**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **32** |
| **Stop Coordinate** | **19867** |
| **Direction (For/Rev)** | **Reverse** |
| **Gap (Overlap) with Previous Gene** | **Yes, 78 bp gap with Gene #33** |
| **Selected Start Coordinate** | **20379** |
| **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.* |
| Is there evidence for coding potential? | *Both GeneMark and GeneMarkS show coding potential in this region, in the third ORF of the complementary sequence.* |
| Is this gene present in other annotated genomes? | *Librie\_30, function unknown, 170, Score:* [*354*](https://phagesdb.org/blast/results/blast.cgi#270942)*, E-value: 5e-98.*  *Pham: Phage Librie, Gene 30, 88242 (12).* |
| Does the gene violate any major guiding principles? | *No overlap.*  *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: 20379.*  *GeneMark Start Coordinate: 20379.* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start 20379:*  *Final RBS score:-2.443*  *Zvalue: 2.995*  *This is the best value.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start site 20379 is the longest ORF, 513 bp.*  *It has a 76 bp gap with upstream gene #33.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *This start site (number 3 at 20379 bp) was called 100% of time when present. Found in 12 of 12 ( 100.0% ) of genes in pham. Manual Annotations of this start: 8 of 8.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master.*  *Best hit: UVK59188 (Librie\_30)*  *Evalue: 0*  *Q#: 1 MFTRTIRGTG*  *S#: 1 MFTRTIRGTG* |
| **DECISION:** | *Based on this information and considering coding potential given by GeneMark, DNA Master predicted start site at 20379 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: Librie\_30 Evalue 5e-98*    *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? | *No quality match. Best HHPred hit:*  *PF19296.4 “RelA\_AH\_RIS ; RelA/SpoT, AH and RIS domains”*  *Probability: 63.07 Evalue: 34* |
| 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 Librie genome, there is a Dna primase/helicase downstream in gene Librie\_31* |
| Is this gene a possible transmembrane protein? | *No, according to DeepTMHMM this is a globular protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *NFK* |
| **DECISION:** | *Hypothetical Protein* |