Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**1**
Stop Coordinate	**359**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Major Gap**
Selected Start Coordinate	**517**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer, bp 517 (strength of 5.74)*
Is there evidence for coding potential?	*Very little coding potential*
Is this gene present in other annotated genomes?	*No other genomes had the gene present.*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**3**
Stop Coordinate	**514**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Major Overlap**
Selected Start Coordinate	**2049**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer, bp 2049 (strength of 10.76)*
Is there evidence for coding potential?	*Very little to none coding potential*
Is this gene present in other annotated genomes?	*Yes – only in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**2**
Stop Coordinate	**625**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**395**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 395*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Didgeridoo and AvocadoMan both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 395*
Does the start site have an associated Ribosome Binding Site with a high score?	*395 – large z-score (3.029) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No it is not the longest ORF, and haven't evaluated the previous gene so we do not know if there is major overlap with the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It was found in 60 of 62 genes in the pham, and called 96.8% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 395, based off of Glimmer/Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical 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?	*No Quality Match* - Clp protease ClpC,Heat shock survival AAA family ATPase ClpK; ATPase associated with diverse cellular activities (AAA), 86.9 % Hypothetical protein
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, terminase small subunit*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, Hypothetical Protein*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**4**
Stop Coordinate	**948**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**625**
Selected Function	**terminase, small subunit**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 625*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyDaisy, BabyYoda, and Avocadoman all had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 625*
Does the start site have an associated Ribosome Binding Site with a high score?	*625 – small z-score (0.913) and appropriate spacer (16 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No it is not the longest ORF, and haven't evaluated the previous gene so we do not know if there is major overlap with the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It was found in 81 of 82 genes in the pham, and called 98.8% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 625, based off of Glimmer/Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, terminase small subunit*
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?	; Terminase_4 ; Phage terminase, small subunit – 95.04%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, terminase small subunit*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, terminase, small subunit*
DECISION:	*terminase, small subunit*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**5**
Stop Coordinate	**2577**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**952**
Selected Function	**terminase, large subunit**

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 bp 2044, Genemark bp 952*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, only in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 2044* *GeneMark Start Coordinate (type NA if not supported):: 952*
Does the start site have an associated Ribosome Binding Site with a high score?	*952 – large z-score (3.029) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (2044) was found in 2 of 2 genes in the pham, and called 100% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 952, based off of Genemark, Genemark coding potential, phamerator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, terminase large subunit*
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?	; TerL_nuclease ; Terminase large subunit, endonuclease domain – 99.02%
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, terminase, large subunit*
DECISION:	*terminase, large subunit*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**7**
Stop Coordinate	**2651**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**3493**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 4270 has strength 6.81 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*Yes – only in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**8**
Stop Coordinate	**3890**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**4084**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original glimmer cell @bp 4094 has strength 5.52 not called by gene mark
Is there evidence for coding potential?	GeneMark did not show coding potential
Is this gene present in other annotated genomes?	Yes- Loviatar Draft had this gene
Does the gene violate any major guiding principles?	The gene does not follow all guiding principles
DECISION:	No

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**9**
Stop Coordinate	**4133**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**4270**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original Glimmer call @bp 4270 has strength 6.81 not called by Genemark
Is there evidence for coding potential?	Genemark did not show coding potential
Is this gene present in other annotated genomes?	Yes-Loviatar_Draft had this gene
Does the gene violate any major guiding principles?	The gene does not follow all guiding principles
DECISION:	No

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**10**
Stop Coordinate	**4270**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**6201**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 6201 has strength 8.81 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*No, the gene is an orphan*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**6**
Stop Coordinate	**4603**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**2588**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 2588*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – both BabyYoda and Avocadoman had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 2588*
Does the start site have an associated Ribosome Binding Site with a high score?	*2588 – large z-score (3.029) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No it is not the longest ORF, and does not overlap with the longest ORF* *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?	*Yes, It was found in 81 of 234 genes in the pham, and called 34.6% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 2588, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Portal protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Portal protein; Archaeal virus, portal, portal capsid interface, Mg ions, VIRUS; HET: MG, HIP; 2.342A {Haloferax tailed -100%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, portal protein*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, portal protein*
DECISION:	*portal protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**11**
Stop Coordinate	**5358**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**4603**
Selected Function	**capsid maturation protease**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Genemark bp 4603*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and Avocadoman both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 4603*
Does the start site have an associated Ribosome Binding Site with a high score?	*4603 – large z-score (3.029) and appropriate spacer (11 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No it is not the longest ORF, and haven't evaluated the previous gene so we do not know if there is major overlap with the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It was found in 153 of 322 genes in the pham, and called 47.5% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 4603, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, capsid maturation protease*
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?	;Peptidase_S78 ; Caudovirus prohead serine protease *99.85%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*YES*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*YES*
DECISION:	capsid maturation protease

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**13**
Stop Coordinate	**6312**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**7052**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 7052 has strength 5.45 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*No, the gene is an orphan*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**12**
Stop Coordinate	**6579**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**5395**
Selected Function	**major capsid protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 5395*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda, and Avocadoman both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 5395*
Does the start site have an associated Ribosome Binding Site with a high score?	*5395 – large z-score (3.262) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (5395) was found in 184 of 348 genes in the pham, and called 52.9% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 5395, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Major capsid 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?	Scaffolding domain delta; Prohead I, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.5A {Escherichia phage HK97} 100%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*YES*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*YES*
DECISION:	*major capsid protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**14**
Stop Coordinate	**6977**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**6648**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 6648*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and Avocadoman both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 6648*
Does the start site have an associated Ribosome Binding Site with a high score?	*6648 – large z-score (3.029) and appropriate spacer (8 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (6648) was found in 60 of 60 genes in the pham, and called 100% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 6648, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*No, Hypothetical 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?	GH94N_like_1; Glycoside hydrolase family 94 N-terminal-like domain of uncharacterized function. *34.14%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**15**
Stop Coordinate	**7563**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**7012**
Selected Function	**head-to-tail adaptor**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 7012*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and Avocadoman both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 7012*
Does the start site have an associated Ribosome Binding Site with a high score?	*7012 – large z-score (3.120) and slightly large spacer (13 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (7012) was found in 82 of 240 genes in the pham, and called 34.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 7012, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, head-to-tail adaptor*
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?	Head completion protein; Neck, Portal, T5, VIRUS, VIRAL PROTEIN; 3.2A {Escherichia phage DT57C} 99.95%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*YES*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*head-to-tail adaptor*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**17**
Stop Coordinate	**7673**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**8554**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original Glimmer cell @bp 8554 has strength 5.40 not called by Genemark
Is there evidence for coding potential?	Genemark did not show coding potential
Is this gene present in other annotated genomes?	Yes- Loviatar_Draft had this gene
Does the gene violate any major guiding principles?	The gene does not follow all guiding prinicples
DECISION:	NO

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**16**
Stop Coordinate	**7892**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**7563**
Selected Function	**head-to-tail stopper**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 7563*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 7563*
Does the start site have an associated Ribosome Binding Site with a high score?	*7563 – large z-score (3.029) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No it is not the longest ORF, but it does not have any major overlap with other genes* *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?	*Yes, It was found in 122 of 156 genes in the pham, and called 78.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved.*
DECISION:	*Start site is bp 7563, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, head-to-tail stopper*
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?	Head-to-tail stopper; Bacteriophage, portal, VIRAL PROTEIN;{Mycobacterium phage Bxb1} 99.65%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*YES*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*head-to-tail stopper*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**18**
Stop Coordinate	**8216**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**7908**
Selected Function	**Hypothetical protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 7908*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles.*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 7908*
Does the start site have an associated Ribosome Binding Site with a high score?	*7908 – low z-score (1.145) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF, but does not overlap with other genes*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (7908) was found in 2 of 363 genes in the pham, and called 0.6% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 7908, based off of Genemark, Genemark coding potential, phamerator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical protein QDW21_gp10*
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?	; DUF5403 ; Family of unknown function (DUF5403) 94.99%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**19**
Stop Coordinate	**8611**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**8216**
Selected Function	**Tail terminator**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 8216*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 8216*
Does the start site have an associated Ribosome Binding Site with a high score?	*8216 – moderate z-score (2.776) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF, but the longest has no overlap with other genomes* *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?	*Yes, It (8216) was found in 77 of 419 genes in the pham, and called 18.4% of the time* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 8216, based off of Genemark, Genemark coding potential, phamerator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, tail terminator*
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?	*Tail terminator; Bacteriophage, portal, VIRAL PROTEIN;{Mycobacterium phage Bxb1} 99.79%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Tail terminator*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**21**
Stop Coordinate	**8640**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**9416**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original Glimmer all @bp 9416 has strength 10.05 not called by Genemark
Is there evidence for coding potential?	Genemark did not show coding potential
Is this gene present in other annotated genomes?	Yes- Lovitar_Draft had this gene
Does the gene violate any major guiding principles?	The gene does not follow all guiding principles
DECISION:	No

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**20**
Stop Coordinate	**9422**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**8625**
Selected Function	**Major tail protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 8625*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Abigail both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 8625*
Does the start site have an associated Ribosome Binding Site with a high score?	*8625 – large z-score (3.262) and appropriate spacer (11 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (8625) was found in 248 of 569 genes in the pham, and called 43.6% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 8625, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, major tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	Major tail protein; Bacteriophage, tail tube, VIRUS, VIRAL PROTEIN;{Mycobacterium phage Bxb1} 100%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*YES*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Major tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**22**
Stop Coordinate	**9826**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**9518**
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)?	*Yes Genemark, bp 9518*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 9518*
Does the start site have an associated Ribosome Binding Site with a high score?	*9518 – small z-score (1.953) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (9518) was found in 116 of 139 genes in the pham, and called 83.5% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 9518, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, tail assembly chaperone*
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?	; GP24_25 ; Mycobacteriophage tail assembly protein 97%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, known functions include tail assembly chaperones*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, tail assembly chaperone*
DECISION:	tail assembly chaperone

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**23**
Stop Coordinate	**9850**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**10074**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 10074*
Is there evidence for coding potential?	*GeneMarkS showed no coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**30**
Stop Coordinate	**12975**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**13718**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original Glimmer cell @bp 13718 has strength 8.47 not called by Genemark
Is there evidence for coding potential?	Genemark did not show coding potential
Is this gene present in other annotated genomes?	Yes-Loviatar_Draft had this gene
Does the gene violate any major guiding principles?	The gene does not follow all guiding principles
DECISION:	NO

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**28**
Stop Coordinate	**13050**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**10420**
Selected Function	**tape measure protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 10420*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 10420*
Does the start site have an associated Ribosome Binding Site with a high score?	*10420 – moderate z-score (2.178) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (10420) was found in 65 of 71 genes in the pham, and called 91.5% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 10420, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, tape measure protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	Tape measure protein (CD1366); Phage tail-like, bacteriocin, baseplate, pre-contraction, VIRUS LIKE PARTICLE;{Clostridio *99.51%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, tape measure protein*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, tape measure protein*
DECISION:	*tape measure protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**32**
Stop Coordinate	**13874**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**15751**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 15751 has strength 9.25 not called by genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes-Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**31**
Stop Coordinate	**13904**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**13047**
Selected Function	Minor Tail Protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 13047*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Abigail both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 13047*
Does the start site have an associated Ribosome Binding Site with a high score?	*13047 – moderate z-score (2.348) and appropriate spacer (12 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (13047) was found in 81 of 378 genes in the pham, and called 21.4% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 13047, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, minor tail protein* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	HYPOTHETICAL PROTEIN 19.1; VIRAL PROTEIN, DISTAL TAIL PROTEIN; 2.95A {BACILLUS PHAGE SPP1} 99.96% *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?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	Minor Tail Protein

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**34**
Stop Coordinate	**15808**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**15966**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 15966 has strength 3.41 not called by Genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes- Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding prinicples*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**35**
Stop Coordinate	**15942**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**19190**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer, bp 19190*
Is there evidence for coding potential?	*GeneMarkS showed no coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**33**
Stop Coordinate	**16030**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**13904**
Selected Function	**Minor tail protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 13904*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and DirtyBubble both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 13904*
Does the start site have an associated Ribosome Binding Site with a high score?	*13904 – low z-score (1.963) and appropriate spacer (12 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (13904) was found in 11 of 11 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 13904, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, minor tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	Minor tail protein; Bacteriophage, tail tip, VIRAL PROTEIN;{Mycobacterium phage Bxb1} 99.96% *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?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Minor Tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**36**
Stop Coordinate	**16880**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**16044**
Selected Function	**Minor Tail protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 16044*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Abigail both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 16044*
Does the start site have an associated Ribosome Binding Site with a high score?	*16044 – large z-score (3.273) and high spacer (13 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (16044) was found in 175 of 178 genes in the pham, and called 98.3% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 16044, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Minor tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	Receptor Binding Protein; beta sandwich domain, phage receptor binding protein, Lactococcus lactis pellicle cell wall polyphosphosaccharide, VIRAL PROTEIN; 1.75A {Lactococcus phage 1358}
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Minor tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akin08**
Gene #	**37**
Stop Coordinate	**17693**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	16890
Selected Function	**Hypothetical**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp* 16890
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Abigail both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: NA* *GeneMark Start Coordinate (type NA if not supported):: 16890*
Does the start site have an associated Ribosome Binding Site with a high score?	*16890 – moderate z-score (2.862) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (16890) was found in 118 of 221 genes in the pham, and called 91.5% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 16890, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Hypothetical Protein QDW21_gp20*
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?	*; DUF859 ; Siphovirus protein of unknown function (DUF859) 99.3%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**39**
Stop Coordinate	**19370**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**20533**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 20533 has strength 3.78 not called by Genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes-Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding prinicples*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**38**
Stop Coordinate	**19409**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**17703**
Selected Function	**Hypothetical**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer (bp 19248 strength of 8.83) and GeneMark (bp 17703)*
Is there evidence for coding potential?	*GeneMarkS showed coding potential at 17703*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 19248* *GeneMark Start Coordinate (type NA if not supported):: 17703*
Does the start site have an associated Ribosome Binding Site with a high score?	*17703 – large z-score (3.273) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (19248) was found in 2 of 2 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 17703, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data. We went with 17703 as the start position instead of 19248 because the coding potential is greatest at the start of 17703. Additionally, the coding potential is really consistent throughout the 18000's up until the end point. Starterator says 19248, however according to our previous gene, it would make more sense for the gene to start at 17703. Lastly, the RBS of the ORF showed the z value at 17703 to be the greatest, a good spacer distance, the lowest final score, and longest ORF length. For all of these reasons, we concluded gene 38 (19409) to start at 17703.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, endolysin*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	No quality match; DUF1949 ; Domain of unknown function (DUF1949) 64.14%
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**40**
Stop Coordinate	**19891**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**19409**
Selected Function	**Hypothetical Protein (Transmembrane?)**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 19409*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Abigail and Arroyo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::19409*
Does the start site have an associated Ribosome Binding Site with a high score?	*19409 – moderate z-score (2.549) and low spacer (6 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (19409) was found in 66 of 66 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 19409, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No quality match ; DUF6264 ; Family of unknown function (DUF6264) 76.8%*
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*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical but includes a transmembrane segment*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**41**
Stop Coordinate	**20243**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**19902**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 19902*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 19902*
Does the start site have an associated Ribosome Binding Site with a high score?	*19902 – moderate z-score (2.708) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, not the longest ORF but does not overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (19902) was found in 87 of 224 genes in the pham, and called 38.8 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 19902, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, membrane 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?	; MTLN ; Mitoregulin probability 95.82
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**42**
Stop Coordinate	**20497**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**20246**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 20246*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 20246*
Does the start site have an associated Ribosome Binding Site with a high score?	*20246 – moderate z-score (2.581) and low spacer (7 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (20246) was found in 55 of 55 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 20246, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, membrane 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?	; Halogen_Hydrol ; 5-bromo-4-chloroindolyl phosphate hydrolysis protein *Probability 73.12*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**43**
Stop Coordinate	**20645**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**21139**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 21139 has strength 5.78 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*Yes – only found in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**44**
Stop Coordinate	**21310**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**20708**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer (bp 21113 strength of 5.76) and GeneMark (bp 20708)*
Is there evidence for coding potential?	*GeneMarkS showed coding potential at 20708*
Is this gene present in other annotated genomes?	*Yes – AvocadoMan and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 21113* *GeneMark Start Coordinate (type NA if not supported):: 20708*
Does the start site have an associated Ribosome Binding Site with a high score?	*20708 – small z-score (1.979) and low spacer (7 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No It is the longest ORF, but the gene does not overlap with the previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (21113) was found in 2 of 2 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 20708, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data. We went with 20708 as the start position instead of 21113 because the coding potential is greatest at the start of 20708. Additionally, the coding potential is really consistent throughout the 21000's up until the end point. Starterator says 21113, however according to our previous gene, it would make more sense for the gene to start at 20708.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical*
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?	Met_tRNA_FMT_C; C-terminal domain of Formyltransferase and other enzymes. C-terminal domain of formyl transferase and ot Probability 57.02
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**46**
Stop Coordinate	**22001**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**21321**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 21321*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Arroyo and Albedo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 21321*
Does the start site have an associated Ribosome Binding Site with a high score?	*21321 – moderate z-score (2.793) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (21321) was found in 59 of 64 genes in the pham, and called 92.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 21321, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical*
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?	; DUF6378 ; Domain of unknown function (DUF6378) *Probability 99.88*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**48**
Stop Coordinate	**22015**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**22500**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 2250 has strength 11.13 not called by Genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes-Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**47**
Stop Coordinate	**22144**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**3 bp overlap**
Selected Start Coordinate	**21998**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 21998*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Annalie and Arroyo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 21998*
Does the start site have an associated Ribosome Binding Site with a high score?	*21998 – moderate z-score (2.690) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (21998) was found in 46 of 46 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 21998, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**49**
Stop Coordinate	**22635**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**22144**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 22144 but glimmer said bp 22462 with a strength of 1.40*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Gene is not present in other annotated genomes*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 22462* *GeneMark Start Coordinate (type NA if not supported):: 22144*
Does the start site have an associated Ribosome Binding Site with a high score?	*22144 – low z-score (1.260) and appropriate spacer (11 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*No, the start site is not conserved in other phage genomes.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved*
DECISION:	*Start site is bp 22144, based off of Genemark, Genemark coding potential, phamerator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**51**
Stop Coordinate	**22638**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Gap**
Selected Start Coordinate	**23432**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 23432*
Is there evidence for coding potential?	*NO coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**52**
Stop Coordinate	**23417**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**23638**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 23638 has strength 6.52 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*Yes – only found in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**50**
Stop Coordinate	**23558**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**3 bp overlap**
Selected Start Coordinate	**22632**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 22632*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Annalie and Arroyo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 22632*
Does the start site have an associated Ribosome Binding Site with a high score?	*22632 – low z-score (1.968) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (22632) was found in 90 of 243 genes in the pham, and called 37.0% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 22632, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**54**
Stop Coordinate	**23632**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**24141**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 24141 has strength 7.67 not called by genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes-Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**53**
Stop Coordinate	**23893**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**23588**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 23588*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and SanaSana both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 23588*
Does the start site have an associated Ribosome Binding Site with a high score?	*23588 – low z-score (1.433) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (23588) was found in 6 of 6 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 23588, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical 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?	*DNA packaging protein Gp17; nucleotide-binding fold, HYDROLASE; HET: ADP; 1.8A {Enterobacteria phage T4} 99.35%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**55**
Stop Coordinate	**24090**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**23899**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 23899*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – WalkingDead_Draft and Laviator_Draft both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 23899*
Does the start site have an associated Ribosome Binding Site with a high score?	*23899 – moderate z-score (2.207) and low spacer (5 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (23899) was found in 5 of 9 genes in the pham, and called 55.6% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 23899, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical 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?	*; HTH_Tnp_Tc3_2_like ; Transposable element Tc3 transposase, HTH* *84.35%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**57**
Stop Coordinate	**24442**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Gap**
Selected Start Coordinate	**25224**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 25224*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**56**
Stop Coordinate	**24466**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**7 bp overlap**
Selected Start Coordinate	**24083**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 24083*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and SanaSana both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 24083*
Does the start site have an associated Ribosome Binding Site with a high score?	*24083 – low z-score (1.944) and appropriate spacer (11 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes has the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (24083) was found in 10 of 79 genes in the pham, and called 12.7% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 24083, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, hypothetical 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?	*c.52.1.35 (A:) stNUC {Salmonella phage SETP3 [TaxId: 424944]} | CLASS: Alpha and beta proteins (a/b), FOLD: Restriction* *84.94%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**58**
Stop Coordinate	**24624**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**No**
Selected Start Coordinate	**24821**
Selected Function	**toxin in toxin/antitoxin system, HicA-like**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 24821*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – only in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 24821*
Does the start site have an associated Ribosome Binding Site with a high score?	*24821 – moderate z-score (2.440) and low spacer (6 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (24821) was found in 32 of 197 genes in the pham, and called 16.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 24821, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, HicA-like toxin*
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?	*; HicA_toxin ; HicA toxin of bacterial toxin-antitoxin, 99.46%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*YES*
DECISION:	*toxin in toxin/antitoxin system, HicA-like*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**60**
Stop Coordinate	**25339**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**25482**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 25482 has strength 8.05 not called by Genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes-Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**61**
Stop Coordinate	**25440**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**26033**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 26033*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**59**
Stop Coordinate	**25521**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**24925**
Selected Function	**dUTPase**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 24925*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albright and Annalie both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 24925*
Does the start site have an associated Ribosome Binding Site with a high score?	*24925 – moderate z-score (2.744) and low spacer (7 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF but it is not overlapping with the previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (24925) was found in 77 of 85 genes in the pham, and called 90.6% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 24925, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, deoxyuridine triphosphatase*
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?	*dUTPase from DI S. aureus phage; Staphylococcus aureus, pathogenicity island, SaPI, dUTPases, signalling, gene transfer, 99.46%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*YES*
DECISION:	*dUTPase*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**63**
Stop Coordinate	**26021**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**26803**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 26803*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**62**
Stop Coordinate	**26024**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**25518**
Selected Function	**thymidylate kinase**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 25518*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 25518*
Does the start site have an associated Ribosome Binding Site with a high score?	*25518 – moderate z-score (2.532) and large spacer (12 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (23588) was found in 77 of 82 genes in the pham, and called 93.9% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 25518, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, thymidylate kinase*
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?	***Thymidylate kinase; Nucleotide monophosphate kinase, TRANSFERASE; HET: TMP; 1.19A {Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)} SCOP: c.37.1.0*** *99.78%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*thymidylate kinase*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**64**
Stop Coordinate	**26788**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**26039**
Selected Function	***Recombination Directionality factor***

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 26039*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Icarion and Sanasana both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 26039*
Does the start site have an associated Ribosome Binding Site with a high score?	*26039 – moderate z-score (2.132) and high spacer (15 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (26039) was found in 81 of 200 genes in the pham, and called 40.5% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 26039, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, recombination directionality factor*
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?	*; Gp3-like ; Recombination directionality factor-like 99.93%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Recombination Directionality factor*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**65**
Stop Coordinate	**26948**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**27604**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 27604 has a strength 5.64 not called by genemark*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**66**
Stop Coordinate	**27043**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**No**
Selected Start Coordinate	**27342**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 27342*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – found in Loviatar_Draft only*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::27342*
Does the start site have an associated Ribosome Binding Site with a high score?	*27342 – moderate z-score (2.062) and high spacer (13 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (27342) was found in 49 of 49 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 27342, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Hypothetical 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?	*No quality match ; RNR_inhib ; Ribonucleotide reductase inhibitor 76.87%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**67**
Stop Coordinate	**27649**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**27416**
Selected Function	**NrdH-like glutaredoxin**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 27416*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – BabyYoda and SanaSana both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 27416*
Does the start site have an associated Ribosome Binding Site with a high score?	*27416 – moderate z-score (2.140) and high spacer (16 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (23588) was found in 81 of 421 genes in the pham, and called 19.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 27416, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, NrdH-like glutaredoxin*
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?	***c.47.1.0 (A:) automated matches {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 559292]} | CLASS: Alpha and beta proteins (a/b), FOLD: Thioredoxin fold, SUPFAM: Thioredoxin-like, FAM: automated matches 99.34%***
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*NrdH-like glutaredoxin*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**68**
Stop Coordinate	**27627**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**28298**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 28298*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**69**
Stop Coordinate	**27975**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**27649**
Selected Function	**Hypothetical**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 27649*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Icarian and Sanasana both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 27649*
Does the start site have an associated Ribosome Binding Site with a high score?	*27649 – moderate z-score (2.167) and high spacer (16 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF but it is not overlapping with the previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (27649) was found in 75 of 353 genes in the pham, and called 21.2% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 27649, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*BlastP/ PhagesDB: Hypothetical protein/ unknown function*
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?	*This protein has a probability of 1.93 % in the HHPred with an alignment of a Hypothetical protein/ unknown 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?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	yes
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**70**
Stop Coordinate	**28313**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**None**
Selected Start Coordinate	**28005**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 28005*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albright and Albedo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 28005*
Does the start site have an associated Ribosome Binding Site with a high score?	*28005 – low z-score (1.797) and high spacer (14 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes It is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (28005) was found in 61 of 61 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 28005, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**71**
Stop Coordinate	**28337**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**28495**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 28495 has strength 7.24 not called by Genemark*
Is there evidence for coding potential?	*GeneMarkS did not show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**72**
Stop Coordinate	**28717**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**28394**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 28394*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 28394*
Does the start site have an associated Ribosome Binding Site with a high score?	*28394 – low z-score (1.889) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (28394) was found in 2 of 2 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 28394, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Hypothetical 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?	*Hypothetical Protein* *40.37%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**73**
Stop Coordinate	**28926**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**None**
Selected Start Coordinate	**28717**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 28717*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albright and Albedo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported):: 28717*
Does the start site have an associated Ribosome Binding Site with a high score?	*28717 – low z-score (1.926) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF. The longest ORF does have an excessive overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (28717) was found in 43 of 63 genes in the pham, and called 68.3% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 28717, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, Hypothetical 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?	*Hypothetical Protein* *63.7%*
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?	*INo*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical Protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**74**
Stop Coordinate	**29007**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**29168**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 29168*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**76**
Stop Coordinate	**29202**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**YES**
Selected Start Coordinate	**29330**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 29330 has strength 15.13 not called by Genemark*
Is there evidence for coding potential?	*GeneMarkS did not show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**77**
Stop Coordinate	**29772**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**31034**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, called by Glimmer bp 31034*
Is there evidence for coding potential?	*Genemark did not show coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*NO deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**78**
Stop Coordinate	**31355**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**33319**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 33319 has strength 9.98 ** not called by GeneMark*
Is there evidence for coding potential?	*No coding potential*
Is this gene present in other annotated genomes?	*Yes, found only in Loviatar_Draft*
Does the gene violate any major guiding principles?	*The gene violates major guiding principles such as overlapping other, stronger genes*
DECISION:	*Not a gene, deleted*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**75**
Stop Coordinate	**31517**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**29043**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 29043*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 29043*
Does the start site have an associated Ribosome Binding Site with a high score?	*29043 – moderate z-score (2.782) and appropriate spacer (7 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (29043) was found in 81 of 198 genes in the pham, and called 40.9% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 29043, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**79**
Stop Coordinate	**33388**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**33272**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 33272*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::33272*
Does the start site have an associated Ribosome Binding Site with a high score?	*27649 – moderate z-score (2.395) and appropriate spacer (12 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes it is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (33272) was found in 2 of 2 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 33272, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**80**
Stop Coordinate	**33585**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**33388**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 33388*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 33388*
Does the start site have an associated Ribosome Binding Site with a high score?	*33388 – moderate z-score (2.479) and high spacer (15 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (33388) was found in 78 of 82 genes in the pham, and called 95.1% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 33388, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**82**
Stop Coordinate	**33623**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**34675**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 34675 has strength 7.00 not called by genemark*
Is there evidence for coding potential?	*GeneMarkS did not show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**81**
Stop Coordinate	**33884**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**33582**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 33582*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albright and Albedo both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported)::33582*
Does the start site have an associated Ribosome Binding Site with a high score?	*33582 – high z-score (3.029) and appropriate spacer (12 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF. It excessively overlaps with the last gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (33582) was found in 101 of 181 genes in the pham, and called 55.8% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 33582, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*BlastP/ PhagesDB unknown function*
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?	*This protein has a probability of 1.93 % in the HHPred with an alignment of a Hypothetical protein/ unknown 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?	*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:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**83**
Stop Coordinate	**34603**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**33884**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 33884*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Albedo and Albright both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 33884*
Does the start site have an associated Ribosome Binding Site with a high score?	*33884 – low z-score (1.581) and appropriate spacer (10 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (33884) was found in 19 of 189 genes in the pham, and called 10.1% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 33884, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**85**
Stop Coordinate	**34854**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**34648**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Genemark, bp 34648*
Is there evidence for coding potential?	*GeneMarkS did not show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*The gene does not follow all guiding principles*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Record where you think the gene should start here and briefly explain your rationale.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**88**
Stop Coordinate	**35280**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**35116**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark, bp 35116*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – Loviatar_Draft both had this gene*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: GeneMark Start Coordinate (type NA if not supported):: 35116*
Does the start site have an associated Ribosome Binding Site with a high score?	*35116 – low z-score (1.797) and high spacer (15 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF, the longest ORF does have excessive overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (35116) was found in 52 of 52 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved, only one blast available.*
DECISION:	*Start site is bp 35116, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**87**
Stop Coordinate	**35702**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**35280**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer called 34863 and GeneMark called calls start at 35820*
Is there evidence for coding potential?	*GeneMarkS did show coding potential*
Is this gene present in other annotated genomes?	*Yes –Loviatar_Draft had this gene*
Does the gene violate any major guiding principles?	*Gene does not violate any major guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 34863* *GeneMark Start Coordinate (type NA if not supported):: 35280*
Does the start site have an associated Ribosome Binding Site with a high score?	*35280 – moderate z-score (2.779) and appropriate spacer (9 bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Not the longest ORF and does not result in excessive gene overlap*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (34863) was found in 2 of 2 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*The reason we went with 35280 as the start position instead of 34863, is because the coding potential is greatest at the start of 35280. Starterator says 34683, however according to other genes it would make sense for the gene to start at 35280. For this reason, we have concluded that moving gene 87 (34683) to start at (35280)*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**91**
Stop Coordinate	**36629**
Direction (For/Rev)	**for**
Gap (Overlap) with Previous Gene	**overlap**
Selected Start Coordinate	**36189**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Genemark, bp 36189*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes – DirtyBubble and Icarian*
Does the gene violate any major guiding principles?	*The gene follows all guiding principles*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: n/a* *GeneMark Start Coordinate (type NA if not supported):: 36189*
Does the start site have an associated Ribosome Binding Site with a high score?	*36189– highest z-score (3.273) and appropriate spacer (13bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it the length of 441. There is overlapping with other genes.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, It (36189) was found in 50/338 genes in the pham, and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 36189, based off of Genemark, Genemark coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

37718

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**96**
Stop Coordinate	**37718**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**37557**
Selected Function	**unknown**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @bp 37557*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, in BAjuniper, BabyDaisy, and Bachaco.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 37557*
Does the start site have an associated Ribosome Binding Site with a high score?	*37557 - Large z-score of 2.444 and a appropriate spacer 9bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be start 1 with a length of 330*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 75/82 the pham and called 91.5 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved*
DECISION:	*Start site is bp 37718, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, an unknown function/ hypothetical 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?	*This protein has a probability of 82.46% in the HHPred with an alignment of uknown/hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*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?	*Yes*
DECISION:	*NKF*

Student Gene Annotation Worksheet

38090

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**98**
Stop Coordinate	**38090**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**38587**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	Original Glimmer call @bp 38587 has strength 1.92 ** not called by GeneMark
Is there evidence for coding potential?	*GeneMark shows no coding potential*
Is this gene present in other annotated genomes?	*Yes, in Loviatar_Draft.*
Does the gene violate any major guiding principles?	*Yes*
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 38587* *GeneMark Start Coordinate (type NA if not supported):: N/A*
Does the start site have an associated Ribosome Binding Site with a high score?	*38587 - Large z-score of 2.281 and a appropriate spacer 7bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be start 1 with a length of 747*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 2/2 the pham and called 100.0 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved*
DECISION:	*Start site is bp 38587, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

38261

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**97**
Stop Coordinate	**38261**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**37776**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @bp 37776*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, in AnnaLie, Arroyo, and BabyDaisy.*
Does the gene violate any major guiding principles?	*No*
DECISION:	

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 37776*
Does the start site have an associated Ribosome Binding Site with a high score?	*37776 - Large z-score of 1.700 and a appropriate spacer 13bp*
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 486*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 75/82 the pham and called 91.5 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved*
DECISION:	*Start site is bp 37776, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

38261

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**97**
Stop Coordinate	**38261**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**37776**
Selected Function	**Unknown function**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @bp 37557*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, in AnnaLie, Arroyo, and BabyDaisy.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 37557*
Does the start site have an associated Ribosome Binding Site with a high score?	*Large z-score of 2.295 and a appropriate spacer 13bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be Start 2*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 2/83 the pham and called 2.4 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved*
DECISION:	*Start site is bp 42028, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, an unknown function/ hypothetical 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?	*This protein has a probability of 60.64% in the HHPred with an alignment of uknown/hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*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?	*yes.*
DECISION:	*NKF*

Student Gene Annotation Worksheet

38584

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**99**
Stop Coordinate	**38584**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**38324**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @bp 38324*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, in Armstrong, Celaena, ChiliPepper.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 38324*
Does the start site have an associated Ribosome Binding Site with a high score?	*38324 - Large z-score of 1.832 and a appropriate spacer 10bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the only start.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 13/39 the pham and called 33.3 % of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved*
DECISION:	*Start site is bp 38324, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes, an unknown function/ hypothetical 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?	*This protein has a probability of 90.55% in the HHPred with an alignment of uknown/hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*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?	*yes*
DECISION:	*NKF*

Student Gene Annotation Worksheet

39149

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**100**
Stop Coordinate	**39149**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**38574**
Selected Function	**Dihydrofolate reductase**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @bp 38574*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, in Bernstein, Brahms, and BubbaBear.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 38574*
Does the start site have an associated Ribosome Binding Site with a high score?	*38574 - Large z-score of 2.132 and a appropriate spacer 10p*
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 576*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 75/75 the pham and called 100.0% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved*
DECISION:	*Start site is bp 38574, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Yes,* Dihydrofolate reductase
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?	*This protein has a probability of 99.91% in the HHPred with an alignment of* Dihydrofolate reductase.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, with* Dihydrofolate reductase
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Yes, this gene should align with* Dihydrofolate reductase

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**113**
Stop Coordinate	**42025**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**42324**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Glimmer, bp 42324, no GeneMark call*
Is there evidence for coding potential?	*No, GeneMark shows no coding potential*
Is this gene present in other annotated genomes?	
Does the gene violate any major guiding principles?	
DECISION:	

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 42324*
Does the start site have an associated Ribosome Binding Site with a high score?	*Large z-score of 1.801 and a appropriate spacer 13bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be Start 1*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 2/2 in the pham and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*There is no BlastP for this gene*
DECISION:	*Start site is bp 42324, based off of Glimmer, no Genemark coding potential, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**114**
Stop Coordinate	**42363**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Overlap**
Selected Start Coordinate	**42028**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @ bp 42363*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	
Does the gene violate any major guiding principles?	*No*
DECISION:	

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 42363*
Does the start site have an associated Ribosome Binding Site with a high score?	*Large z-score of 2.295 and a appropriate spacer 13bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be Start 2*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 54/224 in the pham and called 9.3% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved*
DECISION:	*Start site is bp 42028, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Akino08**
Gene #	**114**
Stop Coordinate	**42491**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**Yes**
Selected Start Coordinate	**42363**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes GeneMark call @ bp 42363*
Is there evidence for coding potential?	*Yes, GeneMark shows coding potential*
Is this gene present in other annotated genomes?	
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes, it is a gene*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: N/A* *GeneMark Start Coordinate (type NA if not supported):: 42363*
Does the start site have an associated Ribosome Binding Site with a high score?	*Large z-score of 2.876 and a appropriate spacer 15bp*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, the longest ORF would be Start 1*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 51/51 in the pham and called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes it is highly conserved*
DECISION:	*Start site is bp 42363, based off Genemark, phamerator, starterator.*

Annotation Decision #3: What is the Function of the Putative Protein?

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