**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **37** |
| **Stop Coordinate** | **23977** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **1bp** |
| **Selected Start Coordinate** | **25155** |
| **Selected Function** | **Cas4 family exonuclease** |

**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? | *Blast hit: phage Zayuliv\_35 E-value: 0*  *Phamerator: phage Zayuliv (EA5), gene 35, pham 230954(498).* |
| 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: 25155*  *GeneMark Start Coordinate: 25155* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Final RBS score:-5.274*  *Zvalue: 1.829*  *It is not the best score, the best is for predicted start site 24291 (RBS final score: -2.765)* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start site is not the longest ORF.*  *The longest ORF is 25227, which has an excessive gene overlap with upstream gene (72bp)* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *This start site (number 61; at 25155 bp) was called by other phage genomes in 38 of 478 ( 7.9% ) of genes in pham, called 94.7% of time when present.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA master:*  *Best hit: Q0P65232 (Zayuliv)*  *Evalue: 0*  *Q#: 1 MPAKALPATG*  *S#: 1 MPAKALPATG* |
| **DECISION:** | *This information and considering coding potential given by GeneMark, DNA master predicted start site is accurate. (Shorter ORF may not include some predicted coding sites)* |

**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 hit: Zayuliv\_35, exonuclease, 392*  *Evalue: 0*    *NCBI and DNA Master 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? | *HHPred match: Q05283, “VG69\_BPML5 Gene 69 protein OS=Mycobacterium phage L5 OX=31757 GN=69 PE=4 SV=1”*  *Probability: 99.85.*  *Evalue: 1.4e-18*  *Score: 155.41* |
| 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 an ASCE ATPase and downstream by a gene from Pham 86939, forming a conserved syntenic arrangement among related phages.* |
| Is this gene a possible transmembrane protein? | *According to SOUI, it’s a SOLUBLE protein.* |
| Is the proposed function found on the SEA-PHAGES approved function list | *YES.* |
| **DECISION:** | *Cas4 family exonuclease* |