MTGIKPSSHGIVVDHKESGTRYAISDKNYDEKVHRKVRDLRAGETVLGYQPKRREQLSGG 60*  *MTGIKPSSHGIVVDHKESGTRYAISDKNYDEKVHRKVRDLRAGETVLGYQPKRREQLSGG*  *Sbjct 1 MTGIKPSSHGIVVDHKESGTRYAISDKNYDEKVHRKVRDLRAGETVLGYQPKRREQLSGG 60*  *Query 61 ESPQSGTGTAEQADGSALNASEGLADGSAGTAQATQTEGANTGTQEKKGK 110*  *ESPQSGTGTAEQADGSALNASEGLADGSAGTAQATQTEGANTGTQEKKGK*  *Sbjct 61 ESPQSGTGTAEQADGSALNASEGLADGSAGTAQATQTEGANTGTQEKKGK 110*  *DNA Master:* |
| 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? | *No Quality Match*  *Probability: 53,72%*  *E-value: 13*  *Name: E3 ubiquitin-protein ligase MIB1; E3 ligase, Notch signaling, LIGASE; HET: SO4; 2.04A {Homo sapiens}* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *Gene #15 is flanked upstream by a tail terminator protein and downstream by a major tail protein.* |
| Is this gene a possible transmembrane protein? | *No, TMHMM result does not show possible transmembrane subunit* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Hypothetical protein* |
| **DECISION:** | *Hypothetical protein* |