**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **38** |
| **Stop Coordinate** | **25155** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **13 bp** |
| **Selected Start Coordinate** | **25319** |
| **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. They differ in the initiation bp* |
| Is there evidence for coding potential? | *No* |
| Is this gene present in other annotated genomes? | *Blast hit: phage QUADZERO\_36 E-value: 1.1 e-31*  *Phamerator: phage QuadZero (EA5), gene 36, pham 86939(22).* |
| Does the gene violate any major guiding principles? | *It’s too short (174 bp) and doesn’t have coding potential.* |
| **DECISION:** | *NO* |

**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: 25328*  *GeneMark Start Coordinate: 25265* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Glimmer start coordinate: 25328*  *Final RBS score: -5.399*  *Z score: 2.099*  *Genemark start coordinate: 25265*  *Final RBS score: -5.815*  *Z score: 1.575*  *Start coordinate predicted by Glimmer has better final score and Z value than GeneMark prediction. However, the site with the best score is 25391, with final RBS score: -2.901 and Zvalue: 2.809.*  *Another alternative, present in starterator, is start coordinate: 25319*  *Final RBS score: -4.555*  *Z score: 2.002* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start sites are not the longest ORF, however the longest ORF (start coordinate 25391) has an excessive gene overlap with upstream gene (85 bp)* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Only the start site predicted by Glimmer is conserved in other phage genomes.*  *Start 4 (25328 bp):*  *• Found in 21 of 21 ( 100.0% ) of genes in pham*  *• Called 9.5% of time when present*  *The start coordinate predicted by GeneMark is not present in the genome*  *The start site predicted by Genemark is not present in this ORF.*  *The consensus start predicted by Starterator is:*  *Start 5 (@25319)*  *• Found in 21 of 21 ( 100.0% ) of genes in pham*  *• Called 85.7% of time when present* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA master:*  *Best hit: QWY84481*  *Evalue: 1.1 E-31*  *Q#: 1 MKPLPNSSAP*  *S#: 1 MKPLPNSSAP* |
| **DECISION:** | *Since this start was called most times in genes of this pham, has decent values and gives the least overlap, the start coordinate selected is 25319* |

**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? | *DNA Master:*  *Quadzero\_36*  *Evalue: 1.1E-31*    *PhagesDB and NCBI show the same result* |
| 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 found.*  *The best hit, 3CJL\_A had 59.84 probability, Evalue 47 and score 21.96, therefore not a quality match.* |
| 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 Actino\_Draft 600 version: In phage Quadzero, this gene is between two genes with assigned functions. Downstream there is a DNA polymerase I (Quadzero\_37) and upstream a Cas4 family exonuclease (Quadzero\_35).* |
| Is this gene a possible transmembrane protein? | *According to DeepTMHMM, it’s a globular protein.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *NFK* |
| **DECISION:** | *Hypothetical Protein* |