**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **34** |
| **Stop Coordinate** | **22100** |
| **Direction (For/Rev)** | **Reverse** |
| **Gap (Overlap) with Previous Gene** | **Yes, overlap in bp 22393 with Gene #35** |
| **Selected Start Coordinate** | **22393** |
| **Selected Function** | **nuclease** |

**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? | *Both GeneMark and GeneMarkS show coding potential in this region, in the first ORF of the complementary sequence.* |
| Is this gene present in other annotated genomes? | *Zepp\_33, nuclease, 97, Score:* [*200*](https://phagesdb.org/blast/results/blast.cgi#535891)*, E-value: 1e-51.*  *Pham: Phage Zepp, Gene 33, 233291 (285).* |
| Does the gene violate any major guiding principles? | *Overlap with Gene #33, few bp (25 bp).*  *Genes before and after this gene in the same direction.*  *Long enough.* |
| **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: 22393.*  *GeneMark Start Coordinate: 22393.* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start 22393:*  *Final RBS score:-6.601*  *Zvalue: 1.040*  *This is not the best value, but the best value has a 30 bp ORF.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *This start site has a 294 ORF, with 1 bp overlap with upstream gene #35.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The Start 40 at 22393 was found in 51 of 285 (17.9% ) of genes in pham. It has 41 of 249 manual annotations. Called 92.2% of time when present. It’s not the most annotated start of the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master (22393)*  *Best hit: QIG58121 (Zepp)*  *Evalue: 0*  *Q#: 1 MDEAEVVRRM*  *S#: 1 MDEAEVVRRM* |
| **DECISION:** | *Based on this information and specially considering the similarity with gene from Zepp, DNA Master predicted start site at 22393 is accurate.* |

**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: Zepp\_33, nuclease*  *Evalue: 1e-51*    *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 hit: 4QBN\_A “Nuclease; Nuclease, HYDROLASE; HET: SO4; 1.85A {Salmonella phage SETP3} SCOP: c.52.1.35”*  *Probability: 99.84*  *Evalue: 8.2e-19* |
| 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 RecA-like DNA recombinase and downstream by a gene from Pham 246632, forming a conserved syntenic block among related phages* |
| Is this gene a possible transmembrane protein? | *According to SOSUI, this is a SOLUBLE protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *nuclease* |