**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **29** |
| **Stop Coordinate** | **19282bp** |
| **Direction (For/Rev)** | **Rev** |
| **Gap (Overlap) with Previous Gene** | **Gap 13bp** |
| **Selected Start Coordinate** | **19467bp** |
| **Selected Function** | **Hypothetical 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 Quaszero\_27 with e value= 1e-27*  *Phamerator phage: pham 229938 (25/4/25)* |
| Does the gene violate any major guiding principles? | *Yes, the gene fails to comply with one key guiding principles* |
| **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: 19282*  *GeneMark Start Coordinate: 19282* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -2.072*  *Z-score: 3.236*  *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? | *The start number called the most often in the published annotations is 29, it was called in 147 of the 175 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QWY84472 (SEA\_QUADZERO\_27)*  *E-value: 6.7E-34*  *Q#1: MAQPTKQVKV*  *S#1: MAQSTKQVKV* |
| **DECISION:** | *Keep DNA Master pre start site 19282* |

**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:***  *QuadZero\_27, function unknown, Evalue:1e-27*  *Query: 1 MAQPTKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL 60*  *MAQ TKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL*  *Sbjct: 1 MAQSTKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL 60*  *Query: 61 P 61*  *P*  *Sbjct: 61 P 61*  ***NCBI:***  [Microbacterium phage QuadZero](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2836050)*; ([QWY84472.1](https://www.ncbi.nlm.nih.gov/protein/QWY84472.1?report=genbank&log$=prottop&blast_rank=1&RID=3HB9SD5H016)); E value:7e-34*  *Query 1 MAQPTKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL 60*  *MAQ TKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL*  *Sbjct 1 MAQSTKQVKVENLNHNDVIIDPEGNEARVIRVRRIDHLRAKLETDIGARIVNLADKFPVL 60*  *Query 61 P 61*  *P*  *Sbjct 61 P 61*  ***DNA Master:***  *idem* |
| 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? | *No, this protein does not align with any known protein of funcional assignment with a probability higher than 90%.* |
| 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, one adjacent genes correspond to proteins of function holin (28).* |
| Is this gene a possible transmembrane protein? | No, the TMHMM result does not indicate the presence of a possible transmembrane domain. |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Hypothetical Protein* |
| **DECISION:** | *Hypothetical Protein* |