Student Gene Annotation Worksheet

Name_____ Date_____

Basic Phage Information	
Phage Name	**DesireeRose**
Gene #	**8**
Stop Coordinate	**3595**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**81**
Selected Start Coordinate	**4275**
Selected Function	**Portal 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*
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.* **Phage:DesireeRose** **Gene #:Protein #8** **E-value: 1e-128** *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 did observe the same gene with a pham of 10322 which is in the same cluster which can be identified as the Badulia_Draft.**
Does the gene violate any major guiding principles?	*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?*

	There was a significant overlap because the gap was 85. It is long enough because the sequence length is 682. The genes are all going in the same direction.
DECISION:	*Yes. It is violated by overlapping 85 and 385.*

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)::**4275*** *GeneMark Start Coordinate (type NA if not supported)::**4275***
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?* ***The predicted start codon is the longest as the length is listed as the longest. It doesn't have a gap that doesn't 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.* ***The start sites don't match each other because the selected one starts at 4275 and the gene beneath that has a start of 3975.*** *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	***The Glimmer starts and the Genemark start at the same place.***

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