**Student Gene Annotation Worksheet**

| **Basic Phage Information** | |
| --- | --- |
| **Phage Name** | **Wardwill** |
| **Gene #** | **17** |
| **Stop Coordinate** | **9631bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap 30 bp** |
| **Selected Start Coordinate** | **9062bp** |
| **Selected Function** | **tail assembly chaperone** |

**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 QuadZero\_16 with e value=e-107*  *Phamerator phage: pham 84979 (25/4/25)* |
| Does the gene violate any major guiding principles? | *Yes, the gene fails to complu whit one key guiding principles* |
| **DECISION:** | *Yes, probably 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: 9062*  *GeneMark Start Coordinate: 9062* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -2.584*  *Z-score: 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)? | *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 163 of the 182 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QWY84462 (QUADZERO)*  *E-value: 0*  *Q#1:MTFSSYEELA*  *S#1:MTFSSYEELA* |
| **DECISION:** | *Keep DNA Master pre start site 9062* |

**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\_16, tail assembly chaperone, 189  E-value: e-107  *NCBI* Microbacterium phage QuadZero  **Sequence ID:** [**QWY84462.1**](https://www.ncbi.nlm.nih.gov/protein/QWY84462.1?report=genbank&log$=protalign&blast_rank=1&RID=2YPU6T23016)  E-value: 5e-134  *DNA Master:* Microbacterium phage QuadZero  Score: 920 |
| 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 Quality Match*  *Probability: 58,34 %*  *E-value: 120*  *Name: NUCLEAR DISTRIBUTION PROTEIN NUDE-LIKE 1; DEVELOPMENTAL PROTEIN, NUCLEAR PROTEIN, NEUROGENESIS, CYTOSKELETON, LIS1 BINDI* |
| 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 “major tail protein” and “tail assembly chaperone”* |
| Is this gene a possible transmembrane protein? | *No, TMHMM result does not show posible transmembrane subunit.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes* |
| **DECISION:** | *tail assembly chaperone* |