**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **4** |
| **Stop Coordinate** | **3543** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 3 4 bp** |
| **Selected Start Coordinate** | **2113** |
| **Selected Function** | **portal 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 GeneMarkS and/or GeneMark-host trained coding potential maps show coding potential for gene 4.* |
| Is this gene present in other annotated genomes? | *BLAST hit Fulton\_4 E value 0.0*  *Phamerator (May 5th)*  *Wardwill\_4 Pham 633*  *Fulton\_4 Pham 633* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes.*  *It is long enough (1431bp, 476 aa)*  *Genes before and after are in the same direction.* |
| **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: 2113*  *GeneMark Start Coordinate: 2113* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *The start site has one possible associated Ribosome Binding site with a high score.*  *RBS 1 (Start position 2113)*  *Raw SD Score -1,559*  *Genomic Z Value3,084*  *Spacer distance 13*  *Final Score -2,605* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start codon is the longest ORF, the length is 1431 bp. This proposed start site has a 4 bp gap with the nearest upstream gene but it does not violate the Guiding Principles.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_4*  *Start 29 2113 bp*  *The candidate start sites are conserved in other phage genomes. The start called by Glimmer and GeneMark matches the start predicted by Starterator.*  *Start 29 was called 91.3 % of time when present. It was called in 17 of the 182 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 34, it was called in 134 of the 182 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *BlastP match (best hit)*  *Fulton E value 0 Accession Number* [*XOE90234*](https://www.ncbi.nlm.nih.gov/protein/XOE90234.1?report=genbank&log$=prottop&blast_rank=1&RID=24YD5CFA016)  *Alignment with Fulton*  *Q#1 MLAIETGTG*  *S#1 MLAIETGTG*  *The start site at 2113 is conserved in other phage genomes as indicated by BlastP.* |
| **DECISION:** | ***Keep DNA Master predicted start site at 2113*** |

**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? | *The most informative BlastP match from each source*  *PhagesDB: portal protein from Fulton*  *E value 0.0*  *Q#1 MLAIETGTGNN*  *S#1 MLAIETGTGNN*  *NCBI: portal protein from Fulton*  *E value 0.0*  *Q#1 MLAIETGTGNN*  *S#1 MLAIETGTGNN*  *DNA Master: portal protein from Fulton*  *E value 0.0*  *Q#1 MLAIETGTGNN*  *S#1 MLAIETGTGNN*  *This protein aligns with a portal protein from Fulton.* |
| 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? | *Selected databases: PDB\_mmCIF70\_30\_Mar, Pfam-A\_v37, UniProt-SwissProt-viral70\_3\_Nov\_2021, NCBI\_Conserved\_Domais(CD)\_v3.19*  *Best match from HHPred*  *HIT* [*O64207*](http://www.uniprot.org/uniprot/O64207)  *NAME PORTL\_BPMD2 Portal protein OS=Mycobacterium phage D29 OX=28369 GN=14 PE=3 SV=1*  *PROBABILITY 99,95*  *E VALUE 1,7e-24*  *SCORE 229,96*  *This protein aligns with a protein having a functional assignment in the databases of HHPred with a 90% probability. This protein is a portal protein.*  *.* |
| 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 not located adjacent to genes of known functions. This gene is in a region of the genome that shows high conservation of gene order.* |
| Is this gene a possible transmembrane protein? | *This gene is not a possible transmembrane protein.*  *According to DeepTMHMM there is not a transmembrane domain in this protein.*  *According to SOSUI this protein is a soluble protein.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes* |
| **DECISION:** | ***portal protein*** |