**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **55** |
| **Stop Coordinate** | **35673** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | ***36539*** |
| **Selected Function** | **aGPT-Pplase1 domain-containing 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? | *GeneMarkS andr GeneMark-M. folio show coding potential.* |
| Is this gene present in other annotated genomes? | *PhagesDB Blast hit: Zayuliv\_54; e-value= e-166*  *Phamerator: phage: Zayuliv; gene: 54; pham: 86072* (05-06-25) |
| Does the gene violate any major guiding principles? | *No* |
| **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 and GeneMark Start Coordinate: 36539* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 36539*  *Final score: -2,523*  *Z-value: 2,995*  *This is the best score.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *It’s the longest ORF and results in gene overlap of 3 pb.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start site is conserve in another phages* [*genomes. It*](http://genomes.it) *is the same predicted by Starterator.*  *You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn’t reveal a consensus start site, you can record that Starterator was not informative.*  *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: Morrigan\_53, aGPT-Pplase1 domain-containing protein*  *E-value: 1e-110*    ***BlastP match from NCBI***  *Best Hit: aGPT-Pplase1 domain-containing protein [Microbacterium phage Morrigan]*  *E-value: 1e-137*    ***DNA Master***  *Best Hit: hypothetical protein SEA\_ZAYULIV\_54 [Microbacterium phage Zayuliv]*  *E-value: 0*  *Q#1: MKTIEDVVKF*  *S#1: MKTIEDVVKF* |
| **DECISION:** | *The gene should start at 36539* |

**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? | *List the most informative BlastP match from each source PhagesDB: Morrigan\_53, aGPT-Pplase1 domain-containing protein*  *NCBI: aGPT-Pplase1 domain-containing protein [Microbacterium phage Morrigan]*  *DNA Master: hypothetical protein SEA\_ZAYULIV\_54 [Microbacterium phage Zayuliv]*  *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.* |
| 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? | *Glycinyltransferase; DNA hypermodification, phage, 5-PmdU, glycinyltransferase, Pseudomonads PaMx11, TRANSFERASE; HET: SO4; 2.0A {Pseudomonas phage PaMx11}*  *Probability: 94.57%, E-value: 0.29*  *List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*  *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.   CDD results showed a aGPT-Pplase1 domain:* |
| 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 aGPT-Pplase1 domain-containing protein is flanked upstream by Pham 8815 and downstream by Pham 4886, forming a conserved syntenic arrangement among related phages* |
| Is this gene a possible transmembrane protein? | *NO* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes, aGPT-Pplase1 domain-containing protein* |
| **DECISION:** | *aGPT-Pplase1 domain-containing protein* |