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	

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 <u>related</u> , <u>annotated</u> 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 <u>Guiding</u> <u>Principles of Genome Annotation</u> 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.

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer Start Coordinate (type NA if not supported)::
Glimmer and GeneMark suggest?	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	Provide the best BlastP match from NCBI, PhagesDB, and
conserved in other	DNA Master with alignment in the format of (Q#:S#), where
phage genomes as	Q (query) is the sequence you are analyzing and S
indicated by BlastP?	(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
DECISION:	briefly explain your rationale.

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
	<u>SEA-PHAGES Official Function List</u> 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.

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

Gathering Evidence	Explain Your Rationale
	Called by glimmer, not genemark
an auto-annotation	

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

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:	

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?	
La Alain mana a manailala	
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 #	12
Stop Coordinate	5763
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	114
Selected Start Coordinate	4648
Selected Function	Hypothetical Protein

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

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

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

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:	

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:	
DECISION.	

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 #	14
Stop Coordinate	6277
Direction (For/Rev)	6277
Gap (Overlap) with Previous Gene	6
Selected Start Coordinate	6086
Selected Function	Hypothetical Protein

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

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	Yes
conserved in other	
phage genomes as	
indicated by Starterator?	
Is this start site	no
conserved in other	
phage genomes as	
indicated by BlastP?	
DECISION:	Yes

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 #	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:	

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark	Glimmer: 6376
suggest?	GeneMark: 6403
Does the start site have	RBS score: -5.308
an associated Ribosome	
Binding Site with a high	
score?	

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

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 #	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:	

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

Gathering Evidence	Explain Your Rationale
Does this protein align	No
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?	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

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:	

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:	

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

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation	Climana y Canana de aglia da atanta et 0000
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:	

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:	

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

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	

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:	

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	yes
conserved in other	
phage genomes as indicated by BlastP?	
DECISION:	Yes

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

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

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:	

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

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:	

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 <u>related</u> , <u>annotated</u> phages contain this gene. In your answer, record the name of the phage, gene #, and evalue 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 <u>Guiding</u> <u>Principles of Genome Annotation</u> 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.	

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.

Gathering Evidence	Explain Your Rationale
Does this protein align with	List the most informative BlastP match from each source
a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB and/or	DNA Master:
GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and e-
coverage?	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	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
probability of 90% or greater with appropriate coverage?	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 <u>SEA-PHAGES Official Function List</u> 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.

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

Gathering Evidence	Explain Your Rationale	
Was the gene called by an		
auto-annotation program	YES BOTH	
(Glimmer, GeneMark)?		
Is there evidence for	Yes	
coding potential?	163	
Is this gene present in	Yes, in 100% of times the gene present it's called	
other annotated genomes?		
Does the gene violate any	No	
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:	

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

Gathering Evidence	Explain Your Rationale	
Was the gene called by an	Both Glimmer & GeneMark agreed on call @bp 519	
auto-annotation program		
(Glimmer, GeneMark)?		
Is there evidence for	Yes, there is evidence for coding potential.	
coding potential?		
Is this gene present in	Yes, 100% of the time start site 519 is conserved.	
other annotated genomes?		
Does the gene violate any	Yes, the Genomic Z value is less than different possible start	
major guiding principles?	codon.	
DECISION:	Yes	

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	There was a finale RBS score of: -6.770
associated Ribosome	The Genomic Z value was: 1.575
Binding Site with a high	1.575 was not the highest/best recorded score.
score?	
Is the predicted start codon	No, the predicted start codon does not have the longest ORF.
the longest ORF? If not,	The longest ORF excessive gene overlaps >30bp.
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	Yes, the start site is conserved in other phage genomes as
in other phage genomes	indicated by Starterator.

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

Gathering Evidence	Explain Your Rationale
Does this protein align with	No, none of the ones it aligns with has a functional assignment
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 #	3
Stop Coordinate	880
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	653
Selected Function	Hypothetical Protein

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

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:	

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

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	Yes, both
(Glimmer, GeneMark)?	
Is there evidence for	Yes
coding potential?	162
Is this gene present in	Yes, it is called 100% of the time when present
other annotated genomes?	
Does the gene violate any	No
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:	

Gathering Evidence	Explain Your Rationale
--------------------	------------------------

D 41-1	
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

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	Yes, both
(Glimmer, GeneMark)?	
Is there evidence for	Yes
coding potential?	163
Is this gene present in	Yes
other annotated genomes?	
Does the gene violate any	No
major guiding principles?	IVO
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:	

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

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

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

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	1174 bur genemark does not call that. When revised 1174 was
and GeneMark suggest?	the best probable option
Does the start site have an	N/A
associated Ribosome	

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

Gathering Evidence	Explain Your Rationale
Does this protein align with	No
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	No
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	No
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	No
transmembrane protein?	
Is the proposed function	No
found on the SEA-	
PHAGES approved	
function list?	No Francisco Nata Ocean
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

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	It was called, but they disagreed
(Glimmer, GeneMark)?	
Is there evidence for	Vos thora is cadina notantial
coding potential?	Yes there is coding potential
Is this gene present in	Yes, the most annotated start is 1527
other annotated genomes?	
Does the gene violate any	Yes
major guiding principles?	163
DECISION:	Yes

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

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

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	Called By Glimmer, not GeneMark
(Glimmer, GeneMark)?	
Is there evidence for	No
coding potential?	NO
Is this gene present in	Yes
other annotated genomes?	163
Does the gene violate any	Yes, it overlaps with gene #10, has little to no coding potential,
major guiding principles?	and the bp is below 200.
DECISION:	No. It is not a 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:	

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

Gathering Evidence	Explain Your Rationale	
Was the gene called by an		
auto-annotation program	Called by Glimmer but Genemark called 2984	
(Glimmer, GeneMark)?		
Is there evidence for	Yes there is a bunch	
coding potential?		
Is this gene present in	Yes	
other annotated genomes?		
Does the gene violate any	Yes the gap is more than 30bp	
major guiding principles?		
DECISION:		

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

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

Gathering Evidence	Explain Your Rationale
Does this protein align with	Phages db: 0.0, portal protein
a protein having a	NCBI: 0.0, portal protein
functional assignment in	
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	HHpred: 100 probability of a portal protein
a protein having a	Above 90%
functional assignment in	100% identical to portal protein
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	no
transmembrane protein?	
Is the proposed function	yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Portal protein

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,	Yes, both
GeneMark)?	
Is there evidence for coding	Yes
potential?	165
Is this gene present in other	Found in 7.6% of genes
annotated genomes?	Found in 7.0 % of genes
Does the gene violate any major	No
guiding principles?	NO
DECISION:	Yes

What start site do	
Glimmer and GeneMark	Glimmer and GeneMark at 9789
suggest?	
Does the start site have	Z value: 2.808
an associated Ribosome	Final score: -3.447
Binding Site with a high	
score?	
Is the predicted start	Yes, it's the longest at 444 ORF
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	
phage genomes as	Yes, in about 7.6% of the other genes
indicated by Starterator?	
Is this start site	N/A
conserved in other	
phage genomes as	
indicated by BlastP?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
Does this protein align with a	PhagesDB: 4e-82, tail terminator
protein having a functional	NCBI: 2e – 101, tail terminator
assignment in BlastP	HHpred: no e values below 10 ⁻⁴
(phagesDB and/or GenBank)	
with an alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with a	HHpred: 100% probability with a gene for a tail
protein having a functional	terminator protein
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	Yes tail terminator
genes of known function and in	
a region of the genome that	
shows high conservation of	
gene order?	
Is this gene a possible	no
transmembrane protein?	
Is the proposed function found	yes
on the SEA-PHAGES approved	
function list?	
DECISION:	Tail terminator

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
-------------------	--------------------

Gathering Evidence	Explain Your Rationale	
Was the gene called by		
an auto-annotation	Yes it was called by both Glimmer and GeneMark	
program (Glimmer,	Tes it was called by both diffiller and deficitiant	
GeneMark)?		
Is there evidence for	Yes	
coding potential?	165	
Is this gene present in		
other annotated	Yes, it is called 96.6% of the time when present	
genomes?		
Does the gene violate		
any major guiding	No	
principles?		
DECISION:	Yes	

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Both Glimmer and GeneMark suggest a start site of 10248
suggest?	
Does the start site have	
an associated Ribosome	Yes, it has the highest Z value: 3.146
Binding Site with a high	
score?	
Is the predicted start	
codon the longest ORF?	Yes, the predicted start codon has the longest ORF with a
If not, does the longest	length of 513 bp.

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

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 5e – 97, Major tail protein
with a protein having a	NCBI: 7e-118 , Major tail protein
functional assignment in	HHped:
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	100% align with this function
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	No synteny observed
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	no
transmembrane protein?	
Is the proposed function	yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	
	Major tail protein

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	Called by Glimmer
an auto-annotation	

program (Glimmer,	
GeneMark)?	
Is there evidence for	Yes
coding potential?	165
Is this gene present in	
other annotated	Yes
genomes?	
Does the gene violate	
any major guiding	Gap violation
principles?	
DECISION:	Yes, it is a gene

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer calls 10835
suggest?	
Does the start site have	The Z value is 1.467
an associated Ribosome	The final score – 6.312
Binding Site with a high	
score?	
Is the predicted start	Vos. it is the langest 180
Is the predicted start	Yes, it is the longest, 489
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	Yes it is by 28.6%
conserved in other	
phage genomes as	
indicated by Starterator?	
Is this start site	N/A
conserved in other	

phage genomes as indicated by BlastP?	
DECISION:	A gene

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	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.

	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.

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	Yes Both
program (Glimmer,	res dotti
GeneMark)?	
Is there evidence for	VOC
coding potential?	yes
Is this gene present in	
other annotated	100% found in other genes
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer calls 11350 and GeneMark is 11320
suggest?	
Does the start site have	
an associated Ribosome	Z value: 2.190
Binding Site with a high	Final score: -4.866
score?	
Is the predicted start	No its not the longest ORF
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	
phage genomes as	Yes 100%
indicated by Starterator?	

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	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.
	νειιεύ τοι αρριοναι.
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.

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	Yes
program (Glimmer,	1 53
GeneMark)?	
Is there evidence for	Yes
coding potential?	res
Is this gene present in	
other annotated	Found in 13 of 34 (38.2%) of genes in pham
genomes?	Found in 13 of 34 (36.2 %) of genes in priam
Does the gene violate	
any major guiding	ORFs length over 120bp
principles?	
DECISION:	Yes, it is a gene

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

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

Gathering Evidence	Explain Your Rationale
Does this protein align	NCBI: 0.0 , tape measure protein
with a protein having a	Phagesdb: 0.0 tape measure protein
functional assignment in	
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	Above 90%
with a protein having a	HHpred: 99.86% identical gene that it is a tape measure
functional assignment in	protein
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	Yes
transmembrane protein?	PDB_mmCIF70_8_Mar
	CATH_S40_v4.3
Is the proposed function	yes
found on the SEA-	

PHAGES approved function list?	
DECISION:	Tape measure protein

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

Gathering Evidence	Explain Your Rationale
Was the gene called by	
an auto-annotation	Both Called
program (Glimmer,	Dotti Called
GeneMark)?	
Is there evidence for	Yes
coding potential?	165
Is this gene present in	
other annotated	Yes 5.9%
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

Gathering	Explain Your Rationale
Evidence	
What start site do	
Glimmer and	Glimmer called 11350 and genemark called 11320
GeneMark suggest?	
Does the start site	Z value: 2.737
have an associated	
Ribosome Binding	Final Score: -3.586
Site with a high	
score?	
Is the predicted start	Yes its the longest ORF with 861
codon the longest	
ORF? If not, does	
the longest ORF	
result in excessive	
gene overlap	
(>30bp)?	
Is this start site	20 called the same start out of 339
conserved in other	
phage genomes as	
indicated by	
Starterator?	
Is this start site	
conserved in other	
phage genomes as	
indicated by BlastP?	
DECISION:	

Gathering Evidence	Explain Your Rationale
--------------------	------------------------

Does this protein align with a	List the most informative BlastP match from each
protein having a functional	source PhagesDB:
assignment in BlastP (phagesDB	NCBI: 0.0, minor tail protein
and/or GenBank) with an	DNA Master: e-163 minor tail protein
alignment of 10 ⁻⁴ or smaller with	
appropriate coverage?	
Does this protein align with a	100% identical to minor tail protein
protein having a functional	HHpred: 99.96%identical to the minor tail protein
assignment in the PDB or other	Above 90%
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	no
transmembrane protein?	
Is the proposed function found on	yes
the SEA-PHAGES approved	
function list?	
DECISION:	Minor tail protein

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	Yes, it was called by both Glimmer and GeneMark
program (Glimmer,	res, it was called by both Gilliller and Genewark
GeneMark)?	
Is there evidence for	Yes
coding potential?	
Is this gene present in	
other annotated	Yes, the start site is called 100% of the time.
genomes?	
Does the gene violate	
any major guiding	No
principles?	
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	They both suggest the start site: 14,693	
suggest?		

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI: 0.0, minor tail protein
BlastP (phagesDB and/or	DNA Master: 0.0, minor tail protein
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	100% identical to minor tail protein
with a protein having a	HHpred: 99.85% identical to a minor tail protein
functional assignment in	Above 90%
the PDB or other	
database in HHPred with	
a probability of 90% or	

greater with appropriate	
coverage?	
Is this gene located	If the answer is YES, evaluate the proposed function in
adjacent to genes of	the gene order. Examine the adjacent genes found in the
known function and in a	most closely related annotated phage (hint: use
region of the genome that	Phamerator) and record the function of the genes found
shows high conservation	on each side of the gene in the same pham in the most
of gene order?	closely related phage. If the answer is NO, enter No
	Synteny Observed.
Is this gene a possible	no
transmembrane protein?	
Is the proposed function	yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Minor tail protein

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

Gathering Evidence	Explain Your Rationale
Was the gene called by	
an auto-annotation	Yes glimmer only
program (Glimmer,	res giiriirier oriiy
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	
other annotated	100% of the time its present
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer Start Coordinate : 16568
Glimmer and GeneMark	GeneMark Start Coordinate NA
suggest?	
Does the start site have	Final Score: -3.190
an associated Ribosome	Z Score: 3.047
Binding Site with a high	
score?	
Is the predicted start	Longest ORF with 309
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.

Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.
	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.

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	Yes, both
program (Glimmer,	res, bour
GeneMark)?	
Is there evidence for	Yes, as shown in GeneMark.
coding potential?	res, as shown in Generalk.
Is this gene present in	
other annotated	Yes, in
genomes?	
Does the gene violate	
any major guiding	No
principles?	
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 suggest that it starts at 16888, it is supported by
Glimmer and GeneMark	genemark
suggest?	genemark
Does the start site have	Final Score: -6.580
an associated Ribosome	Z Score:1.802
Binding Site with a high	
score?	
Is the predicted start	Yes, it is the longest
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB: e-135
functional assignment in	NCBI: 3e-171
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	Above 90%
with a protein having a	HHpred: 99.75% identical to minor tail protein
functional assignment in	100% identical
the PDB or other	
database in HHPred with	
a probability of 90% or	
greater with appropriate	
coverage?	
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	Yes
transmembrane protein?	PDB_mmCIF70_8_Mar

	CATH_S40_v4.3
Is the proposed function	yes
found on the SEA-	
PHAGES approved	
function list?	
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

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

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

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB: e-156 , minor tail protein
functional assignment in	NCBI: 0.0, minor tail protein
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	Above 90%
with a protein having a	HHpred: 96.97% identical to minor tail protein
functional assignment in	100% identical
the PDB or other	
database in HHPred with	
a probability of 90% or	
greater with appropriate	
coverage?	
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	no
transmembrane protein?	
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Minor tail protein

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

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

DECISION:	Yes

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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:

BlastP (phagesDB	DNA Master:
,, ,	DIVA Master.
and/or GenBank) with an	List: you may have already found this information from
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.
	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.

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	No just glimmer	
program (Glimmer,	TVO JUST GIITITIOI	
GeneMark)?		
Is there evidence for	Yes	
coding potential?	103	
Is this gene present in		
other annotated	Is present in 6.5% of other genes	
genomes?		
Does the gene violate		
any major guiding	No	
principles?		
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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB: lysin A
functional assignment in	NCBI: lysin A
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	
with a protein having a	Note: If you believe there is not a quality HHPred match,
functional assignment in	type No Quality Match and list the data for the best match
the PDB or other	available to affirm the poor quality of the result and to
database in HHPred with	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? Is the proposed function found on the SEA-PHAGES approved function list?	If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs. 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.

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	Called by Glimmer	
program (Glimmer,	Called by Cliffine	
GeneMark)?		
Is there evidence for	Yes	
coding potential?	163	
Is this gene present in		
other annotated	Present in about 6.5% of other genes	
genomes?		
Does the gene violate		
any major guiding	It does overlap the gap rule by being more than 30.	
principles?		
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
	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

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	

database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.
	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.

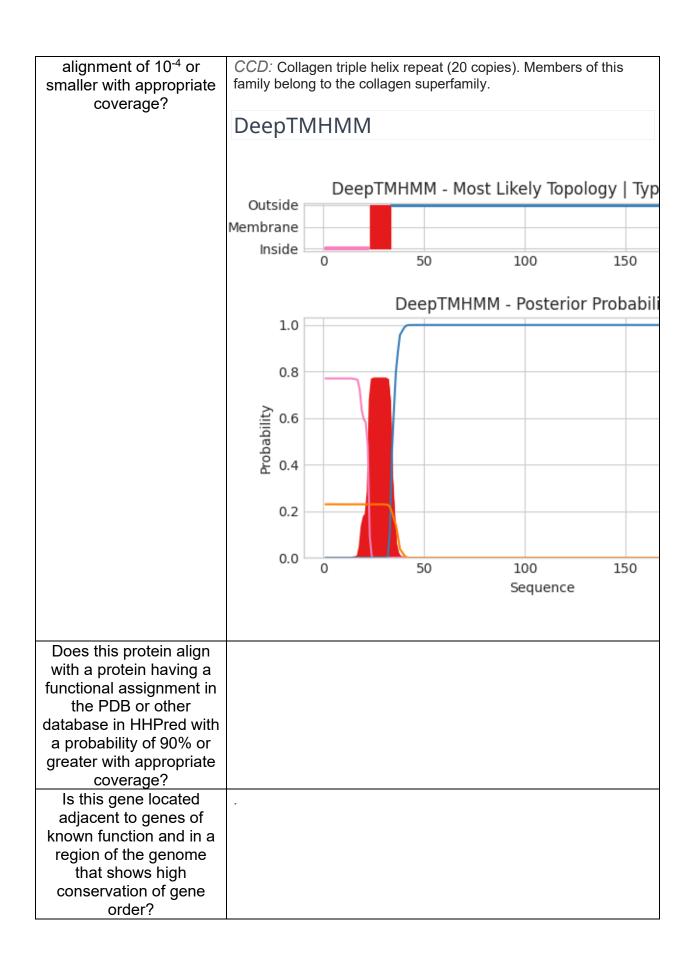
SAVE AS GUIDLINE!

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	

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark	Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::
suggest?	Genewark Start Coordinate (type NA II 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.
DECISION:	Note: if you are considering more than 1 start site, provide the same information for each proposed start site. Record where you think the gene should start here and
DECISION:	briefly explain your rationale.

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	



Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	

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				

Gathering Evidence	Explain Your Rationale			
Was the gene called by				
an auto-annotation	Voc. it was called by both Climmer and CanaMark			
program (Glimmer,	Yes, it was called by both Glimmer and GeneMark			
GeneMark)?				
Is there evidence for	Yes			
coding potential?	165			

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

Gathering Evidence			- - - - -	n Vaur	Pationa	ılo.			
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:								
		Dee	ерТМІ	HMM - N	Most Lik	cely Top	oology	Тур	
	Outside Membrane Inside								
		0	25	50	75	100	125	150	
	1.0			eepilv	IHMM -	Poster	or Proc	abili	
								$\ \ $	
	0.8								
	Probability 6.0 9.0								
	0.2								
	0.0	0	25	50	75 S	100 Sequenc	125 e	150	
Does this protein align with a protein having a	PF16081.9	9	Ph	age_ho	lin_7_1 ;	Mycob	acterial	2 TM	S Phage Holi
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	If the answ	ver is	YES, in	ndicate s	upportin	g data fr	om at le	ast
transmembrane protein?	2 different							
Is the proposed function								
found on the SEA-								
PHAGES approved								
function list?								
		D	еерТМ	HMM - I	Most Li	kely Top	ology	Тур
	Outside							
	Membrane							
	Inside							
		0	25	50	75	100	125	150
				DeepTM	инмм -	Posteri	or Prob	abili
	1.0							
	0.0							
DECISION:	0.8							
DECISION.								
	<u>≩</u> 0.6							
	Probability 6.0 9.0							
	go							
	₹ 0.4							
	0.2							
1								
	0.0	0	25	50	75	100	125	150
	0.0	0	25	50	75	100 Seguence	125	150
	0.0	0	25	50		100 Sequence		150

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				

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

DECISION:	

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?	

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
	annotation decision #2. Provide the alignment (q#:s#) and

11 20 20	
smaller with appropriate	e-value. It is only necessary to provide one match from
coverage?	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.
	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.
	1

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	

Gathering Evidence	Explain Your Rationale
Was the gene called by	
an auto-annotation	Glimmer calls
program (Glimmer,	Silitifici salis
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	
other annotated	Yes, it is
genomes?	
Does the gene violate	
any major guiding	no
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?	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	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.

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.
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.
If the answer is YES, indicate supporting data from at least
2 different transmembrane prediction programs.
Indicate a response with a Yes or No response.
Once you have arrived at a functional decision, check the
<u>SEA-PHAGES Official Function List</u> 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.
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.

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	Called by both Glimmer and GeneMark
program (Glimmer,	Caned by Both Chilling and Geneman
GeneMark)?	
Is there evidence for	yes
coding potential?	yes
Is this gene present in	
other annotated	It is found in 8 out of 8 (100.0%) genes in pham
genomes?	
Does the gene violate	
any major guiding	no
principles?	
DECISION:	Yes, it is a Gene

Explain Your Rationale
Glimmer Start Coordinate 21762
GeneMark StartNA
Z value : 0.939
Final score :-8.065
ORF: 834 it is the highest
No the start site for the ODE did not agree but there was a
No the start site for the ORF did not agree but there was a
better start site called on phamerator
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.
anguinent of the best materi for all tillee blastr sources.
Note: if you are considering more than 1 start site, provide
the same information for each proposed start site.
Record where you think the gene should start here and
briefly explain your rationale.

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	

e-value. It is only necessary to provide one match from
each database.
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.
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.
If the answer is YES, indicate supporting data from at least
2 different transmembrane prediction programs.
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.
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.

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	Glimmer only
program (Glimmer,	Cilifinities of thy
GeneMark)?	
Is there evidence for	yes
coding potential?	ycs
Is this gene present in	
other annotated	100% of the time
genomes?	
Does the gene violate	
any major guiding	No violation
principles?	
DECISION:	yes

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.

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	Hint: you may have already found this information from
smaller with appropriate	annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from
	each database.
Does this protein align	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	

database in HHPred with a probability of 90% or	Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	document that HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the
adjacent to genes of	gene order. Examine the adjacent genes found in the most
known function and in a	closely related annotated phage (hint: use Phamerator)
region of the genome	and record the function of the genes found on each side of
that shows high	the gene in the same pham in the most closely related
conservation of gene	phage. If the answer is NO, enter No Synteny Observed.
order?	
Is this gene a possible	If the answer is YES, indicate supporting data from at least
transmembrane protein?	2 different transmembrane prediction programs.
Is the proposed function	Indicate a response with a Yes or No response.
found on the SEA-	Once you have arrived at a functional decision, check the
PHAGES approved	SEA-PHAGES Official Function List to ensure that you are
function list?	following the guidelines for function naming. Functions that
	are not present on the approved list must be carefully
	vetted for approval.
	Total of approval.
	If you believe this gene should be assigned, please write
	the name of the function here. If the evidence does not
DECISION:	support a functional call, record "NKF" for no known
	function. 50-70% of phage genes fall into the NKF
	category.
	outogory.

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	Called by both Glimmer and GeneMark	
program (Glimmer,	Called by both Gillillile and Genewark	
GeneMark)?		
Is there evidence for	VAS	
coding potential?	yes	
Is this gene present in		
other annotated	yes	
genomes?		
Does the gene violate		
any major guiding	no	
principles?		
DECISION:	Yes, this is a 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.

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	List the most informative HHPred match, including
with a protein having a	database source and probability score. It is only necessary
functional assignment in	to provide the best match.
the PDB or other	
database in HHPred with	Note: If you believe there is not a quality HHPred match,
a probability of 90% or	type No Quality Match and list the data for the best match
greater with appropriate	available to affirm the poor quality of the result and to
coverage?	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.

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	Yes, it is called by both Glimmer and GeneMark
program (Glimmer,	1 cs, it is called by both diffilled and deficilitation
GeneMark)?	
Is there evidence for	Yes
coding potential?	100
Is this gene present in	
other annotated	Yes, it is found in 11 of 771 (1.4%) of genes in pham
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes, it is a gene

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	They suggest the start site "22713"
suggest?	
Does the start site have	Yes, it has a score of:
an associated Ribosome	Genomic Z Value: 1.506
Binding Site with a high	Final Score: -5.854
score?	
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

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	Yes, found in both Glimmer and GeneMark
program (Glimmer,	
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	Found in 9 of 59 (15.3%) of genes in pham; it is called
other annotated	100.0% of the time when present.
genomes?	100.070 of the time when present.
Does the gene violate	
any major guiding	No each gene is moving forward.
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale	
What start site do	Glimmer Start Coordinate : 22106	
Glimmer and GeneMark	Gillinia Start Goordinate . 22100	
suggest?		

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	Not the longest ORF: 870 (MenE) The Longest ORF present is: 1212
ORF result in excessive gene overlap (>30bp)?	

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? Is the proposed function	No. Yes
found on the SEA-	100

PHAGES approved function list?	
DECISION:	Exonuclease

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	Called by both Glimmer and MenE
program (Glimmer,	Canca by Both Chilinia and Mone
GeneMark)?	
Is there evidence for	Yes as shown by Genemark
coding potential?	163 as shown by Ochemark
Is this gene present in	
other annotated	Yes
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	They suggest the start site: 24361
suggest?	
Does the start site have	Finale Score: -2.812
an associated Ribosome	Z value: 3.274
Binding Site with a high	Raw SD score: -1.767
score?	Not the highest score
Is the predicted start	ORF length is 1005 which is the longest.
codon the longest ORF?	
If not, does the longest	

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

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	Yes, the Gene was called by both Glimmer and GeneMark
program (Glimmer,	
GeneMark)?	
Is there evidence for	Yes
coding potential?	
Is this gene present in	Yes, it is annotating in 11 out of 313 (3.5%) of genes in
other annotated	Pham; it is called 90.9% of the time when present.
genomes?	
Does the gene violate	No
any major guiding	NO
principles?	
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 Start Coordinate: 25365
Glimmer and GeneMark	GeneMark Start Coordinate: 25365
suggest?	
Does the start site have	RBS Score: -6.377
an associated Ribosome	Z Score: 1.703
Binding Site with a high	
score?	
	Vac it is the leavest ODE
Is the predicted start	Yes, it is the longest ORF
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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,	Yes, it was called by both Glimmer and GeneMark	
GeneMark)?		
Is there evidence for	Yes	
coding potential?	103	
Is this gene present in	Yes, present in 9 out of 148 of genes in the pham; it is called 100.0% of the time when present.	
other annotated genomes?		
Does the gene violate		
any major guiding	Its spacer distance is above 12: This gene has a spacer	
principles?	distance of 14.	
DECISION:	Yes	

Gathering Evidence	Explain Your Rationale
What start site do	TI 11 1 1 1 0 5 7 0 0
Glimmer and GeneMark suggest?	They suggest the start site: 25720
Does the start site have	Yes, it's score is:
an associated Ribosome	Genomic Z Value: 2.054
Binding Site with a high	Final Score: -5.435
score?	
Is the predicted start	Yes, it has an ORF Length of 378
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes, it is called by both Glimmer and GenEMark
program (Glimmer,	Too, it is sailed by both similarior and sonEwark
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	
other annotated	Yes, it is found in 30 of 25 (85.7%) of genes in pham.
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 26123
suggest?	
Does the start site have	
an associated Ribosome	Z value: 2.797
Binding Site with a high	Final score: -3.469
score?	
Is the predicted start	Longest ORF of 837
codon the longest ORF?	Longest Orti of 637
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes, both
an auto-annotation	100, 5001

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	Both suggest 26956
suggest?	
Does the start site have	ORF: 333
an associated Ribosome	Finale Score: -3.366
Binding Site with a high	Z Value: 2.964
score?	Raw SD score: -2.366
Is the predicted start	Longest ORF length, 333
codon the longest ORF?	Longest Orth Tength, 555
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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

Gathering Evidence	Explain Your Rationale
What start site do	Both call 27285
Glimmer and GeneMark	Botti Call 27200
suggest?	
Does the start site have	
an associated Ribosome	Z value: 1.506
Binding Site with a high	Final Score: -5.918
score?	

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

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

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest? Does the start site have an associated Ribosome Binding Site with a high score?	Glimmer Start Coordinate: 28591 GeneMark Start Coordinate: 28591 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

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	Called by both Glimmer and GeneMark
program (Glimmer,	Called by Both Gillilliller and Genewark
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	
other annotated	Yes, found in 209 of 356 (58.7%) of genes in pham.
genomes?	
Does the gene violate	
any major guiding	It has a spacer distance of 15.
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	GeneMark and Glimmer suggest the start site: 28803
suggest?	
Does the start site have	Yes, It has the highest Z Value.
an associated Ribosome	Genomic Z value: 2.787
Binding Site with a high	Final Score: -4.256
score?	

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

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	Called by Glimmer
program (Glimmer,	Called by Cliffine
GeneMark)?	
Is there evidence for	Yes
coding potential?	100
Is this gene present in	
other annotated	Yes about 9.8%
genomes?	
Does the gene violate	
any major guiding	The gap is large
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark	Glimmer suggests 29392
suggest?	Gillini Suggests 20002
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.

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	Called by both Glimmer and GenEMark
program (Glimmer,	cance by boar committee and continuent
GeneMark)?	
Is there evidence for	Yes
coding potential?	103
Is this gene present in	
other annotated	96.2 % present
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

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

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	Examine the data for the auto-annotation programs and
an auto-annotation	indicate (YES BOTH, YES Glimmer only, YES GeneMark
program (Glimmer,	only, Neither)
GeneMark)?	
Is there evidence for	Discuss whether the GeneMarkS and/or GeneMark-host
coding potential?	trained coding potential map(s) show coding potential.
Is this gene present in other annotated genomes?	Discuss if other <u>related</u> , <u>annotated</u> 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 <u>Guiding</u> <u>Principles of Genome Annotation</u> 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.

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer Start Coordinate (type NA if not supported)::
Glimmer and GeneMark	GeneMark Start Coordinate (type NA if not supported)::
suggest?	
Does the start site have	List the final RBS score and Z-score of the currently
an associated Ribosome	predicted start site using the Kibler6/Karlin Medium scoring
Binding Site with a high	table. Indicate in your response if this is the best score or
score?	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.

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	Called by Glimmer	
program (Glimmer,	Called by GillTillel	
GeneMark)?		
Is there evidence for	Yes	
coding potential?	165	
Is this gene present in		
other annotated	Yes in 100%	
genomes?		
Does the gene violate		
any major guiding	Yes the gap	
principles?		
DECISION:		

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

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	Glimmer only
program (Glimmer,	Gilline only
GeneMark)?	
Is there evidence for	Vec
coding potential?	yes
Is this gene present in	
other annotated	100% found
genomes?	
Does the gene violate	
any major guiding	no
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 31078
suggest?	
Does the start site have	
an associated Ribosome	Final Score: -7.323
Binding Site with a high	Z value: 1.089
score?	
Is the predicted start	Longest ORF 1053
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Glimmer only
program (Glimmer,	Gilliller only
GeneMark)?	
Is there evidence for	Voc
coding potential?	yes
Is this gene present in	
other annotated	1.9% present
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 32276
suggest?	
Does the start site have	
an associated Ribosome	Z value: 3.013
Binding Site with a high	Final: -3.432
score?	
Is the predicted start	ORF 378 longest is 564
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Climanaranhy
program (Glimmer,	Glimmer only
GeneMark)?	
Is there evidence for	Vec
coding potential?	yes
Is this gene present in	
other annotated	10.1% found
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer: 32937
Glimmer and GeneMark	Gillitiller. 32937
suggest?	
Does the start site have	Final Score: -6.744
an associated Ribosome	Z value: 1.473
Binding Site with a high	
score?	
Is the predicted start	ORF 144
codon the longest ORF?	Long is 900
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Glimmer only	
program (Glimmer,	Gilliller only	
GeneMark)?		
Is there evidence for	Voc	
coding potential?	yes	
Is this gene present in		
other annotated	10.9%	
genomes?		
Does the gene violate		
any major guiding	Yes, it has a gap over 30< ; its gap is 272	
principles?		
DECISION:		

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 32937
suggest?	
Does the start site have	
an associated Ribosome	Final value : - 6.744
Binding Site with a high	Z value: 1.473
score?	
Is the predicted start	ORF is 144 not the longest
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes, both	
program (Glimmer,		
GeneMark)?		
Is there evidence for	Yes	
coding potential?	163	
Is this gene present in		
other annotated	Yes, it is present in 7 other genomes	
genomes?		
Does the gene violate	No	
any major guiding	NO	
principles?		
DECISION:	Yes	

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer Start Coordinate: 33077
Glimmer and GeneMark	GeneMark Start Coordinate: 33077
suggest?	
Does the start site have	RBS Score: -6.408
an associated Ribosome	Z Score: 1.760
Binding Site with a high	
score?	
Is the predicted start	Yes it is
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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

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

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	Yes both	
program (Glimmer,	1 es bout	
GeneMark)?		
Is there evidence for	Yes	
coding potential?		
Is this gene present in	Yes, it is present in 9 other genomes	
other annotated		
genomes?		
Does the gene violate	63 Gap	
any major guiding	03 Gap	
principles?		
DECISION:	Yes	

Gathering Evidence	Explain Your Rationale
What start site do	Glimmer Start Coordinate: 33813
Glimmer and GeneMark	GeneMark Start Coordinate: 33813
suggest?	
Does the start site have	RBS Score: -4.133
an associated Ribosome	Z Score: 2.417
Binding Site with a high	
score?	
Is the predicted start	Not the longest, but it is called in 100% of similar genomes
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Glimmer only
program (Glimmer,	
GeneMark)?	
Is there evidence for	Ves
coding potential?	yes
Is this gene present in	
other annotated	100% present
genomes?	
Does the gene violate	
any major guiding	no
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 34205
suggest?	
Does the start site have	Final –5.164
an associated Ribosome	Z value 2.422
Binding Site with a high	
score?	
Is the predicted start	ORF 438
codon the longest ORF?	Longest ORF is 438
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes, called by both GeneMark and Glimmer	
program (Glimmer,	res, called by both Genewark and Gillille	
GeneMark)?		
Is there evidence for	Yes	
coding potential?		
Is this gene present in	Yes, this gene is present in 9 out of 50 (18.0%) of genes in	
other annotated	pham.	
genomes?	priam.	
Does the gene violate		
any major guiding	No	
principles?		
DECISION:		

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

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	Called by Glimmer	
program (Glimmer,	Called by Gillillilei	
GeneMark)?		
Is there evidence for	Voc	
coding potential?	yes	
Is this gene present in		
other annotated	100% present in other genes	
genomes?		
Does the gene violate		
any major guiding	no	
principles?		
DECISION:		

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

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	Called by Glimmer	
program (Glimmer,	Called by Gill little!	
GeneMark)?		
Is there evidence for	Vas supported by Ganamark	
coding potential?	Yes, supported by Genemark	
Is this gene present in		
other annotated	Yes 100%	
genomes?		
Does the gene violate		
any major guiding	No	
principles?		
DECISION:		

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

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	Glimmer only
program (Glimmer,	Gilliller only
GeneMark)?	
Is there evidence for	yes
coding potential?	yes
Is this gene present in	
other annotated	100%present
genomes?	
Does the gene violate	
any major guiding	no
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 36452
suggest?	
Does the start site have	
an associated Ribosome	Z value: 2.155
Binding Site with a high	Final score -5.014
score?	
Is the predicted start	ORF 723 not the longest
codon the longest ORF?	Longest ORF is 1011
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes	
coding potential?		
Is this gene present in		
other annotated	No, it is only found in MenE.	
genomes?		
Does the gene violate		
any major guiding	No	
principles?		
DECISION:	It is a gene!	

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

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	Glimmer only	
program (Glimmer,	Gillille only	
GeneMark)?		
Is there evidence for	Ves	
coding potential?	yes	
Is this gene present in	Not most annotated in other genes	
other annotated		
genomes?		
Does the gene violate		
any major guiding	no	
principles?		
DECISION:	yes	

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 37503
suggest?	
Does the start site have	Final score - 4.699
an associated Ribosome	Z value 2.128
Binding Site with a high	Raw : -5.296
score?	Not the highest
Is the predicted start	ORF 471
codon the longest ORF?	Longest ORF 528
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Glimmer
program (Glimmer,	
GeneMark)?	
Is there evidence for	Voc
coding potential?	yes
Is this gene present in	
other annotated	100% present
genomes?	
Does the gene violate	
any major guiding	yes
principles?	
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer 37970
suggest?	
Does the start site have	
an associated Ribosome	Final score: -4.905
Binding Site with a high	Z value: 2.032
score?	Raw : -4.131
	ORF:306
	Highest score
Is the predicted start	ORF 306 and the longest
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	

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	Yes, both
an auto-annotation	
program (Glimmer,	
GeneMark)?	
Is there evidence for	Yes
coding potential?	163
Is this gene present in	
other annotated	Yes, in 7 other genomes
genomes?	
Does the gene violate	
any major guiding	No
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Both suggest 38272
suggest?	
Does the start site have	RBS Score: -4.548
an associated Ribosome	Z Score: 2.173
Binding Site with a high	
score?	
Is the predicted start	No, but the longest results in a –249 gap
codon the longest ORF?	
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	