Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**Gene #1**
Stop Coordinate	**379**
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-3
Selected Start Coordinate	41
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, both*
Is there evidence for coding potential?	*Yes, there is.*
Is this gene present in other annotated genomes?	*Yes, it is present in other annotated genomes.*
Does the gene violate any major guiding principles?	*There are no significant violations of the major guiding principles in this gene.*
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?	*41*
Does the start site have an associated Ribosome Binding Site with a high score? RBS score: -5.036 Z- Value: 2.054	*RBS score: -5.036* *Z- Value: 2.054* *These can be considered as acceptable scores.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*This is not the longest ORF. The longest ORF length would be 378.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is. The start site in other phage genomes is the same in all of them.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The gene should start at start site 41. We believe this as glimmer, starterator (100%) and genemark agrees that this is the start site.*

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

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10⁻⁴ or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: unnamed protein product* *NCBI: unknown function* *DNA Master: hypothetical 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?	*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	**Yami**
Gene #	**Gene #2**
Stop Coordinate	**1059**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**-3**
Selected Start Coordinate	**376**
Selected Function	**No known 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 BOTH*
Is there evidence for coding potential?	*GeneMark coding potential map shows coding potential.*
Is this gene present in other annotated genomes?	*Yes it has been present in another 126 genomes.*
Does the gene violate any major guiding principles?	*There aren't any significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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?	*376*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -4.334* *Z- Value: 2.405* *Spacer: 13* *These can be considered acceptable scores.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The predicted start codon has the longest ORF length being 684 bp.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start is found in 100% of the genes, but is only called 95.2% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The gene should start here as both glimmer and genemark agree with the start.*

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: Function Unknown* *NCBI: Hypothetical protein* *DNA Master: Hypothetical 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?	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*NO*
Is the proposed function found on the SEA-PHAGES approved function list?	*yes*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**Gene #3**
Stop Coordinate	**2522**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**1062**
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, the gene was called by both*
Is there evidence for coding potential?	*The genemark map shows coding potential.*
Is this gene present in other annotated genomes?	*Yes, a similar gene in the phage TiniBug was found. The gene was presented in many more phages, too.*
Does the gene violate any major guiding principles?	*No violations were found in this gene.*
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: 1068* *GeneMark Start Coordinate: 1062*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score 1068: -6.887* *Z-value 1068: 0.955* *RBS Score 1062: -5.818* *Z-value 1062: 2.140*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length 1062: 1461* *ORF length 1068: 1455* **1062 is the longest ORF length*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start site 1068 is called 8.3% of the time when it is present* *Start site 1062 is called 91% of the time when it is present* *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:	*The start site should be 1062 since it is called 91% of the time and has a viable gap/overlap with the previous gene.*

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

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**Gene #4**
Stop Coordinate	**3774**
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	125
Selected Start Coordinate	2647
Selected Function	Portal protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Yes, there is coding potential*
Is this gene present in other annotated genomes?	*Yes, a phage with a similar gene is Cazares, with the gene having a similar start and stop codon called by Glimmer.*
Does the gene violate any major guiding principles?	*No. There weren't any violations found.*
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: 2647* *Genemark: 2752*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score 2647: -5.756* *Z- value 2647: 1.972* *RBS score 2752: -6.543* *Z- Value 2752: 1.778*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length 2647: 1128* *ORF Length 2752: 1023* **2647 is the longest ORF length**
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start Site 2647 is called 70.3% of the time present* *Start Site 2752: is called 0% of the time present*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site should be 2647, since it is called 70.3% of the time and has a viable gap/ overlap with the previous gene.*

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

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

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**Gene # 5**
Stop Coordinate	**5357**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**-3**
Selected Start Coordinate	**3771**
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 Both*
Is there evidence for coding potential?	*Yes there is coding coding potential.*
Is this gene present in other annotated genomes?	*Yes, the gene is present in 130 other annotated genomes. One of the annotated phages is Charbie, with Charbie and Yami having the same start and stop coordinates.*
Does the gene violate any major guiding principles?	*There are no significant violations of the [Guiding Principles of Genome Annotation](#) with the gene call.*
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: 3771* *GeneMark Start Coordinate: 3771*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS Score: -3.501* *Z-score: 3.094*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length: 3771 to 5357* **Longest ORF length is 1587**
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start site is found in 130 out of 130 of the genes in the phamerator.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site chosen is 3771 since it is called by both genemark and glimmer and is called 100% of the time.*

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	**Yami**
Gene #	**Gene # 6**
Stop Coordinate	**5714**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**4**
Selected Start Coordinate	**5361**
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, both.*
Is there evidence for coding potential?	*Yes, there is adequate evidence to support coding potential for this gene.*
Is this gene present in other annotated genomes?	*Yes, this gene is present in 130 other annotated genomes. One phage that has a similar annotated gene is Lunatic, having the same start and stop coordinates as Yami.*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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?	*Start coordinate called by both: 5361*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -4.553* *Z- value: 2.670*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length: 5361 to 5714* **The longest ORF length is 354**
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start site is called 100.0% of the time when it is present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene should be 5361, since there is a viable gap between the previous gene, and it doesn't violate any rules.*

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	**Yami**
Gene #	**Gene # 7**
Stop Coordinate	**6091**
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	3
Selected Start Coordinate	5711
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, both.*
Is there evidence for coding potential?	*Genemark shows coding potential.*
Is this gene present in other annotated genomes?	*This gene is found in 130/130 other annotated genes.*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
DECISION:	*YES*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS: -6.591* *Z-Value: 1.516*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length: 381* *This is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is found in 100% of the genes and called 99.2% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The gene should start at 5771 as it is suggested by both genemark and glimmer and is called 99.2% of the time.*

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	**Yami**
Gene #	**Gene # 8**
Stop Coordinate	**6567**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**39**
Selected Start Coordinate	**6130**
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, both.*
Is there evidence for coding potential?	*Yes it shows coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is found in 130/194 of the annotated genes. And is called 100% of the time.*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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: 6130* *GeneMark Start Coordinate: 6130*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -4.836* *Z-Value: 1.979*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length: 438* *Yes this is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, this starts site is found in 67% of the phage genomes and is called 100% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The gene should start at 6130 because it is suggested by both gene mark and glimmer and is called 100% of the time.*

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	Yami
Gene #	Gene # 9
Stop Coordinate	6963
Direction (For/Rev)	Foward
Gap (Overlap) with Previous Gene	17
Selected Start Coordinate	6580
Selected Function	nkf

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, the gene was called by both.*
Is there evidence for coding potential?	*Yes, the gene mark map shows coding potential.*
Is this gene present in other annotated genomes?	*Found in 136 other annotated genes.*
Does the gene violate any major guiding principles?	*No violations were found in this gene.*
DECISION:	*Yes.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Both suggest the start site to be 6580.*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score 6580: -4.294* *Z- Value 6580: 2.249*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length 6580: 384* *Yes, 384 is the longest.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start site 6580 is called 71.6% of the time when it's present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site is 6580 because it is suggested by both glimmer and GenMark.*

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	**Yami**
Gene #	**Gene # 10**
Stop Coordinate	**7297**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**14**
Selected Start Coordinate	**6977**
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, both*
Is there evidence for coding potential?	*Yes, there is adequate evidence that*
Is this gene present in other annotated genomes?	*Yes, it is present in 130 other annotated genomes. The gene in the phage Cazares is like Yami. Both have the same start and stop coordinates*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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?	*Start coordinate called by both: 6977*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -3.368* *Z-score: 2.742*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length: 6977 to 7297* **Longest ORF length is 321**
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start site is called in 100.0% of the times it is present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The chosen start site for this gene is*

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	**Yami**
Gene #	**Gene# 11**
Stop Coordinate	**9636**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**234**
Selected Start Coordinate	**7531**
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, both*
Is there evidence for coding potential?	*Yes, there is coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in 129 genomes.*
Does the gene violate any major guiding principles?	*There are no violations.*
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: 7552* *GeneMark Start Coordinate: 7531*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score for 7552: -5.849* *Z-score for 7552: 1.799* *RBS score for 7531: -4.386* *Z-score for 7531: 2.753*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Start Site 7552: It is not predicted to be the longest ORF, and no, it does not overlap.* *Start Site 7531: It is predicted to be the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*7552: The start is found in 92% of genes and is called 2.3 % if the time.* *7531: The start is found in 93.5% of genes and is called 97.7% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene is 7531, since this start is called 97.7% of the time.*

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	**Yami**
Gene #	**#12**
Stop Coordinate	**10139**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**1381**
Selected Start Coordinate	**9633**
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, Glimmer and GeneMark*
Is there evidence for coding potential?	*Yes, the gene mark shows coding potential*
Is this gene present in other annotated genomes?	*Yes, it is found in 22 out of 22 annoted genes.*
Does the gene violate any major guiding principles?	*No, it does not violate any principals.*
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): 9633* *GeneMark Start Coordinate (type NA if not supported): 9633*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS value: -4.924* *Z-value: 2.005*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*9633 ORF length: 507, Yes, it is the longest ORF*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start Site 9633 is called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Yes, the start site for this gene is 9633.*

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	**Yami**
Gene #	**13**
Stop Coordinate	**12196**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**0**
Selected Start Coordinate	**10139**
Selected Function	minor tail protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Both*
Is there evidence for coding potential?	*Yes, Genemark shows evidence for coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in other genomes.*
Does the gene violate any major guiding principles?	*No, there are no violations of the 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?	*Both glimmer and genemark call the start site of 10139*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score-4.089 and Z-score-2.804* *This RBS score is not the highest, the highest score is 8.726.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The ORF length is 2058 bp.* *This length is the longest* *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?	*This start site is in 130/130 of the other phage genomes and 100% of them call the gene.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site is 10139. We chose this because it was called by both Glimmer and Genemaster, has the least amount of overlap, and is called 100% of the time when present.*

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

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *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	**Yami**
Gene #	**Gene #14**
Stop Coordinate	**12749**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**2**
Selected Start Coordinate	**12198**
Selected Function	minor tail protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Yes, there is coding potential.*
Is this gene present in other annotated genomes?	*Yes. There are 159 other members.*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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): 12198* *GeneMark Start Coordinate (type NA if not supported): 10139*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -3.955* *Z-score: 2.489*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length: 552*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Called 99.2% of the times it was present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene should be 12198, since there is a viable gap between the previous gene, and it doesn't violate any rules.*

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	Yami
Gene #	15
Stop Coordinate	13085
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	37
Selected Start Coordinate	12786
Selected Function	nkf

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer calls the start site 12786 and Genemark calls 12762*
Is there evidence for coding potential?	*Yes, there is evidence for coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is annotated in other genomes. More specifically 41.*
Does the gene violate any major guiding principles?	*The gene does not show any violations of these 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 12786 GeneMark Start Coordinate 12762
Does the start site have an associated Ribosome Binding Site with a high score?	RBS score 12762 = 5.692 Z value 12762=1.621 RBS score 12786= 3.589 z-value 1278=2.641 These are not the highest scores, the highest score is 6.963.
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	ORF length 12762- 324 ORF Length 12786- 300 The longest length is 324 bp and there are no major overlaps
Is this start site conserved in other phage genomes as indicated by Starterator?	12786-Yes, the start site is present in 54/55 of the genes in phamerator and called 98.1% of the time when present 12762- Yes, the start site is present in 28/55 of the genes in phamerator and called 3.6% of the time when present
Is this start site conserved in other phage genomes as indicated by BlastP?	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources. Note: if you are considering more than 1 start site, provide the same information for each proposed start site.
DECISION:	The gene should start at 12786. We decided this because it is present in 54/55 of the other genes and called 98.1% of the time when present according to phamerator.

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	**Yami**
Gene #	**16**
Stop Coordinate	**13800**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**20**
Selected Start Coordinate	**13105**
Selected Function	endolysin

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Yes, there is coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is found in 126 other genes.*
Does the gene violate any major guiding principles?	*No. It does not violate any major guidelines.*
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 suggest: 13177* *GeneMark Start Coordinate suggest: 13105*
Does the start site have an associated Ribosome Binding Site with a high score?	*13177 RBS score: -6.290* *13177 Z-score: 1.323* *13105 RBS score: -3.899* *13105 Z- score: 2.477*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*13177 ORF Length: 624, it is not the longest ORF, and it does not overlap.* *13105 ORF Lenth: 697, it is the longest ORD length.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*13177: Start Site is found in 96.2% and called 1.6% of the time.* *13105: Start Site is found in 96.9% of genes and called 96.1% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene is 13105, since the start site is found 96.9 % of the time.*

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 17

Basic Phage Information	
Phage Name	Yami
Gene #	
Stop Coordinate	14030
Direction (For/Rev)	Forward
Gap (Overlap) with Previous Gene	-33
Selected Start Coordinate	13767
Selected Function	nkf

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both*
Is there evidence for coding potential?	*Yes.*
Is this gene present in other annotated genomes?	*Yes, it has 131 members*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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): 13767* *GeneMark Start Coordinate (type NA if not supported): 13767*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -6.724* *Z-score: 1.586*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF length: 264* *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?	*Called 85.9% of the time when it is present.* *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:	*The start site for this gene should be 13767, since there is a viable gap between the previous gene, and it doesn't violate any rules.*

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	**Yami**
Gene #	**#18**
Stop Coordinate	**14251**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**11**
Selected Start Coordinate	**14027**
Selected Function	membrane protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Shows coding potential*
Is this gene present in other annotated genomes?	*Yes, it has 131 members*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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: 14027* *GeneMark Start Coordinate: 14027*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS: -3.171* *Z-Value: 2.841*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF lengths: 225*
Is this start site conserved in other phage genomes as indicated by Starterator?	*It is called 98.5% of the time*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene should be 14027, since there is a viable gap between the previous gene, and it doesn't violate any rules.*

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	**Yami**
Gene #	**19**
Stop Coordinate	**14320**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**69**
Selected Start Coordinate	**14529**
Selected Function	Lsr2-like DNA bridging protein

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Yes, there is coding potential.*
Is this gene present in other annotated genomes?	*Yes, this gene is present in 129 other annotated genomes.*
Does the gene violate any major guiding principles?	*No, there are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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 and GeneMark Start Coordinate: 14529*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score 14529: -3.627* *Z-score14529: 2.652*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Length 14529: 210*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start Site 14529 is called 100% of the time when present*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site should be 14529 since it is called 100% of the time.*

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.*

15038Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**20**
Stop Coordinate	**14532**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**-3**
Selected Start Coordinate	**15038**
Selected Function	helix-turn-helix DNA binding domain

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*There is coding potential*
Is this gene present in other annotated genomes?	*This gene is present in other annotated genomes.*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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 15122* *GeneMark Start Coordinate 15038*
Does the start site have an associated Ribosome Binding Site with a high score?	*Glimmer Start Coordinate 15122: RBS score: -7.298 Z-score:0.820* *GeneMark Start Coordinate 15038: RBS score: -6.1782 Z-score: 1.460* *.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Glimmer Start Coordinate 15122: 591 (longest ORF)* *GeneMark Start Coordinate 15038: 507 (3rd Longest)*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Glimmer Start Coordinate 15122: Found in 47/131 and called 14% of the time when present* *GeneMark Start Coordinate 15038: found in 121/131 and called 66% of the time when present*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site of this gene should be at 15038 because it is called 66% of the time when present in phamerator and is better shown in the coding potential.*

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	**Yami**
Gene #	**21**
Stop Coordinate	**15119**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**-3**
Selected Start Coordinate	**15349**
Selected Function	helix-turn-helix DNA binding domain

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both*
Is there evidence for coding potential?	*Yes, there is coding potential.*
Is this gene present in other annotated genomes?	*Yes, this gene is present in 131 other genes.*
Does the gene violate any major guiding principles?	*No, there are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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 and GeneMark Start Coordinate : 15349*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score 15349: -7.526* *Z-score 15349: 0.899*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*ORF Lenght: 231*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start Site 15349 is called 95.4 % of the time when present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence, you are analyzing, and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site should be 15349 since it is called by both genemark and glimmer and is called 95.4% of the time.*

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	**Yami**
Gene #	**22**
Stop Coordinate	**16245**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**740**
Selected Start Coordinate	**15862**
Selected Function	**nkf**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes Both*
Is there evidence for coding potential?	*Yes there is coding potential*
Is this gene present in other annotated genomes?	*Yes, this gene us present in other annotated genomes*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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 15862 GeneMark Start Coordinate 16042
Does the start site have an associated Ribosome Binding Site with a high score?	The RBS score is 15862-5.399 The Z-score 15862-1.961 The RBS score 16042-6.535 The Z-score 16042-1.201
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	The ORF length 15862 -384 The ORF length 16042-204 The longest is 384, there is no major overlap
Is this start site conserved in other phage genomes as indicated by Starterator?	Start 15862 is present in 94/96 and called 97.9% of the time when present. Start 16042 is present in 2/96 and called 100% of the time when present.
Is this start site conserved in other phage genomes as indicated by BlastP?	Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources. Note: if you are considering more than 1 start site, provide the same information for each proposed start site.
DECISION:	Teh gene should be in start site 15862 since it was present 97.9% of the time when present

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

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10⁻⁴ or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *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	**Yami**
Gene #	**23**
Stop Coordinate	**16554**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**91**
Selected Start Coordinate	**16336**
Selected Function	helix-turn-helix DNA binding domain

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both.*
Is there evidence for coding potential?	*Yes, there is coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in 108 genes.*
Does the gene violate any major guiding principles?	*No, it does not violate any major guiding pronciples*
DECISION:	*YES*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*They both suggest: 16336*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -7.139* *Z-score:1.005*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the ORF length, (219) is the longest.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start is found in 83.7% of the time and called 98.1% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start should be 16336, since they are called by both GenMark and Glimmer call 98.1% of the time.*

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.*

The superscript in the first row reads 10^{-4}.

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Yami**
Gene #	**24**
Stop Coordinate	**16850**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**-3**
Selected Start Coordinate	**16551**
Selected Function	hnh endonuclease

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both*
Is there evidence for coding potential?	*Yes.*
Is this gene present in other annotated genomes?	*Yes, there are 179 members*
Does the gene violate any major guiding principles?	*There are no significant violations of the Guiding Principles of Genome Annotation with the gene call.*
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): 16551* *GeneMark Start Coordinate (type NA if not supported): 16551*
Does the start site have an associated Ribosome Binding Site with a high score?	*RBS score: -6.299* *Z-score: 1.900* *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)?	*ORF length: 300* *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?	*Called 100.0% of the time when present.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*The start site for this gene should be 16551, since there is a viable gap between the previous gene, and it doesn't violate any rules.*

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.*