

Student Gene Annotation Worksheet

Name _____

Date _____

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)?	<i>Examine the data for the auto-annotation programs and indicate (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)</i>
Is there evidence for coding potential?	<i>Discuss whether the GeneMarkS and/or GeneMark-host trained coding potential map(s) show coding potential.</i>
Is this gene present in other annotated genomes?	<i>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.</i>
Does the gene violate any major guiding principles?	<i>Discuss if there are any significant violations of the Guiding Principles of Genome Annotation 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?</i>
DECISION:	<i>Respond here with YES or NO after reviewing the evidence gathered above.</i>

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?	<i>Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::</i>
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	<i>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.</i>
Is this start site conserved in other phage genomes as indicated by Starterator?	<i>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.</i>
DECISION:	<i>Record where you think the gene should start here and briefly explain your rationale.</i>

- 1) **DUPLICATE** worksheet (table 1 and 2) for each gene you are assigned
- 2) Add the information to PECAAN