Basic Phage Information	
Phage Name	Acosta
Gene #	1
Stop Coordinate	522 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	1 bp
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)?	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?	Yes, GenemarkS does show coding potential.
Is this gene present in other annotated genomes?	There were several phages from the same cluster (EA1) that observed this same gene. This start site was annotated by 100 phages in cluster EA1, with only 4 phages annotating another start site.
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? There is no significant overlap with other genes, the ORF is long enough (522 bp), and the genes before and after this one are both forward.
DECISION:	YES

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Both Glimmer and GeneMark say start is at 1 bp	
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS Score: -5.135 Z-score: 1.774 Z-score isn't the lowest z-score on there.	
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 with a length of 522 bp. No, it does not overlap to the nearest stop codon of the upstream ORF.	
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, starterator said that this start site was called in 111 of 155 non-draft genes.	
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	
DECISION:	 same information for each proposed start site. The gene should start at bp 1 due to the following reasons: Glimmer and Genemark agree it starts here No overlap with nearest stop codon Not the lowest z-score Starterator showed other phages calling this the start site for this gene. 	

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 smaller with appropriate coverage?	Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and evalue. It is only necessary to provide one match from each database.
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate	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
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 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
transmembrane protein? 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	Acosta
Gene #	2
Stop Coordinate	1922
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	Yes, there is overlap with previous Gene
Selected Start Coordinate	519
Selected Function	terminase

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 Glimmer and Gene Mark
Is there evidence for coding potential?	Yes, there does seem to be coding potential for this gene.
Is this gene present in other annotated genomes?	Yes, the gene is present in other annotated genomes in the EA1 cluster.
Does the gene violate any major guiding principles?	There is overlap of the genes and they both are heading the same direction as eachother.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	They had a suggest on EdO and atomic value of a suggest of
Glimmer and GeneMark suggest?	They both agreed on 519 and starterator also agrees.
Does the start site have	2.626 was the z value and the final score is –4.066. These
an associated Ribosome	were the best score that were avaliable.
Binding Site with a high	
score?	
Is the predicted start	Yes, it does seem to be, and it does result in some gene
codon the longest ORF?	overlap.
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	No a three start waint days monthly what atoms we water was distant
phage genomes as	Yes, the start point does match what starterator predicted.
indicated by Starterator?	
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.
	anginition the best match for all three blastif 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
DEGIGION.	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: Many e values of 0.0
functional assignment in	NCBI: Also many e values of 0.0
BlastP (phagesDB and/or	DNA Master: Many e values of 0.0
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	There are a few results that have over 95% match.
with a protein having a	
functional assignment in	Note: If you believe there is not a quality HHPred match,
the PDB or other	type No Quality Match and list the data for the best match
database in HHPred with	available to affirm the poor quality of the result and to
a probability of 90% or	document that HHPred was considered.
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, since having no evidence to support this possibility.
transmembrane protein?	
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.
	Yes, it is supported.
DECISION:	Terminase

Basic Phage Information	
Phage Name	Acosta
Gene #	3
Stop Coordinate	3334 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	37
Selected Start Coordinate	1958 bp
Selected Function	Portal protein

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?	Yes, GenemarkS does show coding potential.
Is this gene present in other annotated genomes?	This gene occurred in 104 annotated phages from cluster EA1.
Does the gene violate	There are no significant violations. There is no overlap with
any major guiding principles?	other genes, the gene is long enough, and the genes before and after this one are in the same direction.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Both suggest start site of 1958 bp.
Does the start site have	Final RBS score: -2.156
an associated Ribosome Binding Site with a high score?	Z-score: 3.092 This was the highest Z-score.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Length of ORF is 1377 bp, and there is a gap of 36 bp between the start codon of this gene and the stop codon of the upstream ORF.
Is this start site conserved in other	This start site was called in 116 of 158 non-draft genes.
phage genomes as indicated by Starterator?	
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:	Yes, this should be the starting site for the gene due to: • Glimmer and GeneMark both agree on this site • Highest z-score • No gene overlap • Start site called in 116 of 158 non-draft genes

Gathering Evidence	Explain Your Rationale

Does this protein align	Phospan DP: payaral a values of 0.0 that any nortal protain
with a protein having a	PhagesDB: several e values of 0.0 that say portal protein NCBI: several e values of 0.0 thay say portal protein
	HHpred: 1.4x10 ⁻²⁵ that says portal protein
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
	HHPred: 99.95% probability that says portal protein
	Ncbi: 2 with 90% identical that both say portal protein
functional assignment in the PDB or other	
database in HHPred with	
a probability of 90% or	
greater with appropriate	
coverage?	
_	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
order?	phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible	no
transmembrane protein?	
• •	yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Portal protein

Basic Phage Information	
Phage Name	Acosta
Gene #	4
Stop Coordinate	3437 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	3336 bp
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?	GenemarkS shows coding potential.
Is this gene present in other annotated genomes?	There are no other annotated phages that have observed this same gene.
Does the gene violate any major guiding principles?	No significant overlap, it is not longer than 120 bp, and the genes near this one are in the same direction.
DECISION:	No, it's super short and hasn't been annotated in any annotated phages.

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Genemark says start site of 3336 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS score: -4.541 Z-score: 2.406 Highest z-score
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?	Starterator not helpful.
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 a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10 ⁻⁴ or smaller with appropriate coverage?	List the most informative BlastP match from each source PhagesDB: NCBI: DNA Master: Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from
Door this protein alima	each database.
Does this protein align with a protein having a functional assignment in the PDB or other	List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.
database in HHPred with a probability of 90% or greater with appropriate coverage?	Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible transmembrane protein?	If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.
Is the proposed function found on the SEA-PHAGES approved function list?	Indicate a response with a Yes or No response. Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.
DECISION:	If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.

Basic Phage Information	
Phage Name	Acosta
Gene #	5
Stop Coordinate	4102
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-1
Selected Start Coordinate	3437
Selected Function	Hypothetical protein

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?	There is coding potential in GeneMarkS.
Is this gene present in other annotated genomes?	This gene is present in many other cluster EA1 phages.
Does the gene violate any major guiding principles?	There are no significant violations of the major guiding principles. No significant overlap, the gene is long enough, and the genes before and after this one are both forward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and GeneMark both have start site at 3437 bp.
Does the start site have	Final RBS Score: -3.161
an associated Ribosome	Z-score: 3.047
Binding Site with a high	This is the highest z-score.
score?	
	Length of ODE is CCC, and there is avenue of 1 by with the
Is the predicted start codon the longest ORF?	Length of ORF is 666, and there is overlap of 1 bp with the stop codon of the upstream gene.
If not, does the longest	stop codori or trie upstream gene.
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	Starterator not useful: this gene doesn't contain the most
conserved in other	often annotated starting site, meaning it is not in
phage genomes as	consensus with the majority.
indicated by Starterator?	
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.
55010101	The start site should be at 3437 bp because:
DECISION:	Glimmer and GeneMark agree on it
	This start site has the highest z-score

Gathering Evidence	Explain Your Rationale
	PhagesDB: several e values of e-123 that says MuF-like
with a protein having a	minor capsid protein

functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10 ⁻⁴ or smaller with appropriate coverage?	NCBI: e value that of 6e-157 that says MuF-like minor capsid protein HHpred: e value of 14 that says Transferase
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: 99.95% probabilty match MuF-like minor capsid protein NCBI: 65.68 % probabilty match with Transferase
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible transmembrane protein?	NO
Is the proposed function found on the SEA-PHAGES approved function list?	Yes
DECISION:	Hypothetical protein

Basic Phage Information	
Phage Name	Acosta
Gene #	6
Stop Coordinate	4748
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-82
Selected Start Coordinate	4185
Selected Function	Scaffolding protein

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?	There is coding potential
Is this gene present in other annotated genomes?	104 times for cluster EA1
Does the gene violate any major guiding principles?	No, genes do not overlap and the gene is long enough
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Both agree @bp 4185
Does the start site have an associated Ribosome Binding Site with a high score?	Z value: 2.333 RBS: -4.533
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	723 is the longest
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 186 of 186 (100.0%)
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
DECISION:	same information for each proposed start site. It should start at 4185 because - glimmer and genemark both agree on it - No major overlap

Gathering Evidence	Explain Your Rationale
Does this protein align with	NCBI:4e-94; scaffolding protein
a protein having a	Phagesdb: 2e-104; scaffolding protein
functional assignment in	HHPred: no good e values
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	NCBI: 94.51% identical; scaffolding protein
a protein having a	HHPred: 93.41 probablility; Coronin-1A
functional assignment in	
the PDB or other database	
in HHPred with a	
probability of 90% or	
greater with appropriate	
coverage?	If the consuming VEO conducts the consumer of function in the consum
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	no
transmembrane protein?	Yes
Is the proposed function found on the SEA-	100
PHAGES approved	
function list?	
DECISION:	Scaffolding protein
DECISION.	Scandiding protein

Basic Phage Information	
Phage Name	Acosta
Gene #	7
Stop Coordinate	5819
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	79
Selected Start Coordinate	4827
Selected Function	Major capsid protein

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?	Yes, there is coding potiental
Is this gene present in other annotated genomes?	Start number 8 was manually annotated 94 times for cluster EA1.
Does the gene violate any major guiding principles?	No significant violations, no extreme gene overlap, long enough ORF, and the genes around it are both forward
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 4827 bp
Does the start site have an associated Ribosome Binding Site with a high score?	RBS score= -2.095 Z-score= 3.186 This was the highest Z score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF length is 993 bp and this is the longest length.
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 424 of 453 (93.6%) of genes in pham
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:	I think the start site should be at 4827 bp because: Genemark and Glimmer agree Highest z-score It is found in 93.6% of genes

Gathering Evidence	Explain Your Rationale
Does this protein align with	PhagesDB: an e value of 0.0 that says major capsid protein
a protein having a	NCBI: several e values of 0.0 that say major capsid protein
functional assignment in	HHPred: e value of 3.7e-30 that says major capsid protein
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	HHPred: 100% probability match with major capsid protein
a protein having a	NCBI: 94.48% match that says major capsid protein
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	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:	Major capsid protein

Basic Phage Information	
Phage Name	Acosta
Gene #	8
Stop Coordinate	6115
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	Yes, there is gene overlap with the previous gene
Selected Start Coordinate	5822
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Yes, both Glimmer and GeneMark
Is there evidence for coding potential?	No, they do not seem to have some coding potential.
Is this gene present in other annotated genomes?	Yes, this gene is present within other geromes in the EA1 cluster.
Does the gene violate any major guiding principles?	Yes, there is an overlap with this gene and the previous one, but they are heading the same direction.
DECISION:	No

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark	Thou both agreed on 5000 as the start site
suggest?	They both agreed on 5822 as the start site.
Does the start site have	A z- value of 1.954 and had a final score of –4.747. They
an associated Ribosome	
	were not the longest ones and they do overlap with other
Binding Site with a high score?	genes.
Score?	
Is the predicted start	They were not the longest ones and they do overlap with
codon the longest ORF?	other genes.
If not, does the longest	emer geneer
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	
phage genomes as	Yes, phages dB, Starterator, and DNA Master all say the
indicated by Starterator?	same thing it is 5822.
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
	briefly explain your rationale.

Gathering Evidence Explain Your Rationale

Doos this protein align	Phocoopp:
Does this protein align with a protein having a	PhagesDB: NCBI:
	HHPred:
functional assignment in	ппетеа.
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	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.
L	g,-

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	6371
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	74
Selected Start Coordinate	6189
Selected Function	Hypothetical Protein

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?	GenemarkS does show coding potential.
Is this gene present in other annotated genomes?	3 annotated phages from cluster EA1 also observed this same game.
Does the gene violate any major guiding principles?	There are no significant violations. No overlap, it is longer than 120 bp, and the genes around it also go forward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark	Glimmer and Genemark agree on 6189 start site.
suggest? Does the start site have	Final RBS Score: -2.505
an associated Ribosome	Z-Score: 2.959
Binding Site with a high	This is the highest z-score.
score?	3
Is the predicted start	ORF is 183 bp long, and there is no overlap with stop
codon the longest ORF?	codon of upstream ORF.
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	Starterator not helpful: this gene doesn't contain the
conserved in other	consensus start site.
phage genomes as	
indicated by Starterator? 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
maicated by Black :	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.
	The start site should be at 6189 bp because:
DECISION:	Glimmer and Genemark agree on it
	Z-score is highest out of options

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

functional assignment in	NCBI: hypothetical
BlastP (phagesDB	DNA Master: hypothetical
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	Highest HHPred was 68%
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	<u>SEA-PHAGES Official Function List</u> 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.
DECISION:	Hypothetical Protein

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	6800 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	16 bp
Selected Start Coordinate	6387 bp
Selected Function	Hypothetical Protein

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?	Yes GenemarkS shows coding potential.
Is this gene present in other annotated genomes?	This gene was observed in 104 annotated phages in the cluster EA1.
Does the gene violate any major guiding principles?	No significant violations, it is long enough, there is no gene overlap, and the genes nearby are also forward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on this start coordinate at 6387 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS Score: -2.443 Z-score: 2.959 Z-score is the highest score on the list.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 414 bp long, and there is a gap of 16 bp with the nearest upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This start site is the most annotated start site, being manually annotated in 102 of 158 annotated phages.
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:	Yes, I think the start site is at 6387: Genemark and Glimmer agree Most annotated start on starterator Highest z-score

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

BlastP (phagesDB	DNA Master: Hypothetical Protein
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	Highest match was 75
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.
DECISION:	Hypothetical Protein

Basic Phage Information	
Phage Name	Acosta
Gene #	7179
Stop Coordinate	7179
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	Yes, there is overlap -28
Selected Start Coordinate	6772
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by	
an auto-annotation	Yes, both do
program (Glimmer,	res, both do
GeneMark)?	
Is there evidence for	Yes, it does show coding potential.
coding potential?	res, it does snow coding potential.
Is this gene present in	
other annotated	Yes, it is present within other genomes in the cluster.
genomes?	
Does the gene violate	
any major guiding	No, it does not violate any guiding principles.
principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer Start Coordinate (6772) GeneMark Start Coordinate (6778)
Does the start site have an associated Ribosome Binding Site with a high score?	6772- final score –4.407 & z 2.558 6778- final score –3.497 & z 2.358 Either of them was the largest score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	6772- ORF 408 6778- ORF 402 Yes, both have an overlap and do not seem to break any guidelines.
Is this start site conserved in other phage genomes as indicated by Starterator?	6772 was what starterator presented
Is this start site conserved in other phage genomes as indicated by BlastP?	6772- is conserved and presented in other genomes within the cluster 6778- is not on starterator nor phagesdb
DECISION:	Starterator, glimmer, and Phagedb all argee that start is 6772

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB: unknown
functional assignment in	NCBI: Hypothetical Protein
BlastP (phagesDB and/or	DNA Master: Hypothetical Protein
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	98% to Hypothetical Protein
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 that	and record the function of the genes found on each side of
shows high conservation	the gene in the same pham in the most closely related
of gene order?	phage. If the answer is NO, enter No Synteny Observed.
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.
DECISION:	Hypothetical Protein

Basic Phage Information	
Phage Name	Acosta
Gene #	12
Stop Coordinate	7517
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	7176
Selected Function	Hypothetical Protein

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?	Yes there is coding potiental
Is this gene present in other annotated genomes?	Start number 2 was manually annotated 104 times for cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations. There is no overlap with other genes, the gene is long enough, and the genes before and after this one are in the same direction
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	They suggest the start site is at 7176 bp
and GeneMark suggest?	
Does the start site have an	RBS score= -2.584
associated Ribosome	Z-score= 2.959
Binding Site with a high	This was the highest Z score
score?	
Is the predicted start codon	Length of ORF is 342 bp, this was the longest ORF length.
the longest ORF? If not,	There is a gap of -3 bp between the start codon of this gene and
does the longest ORF	the stop codon of the upstream ORF.
result in excessive gene	
overlap (>30bp)?	Found in 404 of 400 (00 00/) of managing where
Is this start site conserved	• Found in 184 of 186 (98.9%) of genes in pham
in other phage genomes	Manual Annotations of this start: 157 of 159 Colled 100.0% of time when present.
as indicated by Starterator?	Called 100.0% of time when present
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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.
	Yes, this should be the starting site for the gene due to:
	Glimmer and GeneMark both agree on this site Highest 7 pages
DECISION:	Highest z-scoreStart site called in 184 of 186 non-draft genes
	Called 100% of the time
	Canad 10070 of the time

Gathering Evidence	Explain Your Rationale
Does this protein align with	List the most informative BlastP match from each source
a protein having a	PhagesDB: unknown function
functional assignment in	NCBI: Hypothetical Protein
BlastP (phagesDB and/or	DNA Master: Hypothetical Protein
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	99 minor tIL PROTEIN
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
DECISION:	Hypothetical Protein
220,0,0,,,	

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	7876 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	0 bp
Selected Start Coordinate	7517 bp
Selected Function	Tail Terminator

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?	GeneMarkS shows coding potential.
Is this gene present in other annotated genomes?	This gene was manually annotated in 104 other phages from cluster EA1.
Does the gene violate any major guiding principles?	No significant violations, no extreme gene overlap, long enough ORF, and the genes around it are both forward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 7517 bp.
Does the start site have	Final RBS score: -3.917
an associated Ribosome Binding Site with a high score?	Z-score: 2.311 Tied for highest z-score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 360 bp long, and the start site starts exactly where the upstream gene ends.
Is this start site	This gene doesn't contain the consensus start site on
conserved in other phage genomes as indicated by Starterator?	starterator, but this start site I chose was manually annotated in 44 of 158 phages with this gene.
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:	I think the start site should be at 7517 bp because: Genemark and Glimmer agree Highest z-score Not consensus start site, but still annotated by others

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: e of 3e-46, tail terminator
with a protein having a	NCBI: e of 2e-56, tail terminator

fti 1 i	III lavado de a contra la lavo 404
functional assignment in	HHpred: no e values below 10 ⁻⁴
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	HHPred: 97.84 probability with a gene for a tail terminator
with a protein having a	protein
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	No
transmembrane protein?	
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Tail Terminator Protein

Basic Phage Information	
Phage Name	Acosta
Gene #	14
Stop Coordinate	8103 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-4 bp
Selected Start Coordinate	7873 bp
Selected Function	Hypothetical Protein

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?	GeneMarkS shows coding potential.
Is this gene present in other annotated genomes?	Gene manually annotated in 104 other phages in cluster EA1.
Does the gene violate any major guiding principles?	No significant violations, no significant gene overlap, long enough ORF, and the genes surrounding are both forward
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and GenemarkS agree on start site 7873 bp.
Does the start site have an associated Ribosome Binding Site with a high score? Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Final RBS score: -2.175 Z-score: 3.112 Highest z-score ORF is 231 bp, and there is overlap of 4 bp with the upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This start site is the most annotated start site on starterator, called in 132 of 137 annotated genomes.
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:	This gene start site should be at 7873 bp because: Highest z-score Glimmer and Genemark agree Starterator calls this most annotated 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: unknown
functional assignment in	NCBI: hypothetical
BlastP (phagesDB	DNA Master: hypothetical
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	Highest match was 68%
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.
DECISION:	Hypothetical Protein
DEGIGIOIT.	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

Basic Phage Information	
Phage Name	Acosta
Gene #	15
Stop Coordinate	8615
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	Yes, there is a gap 15 bp
Selected Start Coordinate	8118
Selected Function	Major tail protein

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?	Yes, there seems to be coding potenial
Is this gene present in other annotated genomes?	Yes, similar genes have been present in other genomes.
Does the gene violate any major guiding principles?	No, it does not seem to violate any major guidelines
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Both point to 8118
suggest?	
Does the start site have	8118- Final score –2.786 & ORF 498
an associated Ribosome	Not the highest score
Binding Site with a high	
score?	
	The way is 4.4 by
Is the predicted start	The gap is 14 bp
codon the longest ORF?	Yes, it does result in gene overlap
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other phage	
genomes as indicated by	Yes, the start site seems to be the same in multpile other
Starterator?	genes.
Is this start site	8118- Yes, it is present within other genomes.
conserved in other phage	
genomes as indicated by	
BlastP?	
DECISION:	8118- starts here because startetator, glimmer, genemark,
DECISION.	and phagesdb all say the same thing.

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: 4e-90 NCBI: 6e-112 HHPred: No e-values below 10-4
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	Yes, in NCBI 97.58% align with a function.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	No Synteny Observed.
Is this gene a possible transmembrane protein?	No
Is the proposed function found on the SEA-PHAGES approved function list?	Yes
DECISION:	Major tail protein

Basic Phage Information	
Phage Name	Acosta
Gene #	16
Stop Coordinate	9215
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	30 bp
Selected Start Coordinate	8646 bp
Selected Function	Tail Assembly Chaperone

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?	GenemarkS shows coding potential
Is this gene present in other annotated genomes?	Gene observed in 104 annotated phages from cluster EA1.
Does the gene violate any major guiding principles?	No significant violations, not significant overlap, long enough reading frame, and genes surrounding are both forward
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 8646
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. Final RBS score: -2.794 Z-score: 2.959 Highest z-score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 570 bp long, with there being a 30 bp gap with the nearest upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the most annotated start site in starterator, being called in 140 of 158 annotated genes.
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:	The start site should be at 8646 bp because: • Highest z-score • Glimmer and Genemark agree • Starterator says this is most annotated start site

Gathering Evidence	Explain Your Rationale

Deserthic protein alien	List the most informative Block procedule from each course
Does this protein align	List the most informative BlastP match from each source
with a protein having a functional assignment in	PhagesDB: 2e-97 for a gene assigned tail assembly
BlastP (phagesDB	chaperone
and/or GenBank) with an	NCBI: 1e-124 for a gene assigned tail assembly chaperone
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	No HHPred matches with above 90 probability
	· · ·
with a protein having a	NCBI: 91.49% identical gene that is for a tail assembly
functional assignment in the PDB or other	chaperone
database in HHPred with	
a probability of 90% or	
greater with appropriate	
coverage?	If the energy is VEC evaluate the managed function in the
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?	No
Is this gene a possible	IVO
transmembrane protein?	Voe
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	Tail accombly above areas
DECISION:	Tail assembly chaperone

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	9604 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	15 bp
Selected Start Coordinate	9230 bp
Selected Function	Tail Assembly Chaperone

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?	Genemark S and Genemark-host shows coding potential
Is this gene present in other annotated genomes?	This gene was manually annotated in 106 other cluster EA1 phage genomes.
Does the gene violate any major guiding	There are no significant violations, it is long enough, there is no overlap with the upstream gene, and the genes
principles?	around go in the same direction.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and GeneMark	Glimmer and Genemark agree on 9230 bp.
suggest?	
Does the start site have	Final RBS: -7.612
an associated Ribosome	Z-score: .943

Binding Site with a high score?	These aren't very good scores.
Is the predicted start codon the longest ORF? If not, does the longest	The ORF is 375 bp with a gap of 15 bp between the upstream gene.
ORF result in excessive gene overlap (>30bp)?	The longest ORF has 171 bp overlap, so not an option.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the consensus start site, called in 110 of 121 non-draft genes.
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:	I think the start site should be 9230 bp because: Glimmer and GeneMark agree Consensus start site on starterator

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 2e-68; tail assembly chaperone
with a protein having a	NCBI: 3e-83; tail assembly chaperone
functional assignment in	HHpred: no low e value
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	NCBI: 95.97% identical; tail assembly chaperone
with a protein having a	HHPred: 90.66 probability; metal binding protein
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	If the answer is YES, evaluate the proposed function in the 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:	Tail assembly chaperone

Basic Phage Information	
Phage Name	Acosta
Gene #	18
Stop Coordinate	11892
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-24
Selected Start Coordinate	9628
Selected Function	Tape Measure Protein

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?	There is coding potential for both GeneMarkS and GeneMark-host.
Is this gene present in other annotated genomes?	Start number 4 was manually annotated 3 times for cluster EA1.
Does the gene violate any	No significant violations, not significant overlap, long enough
major guiding principles?	reading frame, and genes surrounding are both forward
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	They both suggest a start at 9628 bp
Does the start site have an	final RBS score- (-2.095)
associated Ribosome	Z-score-(3.186)
Binding Site with a high score?	This was the highest Z score
Is the predicted start codon	ORF is 2265 bp long, with there being a -24 bp gap with the
the longest ORF? If not,	nearest upstream gene
does the longest ORF result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	• Found in 69 of 186 (37.1%) of genes in pham
in other phage genomes	Manual Annotations of this start: 58 of 159
as indicated by	Called 100.0% of time when present
Starterator?	
Is this start site conserved in other phage genomes	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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.
	The start site should be at 9628 bp because:
DEGIGION.	Highest z-score Wing a grant Company of the same and the same
DECISION:	 Glimmer and Genemark agree Starterator says this is most annotated start site
	• Starterator says triis is most amnotated start site

Gathering Evidence	Explain Your Rationale
Does this protein align with	List the most informative BlastP match from each source
a protein having a	PhagesDB: 0.0, tape measure protein
functional assignment in	NCBI: e- 0.0
BlastP (phagesDB and/or	HHPred: 1.6e-18
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	HHPred has 99.94% with gene probability of tape measure
a protein having a	protein
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	No
transmembrane protein?	Voc
Is the proposed function	Yes
found on the SEA-	
PHAGES approved function list?	
	Tono Magguro Protoin
DECISION:	Tape Measure Protein

Basic Phage Information	
Phage Name	Acosta
Gene #	19
Stop Coordinate	12662
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	11889
Selected Function	Minor tail protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	Yes Both
(Glimmer, GeneMark)?	
Is there evidence for	GeneMark and GeneMarkS both show coding potential
coding potential?	
Is this gene present in	104 times for cluster EA1
other annotated genomes?	
Does the gene violate any	End and start do not overlap. Gene is long enoughF
major guiding principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Glimmer Start Coordinate 11889
and GeneMark suggest?	GeneMark Start Coordinate 11889
Door the start site house are	DDC 222721 2 240
Does the start site have an associated Ribosome	RBS score: -3.319 Z value: 2.618
Binding Site with a high	2 Value. 2.010
score?	
333.3.	
Is the predicted start codon	ORF: 774
the longest ORF? If not,	Longest ORF in gene overlap
does the longest ORF	
result in excessive gene	
overlap (>30bp)? Is this start site conserved	
in other phage genomes as indicated by	It is found in 160 out of 160 (100.0%)
Starterator?	11 13 10d11d 111 100 0dt 01 100 (100.078)
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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:	Start should be at 11889
DECISION:	Genemark and glimmer show this start
	Does not overlap

Gathering Evidence	Explain Your Rationale
Does this protein align with	PhagesDB: an e value of e –138 that says minor tail protien
a protein having a	NCBI: an e value of 6e-176 that says minor tail protien
functional assignment in	HHPred: e value of 2.3e-25 that says tail protien
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	HHPred: 98.19% probabilty match with
a protein having a	NCBI: 94.94% probabilty match with minor tail protien
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	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

Basic Phage Information	
Phage Name	Acosta
Gene #	15085
Stop Coordinate	15085
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	No
Selected Start Coordinate	12662
Selected Function	Minor tail protein

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?	Yes, it shows coding potential.
Is this gene present in other annotated genomes?	Yes, it is present within other genomes in the same cluster.
Does the gene violate any major guiding principles?	It doesn't seem to be violating any guide lines.
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?	They both point to 12662 and so does phagesdb.
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. Final score: -2.443 and Z-score: 2.959 these are indeed the best scores shown.
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? ORF length is 2424 and it seems to start where it stopped.
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, other genomes have the same start site within this cluster.
Is this start site conserved in other phage genomes as indicated by BlastP?	Yes, there are some other gemones that do start the same .
DECISION:	12662 because Dna master, Glimmer, and Startetator all say the same.

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 0.0; minor tail protein
with a protein having a	NCBI: 0.0; minor tail protein
functional assignment in	HHPred: 2.5e-22; phage tail, tail tip, tape measure protein
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	NCBI: 97.89% identical; minor tail protein
with a protein having a	HHPred: 99.93 probablity; prophage tail protein
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 that	and record the function of the genes found on each side of
shows high conservation	the gene in the same pham in the most closely related
of gene order?	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

Basic Phage Information	
Phage Name	Acosta
Gene #	22
Stop Coordinate	15270 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-1 bp
Selected Start Coordinate	15085 bp
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)?	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?	GenemarkS and Genemark-host show coding potential
Is this gene present in other annotated genomes?	This gene was observed in 104 other cluster EA1 phages.
Does the gene violate any major guiding principles?	Discuss if there are any significant violations of the <u>Guiding 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? No serious violations: no signficant overlap, it is long enough, and the genes surrounding are also forward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on the start site 15085
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS score: -3.566 Z-score: 2.474 Tied for highest z-score.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	The ORF would be 186 bp long with an overlap of 1 bp with the nearest upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This matches the consensus start site on starterator, being called in 74 of the 138 non-draft genes.
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:	Reasons for 15085 bp as the start site: Glimmer and Genemark agree on it Tied for highest z-score Starterator says this is consensus 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:
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.

Basic Phage Information	
Phage Name	Acosta
Gene #	23
Stop Coordinate	15863
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	overlap
Selected Start Coordinate	15267
Selected Function	Hypothetical Protein

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?	Yes, there is some evidence of coding potential
Is this gene present in other annotated genomes?	Yes, the gene is present in other annotated genomes.
Does the gene violate any major guiding principles?	It doesn't seem to violate any guide lines
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?	Both of them say it's 15267
Does the start site have an associated Ribosome Binding Site with a high score?	The final RBS score is –4.602 and Z-score 2.022, both are the best score shown.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	The length of the ORF is 597 and it does have an overlap that doesn't seem to break any guidelines.
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, the start site is reserved in other genomes.
Is this start site conserved in other phage genomes as indicated by BlastP?	Yes, other genomes have the same start site.
DECISION:	15267 Startetator, Phagesdb, Glimmer, and Gene mark all say the same.

Gathering Evidence	Explain Your Rationale

transmembrane protein? Is the proposed function found on the SEA-PHAGES approved function list?	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.
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 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
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match. Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10 ⁻⁴ or smaller with appropriate coverage?	List the most informative BlastP match from each source PhagesDB: NCBI: DNA Master: Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.

Basic Phage Information	
Phage Name	Acosta
Gene #	16348
Stop Coordinate	16348bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	15863bp
Selected Function	Hypothetical Protein

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?	Yes, there is coding potiental
Is this gene present in other annotated genomes?	Start number 11 was manually annotated 103 times for cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations, it is long enough, there is no overlap with the upstream gene, and the genes around go in the same direction.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 15863 bp
Does the start site have an	Final RBS: -3.390
associated Ribosome	Z-score: 2.548
Binding Site with a high	This was the highest Z score
score?	
Is the predicted start codon	The ORF is 486 bp with a gap of 0 bp between the upstream
the longest ORF? If not,	gene.
does the longest ORF	
result in excessive gene	This was the highest ORF length
overlap (>30bp)?	Farmed in 400 of 400 (70 00/) of manage in mlane
Is this start site conserved	Found in 123 of 160 (76.9%) of genes in pham Manual Annotations of this start: 103 of 138
in other phage genomes as indicated by	Manual Annotations of this start. 103 of 136
Starterator?	
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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.
	I think the start site should be 9230 bp because:
DECISION:	Glimmer and GeneMark agree
	Has the longest ORF length I loo the highest 7 ages.
	 Has the highest Z score

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
	If you believe this game about the construct where the
	If you believe this gene should be assigned, please write the
DECISION:	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	Acosta
Gene #	24
Stop Coordinate	18477 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-1 bp
Selected Start Coordinate	16348 bp
Selected Function	Minor tail protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was called in 104 other cluster EA1 phages.
Does the gene violate any major guiding principles?	No significant violations: no significant overlap, it is long enough, and the surrounding genes do go in the same direction.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say start site is at 16348 bp.
Does the start site have an associated Ribosome	Final RBS score: -3.837 Z-score: 2.474
Binding Site with a high score?	Highest z-score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 2130 bp long, with an overlap of 1 bp with the nearest upstream gene.
Is this start site conserved in other	This is the consensus start site on starterator, called in 114 of 138 non-draft genes.
phage genomes as	of 130 non-draft genes.
indicated by Starterator?	
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.
	The start site is at 16348 bp because:
DECISION:	Glimmer and Genemark agree Little and Genemark agree
	 Highest z-score Consensus start site on starterator
	• Constrisus start site on starterator

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 0.0; minor tail protein
with a protein having a	NCBI: 0.0; minor tail protein

functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10 ⁻⁴ or smaller with appropriate coverage?	HHpred: 5.6e-25; chitinase
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?	NCBI: 96.33% identical; minor tail protein HHPred: 99.94 probability; chitinase
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible transmembrane protein?	no
Is the proposed function found on the SEA-PHAGES approved function list?	Yes
DECISION:	Minor tail protein

Basic Phage Information	
Phage Name	Acosta
Gene #	25
Stop Coordinate	19330
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	18509
Selected Function	Lysin A

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program	Examine the data for the auto-annotation programs and indicate (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
(Glimmer, GeneMark)? Is there evidence for coding potential?	Both show coding potential
Is this gene present in other annotated genomes?	Start number 24 was manually annotated 2 times for cluster EA1
Does the gene violate any major guiding principles?	Long enough, does not overlap.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Glimmer Start Coordinate 18509
and GeneMark suggest?	GeneMark Start Coordinate 18509
Does the start site have an	RBS: -2.443
associated Ribosome	Z score: 2.959
Binding Site with a high score?	
score?	
Is the predicted start codon	822 is the longest ORG with 0 bp gap
the longest ORF? If not,	344 34
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	• Found in 4 of 179 (2.2%) of genes in pham
in other phage genomes	Manual Annotations of this start: 2 of 153
as indicated by	Called 100.0% of time when present
Starterator?	Drawide the best Block meetals from NCBI Blocks and BNA
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes as indicated by BlastP?	Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the
as indicated by blastr?	database match. List the e-value and alignment of the best
	match for all three BlastP sources.
	materi for all tillee blasti soulces.
	Note: if you are considering more than 1 start site, provide the
	same information for each proposed start site.
DECISION.	The start should be at 18509 because Genemark and Glimmer
DECISION:	agree.

Gathering Evidence	Explain Your Rationale
Does this protein align with	List the most informative BlastP match from each source
a protein having a	PhagesDB: e-153
functional assignment in	NCBI: 4e-177
BlastP (phagesDB and/or	HHPred: 1.3e-8
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	NCBI: has a 93.41% identical match to lysin A
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	Yes
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:	NCBI: calls 93.41% for Lysin A

Basic Phage Information	
Phage Name	Acosta
Gene #	26
Stop Coordinate	19776
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	gap
Selected Start Coordinate	19363
Selected Function	Hypothetical Protein

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?	Yes, there does seem to have some coding potenial.
Is this gene present in other annotated genomes?	Yes, it is present in other genomes.
Does the gene violate any major guiding principles?	It does not seem to violate any guide lines.
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	Both say 19363 as the start site.
suggest?	
Does the start site have	RBS score1.954
an associated Ribosome	Z-Score- 3.186
Binding Site with a high	Yes, these scores are the highest that are shown.
score?	
Is the predicted start	ORF Length- 414
codon the longest ORF?	Everything seems to line up and the gap does not seem to
If not, does the longest	violate any guidelines.
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	No, other genomes don't have it conserved.
conserved in other phage	
genomes as indicated by	
Starterator?	
Is this start site	No, the start site is not conserved in other genomes.
conserved in other phage	
genomes as indicated by	
BlastP?	
DECISION:	19363 GeneMark, Glimmer, Phages DB, and Startetator all
	say the same 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	List the most informative BlastP match from each source PhagesDB: NCBI: DNA Master: Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and
coverage?	e-value. It is only necessary to provide one match from each database.
Does this protein align with a protein having a functional assignment in the PDB or other	List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.
database in HHPred with a probability of 90% or greater with appropriate coverage?	Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? Is this gene a possible transmembrane protein? Is the proposed function found on the SEA-PHAGES approved function list?	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.
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	Acosta
Gene #	27
Stop Coordinate	20120 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	7 bp
Selected Start Coordinate	19782 bp
Selected Function	Holin

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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was observed in 104 other Cluster EA1 phages.
Does the gene violate any major guiding principles?	No significant violations: no significant overlap, it is long enough, and the genes surrounding are also foward.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both suggest 19782 bp.
Does the start site have an associated Ribosome Binding Site with a high score? Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Final RBS Score: -4.245 Z-score: 2.728 2 nd highest z-score and final score, but the highest z-score start site has ORF that's 48 bp long. ORF is 339 bp long, with a gap of 7 bp to the nearest upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the most annotated start site according to starterator, being called in 112 of 127 non-draft genes.
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:	Start site should be at 19782 because: Genemark and Glimmer agree Starterator says this is consensus High z-score and final score

Gathering Evidence	Explain Your Rationale
Does this protein align	List the most informative BlastP match from each source
with a protein having a	PhagesDB: 8e-51
functional assignment in	NCBI: 1e-62
BlastP (phagesDB	HHPred: 3.6e-16
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?	HHPred calls it Holin with a score of 99.72% that is identical.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	No
Is this gene a possible transmembrane protein?	Yes Query MSA diversity (Neff): 7.42017
transmembrane protein:	Detected sequence features: •Transmembrane segment(s)
Is the proposed function found on the SEA-PHAGES approved function list?	Yes, it is on the approved list of functions.
DECISION:	NCBI calls it with a 91.96% match and HHPred calls it with a 99.72% match both calling holin.

Basic Phage Information	
Phage Name	Acosta
Gene #	29
Stop Coordinate	20654 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	61 bp
Selected Start Coordinate	20845 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was manually annotated in 104 other cluster EA1 genes according to starterator.
Does the gene violate any major guiding principles?	There are no significant violations: there is no overlap, it is long enough, and the gene before this one is the opposite direction (but there is a big enough gap between).
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say the start site is 20845 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS score: -2.443 Z-Score: 2.959 These are the highest z-score and final score.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	The ORF is 192 bp long, and there is a 61 bp gap. Although, this gap is between two opposite direction genes, so it doesn't violate any guidelines.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the most annotated start site on starterator, being called in 147 of 151 non-draft genes.
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
DECISION:	the same information for each proposed start site. Start site should be at 20845 bp because: Glimmer and Genemark Final and Z scores are the highest Most annotated start site on starterator

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.

Basic Phage Information	
Phage Name	Acosta
Gene #	21070
Stop Coordinate	21070
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	There is a Gap
Selected Start Coordinate	21228
Selected Function	Hypothetical Protein

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?	Yes, there is some coding potential.
Is this gene present in other annotated genomes?	Yes, the gene is shown in other annotated genomes.
Does the gene violate any major guiding principles?	I believe it may violate due to going past over 50
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	They both point to 21228
suggest?	
Does the start site have	RBS score = -4.075
an associated Ribosome	Z-score = 2.266
Binding Site with a high	RBS score is best scoring, but the Z-score is the second
score?	best.
Is the predicted start	ORF- 159 is the longest one and does exceed the amount
codon the longest ORF?	needed.
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other phage	No, Starterator does not show that the start site is
genomes as indicated by	conserved in other genomes
Starterator?	
Is this start site	No, the start site is not conserved by other genomes
conserved in other phage	
genomes as indicated by	
BlastP?	
DECISION:	Yes, because Starterator, Glimmer, Gene Mark, and
DECISION.	Phages db say the same.

Gathering Evidence	Explain Your Rationale

Does this protein align with a protein having a	List the most informative BlastP match from each source PhagesDB:
functional assignment in	NCBI:
9	DNA Master:
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
coverage?	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	to provide the best materi.
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 that	and record the function of the genes found on each side of
shows high conservation	the gene in the same pham in the most closely related
of gene order?	phage. If the answer is NO, enter No Synteny Observed.
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	<u>SEA-PHAGES Official Function List</u> 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.

Basic Phage Information	
Phage Name	Acosta
Gene #	21225
Stop Coordinate	21225 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-4
Selected Start Coordinate	22079 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	The gene was annotated in 104 other cluster EA1 genes.
Does the gene violate any major guiding principles?	There is no significant violations: no significant overlap, it is long enough, and the genes surrounding are also reverse.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both say 22079 bp.
Does the start site have	Final: -6.275
an associated Ribosome	Z-score: 1.925
Binding Site with a high score?	Middle z-score and lowest final score. Not very good
Is the predicted start	The ORF is 855 bp long, and there is a small 4 bp overlap
codon the longest ORF?	with the nearest upstream gene.
If not, does the longest ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	This is the most annotated start site on starterator, being
conserved in other	called in 129 of 156 non-draft genes.
phage genomes as	
indicated by Starterator?	
Is this start site	Provide the best BlastP match from NCBI, PhagesDB, and
conserved in other phage genomes as	DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S
indicated by BlastP?	(subject) is the database match. List the e-value and
maicated by Black :	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.
	The start site should be at 22079 bp because:
DECISION:	Glimmer and genemark agree
	Consensus start site on starterator

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	<u>SEA-PHAGES Official Function List</u> 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 haliave this gone should be assigned places write
	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
DECISION.	function. 50-70% of phage genes fall into the NKF
	category.
	category.

Basic Phage Information	
Phage Name	Acosta
Gene #	22189
Stop Coordinate	22189 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-11 bp
Selected Start Coordinate	22545 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	There is a single cluster EA1 phage that manually annotated this gene.
Does the gene violate any major guiding principles?	There are no significant violations: no significant overlap, it is long enough, and the surrounding genes are also reverse.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say 22545 bp as start site.
Does the start site have	Final: -2.214
an associated Ribosome	z-score: 3.103
Binding Site with a high score?	Highest z and final scores.
Is the predicted start	ORF is 357 bp long, and there is an 11 bp overlap with the
codon the longest ORF?	nearest upstream gene.
If not, does the longest ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	This is the consensus start site, but when that only pulls
conserved in other	from a pool of a single annotated gene, that isn't very
phage genomes as	helpful.
indicated by Starterator?	Drawide the heat Plants metals from NCBI Phaseapp and
Is this start site conserved in other	Provide the best BlastP match from NCBI, PhagesDB, and 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.
	The start site is 22545 bp because:
DECISION:	Glimmer and Genemark agree
	Highest z and final scores

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	<u>SEA-PHAGES Official Function List</u> 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.

Basic Phage Information	
Phage Name	Acosta
Gene #	22555
Stop Coordinate	22555
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	Gap
Selected Start Coordinate	22806
Selected Function	Hypothetical Protein

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?	Yes, it does seem to have some coding potential.
Is this gene present in other annotated genomes?	No, it does not seem to be
Does the gene violate any major guiding principles?	No, it does not seem to
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?	Both call 22806
Does the start site have an associated Ribosome Binding Site with a high score?	RBS3.262 Z- score- 2.616 They are the best scores shown
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF- 252 Yes, it has a gap and it doesn't violate any guidelines
Is this start site conserved in other phage genomes as indicated by Starterator?	No, it does not seem to have any other genomes that are similar.
Is this start site conserved in other phage genomes as indicated by BlastP?	No, it does not seem to be conserved in other genomes.
DECISION:	22806 is the start because glimmer, Genemark, starterator, and phagesdb call as that.

Gathering Evidence	Explain Your Rationale

functional assignment in the PDB or other database in HHPred with a probability of 90% or	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
greater with appropriate coverage? Is this gene located	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
adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	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

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	22806 bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-1 bp
Selected Start Coordinate	24488 bp
Selected Function	RecA-like DNA Recombinase

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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was observed in 104 annotated phages in cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations: no significant overlap, it is long enough, and the surrounding genes are also reverse.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Glimmer says: 24488 bp Genemark says: 24419
Does the start site have	24488:
an associated Ribosome	Z: 3.029
Binding Site with a high	Final: -2.433
score?	Highest z score and final score
	3
	24419:
	Z: 2.179
	Final: -4.263
	Average scores
Is the predicted start	24488: ORF is 1683 bp long with a 1 bp overlap
codon the longest ORF?	
If not, does the longest	24419: ORF is 1614 bp long with a 1 bp overlap.
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	24488 bp is the consensus start site according to
conserved in other	starterator, being called in 144 of 211 non-draft genes.
phage genomes as	
indicated by Starterator?	
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.
	The start site should be at 24488 bp because:
DECISION:	Glimmer say so
223.3.3.1	Highest z and final scores
	Consensus start site on starterator

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 0.0; DNA Recombinase
with a protein having a	NCBI: 0.0; DNA Recombinase
functional assignment in	HHpred: 1.3e-18; Regulatory protein repA
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?	NCBI: 95.36% identical; DNA Recombinase HHPred: Regulatory protein repA
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible transmembrane protein?	No
Is the proposed function found on the SEA-PHAGES approved function list?	Sort of; no DNA Recombinase, but RecA-like DNA Recombinase is
DECISION:	RecA-like DNA Recombinase

Basic Phage Information	
Phage Name	Acosta
Gene #	24759
Stop Coordinate	24759
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	Overlap
Selected Start Coordinate	25604
Selected Function	Hypothetical

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?	Yes, there is coding potential
Is this gene present in other annotated genomes?	It does not seem to be present in other genomes
Does the gene violate any major guiding principles?	No, it does not seem to
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?	Both suggest 25604
Does the start site have an associated Ribosome Binding Site with a high score?	RBS= -4.871 Z-score = 1.831 Both are not the highest scoring
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF- 846 It is not the longest and the other do seem to result in overlap
Is this start site conserved in other phage genomes as indicated by Starterator?	Yes, it is conserved in the cluster
Is this start site conserved in other phage genomes as indicated by BlastP?	No, it does not seem to be
DECISION:	Starterator, glimmer, Gene mark, and Phages db all suggest the same start.

Gathering Evidence	Explain Your Rationale

D (1) () !!	List the most information Disagram
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 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
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 that	and record the function of the genes found on each side of
shows high conservation	the gene in the same pham in the most closely related
of gene order?	phage. If the answer is NO, enter No Synteny Observed.
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.

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	25637 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	34 bp
Selected Start Coordinate	26305 bp
Selected Function	AAA-ATPase

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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	Gene observed in 104 annotated genomes in cluster EA1.
Does the gene violate any major guiding principles?	Discuss if there are any significant violations of the <u>Guiding 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? There is no significant violations: no overlap, it is long enough, and the surrounding genes are also reverse.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both say 26305 bp as start site.
Does the start site have an associated Ribosome Binding Site with a high score?	Final: -4.395 Z: 2.090 2 nd highest scores
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 669 bp long, with a 34 bp gap.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is consensus start site, being called in 103 of 158 non-draft genes.
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:	Start site should be at 26305 bp because: • Glimmer and Genemark agree • Pretty high z and final scores • Consensus start site

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: e-118; AAA-ATPase
with a protein having a	NCBI: 5e-155; AAA-ATPase
functional assignment in	HHPred: 6.7e-9; RecA superfamily ATPase
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	NCBI: 95.95% identical; AAA-ATPase
functional assignment in	HHPred: 99.13; Reca superfamily ATPase
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 known function and in a	the gene order. Examine the adjacent genes found in the 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
Is this gene a possible	Synteny Observed. No
transmembrane protein?	740
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	444.470
DECISION:	AAA-ATPase

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	26302 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-3 bp
Selected Start Coordinate	27483 bp
Selected Function	exonuclease

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	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?	GeneMark S and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was manually annotated in 104 other cluster EA1 phages.
Does the gene violate any major guiding principles?	There are no significant violations. No significant overlap, gene is long enough, and surrounding genes are reverse.
DECISION:	YES

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	Glimmer: 27483 bp Genemark: 27492 bp
Does the start site have an associated Ribosome Binding Site with a high score?	27483 bp: Final: -3.671 Z-Score: 2.389 Second highest final and z scores 27492: Final: -4.802
	Z-Score: 1.929 3 rd highest scores
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	With a start site at 27483 bp, the ORF is 1182 bp long, and with a start site of 27492 bp, the ORF is 1191 bp long. Neither would lead to excessive overlap.
Is this start site conserved in other phage genomes as indicated by Starterator?	This gene does contain the consensus start site, being called in 113 of 383 non-draft genes. This start site is at 27483 bp.
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:	Start site should be at 27483 bp because: • Glimmer says • Second highest z-score and final RBS score • Consensus start site on starterator

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: 0.0; exonuclease
with a protein having a	NCBI: 0.0; exonuclease
functional assignment in	HHPred: 3.2e-19; Restriction endonuclease-like
BlastP (phagesDB	
and/or GenBank) with an	

alignment of 10 ⁻⁴ or smaller with appropriate	
coverage?	
Does this protein align	NCBI: 97.46% identical; exonuclease
with a protein having a	HHPred: 99.87 probability; Restriction endonuclease-like
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	No
transmembrane protein?	
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Exonuclease

Basic Phage Information	
Phage Name	Acosta
Gene #	40
Stop Coordinate	27470
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-13
Selected Start Coordinate	29335
Selected Function	DNA Polymerase

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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	(Start: 277 @29335 has 129 MA's •Start number 277 was manually annotated 95 times for cluster EA1. • Found in 190 of 1729 (11.0%) of genes in pham • Manual Annotations of this start: 129 of 1532 • Called 78.4% of time when present
Does the gene violate any	There is no significant violations: no overlap, it is long enough,
major guiding principles?	and the surrounding genes are also reverse
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Glimmer Start Coordinate:29335 bp
and GeneMark suggest?	GeneMark Start Coordinate:29326
Does the start site have an	Final: -2.664
associated Ribosome	Z: 2.885
Binding Site with a high score?	This was the highest Z score
Is the predicted start codon the longest ORF? If not,	ORF is 1866 bp long, with a -13 bp gap.
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	• Found in 190 of 1729 (11.0%) of genes in pham
in other phage genomes	Manual Annotations of this start: 129 of 1532
as indicated by Starterator?	Called 78.4% of time when present
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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.
	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.
	The selected start site is at 29335 bp
DECISION:	 Has the highest Z score
	 Has more manual annotation then the other start site

Gathering Evidence	Explain Your Rationale
Does this protein align with	List the most informative BlastP match from each source
a protein having a	PhagesDB: 0.0 e value
functional assignment in	NCBI: 0.0 e value
BlastP (phagesDB and/or	HHPred: 4.6e-66 e value
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	HHPred: DNA polymerase with a 100%
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	Yes
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, Coiled coil segment
transmembrane protein?	
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	HHPred: 100% was called for DNA Polymerase

Basic Phage Information	
Phage Name	Acosta
Gene #	29535
Stop Coordinate	29535
Direction (For/Rev)	Rev
Gap (Overlap) with Previous Gene	20
Selected Start Coordinate	29966
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program (Glimmer, GeneMark)?	Both Glimmer and GeneMark indicate 29966as the Start.
Is there evidence for coding potential?	GeneMark shows coding potential
Is this gene present in other annotated genomes?	104 annotated phages from cluster EA1 have observed this gene.
Does the gene violate any major guiding principles?	No major violations
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Both indicate a start of bp@ 29966
and GeneMark suggest?	
Does the start site have an	Highest RBS Score: -2.443
associated Ribosome	Z value: 2.959
Binding Site with a high	These are the highest scores
score?	
Is the predicted start codon	ORF is 432 bp long, and this causes no overlap. This is not the
the longest ORF? If not,	longest ORF.
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	This is the appropriate start site, called in 400 of 450 approteted
Is this start site conserved	This is the consensus start site, called in 133 of 158 annotated
in other phage genomes	phages.
as indicated by Starterator?	
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	is the sequence you are analyzing and S (subject) is the
as indicated by Blasti :	database match. List the e-value and alignment of the best
	match for all three BlastP sources.
	mater for all till of Black Courses.
	Note: if you are considering more than 1 start site, provide the
	same information for each proposed start site.
	I think it should start at 29966
DECISION:	- Both Glimmer and GeneMark indicate this
DECISION:	 Highest z and final rbs scores
	- Consensus start site

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	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
DECISION:	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	Acosta
Gene #	30141
Stop Coordinate	30141
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	Gap
Selected Start Coordinate	31547
Selected Function	DNA Helicase

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?	Yes, it does show some coding potenial.
Is this gene present in other annotated genomes?	It does not to be in the same pham track
Does the gene violate any major guiding principles?	No it does not seem to violate any guiding principles.
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	Both say 31547
suggest?	
Does the start site have	RBS score: -3.199
an associated Ribosome	Z-Score: 2.608
Binding Site with a high	Both are the best scores shown.
score?	
Is the predicted start	ORF Length: 1407
codon the longest ORF?	It was not the longest ORF results and yes it does have
If not, does the longest	excessive overlap.
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	No, it does not seem to be conserved in other phage
conserved in other	genomes.
phage genomes as	
indicated by Starterator?	
Is this start site	No, it does not seem to be conserved in other phage
conserved in other	genomes.
phage genomes as	
indicated by BlastP?	
DECISION:	31547 Glimmer, Genemark, and phagesdb all call it.

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: several e values of 0.0 that say DNA Helicase
with a protein having a	NCBI: several e values of 0.0 thay say DNA Helicase

functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10 ⁻⁴ or smaller with appropriate coverage?	HHpred: 7.9e-39 that says DNA Binding protein
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	HHPred: 100% probability that says DNA Binding protein Ncbi: 99% probabilty that says DNA 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?	If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.
Is this gene a possible transmembrane protein?	NO
Is the proposed function found on the SEA-PHAGES approved function list?	YES DNA Helicase

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	31541 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-6 bp
Selected Start Coordinate	31657 bp
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)?	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?	GenemarkS and Genemark-host show coding potential.
Is this gene present in other annotated genomes?	6 annotated phages in cluster EA1 observed the same gene.
Does the gene violate any major guiding principles?	The only real violation is that the ORF is below 120 bp long. The genes surrounding are also reverse, and there is no significant overlap.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale

What start site do Glimmer and GeneMark suggest?	They both suggest 31657 bp.
Does the start site have an associated Ribosome Binding Site with a high score? Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive	Final RBS score: -3.649 Z-score: 2.464 Highest z-score and Final RBS score The ORF is 117 bp long, with an overlap of 7 bp.
gene overlap (>30bp)? Is this start site conserved in other phage genomes as indicated by Starterator? Is this start site conserved in other phage genomes as indicated by BlastP?	This is the consensus start site, which isn't too helpful as only 6 annotated phages have observed this gene, but it is notable that all 6 of them agreed on this start site. 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:	Start site should be at 31657 bp because: Glimmer and Genemark agree Highest z and final rbs scores Consensus start site The only other 2 options would have had ORF's of 87 bp and 75 bp length, so this is only choice

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: several e values over 0 that say function unkown NCBI: e value of 3e-12 that says hypothetical protien HHpred: several e values over 0

Does this protein align with a protein having a	HHPred: 92.22% probability that says unkown Ncbi: 100% probability that says hypothetica protien
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	NO
transmembrane protein?	
Is the proposed function	NO
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	Hypothetical protien

Basic Phage Information	
Phage Name	Acosta
Gene #	31654
Stop Coordinate	31654 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-4 bp
Selected Start Coordinate	31857 bp
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)?	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?	GenemarkS and Genemark-host show coding potential
Is this gene present in other annotated genomes?	This gene was observed in 12 annotated phages of cluster EA1.
Does the gene violate any major guiding principles?	There is no significant overlap, the ORF is long enough, and the surrounding genes are also reversed.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
Gathering Evidence	Explain Four Nationale

What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both suggest 31857 bp as start.
Does the start site have an associated Ribosome Binding Site with a high score? Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Final RBS score: -6.527 Z-score: 1.128 These scores are very low, and of the 4 potential start sites, these are the second lowest scores. ORF is 204 bp long, and there is a 4 bp overlap with the upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the consensus start site, which doesn't mean much when there's only 12 other phages that have observed this gene. All 12 of those share this site, though.
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:	The start site should be at 31857 bp because: • Glimmer and Genemark both say • Consensus 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:
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	<u>SEA-PHAGES Official Function List</u> 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.
	category.

Basic Phage Information		
Phage Name	Acosta	
Gene #	31854	
Stop Coordinate	31854	
Direction (For/Rev)	Reverse	
Gap (Overlap) with Previous Gene	Overlap	
Selected Start Coordinate	32150	
Selected Function	Hypothetical Protein	

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?	Yes, it does seem to have some coding potenial.
Is this gene present in other annotated genomes?	No, it does seem to be presented in other annotated genomes.
Does the gene violate any major guiding principles?	It does not seem to be violating any major guidelines.
DECISION:	Respond here with YES or NO after reviewing the evidence gathered above.

Gathering Evidence	Explain Your Rationale
What start site do	Both call 31250
Glimmer and GeneMark	Both dan 01200
suggest?	
Does the start site have	
an associated Ribosome	RBS score2.523
Binding Site with a high	Z-score- 2.959
score?	Both are the best scores shown
Is the predicted start	ORF= 297
codon the longest ORF?	It's the longest on and has does not result in a gene
If not, does the longest	overlap.
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	Yes, the start site is conserved in other phage genomes.
conserved in other	
phage genomes as	
indicated by Starterator?	
Is this start site	Yes, it does seem to be conserved in other genomes.
conserved in other	
phage genomes as	
indicated by BlastP?	
DECISION:	PhagesDB, Glimmer, GeneMark, and starterator all can
DEGIGIOI4.	the same thing.

Gathering Evidence	Explain Your Rationale

List the most informative BlastP match from each source	
PhagesDB:	
NCBI:	
DNA Master:	
Hint: you may have already found this information from	
annotation decision #2. Provide the alignment (q#:s#) and	
e-value. It is only necessary to provide one match from	
each database.	
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 III Dred motels	
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.	

Basic Phage Information		
Phage Name	Acosta	
Gene #		
Stop Coordinate	33262 bp	
Direction (For/Rev)	Reversed	
Gap (Overlap) with Previous Gene	-4	
Selected Start Coordinate	33795 bp	
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)?	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?	GeneMarkS and GeneMark-host both show coding potential.
Is this gene present in other annotated genomes?	There are no annotated genomes with this observed gene, only 2 other cluster EA1 draft genomes.
Does the gene violate any major guiding principles?	No significant violations, it is long enough, the surrounding genes are also reversed, and there is no significant overlap
DECISION:	Yes

Gathering Evidence	Explain Your Rationale	
What start site do	Glimmer: 33795 bp	
Glimmer and GeneMark	Genemark: 33936 bp	
suggest?		
Does the start site have	33795 bp	
an associated Ribosome	Final RBS: -2.443	
Binding Site with a high	Z-score: 2.959	
score?	Best scores	
	33936 bp:	
	Final RBS: -6.887	
	Z-score: .895	
	2 nd lowest scores	
Is the predicted start	For 33705, the ORF is 534 bp long, and for 33936 bp, the	
codon the longest ORF?	ORF is 675 bp long. Neither result in excessive gene	
If not, does the longest	overlap.	
ORF result in excessive		
gene overlap (>30bp)?		
Is this start site	No annotated genomes has this gene on starterator.	
conserved in other		
phage genomes as indicated by Starterator?		
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.	
	The start site should be at 33795 bp because:	
DECISION:	Glimmer says so Boot = coord DBS final coord	
	Best z-score and RBS final score The Grant Score The Grant Score and RBS final score The Grant Score and RBS final sco	
	The Genemark one isn't a very good option.	

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

Basic Phage Information		
Phage Name	Acosta	
Gene #	33953	
Stop Coordinate	33953	
Direction (For/Rev)	Reverse	
Gap (Overlap) with Previous Gene	Gap	
Selected Start Coordinate	34534	
Selected Function	thymidylate kinase	

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?	Yes, there is some evidence of coding potenial.
Is this gene present in other annotated genomes?	No, it not in other annotated genomes.
Does the gene violate any major guiding principles?	It may have a bigger gap than allowed.
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	Both call 34534 as the startsite
suggest?	
Does the start site have	RBS Score2.253
an associated Ribosome	Z Score- 3.112
Binding Site with a high	They are both the best scores.
score?	
Is the predicted start	ORF- 582
codon the longest ORF?	Yes, it does have a gap that does not violate any
If not, does the longest	guidelines
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	
conserved in other	No, it does not seem to be.
phage genomes as	
indicated by Starterator?	
Is this start site	No, it does not seem to be conserved in other genomes.
conserved in other	
phage genomes as	
indicated by BlastP?	
DECISION:	34534 Glimmer, GeneMark, PhagesDB, and Starterator all
DECISION.	call the same.

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: e-100; thymidylate kinase
with a protein having a	NCBI: 4e-127; thymidylate kinase
functional assignment in	HHpred: 2.1e-15; thymidylate kinase
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	NCBI: no above 90% scores
with a protein having a	HHPred: 99.73 probability; Thymidylate kinase
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
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	Phamerator) and record the function of the genes found
that shows high	on each side of the gene in the same pham in the most
conservation of gene	closely related phage. If the answer is NO, enter No
order?	Synteny Observed.
Is this gene a possible	no
transmembrane protein?	
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	thymidylate kinase

Basic Phage Information	
Phage Name	Acosta
Gene #	50
Stop Coordinate	34550
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	16
Selected Start Coordinate	35485
Selected Function	glycosyltransferase

Gathering Evidence	Explain Your Rationale
Was the gene called by an	Examine the data for the auto-annotation programs and indicate
auto-annotation program (Glimmer, GeneMark)?	(YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
Is there evidence for coding potential?	GeneMark S and Genemark-host both show coding potential
Is this gene present in other annotated genomes?	It is not annotated in any other genomes for cluster EA1
Does the gene violate any	There are no significant violations. No significant overlap, gene
major guiding principles?	is long enough, and surrounding genes are reverse.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	They both suggest 35485 bp as the start site
Does the start site have an	RBS: -3.837
associated Ribosome	Z score: 2.474
Binding Site with a high	This was the highest Z score
score?	
Is the predicted start codon	The ORF length is 936 bp with a gap of 16. This was not the
the longest ORF? If not,	longest ORF but the longest had too much of an overlap.
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	• Found in 2 of 175 (1.1%) of genes in pham
in other phage genomes	Manual Annotations of this start: 1 of 146
as indicated by Starterator?	Called 100.0% of time when present
Is this start site conserved in other phage genomes	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	is the sequence you are analyzing and S (subject) is the
as maisaisa sy Biasti .	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:	The start site should be at 35485 bp because it was called
DECISION.	100% of the time and its Z score and ORF length were good.

Gathering Evidence	Explain Your Rationale
Does this protein align with	PhagesDB: an e value of e-177 that says glycosyltransferase
a protein having a	NCBI: an e vlaue of 0 that says glycosyltransferase
functional assignment in	HHPred: e value of 3.7e-7 that says glycosyltransferase
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	HHPred: 98.9% probabilty match with glycosyltransferase
a protein having a	NCBI: 100% probabilty match with glycosyltransferase
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	different transmembrane prediction programs.
Is the proposed function	YES
found on the SEA-	
PHAGES approved	
function list?	
DECISION:	glycosyltransferase

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	35482 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-4 bp
Selected Start Coordinate	35700 bp
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)?	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?	Both Genemark-S and Genemark-host show coding potential
Is this gene present in other annotated genomes?	88 other annotated phages of Cluster EA1 have observed this gene.
Does the gene violate any major guiding principles?	No significant violations of any guiding principles.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say 35700 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	There's only one start site, so best scores.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 219 bp long. No significant gap with upstream gene.
Is this start site conserved in other phage genomes as indicated by Starterator?	This start site is the consensus start site, with 107 of 128 annotated phages choosing this start site for this gene.
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:	The start site should be at 35700 bp: • Glimmer and Genemark agree • Only option for start site • Consensus start site on starterator

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	<u>SEA-PHAGES Official Function List</u> 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.
	W F

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	35763 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	64 bp
Selected Start Coordinate	36041 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	104 annotated phages of cluster EA1 have observed this gene.
Does the gene violate any major guiding principles?	There are no significant violations.
DECISION:	yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 36041 bp.
Does the start site have an associated Ribosome	RBS final: -2.253 Z: 3.112
Binding Site with a high score?	Highest scores
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF is 279 bp long, and there is no significant overlap with this start site.
Is this start site	This is the consensus start site on starterator, chosen in
conserved in other phage genomes as indicated by Starterator?	104 of 126 annotated phages.
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:	The start site should be at 36041 bp because: • Glimmer and Genemark say • Best scores • Consensus 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:
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.

Basic Phage Information	
Phage Name	Acosta
Gene #	36038
Stop Coordinate	36038
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	Gap
Selected Start Coordinate	36805
Selected Function	thymidylate synthase

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?	Yes, it does seem to have some coding potential.
Is this gene present in other annotated genomes?	It does seem to be present in other annotated genomes.
Does the gene violate any major guiding principles?	No it does not violate any guidelines.
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 36805
Glimmer and GeneMark	GeneMark Start Coordinate 36883
suggest?	
Does the start site have	36805
an associated Ribosome	RBS: -2.443
Binding Site with a high	Z-score: 2.959
score?	36883
	RBS: -7.044
	Z-Score: 1.125
	36805 is the highest scoring
Is the predicted start	36805 ORF- 768
codon the longest ORF?	36883 ORF- 846
If not, does the longest	Either are the longest
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	You will also need to provide the following information
conserved in other	from Starterator: does the start match the consensus start
phage genomes as	site predicted from Starterator? If no, is the consensus
indicated by Starterator?	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.
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#),
phage genomes as	where 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:	36805:
	Glimmer say so

Gathering Evidence	Explain Your Rationale
Does this protein align	PhagesDB: e-146; thymidylate synthase
with a protein having a	NCBI: 0.0; thymidylate synthase
functional assignment in	HHpred: 5.3e-39; CMP hydroxymethylase
BlastP (phagesDB	
and/or GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	NCBI: 97.25% identical: Thymidylate synthase
with a protein having a	HHPred: several 100 probabilities; some "CMP
functional assignment in	hydroxymethylase" and some "thymidylate synthase"
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	Phamerator) and record the function of the genes found
that shows high	on each side of the gene in the same pham in the most
conservation of gene	closely related phage. If the answer is NO, enter No
order?	Synteny Observed.
Is this gene a possible	No
transmembrane protein?	
Is the proposed function	yes
found on the SEA-	

PHAGES approved function list?	
DECISION:	thymidylate synthase

Basic Phage Information	
Phage Name	Acosta
Gene #	54
Stop Coordinate	36935
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	360
Selected Start Coordinate	37147
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an	Examine the data for the auto-annotation programs and indicate
auto-annotation program	(YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
(Glimmer, GeneMark)?	(120 Botti, 120 ciminal city, 120 continuit city, tradition)
Is there evidence for	GeneMark S and Genemark-host both show coding potential
coding potential?	Generalik e and Generialik-nest beth show coding potential
Is this gene present in	Start number 2 was manually annotated 62 times for cluster
other annotated genomes?	EA1.
Does the gene violate any	The only real violation is that overlap is over 30 bp. The genes
major guiding principles?	surrounding are also reverse, and the gene is long enough
DECISION:	YES

Gathering Evidence	Explain Your Rationale	
What start site do Glimmer	They both suggest 37147 bp as the start site however start	
and GeneMark suggest?	37165 is a better choice.	
Does the start site have an	RBS: -6.152	
associated Ribosome	Z score: 1.399	
Binding Site with a high	This was the highest Z score	
score?		
Is the predicted start codon	The ORF length is 231 bp with a gap of 360. This was the	
the longest ORF? If not,	longest ORF length.	
does the longest ORF		
result in excessive gene		
overlap (>30bp)?		
Is this start site conserved	• Found in 101 of 127 (79.5%) of genes in pham	
in other phage genomes	Manual Annotations of this start: 62 of 110 Online 1.01 40/4 of this start are and the start and the start are a start are a start and the start are a start are a start and the start are a start are a start and the start are a start are a start a	
as indicated by	Called 61.4% of time when present	
Starterator?	Dravida the heat Bloot Branch from NCBI DhagaaDB and DNA	
Is this start site conserved in other phage genomes	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query)	
as indicated by BlastP?	is the sequence you are analyzing and S (subject) is the	
do maiodica by Blasti !	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.	
	The start site should be at 37165 bp.	
	Has the longest ORF length	
DECISION:	Highest Z score	
	Has more manual annotations	
	 Is called 61.4 % of the time 	

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
	If you halfang this was a should be seed to be a seed of the
	If you believe this gene should be assigned, please write the
DECISION:	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	Acosta
Gene #	55
Stop Coordinate	37224
Direction (For/Rev)	Rev
Gap (Overlap) with Previous Gene	77
Selected Start Coordinate	37463
Selected Function	Hypothetical Protein

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?	Yes
Is this gene present in other annotated genomes?	Annotated104 times in cluster EA1
Does the gene violate any major guiding principles?	No major violations
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Both suggest bp @ 37463
and GeneMark suggest?	
Does the start site have an	RBS Highest Score: -2.443
associated Ribosome	Z value: 2.959
Binding Site with a high	
score?	
Is the predicted start codon	ORF is 240 bp long, this is the longest ORF, and this doesn't
the longest ORF? If not,	result in excessive overlap.
does the longest ORF	·
result in excessive gene	
overlap (>30bp)?	
Is this start site conserved	This phage doesn't have the consensus start site.
in other phage genomes	
as indicated by	
Starterator?	
Is this start site conserved	Provide the best BlastP match from NCBI, PhagesDB, and DNA
in other phage genomes	Master with alignment in the format of (Q#:S#), where Q (query)
as indicated by BlastP?	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.
	Start should be at 37463
DECISION:	- Both Glimmer and GeneMark indicate this

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
	If you believe this gone about the construct of places with the
	If you believe this gene should be assigned, please write the
DECISION:	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	Acosta
Gene #	
Stop Coordinate	37555 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	93 bp
Selected Start Coordinate	37827 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	101 annotated phages from cluster EA1 have observed this gene.
Does the gene violate any major guiding principles?	No significant violations.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
5	

What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both say 37827 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	RBS final: -2.253 z-score: 3.047 Best Scores
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF with this start site is 273 bp long. There is no significant overlap with this start site.
Is this start site conserved in other phage genomes as indicated by Starterator?	This is the consensus start site, called in 110 of 128 non-draft genes.
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:	The start site should be at 37827 bp: • Glimmer and Genemark • Scores are the highest • Consensus 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:
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.

Basic Phage Information	
Phage Name	Acosta
Gene #	57
Stop Coordinate	37897
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	70
Selected Start Coordinate	38163
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an	Examine the data for the auto-annotation programs and indicate
auto-annotation program	(YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
(Glimmer, GeneMark)?	(120 20 11), 120 ciminor ciny, 120 contains ciny, 110 cinor)
Is there evidence for	GeneMark S and Genemark-host both show coding potential
coding potential?	Concinant o and Concinant host both show coding potential
Is this gene present in	•Start number 8 was manually annotated 95 times for cluster
other annotated genomes?	EA1.
Does the gene violate any	The only real violation is that overlap is over 30 bp. The genes
major guiding principles?	surrounding are also reverse, and the gene is long enough
DECISION:	YES

Gathering Evidence	Explain Your Rationale	
What start site do Glimmer and GeneMark suggest?	They both suggest 38163 bp as the start site	
Does the start site have an	RBS: -2.584	
associated Ribosome	Z score: 2.959	
Binding Site with a high score?	This was the highest Z score	
Is the predicted start codon	The ORF length is 267 bp with a gap of 70. This was the longest	
the longest ORF? If not,	ORF length.	
does the longest ORF result in excessive gene		
overlap (>30bp)?		
Is this start site conserved	• Found in 114 of 114 (100.0%) of genes in pham	
in other phage genomes	Manual Annotations of this start: 96 of 96	
as indicated by	Called 100.0% of time when present	
Starterator?	Duranida tha back Black Buratak frans NOBL Blacks B. and BNA	
Is this start site conserved in other phage genomes	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query)	
as indicated by BlastP?	is the sequence you are analyzing and S (subject) is the	
as maisured by Diasti .	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.	
	The start site should be at 38163 bp.	
DECISION:	Has the longest ORF lengthHighest Z score	
DECISION.	Has more manual annotations	
	Is called 100.0 % of the time	

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	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
DECISION:	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	Acosta
Gene #	38220
Stop Coordinate	38220
Direction (For/Rev)	reverse
Gap (Overlap) with Previous Gene	Gap
Selected Start Coordinate	38714
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)?	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?	Yes, there seems to be some coding potential.
Is this gene present in other annotated genomes?	Yes, it does seem to be present in other annotated genomes.
Does the gene violate any major guiding principles?	It might have too big of a gap between it and the earlier gene.
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	Both call 38714
suggest?	
Does the start site have	RBS4.271
an associated Ribosome	Z-Score- 2.110
Binding Site with a high	They are not the best scores shown.
score?	
Is the predicted start	ORF- 495
codon the longest ORF?	It has a large gap between it and the previous gene.
If not, does the longest	
ORF result in excessive	
gene overlap (>30bp)?	
Is this start site	No, it does not seem to be conserved in other genomes.
conserved in other	
phage genomes as	
indicated by Starterator?	
Is this start site	No, it does not seem to be conserved in other phage
conserved in other	genomes.
phage genomes as	
indicated by BlastP?	
DECISION:	Glimmer, genemark, and phagesdb all call 38714

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/ar Can Davidsth	
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
functional assignment in	necessary 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
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	Phamerator) and record the function of the genes found
that shows high	on each side of the gene in the same pham in the most
conservation of gene	closely related phage. If the answer is NO, enter No
order?	Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at
· ·	
transmembrane protein?	least 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.

Basic Phage Information	
Phage Name	Acosta
Gene #	59
Stop Coordinate	38711
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	38890
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program	Examine the data for the auto-annotation programs and indicate (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
(Glimmer, GeneMark)?	(120 BOTTI, 120 Cililinial City, 120 Concident City, Notifici
Is there evidence for coding potential?	Both Genemark-S and Genemark-host show coding potential
Is this gene present in	•Start number 20 was manually annotated 2 times for cluster
other annotated genomes?	EA1.
Does the gene violate any major guiding principles?	No significant violations of any guiding principles.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say 38890 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	RBS Score: -5.783 Z Score: 1.437
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF length is 180 bp
Is this start site conserved in other phage genomes as indicated by	 Found in 15 of 141 (10.6%) of genes in pham Manual Annotations of this start: 11 of 119 Called 100.0% of time when present
Starterator? 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
DECISION:	same information for each proposed start site. The start site is at 38890 bp • Was called 100% of the time

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

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	38892 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	3 bp
Selected Start Coordinate	39620 bp
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)?	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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	This gene was observed in 104 annotated phages from cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations of any of the major guiding principles.
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark both say start site at 39620 bp.
Does the start site have an associated Ribosome Binding Site with a high score? Is the predicted start codon the longest ORF?	Z-score: 3.047 Final RBS: -2.253 Highest scores available ORF is 729 bp long, and there would be no significant overlap with this start site.
If not, does the longest ORF result in excessive gene overlap (>30bp)?	
Is this start site conserved in other phage genomes as indicated by Starterator?	The consensus start site isn't found in this ORF, but this start site was chosen in 156 of the 426 annotated genomes (the consensus has 218 of 426).
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:	 The start site should be at 39620 bp because: Glimmer and genemark agree Highest z score and final RBS score Not consensus, but start site still has many manual annotations

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
13	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.

Basic Phage Information	
Phage Name	Acosta
Gene #	39716
Stop Coordinate	39716
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	Gap
Selected Start Coordinate	40051
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)?	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?	Yes, it does show coding potential.
Is this gene present in other annotated genomes?	It does not seem to be present in other genomes.
Does the gene violate any major guiding principles?	No, it does not seem to violate any guidelines.
DECISION:	Respond here with YES or NO after reviewing the evidence gathered above.

Gathering Evidence	Explain Your Rationale
What start site do	
Glimmer and	Both call 40051
GeneMark suggest?	
Does the start site	RBS2.253
have an associated	Z-Score- 3.047
Ribosome Binding Site	Both are the best scores shown.
with a high score?	
Is the predicted start	ORF- 336
codon the longest	The longest score was 918 and it results in major gene
ORF? If not, does the	overlap.
longest ORF result in	
excessive gene	
overlap (>30bp)?	
Is this start site	No, the start site does not seem to be conserved in other
conserved in other	genomes.
phage genomes as	
indicated by	
Starterator?	
Is this start site	No, the site does not seem to be conserved in other
conserved in other	genomes.
phage genomes as	
indicated by BlastP?	
DECISION:	40051 Glimmer, Phagesdb, and startaretor all call it.

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

	NODI
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
functional assignment in	necessary 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
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	Phamerator) and record the function of the genes found
that shows high	on each side of the gene in the same pham in the most
conservation of gene	closely related phage. If the answer is NO, enter No
order?	Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at
transmembrane protein?	least 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.

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	40096 bp
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	46 bp
Selected Start Coordinate	40437 bp
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)?	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?	GenemarkS and Genemark-host show coding potential.
Is this gene present in other annotated genomes?	This gene has been observed in 104 annotated phages of cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations of the major guiding principles.
DECISION:	Yes

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

What start site do Glimmer and GeneMark suggest?	Glimmer: 40437 bp Genemark: 40455 bp
Does the start site have an associated Ribosome Binding Site with a high score?	40437: Z: 1.981 Final RBS: -4.548 Average scores 40455: Z: 2.002 Final RBS: -4.644 Average scores
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	40437: 342 bp 40455: 360 bp Neither result in excessive overlap
Is this start site conserved in other phage genomes as indicated by Starterator?	40455 bp is the consensus start site, having 134 of 138 non-draft genes choosing this start. 40437 bp only has 2 annotated genomes with this 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
DECISION:	the same information for each proposed start site. The start site should be at 40455 bp because: Genemark says this site The z-score and final RBS score isn't the lowest This is the consensus 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:
functional assignment in	NCBI:
BlastP (phagesDB	DNA Master:

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

Basic Phage Information	
Phage Name	Acosta
Gene #	63
Stop Coordinate	40452
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	15
Selected Start Coordinate	40652
Selected Function	Hypothetical Protein

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program	Examine the data for the auto-annotation programs and indicate (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)
(Glimmer, GeneMark)? Is there evidence for coding potential?	Both Genemark-S and Genemark-host show coding potential
Is this gene present in other annotated genomes?	•Start number 17 was manually annotated 20 times for cluster EA1.
Does the gene violate any major guiding principles?	No significant violations of any guiding principles.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark say 40652 bp.
Does the start site have an associated Ribosome Binding Site with a high score?	RBS Score: -4.138 Z Score: 2.474
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF length is 174 bp. This was not the highest ORF length but the highest on had an overlap over 30bp.
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 124 of 124 (100.0%) of genes in pham Manual Annotations of this start: 20 of 104 Called 29.8% of time when present
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
DECISION:	same information for each proposed start site. The start site is at 40652 bp • Has the only manual annotations • Good orf length • Good overlap length

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	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
DECISION:	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	Acosta	
Gene #	64	
Stop Coordinate	41746	
Direction (For/Rev)	For	
Gap (Overlap) with Previous Gene	528	
Selected Start Coordinate	41180	
Selected Function	Hypothetical Protein	

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?	YES
Is this gene present in other annotated genomes?	Start 28 is annotated 1 time for cluster EA1
Does the gene violate any major guiding principles?	Overlap is over <30 bp, but there is no overlap
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Both indicate 41180
Does the start site have an associated Ribosome Binding Site with a high score?	Highest RBS Score: -4.820 Z value: 1.920
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF Length: 567 Not the longest and does result in excessive gene overlap
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 3 of 62 (4.8%) of genes in pham
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:	Start should be bp@ 41180 because - Glimmer and GeneMark indicate - No overlap

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

Basic Phage Information	
Phage Name	Acosta
Gene #	
Stop Coordinate	20593bp
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-60
Selected Start Coordinate	20180 bp
Selected Function	Hypothetical Protein

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?	Yes, there is coding potiental
Is this gene present in other annotated genomes?	Start number 3 was manually annotated 2 times for cluster EA1.
Does the gene violate any major guiding principles?	There are no significant violations, it is long enough, there is no overlap with the upstream gene however, the gene afterward goes in reverse direction.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 20180 bp
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS: -4.395 Z-score: 2.297 This was the highest Z score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	The ORF is 414 bp with a gap of -60 bp between the upstream gene. This was the highest ORF length
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 3 of 71 (4.2%) of genes in pham Manual Annotations of this start: 2 of 60 Called 100.0% of time when present
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:	I think the start site should be 20180 bp because: Glimmer and GeneMark agree Has the longest ORF length Has the highest Z score Called in 100% when present

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	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
DECISION:	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	Acosta
Gene #	30
Stop Coordinate	20903
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	58
Selected Start Coordinate	21073
Selected Function	Hypothetical Protein

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?	Yes, there is coding potential
Is this gene present in other annotated genomes?	Start number 13 was manually annotated 77 times for cluster EA1
Does the gene violate any major guiding principles?	There are no significant violations, it is long enough, there is no overlap with the upstream gene, and the genes around go in the same direction.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer	Glimmer and Genemark agree on 20903 bp however the best
and GeneMark suggest?	possible start site will be 21073 bp
Does the start site have an	21073 bp
associated Ribosome	Final RBS Score- (-3.692)
Binding Site with a high	Final Z Score- 2.417
score?	
	This was the highest Z score
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	21703 bp had an ORF length of 171 and this was the longest length.
Is this start site conserved	Found in 123 of 139 (88.5%) of genes in pham
in other phage genomes	Manual Annotations of this start: 77 of 119
as indicated by Starterator?	Called 62.6% of time when present
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:	I think the starts site should be 21073 bp because it has the highest Z score and ORF length. On Genemark it had the most coding potential.

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
	If you believe this gone about the construct of places with the
	If you believe this gene should be assigned, please write the
DECISION:	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	Acosta
Gene #	36
Stop Coordinate	24464
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	-24
Selected Start Coordinate	24757
Selected Function	Nuclease

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?	Yes, there is coding potential
Is this gene present in other annotated genomes?	Start number 36 was manually annotated 103 times for cluster EA1
Does the gene violate any major guiding principles?	There are no significant violations, it is long enough, there is no overlap with the upstream gene, and the genes around go in the same direction.
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and Genemark agree on 24757 bp
Does the start site have an associated Ribosome Binding Site with a high score?	Final RBS Score- (-6.742) Final Z Score- 1.028 This was the highest RBS score.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	24757 bp had an ORF length of 294 and this was not the longest length. The longest length had an overlap greater than 30 bp.
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 182 of 255 (71.4%) of genes in pham Manual Annotations of this start: 152 of 213 Called 95.6% of time when present
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
DECISION:	same information for each proposed start site. I think this should be the start site because it has the highest RBS score and both glimmer and genemark agree on this start. This start was also called in 95.6 % of the time when 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: 2e-51
functional assignment in	NCBI: 2e-65
BlastP (phagesDB and/or	DNA Master: 3.5e-19
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align	HHPred: Nuclease has 99.85% that matches.
with a protein having a	
functional assignment in	Note: If you believe there is not a quality HHPred match, type
the PDB or other database	No Quality Match and list the data for the best match available to
in HHPred with a	affirm the poor quality of the result and to document that HHPred
probability of 90% or	was considered.
greater with appropriate	
coverage?	
Is this gene located	Yes
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?	W
Is the proposed function	Yes
found on the SEA-	
PHAGES approved	
function list?	LILIBrad above 00 959/ match to Niveleges
DECISION:	HHPred shows 99.85% match to Nuclease

Basic Phage Information	
Phage Name	Acosta
Gene #	46
Stop Coordinate	32150
Direction (For/Rev)	Reverse
Gap (Overlap) with Previous Gene	0
Selected Start Coordinate	32413
Selected Function	Hypothetical Protein

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?	GenemarkS and Genemark-host both show coding potential.
Is this gene present in other annotated genomes?	•Start number 16 was manually annotated 7 times for cluster EA1.
Does the gene violate any	The only violation is that the ORF length is below 120bp. It is
major guiding principles?	long enough, and the surrounding genes are also reverse
DECISION:	YES

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	They both suggest start at 32413 bp
Does the start site have an	Final: -5.833
associated Ribosome	Z: 1.872
Binding Site with a high score?	This was not the highest Z score
Score?	
Is the predicted start codon	ORF is 39 bp long, with a 0 bp gap.
the longest ORF? If not,	
does the longest ORF	
result in excessive gene	
overlap (>30bp)?	F 1: 00 (440/4400/) (
Is this start site conserved	• Found in 22 of 148 (14.9%) of genes in pham
in other phage genomes	Manual Annotations of this start: 8 of 125
as indicated by Starterator?	Called 45.5% of time when present
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:	The selected start site is at 32413 bp
DECISION:	Has more manual annotation then the other start sites

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	List the most informative HHPred match, including database
a protein having a	source and probability score. It is only necessary to provide the
functional assignment in	best match.
the PDB or other database	
in HHPred with a	Note: If you believe there is not a quality HHPred match, type
probability of 90% or	No Quality Match and list the data for the best match available
greater with appropriate	to affirm the poor quality of the result and to document that
coverage?	HHPred was considered.
Is this gene located	If the answer is YES, evaluate the proposed function in the gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	NO, enter No Synteny Observed.
Is this gene a possible	If the answer is YES, indicate supporting data from at least 2
transmembrane protein?	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-</u>
PHAGES approved	PHAGES Official Function List to ensure that you are following
function list?	the guidelines for function naming. Functions that are not
	present on the approved list must be carefully vetted for
	approval.
	If you believe this gone about the construct of places with the
	If you believe this gene should be assigned, please write the
DECISION:	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	Acosta	
Gene #	47	
Stop Coordinate	33265	
Direction (For/Rev)	Rev	
Gap (Overlap) with Previous Gene	4	
Selected Start Coordinate	32417	
Selected Function	MazG-like nucleotide pyrophosphohydrolase	

Gathering Evidence	Explain Your Rationale
Was the gene called by an	
auto-annotation program	Both Glimmer and GeneMark
(Glimmer, GeneMark)?	
Is there evidence for	Coding potential in GeneMark
coding potential?	
Is this gene present in	104 times for cluster EA1
other annotated genomes?	104 times for cluster LAT
Does the gene violate any	No major violations. An overlap of 4
major guiding principles?	
DECISION:	Yes

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	Glimmer and GeneMark both indicate 32417
Does the start site have an associated Ribosome Binding Site with a high score?	Highest RBS Score: -7.128 Z score: 0.946
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	Overlap of 4 Longest ORF length: 42
Is this start site conserved in other phage genomes as indicated by Starterator?	Found in 125 of 161 (77.6%)
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:	Start should be at 32417 bp because - Both Glimmer and GeneMark Indicate this - No major overlap

Gathering Evidence	Explain Your Rationale
Does this protein align with	PhagesDB: e-121; MazG-like nucleotide pyrophosphohydrolase
a protein having a	NCBI: 1e-152; nucleotide pyrophosphohydrolase
functional assignment in	HHpred: 8.3e-31; putative NTP pyrophosphohydrolase
BlastP (phagesDB and/or	
GenBank) with an	
alignment of 10 ⁻⁴ or	
smaller with appropriate	
coverage?	
Does this protein align with	NCBI: no high % identical
a protein having a	HHPred: 99.97 probability; putative NTP pyrophosphohydrolase
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 gene
adjacent to genes of	order. Examine the adjacent genes found in the most closely
known function and in a	related annotated phage (hint: use Phamerator) and record the
region of the genome that	function of the genes found on each side of the gene in the
shows high conservation of	same pham in the most closely related phage. If the answer is
gene order?	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:	MazG-like nucleotide pyrophosphohydrolase