**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **39** |
| **Stop Coordinate** | **25306** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **NO** |
| **Selected Start Coordinate** | **27171** |
| **Selected Function** | **DNA Polymerase I** |

**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? | *Yes* |
| Is this gene present in other annotated genomes? | *Blast hit: phage Zepp\_38 E-value: 0*  *Phamerator: phage Zepp (EA5), gene 38, pham 229768(1446).* |
| 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: 27171*  *GeneMark Start Coordinate: 27171* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Final RBS score for (27171 bp): -2.794*  *Zvalue: 2.995* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The proposed start is the longest ORF (1866bp).* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start coordinate 27171 is found in 169 of 1905 ( 8.9% ) of genes in pham and called 94.1% of time when present.*  *The most common start site in pham was not present in Wardwill genome.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA master:*  *Best hit: QIG58126 (DNA pol I, Zepp)*  *Evalue: 0*  *Q#: 1 MDLIITEDWS*  *S#: 1 MDLIITEDWS* |
| **DECISION:** | *Keep DNA master predicted start site, 27171* |

**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? | *Phages DB hit: Zepp\_38 Evalue: 0*    *DNA master and NCBI show the same results.* |
| 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? | *HHPred hit: P19822 “DPOL\_BPT5 DNA polymerase OS=Escherichia phage T5 OX=10726 GN=T5.122 PE=1 SV=3”*  *Probability: 100.*  *Evalue: 2.3E-70*  *Score: 624.73* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *This gene is flanked upstream by a gene from Pham 86939 and downstream by a tRNA annotated as 'tRNA-OTHER', forming a syntenic arrangement observed in related phages* |
| Is this gene a possible transmembrane protein? | *No, this proteins appears to be a globular protein according to:* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *DNA Polymerase I* |