**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **61** |
| **Stop Coordinate** | **38510** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 60 272 bp** |
| **Selected Start Coordinate** | **38782** |
| **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 61.* |
| Is this gene present in other annotated genomes? | *BLAST hit Zayuliv\_60 e value 2e-46*  *Phamerator (May 5th)*  *Wardwill\_61 Pham 2038. Gene 61 not shown in Phamerator Genome Map. Pham number from PhagesDB.*  *Zayuliv\_60 Pham 2038.* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is an overlap of 1 bp with the previous gene.*  *It is long enough (273 bp, 90 aa)*  *Genes before and after are not in the same direction. The direction of gene 62 is forward while gene 60 direction is reverse. But there is a 405 bp gap between one another so the gene does not 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 site do Glimmer and GeneMark suggest? | *Glimmer Start Coordinate 38782*  *GeneMark Start Coordinate 38800* |
| 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 38800)*  *Raw SD Score -6,022*  *Z Value 0,986*  *Spacer Distance 7*  *Final Score -7,545*  *RBS 2 (Start site at 38782)*  *Raw SD Score -1,748*  *Z Value 2,996*  *Spacer Distance 9*  *Final Score -2,523*  *Although GeneMark predicted start coordinate is 38800 bp, it does not have an appropriate Raw SD Score and Final Score. RBS 2, called by Glimmer, has the appropriate SD Score and 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 291 bp. This proposed start site has a 290 bp gap with the nearest upstream gene.*  *For RBS 2: The predicted start codon is not the longest ORF, the length is 273 bp. This proposed start site has a 272 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\_61*  *Start 17 38800 bp*  *Start 21 38782 bp*  *Start 23 38773 bp*    *The candidate start sites are conserved in other phage genomes.*  *Start 17 was called 8.3 % of time when present. It was called in 1 of the 58 non-draft genes in the pham.*  *Start 21 was called 81.5 % of time when present. It was called in 18 of the 58 non-draft genes in the pham.*  *Start 23 was called 48.4 % of time when present. It was called in 14 of the 58 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 21, it was called in 18 of the 58 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 2e-59 Accession Number* [*QOP65256*](https://www.ncbi.nlm.nih.gov/protein/QOP65256.1?report=genbank&log$=prottop&blast_rank=1&RID=2YBT0W4D013)  *Alignment with Zayuliv*  *Q#1 MRNMTKVELS*  *S#1 MRNMTKVELS*  *Start site conserved* |
| **DECISION:** | ***Keep DNA Master predicted start 38782*** |

**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-59*  *Q#1 MRNMTKVELSNHVKV*  *S#1 MRNMTKVELSNHVKV*  *NCBI: hypothetical protein from Zayuliv*  *E value 2e-46*  *Q#1 MRNMTKVELSNHVKV*  *S#1 MRNMTKVELSNHVKV*  *DNA Master: hypothetical protein from Zayuliv*  *E value 0.0*  *Q#1 MRNMTKVELSNHVKV*  *S#1 MRNMTKVELSNHVKV*  *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 a HHPred match.*  *HIT* [*PF19881.4*](https://www.ebi.ac.uk/interpro/entry/pfam/PF19881)  *NAME; DUF6354 ; Family of unknown function (DUF6354)*  *PROBABILITY 97,94*  *E VALUE 0,00013*  *SCORE 53,82*    *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 2038 is flanked upstream by Pham 243605 and downstream by a membrane protein belonging to Pham 247341, 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*** |