**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **6** |
| **Stop Coordinate** | **3926** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 5 of 3 bp** |
| **Selected Start Coordinate** | **3708** |
| **Selected Function** | **NKF** |

**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 6.* |
| Is this gene present in other annotated genomes? | *BLAST hit QuadZero\_5 e value 3e-36*  *Phamerator (May 5th)*  *Wardwill\_6 Pham 103828*  *QuadZero\_5 Pham 103828* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes.*  *It is long enough (219 bp, 72 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: 3708*  *GeneMark Start Coordinate: 3708* |
| 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 3708)*  *Raw SD Score -1,582*  *Genomic Z Value 3,073*  *Spacer distance 12*  *Final Score -2,417* |
| 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 2191 bp. This proposed start site has a 3 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\_6*  *Start 7 3708 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 7 was called 100.0 % of time when present. It was called in 18 of the 22 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 7, it was called in 18 of the 22 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *BlastP match (best hit)*  *QuadZero E value 1e-170 Accession Number* [*QWY84452.1*](https://www.ncbi.nlm.nih.gov/protein/QWY84452.1?report=genbank&log$=prottop&blast_rank=1&RID=24YZ7XVW013)  *Alignment with QuadZero*  *Q#1 MAGRIIFTPG*  *S#1 MAGRIIFTPG*  *The start site at 3708 is conserved in other phage genomes as indicated by BlastP.* |
| **DECISION:** | ***Keep DNA Master predicted start site at 3708*** |

**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: hypothetical protein from QuadZero*  *E value 3e-36*  *Q#1 MAGRIIFTPGD*  *S#1 MAGRIIFTPGD*  *NCBI: hypothetical protein from QuadZero*  *E value 3e-41*  *Q#1 MAGRIIFTPGD*  *S#1 MAGRIIFTPGD*  *DNA Master: hypothetical protein from QuadZero*  *E value 3,2e-41*  *Q#1 MAGRIIFTPGD*  *S#1 MAGRIIFTPGD*  *This protein aligns with a hypothetical protein from QuadZero.* |
| 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* [*PF08274.17*](https://www.ebi.ac.uk/interpro/entry/pfam/PF08274)  *NAME ; YjdM\_Zn\_Ribbon ; PhnA Zinc-Ribbon*  *PROBABILITY 91,32*  *E VALUE 0,74*  *SCORE 22,4*  *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 (just a scaffolding protein downstream). This gene is in a region of the genome that shows high conservation of gene order.*  *According to Phamerator, downstream of gene #6 there is a conserved gene annotated as MuF–minor capsid protein or . Upstream of gene #6, Wardwill contains an additional gene found in only three other phages from the EA cluster, although located in a different genomic position (gene #42).* |
| 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? | *Non known function* |
| **DECISION:** | ***Hypothetical Protein*** |