**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **60** |
| **Stop Coordinate** | **38034** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 59 3 bp** |
| **Selected Start Coordinate** | **38510** |
| **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 60.* |
| Is this gene present in other annotated genomes? | *BLAST hit Zayuliv\_59 e value 2e-59*  *Phamerator (May 5th)*  *Wardwill\_60 Pham 233814. Gene 60 not shown in Phamerator Genome Map. Pham number from PhagesDB.*  *Zayuliv\_59 Pham 233814* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is a gap of 3 bp with the previous gene.*  *It is long enough (477 bp, 158 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 38510*  *GeneMark Start Coordinate 38510* |
| 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 2 (Start site at 38510)*  *Raw SD Score -6,282*  *Z Value 0,864*  *Spacer Distance 13*  *Final Score -7,327*  *RBS 2 (Start site at 38147)*  *Raw SD Score -5,249*  *Z Value 1,349*  *Spacer Distance 10*  *Final Score -5,943*  *RBS 3 (Start site at 38408)*  *Raw SD Score -3,642*  *Z Value 2,105*  *Spacer Distance 7*  *Final Score -5,164* |
| 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 477 bp. This proposed start site has a 1 bp overlap with the nearest upstream gene but it does not violate the Guiding Principles.*  *For RBS 2: The predicted start codon is not the longest ORF, the length is 114 bp. This proposed start site has a 363 bp overlap with the nearest upstream gene that violates the Guiding Principles.*  *For RBS 3: The predicted start codon is not the longest ORF, the length is 375 bp. This proposed start site has a 102 bp overlap with the nearest upstream gene that violates the Guiding Principles.*  *According to this information, although RBS 2 and RBS 3 have better Raw SD Score and Final Score, both violate the Guiding Principles because of an excessive gene overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_60*  *Start 2 38510 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 2 was called 100.0 % of time when present. It was called in 15 of the 16 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 2, it was called in 15 of the 16 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *Best BlastP hit*  *QuadZero e value 2e-74 Accession Number QWY84504*  *Alignment*  *Q#1 MSLYECRVPN*  *S#1 MSLYECQDSV*  *Start site conserved* |
| **DECISION:** | ***Keep DNA Master predicted start site 38510*** |

**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 MSLYECRVPNT*  *S#1 MSLYECQDSVS*  *NCBI: hypothetical protein from QuadZero*  *E value 2e-74*  *Q#1 MSLYECRVPNTDKW*  *S#1 MSLYECQDSVSGKW*  *DNA Master: hypothetical protein from QuadZero*  *E value 0.0*  *Q#1 MSLYECRVPNTDKW*  *S#1 MSLYECQDSVSGKW*  *This protein aligns with hypothetical proteins from QuadZero and 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 are a number of HHPred matches.*  *Match 1*  *HIT* [*PF05823.17*](https://www.ebi.ac.uk/interpro/entry/pfam/PF05823)  *NAME ; Gp-FAR-1 ; Nematode fatty acid retinoid binding protein (Gp-FAR-1)*  *PROBABILITY 92,25*  *E VALUE 3,8*  *SCORE 30,96*  *Match 2*  *HIT* [*6V7M\_B*](http://www.rcsb.org/pdb/explore/explore.do?structureId=6V7M)  *NAME Apolipoprotein E; proteolysis, Alzheimers, lipid, lipoprotein, lipids, disease, LIPID TRANSPORT; 2.0A {Homo sapiens}*  *PROBABILITY 92,25*  *E VALUE 3,8*  *SCORE 30,9*  *Match 3*  *HIT* [*2N5E\_A*](http://www.rcsb.org/pdb/explore/explore.do?structureId=2N5E)  *NAME Apolipoprotein A-I; nanodisc, HDL, lipoproteins, cardiovascular disease, LIPID BINDING PROTEIN; NMR {Homo sapiens}*  *PROBABILITY 90,17*  *E VALUE 5,7*  *SCORE 30,52*    *This protein does align with a protein having a functional assignment in the databases of HHPred with a 90% probability. This protein is probably an apolipoprotein or a fatty acid retinoid binding 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? | *According to the July 30, 2025 Phamerator database version, Pham 243605 is flanked upstream by Pham 87511 and downstream by Pham 2038, 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*** |