**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **33** |
| **Stop Coordinate** | **20457** |
| **Direction (For/Rev)** | **Reverse** |
| **Gap (Overlap) with Previous Gene** | **Yes, overlap bp 22124-22100 with Gene #34** |
| **Selected Start Coordinate** | **22124** |
| **Selected Function** | **RecA-like DNA Recombinase** |

**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 starts.* |
| Is there evidence for coding potential? | *Both GeneMark and GeneMarkS show coding potential in this region, in the second ORF of the complementary sequence. Start site seems to be downstream of bp 22100.* |
| Is this gene present in other annotated genomes? | *Zepp\_32, RecA-like DNA recombinase, 555, Score:* [*1131*](https://phagesdb.org/blast/results/blast.cgi#535890)*, E- value: 0.0.*  *Pham: Phage Zepp, Gene 32, 210763 (279).* |
| Does the gene violate any major guiding principles? | *Overlap with Gene #34, few bp (25 bp).*  *Genes before and after this gene in the same direction.*  *Long enough.* |
| **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: 22124.*  *GeneMark Start Coordinate: 22055.* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *Start 22124:*  *Final RBS score:-2.433*  *Zvalue: 3.066*  *This is the best value.*  *Start 22055:*  *Final RBS score:-4.416*  *Zvalue: 2.134* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The start site 22124 has the longest ORF, 1668 bp. It has a 23 bp overlap with upstream gene #34.*  *The start site 22055 has a 1599 ORF.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *The start 14 at 22124 bp has 168 manual annotations of 243 non-draft genes in the pham.*  *The start 20 at 22091 has 5 manual annotations of 243 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master (22124)*  *Best hit: QIG58120 (Zepp)*  *Evalue: 0*  *Q#: 1 MDSNPLDVVR*  *S#: 1 MDSNPLDVVR* |
| **DECISION:** | *Based on this information and specially considering the similarity with gene from Zepp, Glimmer predicted start site at 22124 is accurate.* |

**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: Zepp\_32, RecA-like DNA Recombinase*  *Evalue: 0*    *NCBI and DNA Master show the same result* |
| 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 hit:* [*1NLF\_A*](http://www.rcsb.org/pdb/explore/explore.do?structureId=1NLF) *“Regulatory protein repA; replicative DNA helicase structural changes, REPLICATION; HET: SO4; 1.95A {Escherichia coli} SC”*  *Probability: 99.86*  *Evalue: 1.2e-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? | *This gene is flanked upstream by a gene from Pham 88242 and downstream by a nuclease, a conserved syntenic arrangement observed across related phages* |
| Is this gene a possible transmembrane protein? | *No, according to SOSUI, this is a SOLUBLE protein* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | *RecA-like DNA Recombinase* |