Student Gene Annotation Worksheet

Name_____   Date_____

Basic Phage Information	
Phage Name	**DesireeRose**
Gene #	**7**
Stop Coordinate	**1547**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**-4**
Selected Start Coordinate	**3598**
Selected Function	**Dna primase/polymerase**

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 #7*** ***E-value:0*** ***Cluster:GE*** *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 observed the same gene in the same cluster that can be identified as MuffinTheCat.*** ***Pham #: 4236*** ***Gene #: Protein 7***

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?* **I don't see an overlap with any other genes significantly besides the ones listed in the cluster ; GE. All the genes before and after are in the same direction. It is considered long because the sequence length is 683.**
DECISION:	**YES. It was violated by overlapping the bp number from -4 to 218.**

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):: **3598*** *GeneMark Start Coordinate (type NA if not supported):: **3598***
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.* ***The predicted start codon is the longest as it has the biggest length****. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles? I**t 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? **Both of the start sites match one another.** 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.*
DECISION:	***I think the gene should start at 3598 considering the fact that the Glimmer start coordinate and the GeneMark are the same.***

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