**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **20** |
| **Stop Coordinate** | **12482bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap 21bp** |
| **Selected Start Coordinate** | **10041bp** |
| **Selected Function** | **tape measure 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, Glimmer and GeneMark call it gene.* |
| Is there evidence for coding potential? | *GeneMarkS and GeneMark-host show coding potencial.* |
| Is this gene present in other annotated genomes? | *PhagesDB blast hit GreenIvy\_18 with e value= 0*  *Phamerator phage: pham 59270 (25/4/25)* |
| Does the gene violate any major guiding principles? | *No, the gene does not violate any key guiding principles* |
| **DECISION:** | *Yes, it is a gene.* |

**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:10041*  *GeneMark Start Coordinate:10041* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -1.954*  *Z-score: 3.225*  *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)? | *Yes.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Yes, is the start number called the most often in the published annotations. It was called in 115 of the 182 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: XEN16670 (GreenIvy)*  *E-value: 0*  *Q#1: MADNDVKIKV*  *S#1: MADNDVKIKV* |
| **DECISION:** | *Keep DNA Master pre start site 10041* |

**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\_18, tape measure protein*  *e-vaue: 0.0*  *NCBI:*  *Name: tape measure protein [Microbacterium phage GreenIvy]*  *Sequence ID: XEN16670.1*  *e-value: 0.0*  *DNA Master: tape measure protein [Microbacterium phage GreenIvy]*  *Score: 3472* |
| 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? | *Yes*  *Name:Tape Measure Protein, gp57; phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A {Staphylococcus virus 80alph*  *Probability:99.95*  *e-value:2.1e-19* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *YES, it is between “tail assembly chaperone” and “minor tail protein”* |
| Is this gene a possible transmembrane protein? | *YES, TMHMM:*  *# WEBSEQUENCE Length: 814*  *# WEBSEQUENCE Number of predicted TMHs: 9*  *# WEBSEQUENCE Exp number of AAs in TMHs: 233.27762*  *# WEBSEQUENCE Exp number, first 60 AAs: 14.2707*  *# WEBSEQUENCE Total prob of N-in: 0.61596*  *# WEBSEQUENCE POSSIBLE N-term signal sequence*  *SOSUI: your query is MEMBRANE protein*  *No. region transmembrane seq. type*  *1 292-314 ESIGASMGILGPQLGSMMSTVVE Secondary*  *2 322-344 AVLEAGVALIQGLIQGITTALPT Primary*  *3 369-391 AAGAQAVVALVTGLASAAPQIIT Primary*  *4 401-423 VQEVINAAPALLEGGIALIMGLA Secondary*  *5 439-461 YLIEGIVAFIQTGVPMLLDAGIQ Primary*  *6 482-504 PGMITSIIGALVGMIPLIVQAGV Primary*  *7 526-548 PQIISSVLTAVLGAIPVIIQGGI Primary*  *8 570-592 PQIITGVISAVIGAIPQIIQAGI Primary*  *9 622-644 SAIVGAVPQLAQAGLQLIQGLWN Secondary* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES, tape measure protein* |
| **DECISION:** | *it is a “tape measure protein”* |