Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	
Gene #	
Stop Coordinate	
Direction (For/Rev)	
Gap (Overlap) with Previous Gene	**(BIGGER START – SMALLER START)**
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 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-	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are*

PHAGES approved function list?	*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.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**11**
Stop Coordinate	**4646**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**13**
Selected Start Coordinate	**N/A**
Selected Function	Hypothetical protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation	*Called by glimmer, not genemark*

program (Glimmer, GeneMark)?	
Is there evidence for coding potential?	*No*
Is this gene present in other annotated genomes?	*No*
Does the gene violate any major guiding principles?	Yes
DECISION:	*No*

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: 4521 Gene Mark:N/A*
Does the start site have an associated Ribosome Binding Site with a high score?	N/A
Is the predicted start codon the longest ORF? If not, does the longest	*N/A*

ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	*N/A*
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a	

region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**12**
Stop Coordinate	**5763**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**114**
Selected Start Coordinate	**4648**
Selected Function	Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	No
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?	*4648*
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, -5.308
Is the predicted start codon the longest ORF? If not, does the longest	*Yes*

ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 37.0% of phages and called 100% of the time when present*
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	Yes

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?	
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?	
Is this gene located adjacent to genes of known function and in a	

region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**13**
Stop Coordinate	**6080**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**5793**
Selected Function	Hypothetical protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer called 5760, Genemark called 5793*
Is there evidence for coding potential?	*Yes, there is*

Is this gene present in other annotated genomes?	*Yes, it is*
Does the gene violate any major guiding principles?	There is no violation.
DECISION:	

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 suggests 5760 and genemark calls 5793*
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it is -5.773
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest, the longest was 789.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is*

Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	

Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**14**
Stop Coordinate	**6277**
Direction (For/Rev)	**6277**
Gap (Overlap) with Previous Gene	**6**
Selected Start Coordinate	**6086**
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer called 6086 and GeneMark called 6128.*
Is there evidence for coding potential?	*Yes there is coding potential.*
Is this gene present in other annotated genomes?	*Yes in 9 other genomes.*
Does the gene violate any major guiding principles?	No
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 called 6086 and Genemark called 6128.*
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 1.994. The final score is –6.001.
Is the predicted start codon the longest ORF?	*No, it's not the longest.*

If not, does the longest ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes*
Is this start site conserved in other phage genomes as indicated by BlastP?	*no*
DECISION:	Yes

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?	
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?	
Is this gene located adjacent to genes of	

known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**15**
Stop Coordinate	**7182**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**99**
Selected Start Coordinate	**6376**
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*It was called, but they both disagree on the start*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	Yes, the gap is more than 30
DECISION:	

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: 6376* *GeneMark: 6403*
Does the start site have an associated Ribosome Binding Site with a high score?	RBS score: -5.308

Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes*
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**16**
Stop Coordinate	**8077**

Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**17**
Selected Start Coordinate	**7199**
Selected Function	Major Head Capsid

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Called by both Original Glimmer and GeneMark*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	ORF Length Over 120 bp
DECISION:	

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?	*They both suggest a start site of 7199*
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has a Z value: 2.313
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it's the longest ORF Length*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes*
Is this start site conserved in other phage genomes as indicated by BlastP?	*yes*
DECISION:	

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	No

smaller with appropriate coverage?	
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?	Yes
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	Yes
Is this gene a possible transmembrane protein?	Yes
Is the proposed function found on the SEA-PHAGES approved function list?	Yes
DECISION:	Major Head Capsid

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**

Gene #	**17**
Stop Coordinate	**8551**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**4**
Selected Start Coordinate	**8081**
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Called by both Glimmer and GeneMark*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	No
DECISION:	

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 and GeneMark suggests a start site of 8081=-*
Does the start site have an associated Ribosome Binding Site with a high score?	No, it has a Z - score of 1.412 The final Score is –6.038
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon has an ORF of 471*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved by 77.8% of genes as indicated by Starterator*
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**18**
Stop Coordinate	**9116**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**111**
Selected Start Coordinate	**8622**
Selected Function	Head to Tail Adapter

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer, Genemark called start at 8622*
Is there evidence for coding potential?	*Yes, there is coding potential shown in genemark*
Is this gene present in other annotated genomes?	*Yes, it barely overlaps with gene 19*
Does the gene violate any major guiding principles?	Violation of gap

DECISION:	

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?	*8622*
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 2.727 The final score – 3.897
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The ORF is 495 which is the longest*
Is this start site conserved in other phage genomes as indicated by Starterator?	*It is conserved in 44.2 percent of other genes.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes*
DECISION:	

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?	Yes
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?	Yes
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	Yes
Is this gene a possible transmembrane protein?	Yes
Is the proposed function found on the SEA-PHAGES approved function list?	Yes
DECISION:	Head to tail adapter

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**19**
Stop Coordinate	**9436**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**9113**
Selected Function	Head to tail stopper

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes*
Is there evidence for coding potential?	*Yes*

Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	No
DECISION:	

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 suggests:7199 and gene Mark does not call anything.*
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 2.315 The final score – 4.271
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes its the longest with 324*
Is this start site conserved in other phage genomes as indicated by Starterator?	*It is conserved at 16.1% in other genes*

Is this start site conserved in other phage genomes as indicated by BlastP?	*yes*
DECISION:	Yes

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?	yes
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?	yes
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	yes
Is this gene a possible transmembrane protein?	yes

Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	Head to tail stopper

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**20**
Stop Coordinate	**9792**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**51**
Selected Start Coordinate	**9487**
Selected Function	Hypothetical 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*
Is there evidence for coding potential?	*Yes, there is*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	The gap is a guideline violation
DECISION:	

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 called 9487, but genemark called 9433*
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 1.590 The final score – 6.099
Is the predicted start codon the longest ORF?	*The ORF length is 306 but the longest is 360*

If not, does the longest ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	*yes*
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of	

known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	
Gene #	
Stop Coordinate	
Direction (For/Rev)	
Gap (Overlap) with Previous Gene	**(BIGGER START – SMALLER START)**
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 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 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.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**MenE**
Gene #	**1**
Stop Coordinate	**522**
Direction (For/Rev)	**Forward**

Gap (Overlap) with Previous Gene	**-----**
Selected Start Coordinate	**1**
Selected Function	**Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	*Yes, in 100% of times the gene present it's called*
Does the gene violate any major guiding principles?	*No*
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?	*Both Glimmer and Genemark suggests a start site @bp 1*
Does the start site have an associated Ribosome Binding Site with a high score?	*The Final RBS Score is predicted to be: -3.356* *The Genomic Z value was: 2.819* *2.819 was the highest Z value recorded.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon has the longest ORF length of 522.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, 100% of the time start site 1 is conserved.*
Is this start site conserved in other phage genomes as indicated by BlastP?	--
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**2**
Stop Coordinate	**656**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**519**
Selected Function	**Hypothetical protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Both Glimmer & GeneMark agreed on call @bp 519*
Is there evidence for coding potential?	*Yes, there is evidence for coding potential.*
Is this gene present in other annotated genomes?	*Yes, 100% of the time start site 519 is conserved.*
Does the gene violate any major guiding principles?	*Yes, the Genomic Z value is less than different possible start codon.*
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?	*Both Glimmer and GeneMark Suggest the Start Coordinate: 519*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a finale RBS score of: -6.770* *The Genomic Z value was: 1.575* *1.575 was not the highest/best recorded score.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the predicted start codon does not have the longest ORF. The longest ORF excessive gene overlaps >30bp.*
Is this start site conserved in other phage genomes	*Yes, the start site is conserved in other phage genomes as indicated by Starterator.*

as indicated by Starterator?	
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	*No, none of the ones it aligns with has a functional assignment*
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**3**
Stop Coordinate	**880**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**653**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Both Glimmer and Genemark call bp 653*
Is there evidence for coding potential?	*Yes, there is coding potential*
Is this gene present in other annotated genomes?	*Yes 100% called starting point in 9/9*
Does the gene violate any major guiding principles?	*Does not violate principles*
DECISION:	

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?	*653*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a finale RBS score of: -5.621* *The Genomic Z value was: 1.697* *1.697 was not the highest/best recorded score.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes the predicted start codon has the longest ORF of 228*
Is this start site conserved in other phage genomes	*Yes it is conserved*

as indicated by Starterator?	
Is this start site conserved in other phage genomes as indicated by BlastP?	--
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**4**
Stop Coordinate	**1059**
Direction (For/Rev)	**Forward**

Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**883**
Selected Function	**Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is called 100% of the time when present*
Does the gene violate any major guiding principles?	*No*
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?	*883*
Does the start site have an associated Ribosome Binding Site with a high score?	*The RBS is –2.034* *The genomic z value is 3.499* *3.499 is the highest Z value*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it's not the longest.* *The longest does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes. Start 1 was called in 100% of the time when present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**5**
Stop Coordinate	**1172**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	**1059**
Selected Function	**Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes, it is a gene*

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?	*1059*
Does the start site have an associated Ribosome Binding Site with a high score?	*The RBS is –3.474* *The Z value is 2.763* *2.763 is the highest Z value*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*It is conserved in 58.3% of the other phage genomes*
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**6**
Stop Coordinate	**1405**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	**1172**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Both*
Is there evidence for coding potential?	*Yes, because*
Is this gene present in other annotated genomes?	*Yes, most of the time the start site 1172 is conserved.*

Does the gene violate any major guiding principles?	*No*
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?	*Both Glimmer and GeneMark suggest the start codon to be 1172*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a final RBS score of : 2.780* *The Genomic Z value was : 1.698* *1.698 was not the highest/best recorded score.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No the predicted start codon does not have the longest ORF. The longest ORF excessive gene overlaps >30bp.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Glimmer, GeneMark, and phagedb called the most annotated start site to be 1172, 7 out of 8 drafted.*
Is this start site conserved in other phage genomes as indicated by BlastP?	--
DECISION:	*Yes*

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?	
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?	
Is this gene located adjacent to genes of	

known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**7**
Stop Coordinate	**1521**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**1174**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*The gene was called Glimmer, not Genemark.*
Is there evidence for coding potential?	*No*
Is this gene present in other annotated genomes?	*No, it is the only one in its category*
Does the gene violate any major guiding principles?	*No*
DECISION:	

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?	*1174 bur genemark does not call that. When revised 1174 was the best probable option*
Does the start site have an associated Ribosome	*N/A*

Binding Site with a high score?	
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No*
Is this start site conserved in other phage genomes as indicated by Starterator?	*No*
Is this start site conserved in other phage genomes as indicated by BlastP?	---
DECISION:	

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?	*No*
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?	*No*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*No Function, Not a Gene*

Basic Phage Information	
Phage Name	**MenE**
Gene #	**8**
Stop Coordinate	**2885**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**122**
Selected Start Coordinate	**1527**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*It was called, but they disagreed*
Is there evidence for coding potential?	*Yes there is coding potential*
Is this gene present in other annotated genomes?	*Yes, the most annotated start is 1527*
Does the gene violate any major guiding principles?	*Yes*
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 suggests 1527, and Genemark suggests 1518*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a finale RBS score of : -6.565* *The Genomic Z value was: 1.143*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, and the longest ORF results in an excess gene overlap of 113.*

Is this start site conserved in other phage genomes as indicated by Starterator?	*It is conserved in 60% of other phage genomes*
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**9**

Stop Coordinate	**2964**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**2836**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Called By Glimmer, not GeneMark*
Is there evidence for coding potential?	*No*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	*Yes, it overlaps with gene #10, has little to no coding potential, and the bp is below 200.*
DECISION:	*No. It is not a gene*

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 calls @bp 2836*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a final RBS score of : -6.038* *The Genomic Z value was : 1.412* *1.412 highest*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes*
Is this start site conserved in other phage genomes as indicated by Starterator?	*No*
Is this start site conserved in other phage genomes as indicated by BlastP?	---
DECISION:	

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?	
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	
Is the proposed function found on the SEA-PHAGES approved function list?	
DECISION:	

Basic Phage Information	
Phage Name	**MenE**
Gene #	**10**
Stop Coordinate	**4534**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**66**
Selected Start Coordinate	**2921**
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)?	*Called by Glimmer but Genemark called 2984*
Is there evidence for coding potential?	*Yes there is a bunch*
Is this gene present in other annotated genomes?	*Yes*
Does the gene violate any major guiding principles?	*Yes the gap is more than 30bp*
DECISION:	

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 suggest 2921 and Phamerator supports*
Does the start site have an associated Ribosome Binding Site with a high score?	*There was a final RBS score of : -4.475* *The Genomic Z value was : 2.252* *2.252 was not the highest/best recorded score.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes*
Is this start site conserved in other phage genomes	*Yes, it is present in 25.0 % of other phage genomes*

as indicated by Starterator?	
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	*Phages db: 0.0, portal protein* *NCBI: 0.0, portal protein*
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: 100 probability of a portal protein* *Above 90%* *100% identical to portal protein*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*yes*
DECISION:	*Portal protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	MenE
Gene #	21
Stop Coordinate	10232
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	9789
Selected Function	Tail terminator

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?	Yes
Is this gene present in other annotated genomes?	Found in 7.6% of genes
Does the gene violate any major guiding principles?	No
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 and GeneMark at 9789
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.808 Final score: -3.447
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Yes, it's the longest at 444 ORF
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, in about 7.6% of the other genes
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	*Yes*

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?	*PhagesDB: 4e-82, tail terminator* *NCBI: 2e – 101, tail terminator* *HHpred: no e values below* 10^{-4}
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a	*HHpred: 100% probability with a gene for a tail terminator protein*

probability of 90% or greater with appropriate coverage?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes tail terminator*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*Tail terminator*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	22
Stop Coordinate	10760
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**16 bp**
Selected Start Coordinate	10248

Selected Function	Major tail 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 it was called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, it is called 96.6% of the time when present
Does the gene violate any major guiding principles?	No
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?	Both Glimmer and GeneMark suggest a start site of 10248
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has the highest Z value: 3.146
Is the predicted start codon the longest ORF? If not, does the longest	*Yes, the predicted start codon has the longest ORF with a length of 513 bp.*

ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in 30.2% of genes as indicated by Starterator (called 96.6% when present)
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	*Yes*

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?	*PhagesDB: 5e – 97, Major tail protein* *NCBI: 7e-118 , Major tail protein* *HHped:*
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?	*100% align with this function*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No synteny observed*

Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*Major tail protein*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	23
Stop Coordinate	11323
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	**75**
Selected Start Coordinate	10835
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation	Called by Glimmer

program (Glimmer, GeneMark)?	
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes
Does the gene violate any major guiding principles?	Gap violation
DECISION:	Yes, it is a gene

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 calls 10835
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 1.467 The final score – 6.312
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest, 489*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes it is by 28.6%
Is this start site conserved in other	*N/A*

phage genomes as indicated by BlastP?	
DECISION:	*A gene*

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?	
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?	
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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	24
Stop Coordinate	11610
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**-16**
Selected Start Coordinate	11350
Selected Function	Hypothetical 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?	yes
Is this gene present in other annotated genomes?	100% found in other genes
Does the gene violate any major guiding principles?	No
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 calls 11350 and GeneMark is 11320
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.190 Final score: -4.866
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No its not the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes 100%

Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*yes*

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-	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the [SEA-PHAGES Official Function List](#) to ensure that you are*

PHAGES approved function list?	*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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	25
Stop Coordinate	13833
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	**25**
Selected Start Coordinate	11635
Selected Function	Tape measure 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
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Found in 13 of 34 (38.2%) of genes in pham
Does the gene violate any major guiding principles?	ORFs length over 120bp
DECISION:	Yes, it is a gene

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?	Both Glimmer and GeneMark suggest a start site of 11635
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has a Z Score of: 2.301, but it is not the highest Z score. The final Score was : -4.299
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon has the longest ORF Length.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, conserved 100% of the times

Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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?	*NCBI: 0.0 , tape measure protein* *Phagesdb: 0.0 tape measure protein*
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?	*Above 90%* *HHpred: 99.86% identical gene that it is a tape measure protein*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	*Yes* *PDB_mmCIF70_8_Mar* *CATH_S40_v4.3*
Is the proposed function found on the SEA-	yes

PHAGES approved function list?	
DECISION:	*Tape measure protein*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	26
Stop Coordinate	14693
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	13833
Selected Function	Minor tail protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Both Called
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes 5.9%
Does the gene violate any major guiding principles?	No
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 called 11350 and genemark called 11320
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.737 Final Score: -3.586
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes its the longest ORF with 861*
Is this start site conserved in other phage genomes as indicated by Starterator?	20 called the same start out of 339
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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: 0.0, minor tail protein* *DNA Master: e-163 minor tail protein*
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?	*100% identical to minor tail protein* *HHpred: 99.96%identical to the minor tail protein* *Above 90%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*Minor tail protein*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE

Gene #	27
Stop Coordinate	16561
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	14693
Selected Function	Minor tail 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, it was called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, the start site is called 100% of the time.
Does the gene violate any major guiding principles?	No
DECISION:	Yes, it is a gene

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?	They both suggest the start site: 14,693

Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has the highest Z-value of “2.335” The final Score: -4.374
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*:* *No, it is not the longest ORF length, and the gene overlap does not exceed 30.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved 100% of the time
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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: 0.0, minor tail protein* *DNA Master: 0.0, minor tail protein*
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	*100% identical to minor tail protein* *HHpred: 99.85% identical to a minor tail protein* *Above 90%*

greater with appropriate coverage?	
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?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*Minor tail protein*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	28
Stop Coordinate	16876
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**7**
Selected Start Coordinate	16568
Selected Function	Hypothetical 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 glimmer only
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	100% of the time its present
Does the gene violate any major guiding principles?	No
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 : 16568* *GeneMark Start Coordinate NA*
Does the start site have an associated Ribosome Binding Site with a high score?	Final Score: -3.190 Z Score: 3.047
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF with 309*
Is this start site conserved in other	

phage genomes as indicated by Starterator?	27 out of 27 called the same start
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes*

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](#) 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	29
Stop Coordinate	17621
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**12**
Selected Start Coordinate	16887
Selected Function	Minor tail 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?	Yes, as shown in GeneMark.
Is this gene present in other annotated genomes?	Yes, in
Does the gene violate any major guiding principles?	No
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 suggest that it starts at 16888, it is supported by genemark
Does the start site have an associated Ribosome Binding Site with a high score?	Final Score: -6.580 Z Score:1.802
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest*

Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in 97.9% of genes
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	*Yes*

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: e-135* *NCBI: 3e-171* *.*
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?	*Above 90%* *HHpred: 99.75% identical to minor tail protein* *100% identical*
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?	*Yes* *PDB_mmCIF70_8_Mar*

	CATH_S40_v4.3
Is the proposed function found on the SEA-PHAGES approved function list?	*yes*
DECISION:	*Minor tail protein*

Basic Phage Information	
Phage Name	MenE
Gene #	30
Stop Coordinate	18459
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**17**
Selected Start Coordinate	17638
Selected Function	Minor tail protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer 17638
Is there evidence for coding potential?	yes

Is this gene present in other annotated genomes?	Most annotated
Does the gene violate any major guiding principles?	No, there were no major guiding principles violated.
DECISION:	Yes, this is a gene

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?	They suggest a start site of 17638
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 1.732 Final Score: -5.492
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF with 822*
Is this start site conserved in other phage genomes as indicated by Starterator?	Most annotated 68.7%
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes, this is the best possible start site*

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: e-156 , minor tail protein* *NCBI: 0.0, minor tail protein*
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?	*Above 90%* *HHpred: 96.97% identical to minor tail protein* *100% identical*
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?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Minor tail protein*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	31
Stop Coordinate	18758
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	**39**
Selected Start Coordinate	18498
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes most annotated
Does the gene violate any major guiding principles?	Gap violation

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 calls 18498
Does the start site have an associated Ribosome Binding Site with a high score?	The Z value is 1.473 The final score – 6.446
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest, 402*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes it is the most annotated
Is this start site conserved in other phage genomes as indicated by BlastP?	*N/A*
DECISION:	*Yes, this is the best possible start site present*

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	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:*

BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	32
Stop Coordinate	19644
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**-7**
Selected Start Coordinate	18751
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	No just glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Is present in 6.5% of other genes
Does the gene violate any major guiding principles?	No
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- 18751
Does the start site have an associated Ribosome Binding Site with a high score?	Final score : -3.893 Z value: 2.540
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No its not the longest* *ORF 894*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes in 9 out of 138 genomes
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes*

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: lysin A* *NCBI: lysin A* *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	*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.*

a probability of 90% or greater with appropriate coverage?	
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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	33
Stop Coordinate	20017
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	**69**
Selected Start Coordinate	19703
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Present in about 6.5% of other genes
Does the gene violate any major guiding principles?	It does overlap the gap rule by being more than 30.
DECISION:	Yes, this is a gene

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 suggest 19703, Genemark does not call
Does the start site have an associated Ribosome Binding Site with a high score?	Finale score: -3.749 Raw SD score: -2.624 Z value: 2.831 Does have the highest score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF score is 315 which is the highest.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes by 6.5%
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes a Gene*

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	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*

database in HHPred with a probability of 90% or greater with appropriate coverage?	*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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	34
Stop Coordinate	20679
Direction (For/Rev)	forward
Gap (Overlap) with Previous Gene	**15**
Selected Start Coordinate	20032
Selected Function	Membrane protein

Annotation Decision #1: Is this a Gene?

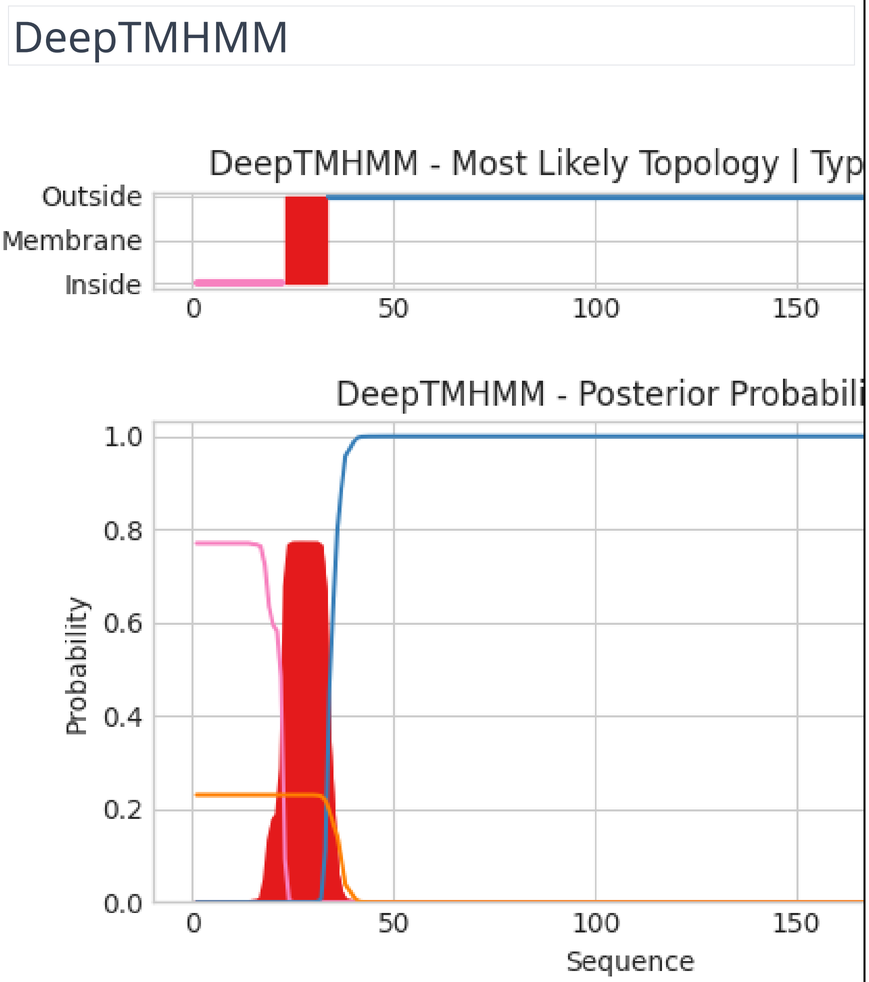
Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Yes both
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	Found 100% of the time
Does the gene violate any major guiding principles?	No
DECISION:	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 (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?	*Z value : 2.167* *Final score : -4.641*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 747 longest*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes 100% of them
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	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:* *DNA Master:*

alignment of 10^{-4} or smaller with appropriate coverage?	*CCD:* Collagen triple helix repeat (20 copies). Members of this family belong to the collagen superfamily. 
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?	
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	.

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?	*Yes*
DECISION:	

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	35
Stop Coordinate	21257
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**-15**
Selected Start Coordinate	20664
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

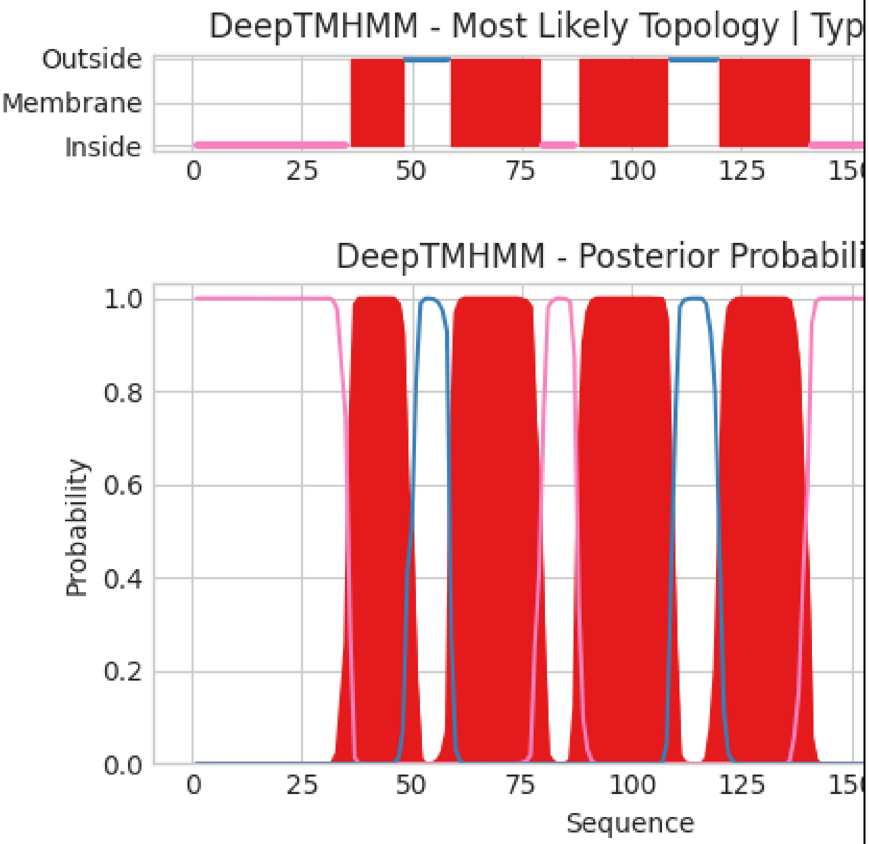
Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Yes, it was called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes

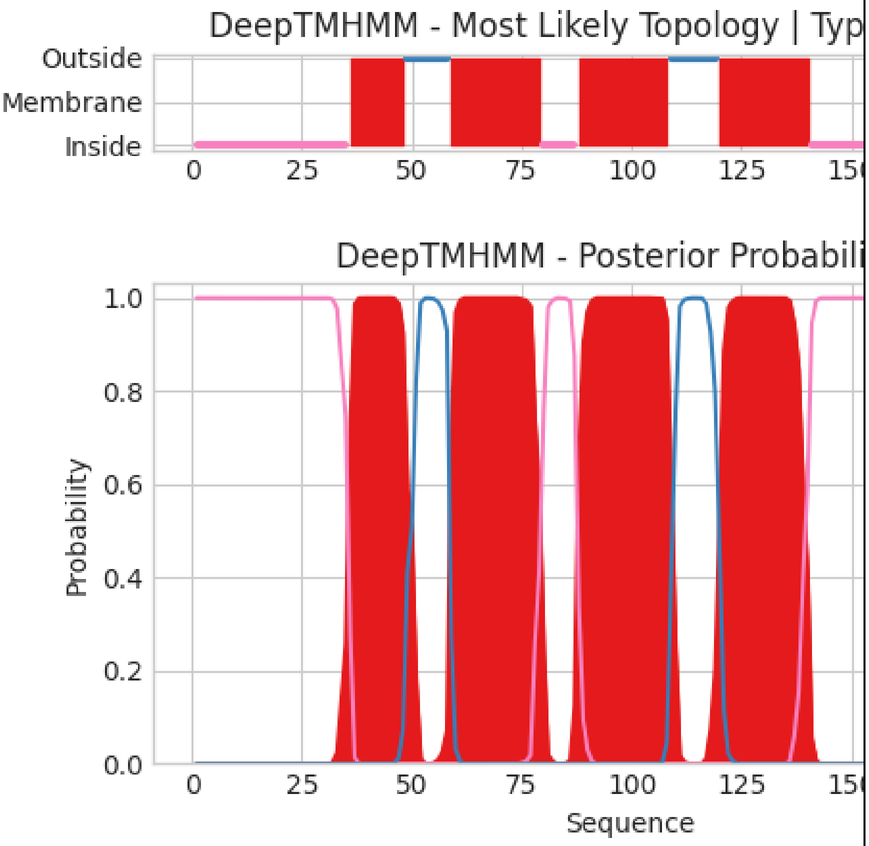
Is this gene present in other annotated genomes?	Yes, found in 64.3% of genes in the pham.
Does the gene violate any major guiding principles?	No there are not any major guiding principles violations.
DECISION:	Yes, it is a gene

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?	Both call the start site 20664
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, its score is: “2.190”, but it is not the highest score of the start site.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the ORF is not the longest, and there’s not an excessive gene overlap. It has an excessive gene overlap of 15.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in 9 out 14 (64.3%) genes in pham.
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes, this is the best possible start site present.*

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: holin* *NCBI: holin* *DNA Master:* *Ccd blank* *Deep:* 
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?	PF16081.9 Phage_holin_7_1 ; Mycobacterial 2 TMS Phage Holi
Is this gene located adjacent to genes of	.

known function and in a region of the genome that shows high conservation of gene order?	
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?	
DECISION:	

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	36
Stop Coordinate	21441
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**125**
Selected Start Coordinate	21310
Selected Function	N/A

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*It was called by Glimmer, but not GeneMark*
Is there evidence for coding potential?	*No, there is not coding potential*
Is this gene present in other annotated genomes?	*Yes in 6 other genomes*
Does the gene violate any major guiding principles?	Yes

DECISION:	

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 suggests 21310, GeneMark doesn't call it.*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score:- 5.284* *Z Score: 1.458*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length is 126 and it is not the longest.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	

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	*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*

smaller with appropriate coverage?	*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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	37
Stop Coordinate	21765
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	**5**
Selected Start Coordinate	21466
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer calls
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, it is
Does the gene violate any major guiding principles?	no
DECISION:	

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 Calls 21441, genemark doesn't call
Does the start site have an associated Ribosome Binding Site with a high score?	Finale score: -4.463 Raw SD score: -2.366 Z value:2.964 Not the highest score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The ORF is 300 it is not the highest codon, does not result in a gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in other genomes by 62.8%
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes, it is a Gene*

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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	38
Stop Coordinate	22106
Direction (For/Rev)	forward
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	21762
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by both Glimmer and GeneMark
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	It is found in 8 out of 8 (100.0%) genes in pham
Does the gene violate any major guiding principles?	no
DECISION:	Yes, it is a Gene

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 21762* *GeneMark StartNA*
Does the start site have an associated Ribosome Binding Site with a high score?	Z value : 0.939 Final score :-8.065
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF : 834 it is the highest*
Is this start site conserved in other phage genomes as indicated by Starterator?	No the start site for the ORF did not agree but there was a better start site called on phamerator
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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	39
Stop Coordinate	22552
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	22106
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	100% of the time
Does the gene violate any major guiding principles?	No violation
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 : 22106
Does the start site have an associated Ribosome Binding Site with a high score?	Final score: -5.289 Z value: 2.128
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF of 882*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is found in 9 out of 9 (100.0%) of genes in pham
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes, it has the best possible start site recommended.*

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	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*

database in HHPred with a probability of 90% or greater with appropriate coverage?	*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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	40
Stop Coordinate	22716
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	22549
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by both Glimmer and GeneMark
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	yes
Does the gene violate any major guiding principles?	no
DECISION:	Yes, this is a gene

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 suggests 22549
Does the start site have an associated Ribosome Binding Site with a high score?	Finale Score: -6.066 Raw SD score: -5.020 Z value: 1.594 Not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF codon is 168 and it is not the highest, and no, it does NOT result in an excessive gene overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in 9 out of 9, 100.0%, phages.
Is this start site conserved in other phage genomes as indicated by BlastP?	
DECISION:	*Yes, this is the best possible start site.*

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?	
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 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	41
Stop Coordinate	23276
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	22713
Selected Function	Hypothetical 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, it is called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, it is found in 11 of 771 (1.4%) of genes in pham
Does the gene violate any major guiding principles?	No
DECISION:	Yes, it is a gene

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?	They suggest the start site “22713”
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has a score of: Genomic Z Value: 1.506 Final Score: -5.854
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it has an ORF Length of 564*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	42
Stop Coordinate	24364
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	219
Selected Start Coordinate	23495
Selected Function	exonuclease

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, found in both Glimmer and GeneMark*
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Found in 9 of 59 (15.3%) of genes in pham; it is called 100.0% of the time when present.
Does the gene violate any major guiding principles?	No each gene is moving forward.
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 : 22106*

Does the start site have an associated Ribosome Binding Site with a high score?	Final score: -6.968 Z value: 1.473
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF: 870 (MenE)* *The Longest ORF present is: 1212*

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: 0.0* *NCBI: 0.0* *HHpred: [e-value: 5.7e-23][Alignment #: 262]*
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?	*Yes, it has a probability of 99.89% according to HHpred.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function is mostly exonuclease, with the exception of >5 Cas4 family exonuclease.*
Is this gene a possible transmembrane protein?	*No.*
Is the proposed function found on the SEA-	*Yes*

PHAGES approved function list?	
DECISION:	*Exonuclease*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	43
Stop Coordinate	25365
Direction (For/Rev)	For

Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	24361
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by both Glimmer and MenE
Is there evidence for coding potential?	Yes as shown by Genemark
Is this gene present in other annotated genomes?	Yes
Does the gene violate any major guiding principles?	No
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?	They suggest the start site: 24361
Does the start site have an associated Ribosome Binding Site with a high score?	Finale Score: -2.812 Z value: 3.274 Raw SD score: -1.767 Not the highest score
Is the predicted start codon the longest ORF? If not, does the longest	*ORF length is 1005 which is the longest.*

ORF result in excessive gene overlap (>30bp)?	

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	44
Stop Coordinate	25727
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	25365
Selected Function	Hypothetical 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, the Gene was called by both Glimmer and GeneMark*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is annotating in 11 out of 313 (3.5%) of genes in Pham; it is called 90.9% of the time when present.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes, this is a gene*

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: 25365* *GeneMark Start Coordinate: 25365*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score: -6.377* *Z Score: 1.703*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	45
Stop Coordinate	26097
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-7
Selected Start Coordinate	25720
Selected Function	Hypothetical 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, it was called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, present in 9 out of 148 of genes in the pham; it is called 100.0% of the time when present.
Does the gene violate any major guiding principles?	Its spacer distance is above 12: This gene has a spacer distance of 14.
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?	They suggest the start site: 25720
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it's score is: Genomic Z Value: 2.054 Final Score: -5.435
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it has an ORF Length of 378*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	46
Stop Coordinate	26959
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	27
Selected Start Coordinate	26123
Selected Function	Hypothetical 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, it is called by both Glimmer and GenEMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, it is found in 30 of 25 (85.7%) of genes in pham.
Does the gene violate any major guiding principles?	No
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 26123
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.797 Final score: -3.469
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF of 837*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	47
Stop Coordinate	27288
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	26956
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation	Yes, both

program (Glimmer, GeneMark)?	
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, it is found in 29 of 29 (100.0%) of genes in pham.
Does the gene violate any major guiding principles?	No
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?	Both suggest 26956
Does the start site have an associated Ribosome Binding Site with a high score?	ORF: 333 Finale Score: -3.366 Z Value: 2.964 Raw SD score: -2.366
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF length, 333*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	48

Stop Coordinate	28589
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	27285
Selected Function	DnaB-like dsDNA helicase

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?	Yes
Is this gene present in other annotated genomes?	Yes, it is found in 17 of 258 (6.6%) of genes in pham
Does the gene violate any major guiding principles?	No
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?	*Both call 27285*
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 1.506 Final Score: -5.918

Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not longest ORF 1305* *Longest is 1551, but results in a –249 gap*

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?	*PhagesDB: 0.0, DnaB-like dsDNA helicase* *NCBI: 0.0, DnaB-like dsDNA helicase*
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: 100% probability with a gene for a DnaB-like dsDNA helicase*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes DnaB-like dsDNA helicase*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*DnaB-like dsDNA helicase*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	49
Stop Coordinate	28800
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	2
Selected Start Coordinate	28591
Selected Function	Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	It is only present in one other genome
Does the gene violate any major guiding principles?	*No*
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: 28591* *GeneMark Start Coordinate: 28591*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score: -7.487* *Z Value: 1.235*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*It's not the longest ORF, but the longest one results in a gap of 19 bp*

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?	*PhagesDB: 0.0, DnaB-like dsDNA helicase* *NCBI: 0.0, DnaB-like dsDNA helicase*
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: 100% probability with a gene for a DnaB-like dsDNA helicase*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes DnaB-like dsDNA helicase*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*DnaB-like dsDNA helicase*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE

Gene #	50
Stop Coordinate	29294
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	28803
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by both Glimmer and GeneMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, found in 209 of 356 (58.7%) of genes in pham.
Does the gene violate any major guiding principles?	It has a spacer distance of 15.
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?	GeneMark and Glimmer suggest the start site: 28803
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, It has the highest Z Value. Genomic Z value: 2.787 Final Score: -4.256

Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it has the largest ORF Length:* *Its ORF Length is: 492*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	51
Stop Coordinate	29925
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	98
Selected Start Coordinate	29392
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes about 9.8%
Does the gene violate any major guiding principles?	The gap is large
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 suggests 29392
Does the start site have an associated Ribosome Binding Site with a high score?	ORF: 534 Finale Score: -5.913 Z Value: 1.673 Raw SD scores: -4.868 Not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF is 534 it is not the longest, does not result in excessive gene overlap.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Phage Name	MenE
Gene #	52
Stop Coordinate	30544
Direction (For/Rev)	for
Gap (Overlap) with Previous Gene	-7
Selected Start Coordinate	29918
Selected Function	Hypothetical protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by both Glimmer and GenEMark
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	96.2 % present
Does the gene violate any major guiding principles?	No
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 29918*
Does the start site have an associated Ribosome Binding Site with a high score?	Final : -4.323 Z value: 2.468
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF 627*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

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

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]() 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.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	53
Stop Coordinate	31028
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	158
Selected Start Coordinate	30702
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes in 100%
Does the gene violate any major guiding principles?	Yes the gap
DECISION:	

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 suggests 30702
Does the start site have an associated Ribosome Binding Site with a high score?	ORF: 327 Finale score: -2.661 Z value: 3.352 Raw SD score: - 1.615 Not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length is 327* *Not the longest there is an excessive bp*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	
Gene #	54
Stop Coordinate	32130
Direction (For/Rev)	for
Gap (Overlap) with Previous Gene	50
Selected Start Coordinate	31078
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	*100% found*
Does the gene violate any major guiding principles?	no
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 31078
Does the start site have an associated Ribosome Binding Site with a high score?	Final Score: -7.323 Z value: 1.089
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Longest ORF 1053*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	55
Stop Coordinate	32653
Direction (For/Rev)	for
Gap (Overlap) with Previous Gene	146
Selected Start Coordinate	32276
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	1.9% present
Does the gene violate any major guiding principles?	No
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 32276
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 3.013 Final: -3.432
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 378 longest is 564*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	56
Stop Coordinate	32940
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	32665
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	10.1% found
Does the gene violate any major guiding principles?	No
DECISION:	

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: 32937
Does the start site have an associated Ribosome Binding Site with a high score?	Final Score: -6.744 Z value: 1.473
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 144* *Long is 900*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	57
Stop Coordinate	33080
Direction (For/Rev)	for
Gap (Overlap) with Previous Gene	272
Selected Start Coordinate	32937
Selected Function	winged helix-turn-helix, (turn helix Turn)

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	10.9%
Does the gene violate any major guiding principles?	Yes, it has a gap over 30< ; its gap is 272
DECISION:	

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 32937
Does the start site have an associated Ribosome Binding Site with a high score?	Final value : - 6.744 Z value: 1.473
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF is 144 not the longest*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	58
Stop Coordinate	33427
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	33077
Selected Function	Hypothetical 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?	Yes
Is this gene present in other annotated genomes?	Yes, it is present in 7 other genomes
Does the gene violate any major guiding principles?	*No*
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: 33077* *GeneMark Start Coordinate: 33077*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score: -6.408* *Z Score: 1.760*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	59
Stop Coordinate	33750
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	33424
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	yes
Does the gene violate any major guiding principles?	no
DECISION:	

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 suggests 33400 but with the evidence shown 33424 is a better choice for start codon
Does the start site have an associated Ribosome Binding Site with a high score?	ORF: 327 Finale Score: -3.606 z-value: 1.052 Raw SD score: -6.070 It is not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The ORF length is 327 it is not the highest and the overlap is not over 30*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	60
Stop Coordinate	34208
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	63
Selected Start Coordinate	33813
Selected Function	Hypothetical 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?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is present in 9 other genomes*
Does the gene violate any major guiding principles?	*63 Gap*
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: 33813* *GeneMark Start Coordinate: 33813*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score: -4.133* *Z Score: 2.417*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest, but it is called in 100% of similar genomes*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	61
Stop Coordinate	34642
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	34205
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	100% present
Does the gene violate any major guiding principles?	no
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 34205
Does the start site have an associated Ribosome Binding Site with a high score?	Final –5.164 Z value 2.422
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 438* *Longest ORF is 438*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	62
Stop Coordinate	35135
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-25
Selected Start Coordinate	34617
Selected Function	Hypothetical 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, called by both GeneMark and Glimmer
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	Yes, this gene is present in 9 out of 50 (18.0%) of genes in pham.
Does the gene violate any major guiding principles?	No
DECISION:	

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?	They both suggest the start site: 34617
Does the start site have an associated Ribosome Binding Site with a high score?	Yes, it has: Genomic Z Value: 3.418 Final Score: -2.112
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, and it has an ORF Length of 519*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	63
Stop Coordinate	35608
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	35132
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	100% present in other genes
Does the gene violate any major guiding principles?	no
DECISION:	

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 suggests 35132
Does the start site have an associated Ribosome Binding Site with a high score?	ORF:477 Finale score: -5.589 Raw SD score: -4.543 z- value: 1.840 Not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF: 477* *Highest score*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	64
Stop Coordinate	36455
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	17
Selected Start Coordinate	35625
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Called by Glimmer
Is there evidence for coding potential?	Yes, supported by Genemark
Is this gene present in other annotated genomes?	Yes 100%
Does the gene violate any major guiding principles?	No
DECISION:	

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 suggests 35625
Does the start site have an associated Ribosome Binding Site with a high score?	ORF:831 Finale score: -2.661 Raw SD score: -1.615 z- value: 3.352 Not the highest score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length is 831, it is the highest.*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	
Gene #	65
Stop Coordinate	37174
Direction (For/Rev)	forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	36452
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	100%present
Does the gene violate any major guiding principles?	no
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 36452
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.155 Final score –5.014
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 723 not the longest* *Longest ORF is 1011*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	66
Stop Coordinate	37503
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-15
Selected Start Coordinate	37159
Selected Function	Hypothetical 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, but they both called different start sites. Glimmer called a start site at: 37159 GeneMark called a start site at: 37171
Is there evidence for coding potential?	Yes
Is this gene present in other annotated genomes?	No, it is only found in MenE.
Does the gene violate any major guiding principles?	No
DECISION:	It is a gene!

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 called a start site at: 37159 GeneMark called a start site at: 37171
Does the start site have an associated Ribosome Binding Site with a high score?	When comparing the RBS score to the other suggested start sites, it is in the middle. Genomic Z Value: 1.437 Final Score: -6.070
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it has an ORF Length of 345*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	
Gene #	67
Stop Coordinate	37973
Direction (For/Rev)	forward
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	37503
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer only
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	Not most annotated in other genes
Does the gene violate any major guiding principles?	no
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 37503
Does the start site have an associated Ribosome Binding Site with a high score?	Final score – 4.699 Z value 2.128 Raw : -5.296 Not the highest
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 471* *Longest ORF 528*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	68
Stop Coordinate	38275
Direction (For/Rev)	forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	37970
Selected Function	Hypothetical Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Glimmer
Is there evidence for coding potential?	yes
Is this gene present in other annotated genomes?	100% present
Does the gene violate any major guiding principles?	yes
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 37970
Does the start site have an associated Ribosome Binding Site with a high score?	Final score: -4.905 Z value: 2.032 Raw : -4.131 ORF:306 Highest score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF 306 and the longest*

Student Gene Annotation Worksheet

SAVE AS GUIDLINE!

Basic Phage Information	
Phage Name	MenE
Gene #	69
Stop Coordinate	38427
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	38272
Selected Function	Hypothetical 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?	Yes
Is this gene present in other annotated genomes?	Yes, in 7 other genomes
Does the gene violate any major guiding principles?	No
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?	Both suggest 38272
Does the start site have an associated Ribosome Binding Site with a high score?	RBS Score: -4.548 Z Score: 2.173
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, but the longest results in a –249 gap*