**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** |  |
| **Gene #** |  |
| **Stop Coordinate** |  |
| **Direction (For/Rev)** |  |
| **Gap (Overlap) with Previous Gene** |  |
| **Selected Start Coordinate** |  |
| **Selected Function** |  |

**Annotation Decision #1: Is this a Gene?**

|  |  |
| --- | --- |
| **Gathering Evidence** | **Explain Your Rationale** |
| Was the gene called by an auto-annotation program (Glimmer, GeneMark)? | *Examine the data for the auto-annotation programs and indicate (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)* |
| Is there evidence for coding potential? | *Discuss whether the GeneMarkS and/or GeneMark-host trained coding potential map(s) show coding potential.* |
| Is this gene present in other annotated genomes? | *Discuss if other* ***related, annotated*** *phages contain this gene. In your answer, record the name of the phage, gene #, and e-value of the PhagesDB Blast hit. Listing the best match is sufficient.*  *Did you observe the same gene (similar pham) in an annotated phage of the same cluster in Phamerator. Indicate the phage name, gene number, and pham of the similar gene.* |
| Does the gene violate any major guiding principles? | *Discuss if there are any significant violations of the* [*Guiding Principles of Genome Annotation*](https://seaphagesbioinformatics.helpdocsonline.com/article-27) *with the gene call. Do you see significant overlap with other genes? Is it long enough? Are the genes before and after this gene in the same direction?* |
| **DECISION:** | *Respond here with YES or NO after reviewing the evidence gathered above.* |

**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 (type NA if not supported)::*  *GeneMark Start Coordinate (type NA if not supported)::* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.*  *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?*  *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn’t reveal a consensus start site, you can record that Starterator was not informative.*  *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.*  *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* |
| **DECISION:** | *Record where you think the gene should start here and briefly explain your rationale.* |

**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:*  *NCBI:*  *DNA Master:*  *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.* |
| 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? | *List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*  *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.* |
| Is this gene a possible transmembrane protein? | *If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Indicate a response with a Yes or No response.*  *Once you have arrived at a functional decision, check the* [*SEA-PHAGES Official Function List*](https://seaphages.org/blog/2017/10/30/official-function-list/) *to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.* |
| **DECISION:** | *If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record “NKF” for no known function. 50-70% of phage genes fall into the NKF category.* |