**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **63** |
| **Stop Coordinate** | **39722** |
| **Direction (For/Rev)** | **Forward** |
| **Gap (Overlap) with Previous Gene** | **Overlap with gene 62 1 bp** |
| **Selected Start Coordinate** | **39387** |
| **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 GeneMarkS and/or GeneMark-host trained coding potential maps show coding potential for gene 63.* |
| Is this gene present in other annotated genomes? | *BLAST hit Zayuliv\_62 e value 4e-34.*  *Phamerator (May 5th)*  *Wardwill\_63 Pham 209340.*  *Zayuliv\_62 Pham 208791.* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is a 1 bp overlap with the previous gene.*  *It is long enough (336 bp, 111 aa)*  *The gene before is in the same direction. If we take gene 1 as the gene after it is also in the same direction (forward)* |
| **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 39387*  *GeneMark Start Coordinate 39387* |
| 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 39306)*  *Raw SD Score -4,283*  *Z Value 1,803*  *Spacer Distance 10*  *Final Score -4,977*  *RBS 2 (Start site at 39387)*  *Raw SD Score -4,623*  *Z Value 1,643*  *Spacer Distance 10*  *Final Score -5,318*  *RBS 3 (Start site 39513)*  *Raw SD Score -1,748*  *Z Value 2,995*  *Spacer Distance 13*  *Final Score -2,794* |
| 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 417 bp. This proposed start site has a 81 bp gap with the nearest upstream gene, which violates the Guiding Principles because of an excessive gene overlap.*  *For RBS 2: The predicted start codon is not the longest ORF, the length is 336 bp. This proposed start site has a 1 bp overlap with the nearest upstream gene.*  *For RBS 3: The predicted start codon is not the longest ORF, the length is 210 bp. This proposed start site has a 126 bp gap with the nearest upstream gene.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_63*  *Start 9 39387*  *The candidate start sites are conserved in other phage genomes. The start called by Glimmer and GeneMark matches the start predicted by Starterator.*  *Start 9 was called 100.0 % of time when present. It was called in 2 of the 4 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 8, it was called in 2 of the 4 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 3e-42 Accesion 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 MSKEQELMAR*  *S#1 MSEEQELMAR*  *Start site is conserved* |
| **DECISION:** | ***Keep DNA Master predicted start site at 39387*** |

**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 4e-34*  *Q#1 MSKEQELMAREMSQPK*  *S#1 MSEEQELMAREMSQPK*  *NCBI: hypothetical protein from Zayuliv*  *E value 3e-42*  *Q#1 MSKEQELMAREMSQPK*  *S#1 MSEEQELMAREMSQPK*  *DNA Master: hypothetical protein from Zayuliv*  *E value 2,9e-42*  *Q#1 MSKEQELMAREMSQPK*  *S#1 MSEEQELMAREMSQPK*  *This protein aligns with a hypothetical 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* [*1OU0\_A*](http://www.rcsb.org/pdb/explore/explore.do?structureId=1OU0)  *NAME precorrin-8X methylmutase related protein; structural genomics, methylmutase, precorrin-8X, PSI, Protein Structure Initi*  *PROBABILITY 41,16*  *E VALUE 140*  *SCORE 235,04*    *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, Pham 209340, the last annotated gene, is flanked upstream by the membrane protein from Pham 247341 and, due to the circular nature of the genome, downstream by Pham 84998 (gene #1), forming a conserved syntenic arrangement among related phages.* |
| 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*** |