**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **8** |
| **Stop Coordinate** | **5320** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 7 of 115 bp** |
| **Selected Start Coordinate** | **4754** |
| **Selected Function** | **SCAFFOLDING 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 8.* |
| Is this gene present in other annotated genomes? | *BLAST hit QuadZero\_7 e value 3e-104*  *Phamerator (May 5th)*  *Wardwill\_8 Pham 67102*  *QuadZero\_7 Pham 67102* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is a gap of 115 bp with the previous gene.*  *It is long enough (567 bp, 188 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 sites do Glimmer and GeneMark suggest? | *Glimmer Start Coordinate: 4754*  *GeneMark Start Coordinate: 4754* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *The start site has just one possible associated Ribosome Binding site with a high score.*  *RBS 1*  *Raw SD Score -3,014*  *Z Value 2,400*  *Spacer Distance 10*  *Final Score -3,709* |
| 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 567 bp. This proposed start site has a 115 bp gap with the nearest upstream gene.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start site is conserved in other phage genomes. Yes, the start matches the consensus start site predicted from Starterator.*  *Start 11 was the most called start site (95.5%) according to Starterator*  *Yes, the start number called is the most often in the published annotations, it was called in 174 of the 182 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *The start site is conserved in other phages genomes*  *BlastP match (best hit)*  *QuadZero E value 0 Accession Number QWY84454*  *Alignment with QuadZero*  *Q#1 MAPESQEQQA*  *S#1 MAPESQEQQA*  *The start site at 4754 is conserved in other phage genomes as indicated by BlastP.* |
| **DECISION:** | *Keep DNA Master predicted start site* ***4754*** |

**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: scaffolding protein from QuadZero*  *E value 1e-134*  *Q#1 MAPESQEQQ*  *S#1 MAPESQEQQ*  *NCBI: scaffolding protein from QuadZero*  *E value 9e-133*  *Q#1 MAPESQEQQ*  *S#1 MAPESQEQQ*  *DNA Master: scaffolding protein from QuadZero*  *E value 0.0*  *Q#1 MAPESQEQQ*  *S#1 MAPESQEQQ*  *This protein aligns with a scaffolding 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*  *There is a match from HHPred*  *HIT* [*PF10805.13*](https://www.ebi.ac.uk/interpro/entry/pfam/PF10805)  *NAME ; DUF2730 ; Protein of unknown function (DUF2730)*  *PROBABILITY 94,04*  *E VALUE 0,83*  *SCORE 34,48*  *This protein does not align with a protein having a functional assignment in the databases of HHPred with a 90% probability.*  *Other matches from HHPRED*  *HIT* [*Q9T1W9*](http://www.uniprot.org/uniprot/Q9T1W9)  *NAME GP25\_BPMU Uncharacterized protein gp25 OS=Escherichia phage Mu OX=10677 GN=Mup25 PE=2 SV=1*  *PROBABILITY 93,56*  *E VALUE 1,2*  *SCORE 34,47*  *HIT* [*2AKF\_B*](http://www.rcsb.org/pdb/explore/explore.do?structureId=2AKF)  *NAME ; Coronin-1A; coiled coil, coronin 1, PROTEIN BINDING; 1.2A {N/A}*  *PROBABILITY 93,22*  *E VALUE 0,27*  *SCORE 33,75*  *HIT* [*PF08954.16*](https://www.ebi.ac.uk/interpro/entry/pfam/PF08954)  *NAME ; Trimer\_CC ; Trimerisation motif*  *PROBABILITY 92,3*  *E VALUE 0,75*  *SCORE 34,26* |
| 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 located adjacent to genes of known functions (minor and major capsid proteins).This gene is in a region of the genome that shows high conservation of gene order.*  *According to Phamerator, this gene is flanked upstream by the MuF-like minor capsid protein and downstream by the major capsid protein.* |
| 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:** | ***scaffolding protein*** |