**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **51** |
| **Stop Coordinate** | **34330** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **No** |
| **Selected Start Coordinate** | **35085** |
| **Selected Function** | ***thymidylate synthase*** |

**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: GreenIvy\_49, e-value= e-146*  *QuadZero\_50, e-value= e-145*  *Librie\_50, e-value= e-142*  *Phamerator: phage: QuadZero, Librie*  *Similarities found with gene 50 and 49 (GreenIvy) pham: 654* (5-6-25)  *No similarities found with gene 51* |
| 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 and GeneMark: 35085* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 35085*  *Final score: -2,584*  *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 doesn’t result in gene overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start site 35085 it is conserve in another phages* [*genomes. It*](http://genomes.it) *is the same predicted by Starterator.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | ***PhagesDB***  *Best Hit: >GreenIvy\_49, thymidylate synthase*  *E-value: e-146*  *Q#1: MIHKNTIDDVVLALAAAPIDERTGNRVLYSVDWLA*  *S#1: MIHKNTIDDVVLALAAAPIDERTGNRV*  ***BlastP match from NCBI***  *Best Hit:* thymidylate synthase [Microbacterium phage GreenIvy]  *E-value: 0*  *Q#1: MIHKNTIDDVVLALAAAPIDERTGNRVL*  *S#1: MIHKNTIDDVVLALAAAPIDERTGNRVLYSVDW*  ***DNA Master***  *Best Hit: XEN16700 thymidylate synthase [Mycobacterium phage GreenIvy]*  *E-value: 0*  *Q#1: MIHKNTIDDV*  *S#1: MIHKNTIDDV* |
| **DECISION:** | *The gene should start at 35085* |

**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: GreenIvy\_49, thymidylate synthase*  *NCBI:* thymidylate synthase [Microbacterium phage GreenIvy]  *DNA Master: XEN16700 thymidylate synthase [Mycobacterium phage GreenIvy]* |
| 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? | *The best mach is PROTEÍNA 1B5E\_B (DESOXICITIDILATO HIDROXIMETILASA); HIDROXIMETILASA, COMPLEJO SINTETIZADOR DE DNTP, TRANSFERASA; HET: DCM; 1.6A {fago T4 de Enterobacteria}*  *Probabilidad: 100%, Valor E: 5,2e-41* |
| 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 4923 and downstream by a gene from Pham 7069, 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? | *SI* |
| **DECISION:** | *Yes, it is a thymidylate synthase* |