**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **30** |
| **Stop Coordinate** | **19480** |
| **Direction (For/Rev)** | **Reverse** |
| **Gap (Overlap) with Previous Gene** | **Yes, 20 bp gap with Gene #31** |
| **Selected Start Coordinate** | **19665** |
| **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? | *Fulton\_28, function unknown, 61, Score:* [*120*](https://phagesdb.org/blast/results/blast.cgi#171116)*, E-value: 1e-27.*  *Phamerator: Phage Fulton, Gene #28, 6231 (15).* |
| Does the gene violate any major guiding principles? | *No overlap.*  *Genes before and after this gene in the same direction.*  *Maybe not long enough (186 bp).* |
| **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: 19665.*  *GeneMark Start Coordinate: 19665.* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Final RBS score:-5.512*  *Zvalue: 1.590*  *It is not the best score, the best is for predicted start site 19545 (RBS final score: -5.348), but ORF length of this start site is low, 66 bp.* |
| 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 the longest ORF, 186 bp.*  *It has a gene gap of 19 bp with upstream gene #31.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *This start site (number 2 at 19665 bp) was called 100% of time when present. Manual Annotations of this start: 8 of 11.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master.*  *Best hit: X0E90257 (Fulton\_28)*  *Evalue: 1.2E-35*  *Q#: 1 MNKVRIPGNT*  *S#: 1 MNKVRIPGNT* |
| **DECISION:** | *Based on this information and considering coding potential given by GeneMark, DNA Master predicted start site at 19665 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? | *List the most informative BlastP match from each source PhagesDB hit: Fulton\_28, Evalue: 1e-27*    *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? | *There is no quality match with HHPred, the best hit was:*  *PF22299.1 “BRISC\_FAM175B\_helical ; BRISC complex subunit FAM175B, helical domain” with 70.27 of probability and Evalue of 8.3* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *A conserved gene assigned to Pham 233367 was located upstream, and an Orpham gene downstream; beyond that point, synteny is conserved (July 31st, 2025).* |
| 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* |