**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **41** |
| **Stop Coordinate** | **27828** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **3 bp** |
| **Selected Start Coordinate** | **29255** |
| **Selected Function** | **DNA helicase** |

**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: LilTerminator, gene 40, function unknown, 475 Score 972 Evalue 0.0*  *Phamerator: phage: LilTerminator, gene: 40, pham: 229846* |
| 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: 29255*  *GeneMark Start Coordinate: 29255* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start site 29255*  *RBS final score: -4.256*  *Zvalue: 2.569* |
| 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 (1428 bp)*  *The overlap with downstream gene (gene 42) is: 3bp* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start site is conserved in other phage genomes indicated by Starterator (80). This start was found in 56 of 285 ( 19.6% ) of genes in pham.*  *Called 55.4% of time when present.*  *(29255)*  *The start site called most often was not present in Wardwill genome* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *the best BlastP match from NCBI:*  *XEN16691.1 [GreenIvy]*  *e. value: 0,0*  *Q#: 1 MTLQLRDSQR*  *S#: 1 MTLQLRDSQR* |
| **DECISION:** | *Keep DNA master predicted site: 29255* |

**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\_39, DNA helicase*    *evalue: 0.0*  *NCBI: DNA helicase [GreenIvy\_39]*    *evalue: 0.0* |
| 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? | *HHPred best match: 8V6V\_X (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5; nucleosome, chromatin rem)*  *Probability: 100*  *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? | *This gene is flanked upstream by a gene from Pham 652 and downstream by a gene from Pham 7632, forming a conserved syntenic block among related phages.* |
| Is this gene a possible transmembrane protein? | *No, it appears to be a globular protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *DNA helicase* |