**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **31** |
| **Stop Coordinate** | **19685** |
| **Direction (For/Rev)** | **Reverse** |
| **Gap (Overlap) with Previous Gene** | **No gap; zero overlap.** |
| **Selected Start Coordinate** | **19867** |
| **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, at different start sites.* |
| Is there evidence for coding potential? | *Both GeneMark and GeneMarkS show coding potential in this region, in the first ORF of the complementary sequence.* |
| Is this gene present in other annotated genomes? | *Zayuliv\_29, function unknown, 60, Score: 99, E-value: 4e-21.*  *FALTA ESTO: No si si está bien. Pero se encuentra el fago Zayuliv, Gene 29, 85118 (149).* |
| Does the gene violate any major guiding principles? | *No overlap.*  *Genes before and after this gene in the same direction.*  *Maybe not long enough (132 bp for start site 19816, 183 bp for start site 19867).* |
| **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: 19816.*  *GeneMark Start Coordinate: 19867.* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start 19816:*  *Final RBS score:-6.831*  *Zvalue: 1.321*  *Start 19867*  *Final RBS score:-5.709*  *Zvalue: 1.625*  *This start has better score and longer ORF* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start site 19867 is the longest ORF, 183 bp.*  *It comes right after stop codon of upstream gene #32, with 1 bp overlap.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | Orpham, no data. |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master (19867)*  *Best hit: Q0P65226 X0E90257 (Zayuliv\_29)*  *Evalue: 4.1E-12*  *Q#: 1 MEEGSWDRTA*  *S#:18 MEEGSWDRTA* |
| **DECISION:** | *Based on this information and considering coding potential given by GeneMark, start site is probably at 19867, as predicted by GeneMark. This has better RBS score, less gap with upstream gene, and a longer ORF more similar to gene #29 from Zayuliv.* |

**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? | *List the most informative BlastP match from each source PhagesDB hit: Zayuliv\_29, unknown function*    *NCBI and DNA master show the same results* |
| 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, best HHPred hit:* [*1QYP\_A*](http://www.rcsb.org/pdb/explore/explore.do?structureId=1QYP) *“RNA POLYMERASE II; TRANSCRIPTION, RNA POLYMERASE II SUBUNIT, RPB9, ZN RIBBON, HYPERTHERMOPHILIC, EXTREMOPHILE; HET: ZN”*  *Probability of 73.84*  *Evalue of 1.2* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *The annotated gene labeled as 31 (draft version) is an Orpham and appears to have replaced a gene belonging to Pham 85118. It is located between a gene from Pham 6231 upstream and another from Pham 88242 downstream. Pham database updated as of July 31st, 2025.* |
| Is this gene a possible transmembrane protein? | *No, according to DeepTMHMM, this is a globular protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *NKF* |
| **DECISION:** | *Hypothetical Protein* |