**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **62** |
| **Stop Coordinate** | **39387** |
| **Direction (For/Rev)** | **Forward** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 61 405 bp** |
| **Selected Start Coordinate** | **39187** |
| **Selected Function** | **MEMBRANE 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 62.* |
| Is this gene present in other annotated genomes? | *BLAST hit Zayuliv\_61 e value 2e-32*  *Phamerator (May 5th)*  *Wardwill\_62 Pham 233813.*  *Zayuliv\_61 Pham 233813.* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is a 405 bp gap between gene 61 and gene 62.*  *It is long enough (201 bp, 66 aa)*  *Genes before and after are not in the same direction. The direction of gene 61 is reverse while gene 63 direction is forward. Anyways, there is a 405 bp gap between gene 61 and 62, bigger than 30 bp so it doesn’t violate any major guiding principles.* |
| **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 39187*  *GeneMark Start Coordinate 39187* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *The start site has more than just one possible associated Ribosome Binding site with a high score.*  *RBS 1 (Start site at 39187)*  *Raw SD Score -4,013*  *Z Value 1,930*  *Spacer Distance 13*  *Final Score -5,059*  *RBS 2 (Start site at 39346)*  *Raw SD Score -2,908*  *Z Value 2,450*  *Spacer Distance 11*  *Final Score -3,665*  *According to this information, RBS 2 has the highest Raw SD Score and the less negative Final Score.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *For RBS 1: The predicted start codon is the longest ORF, the length is 201 bp. This proposed start site has a 405 bp gap with the nearest upstream gene.*  *For RBS 2: The predicted start codon is not the longest ORF, the length is 42 bp. This proposed start site has a 41 bp gap with the nearest upstream gene. Anyways, this gene violates the Guiding Principles that says \*Protein coding genes are at least 120 bp\*.*  *According to this information, RBS 1 has the appropriate start site.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_62*  *Start 19 39187 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 19 was called 91.7 % of time when present. It was called in 9 of the 19 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 19, it was called in 9 of the 19 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *Best BlastP hit*  *Zayuliv e value 5e-35 Accession Number* [*QOP65257*](https://www.ncbi.nlm.nih.gov/protein/QOP65257.1?report=genbank&log$=prottop&blast_rank=1&RID=2YC9H9RP016)  *Alignment with Zayuliv*  *Q#1 MEDQV-----*  *S#1 MEDQEGNSPQ*  *Start site conserved* |
| **DECISION:** | ***Keep DNA Master predicted site at 39187*** |

**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 Zayuliv*  *E value 2e-32*  *Q#1 MEDQV-----PHQPVK*  *S#1 MEDQEGNSPQAHQPVK*  *NCBI: membrane protein from Zayuliv*  *E value 5e-35*  *Q#1 MEDQV-----PHQPVK*  *S#1 MEDQEGNSPQAHQPVK*  *DNA Master: membrane protein from Zayuliv*  *E value 5e-35*  *Q#1 MEDQV-----PHQPVK*  *S#1 MEDQEGNSPQAHQPVK*  *This protein aligns with a membrane protein from Zayuliv.* |
| 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 not a quality HHPred match.*  *HIT* [*cd20247*](http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd20247)  *NAME DWORF; DWarf Open Reading Frame (DWORF). DWarf Open Reading Frame (DWORF) is a small protein that plays a key role in he*  *PROBABILITY 68,42*  *E VALUE 10*  *SCORE 23,29*    *This protein does align with a protein having a functional assignment in the databases of HHPred with a 90% probability.* |
| 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 the July 30, 2025 Phamerator database version, the membrane protein from Pham 247341 is flanked upstream by Pham 2038 and downstream by Pham 209340, forming a conserved syntenic arrangement among related phages.* |
| Is this gene a possible transmembrane protein? | *This gene is a possible transmembrane protein.*  *According to DeepTMHMM there is a transmembrane domain in this protein.*  *According to SOSUI this protein is a membrane protein.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | ***membrane protein*** |