**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **46** |
| **Stop Coordinate** | **31383** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **Yes, with gene 45 (22 bp)** |
| **Selected Start Coordinate** | ***31976*** |
| **Selected Function** | ***thymidylate kinase*** |

**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:  Zepp\_45; e-value= e-113  GreenIvye\_44; e-value= e-113  QuadZero\_45; e-value=e-114  Phamerator: pham: 657 (05-05-25)  No similarities found |
| 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 Start Coordinate and GeneMark Start Coordinate: 31976* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Final RBS\_start: -2,214*  *Z-score: -1,439*  *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.*  *The proposed start site don’t have a gap/overlap with the nearest upstream gene.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start match the consensus start site predicted from Starterator: number 9. It’s the first start available in gene 46. It`s conserved in other phage genomes.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***BlastP match from NCBI:***  *Best Hit: Zepp*  *E-value: 3e-143*  *Q#1: MLIAFEGPDNVGKSHSAAAL*  *S#1: MLIAFEGPDNVGKSHSAAAL*  ***PhagesDB***  *Best Hit: Zepp\_45*  *E-value: e-113*  *Q#1: MLIAFEGPDNVGKSHSAAAL*  *S#1: MLIAFEGPDNVGKSHSAAAL*  ***DNA Master***  *Best Hit: QIG58133 (Zepp)*  *E-value: 0*  *Q#1: MLIAFEGPDN*  *S#1: MLIAFEGPDN* |
| **DECISION:** | *The start site is 31976* |

**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? | *PhagesDB: Zepp\_45 thymidylate kinase*  *NCBI: Zepp thymidylate kinase*  *DNA Master: QIG58133 (Zepp) thymidylate kinase* |
| 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? | *Best match (HHPred, NCBI conserved domains database):*  *THYMIDYLATE KINASE; TRANSFERASE, MALARIA, INHIBITOR*  *Probability: 99.36%, E-value: 3.6e-11,* |
| 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, this gene is flanked upstream by a gene from Pham 242357 and downstream by a glycosyltransferase, forming a conserved syntenic arrangement observed in related phages.* |
| Is this gene a possible transmembrane protein? | *NO* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *thymidylate kinase* |