**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **44** |
| **Stop Coordinate** | **29751** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **3bp** |
| **Selected Start Coordinate** | **30626** |
| **Selected Function** | **MazG-like nucleotide pyrophosphohydrolase** |

**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 and GeneMark-M show coding potential.* |
| Is this gene present in other annotated genomes? | *PhagesDB Blast hit: GreenIvy, gene 42, MazG-like nucleotide pyrophosphohydrolase, 291 Score 595 Evalue e-170*  *Phamerator: phage: GreenIvy, gene: 42, pham: 225020* |
| 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: 30626*  *GeneMark Start Coordinate: 30626* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Predicted site 30626.*  *RBS final score: -2.666*  *Zvalue: 3.055* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *This start site is the longest ORF (876bp).*  *There is a 3bp overlap with downstream gene (gene 45)* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The predicted site (30626) was found in 11 of 12 ( 91.7% ) of genes in pham.*  *Called 100.0% of time when present.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *The best match from NCBI:*  *XEN16693.1 [GreenIvy]*  *e-value: 0.0*  *Q#: MTVTPPEPKF*  *S#: MTVTPPEPKF* |
| **DECISION:** | *Keep DNA master predicted site, 30626* |

**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: GreenIvy, MazG-like nucleotide pyrophosphohydrolase*    *NCBI and DNA master show the same result* |
| 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? | *List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*  *HHPred best match: 3NL9\_A (putative NTP pyrophosphohydrolase)*  *Probability: 99,97*  *Database: PDB* |
| 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 membrane protein from Pham 247197 and downstream by a gene from Pham 242357, forming a conserved syntenic arrangement among related phages.* |
| Is this gene a possible transmembrane protein? | *No, according to DeepTMHMM, it is a Globular protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *Function assigned: MazG-like nucleotide pyrophosphohydrolase* |