Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**1**
Stop Coordinate	**379**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**41**
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, bp 41*
Is there evidence for coding potential?	*Both genemark programs show coding potential. Note that bp 41 start site would capture all of the coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Yami, Niffler, and Cazares.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes, this is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 41* *GeneMark Start Coordinate (type NA if not supported):: 41*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering start site base pair 41 and 35*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Bp 41 has the ORF length of 339 and bp 35 has ORF length of 345, however, z score and final scores of bp 41 are more desirable.*
Is this start site conserved in other phage genomes as indicated by Starterator?	The start number called the most often in the published annotations is 3, it was called in 127 of the 127 non-draft genes in the pham. *Site 3 is bp 41 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical protein for the microbacterium phages BurtonThePup, VitulaEligans, and Snuffles.*
DECISION:	*Bp 41 based on glimmer/genemark, coding potential, starterator, phamerator, and Blast.*

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?	*DNA Master: hypothetical protein (BurtonThePup)* *E-value: 0.0E0*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No Quality Match. The best match is DUF3151, but it only had a probability of 38.01 and database score of 20.86, meaning it is a hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed; hypothetical protein.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, Hypothetical protein is on the SEA-PHAGES approved function list.*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**2**
Stop Coordinate	**1173**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**4**
Selected Start Coordinate	**376**
Selected Function	**Hypothetical protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer calls bp 376 and GeneMark calls bp 361.*
Is there evidence for coding potential?	*Yes, because the bp values given are very similar for both programs.*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Concrete, Danimal, and PoRanda.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 376* *GeneMark Start Coordinate bp 361*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 376 and bp 361, the first value has a RBS score of -4.175 and a Z value of 1.242, which are the highest. The second value has a RBS score of –5.438 and a Z value of 1.719.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon has a bp length of 798, which was the second longest ORF. The overlap with the previous gene is not excessive. The length of the ORF with the second most likely probability is 813 bp, the highest length, with a slightly greater overlap, yet still not excessive.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 2, it was called in 122 of the 124 non-draft genes in the pham.* *Site 2 is bp 376 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical protein for Hernandez44, Azizam, and McShie.*
DECISION:	*Bp 376 based on glimmer/genemark, coding potential, starterator, 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* *DNA Master:* *Hypothetical protein [Hernandez44]* *e-value: 0.0E0*
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?	*According to HHPred, there is no quality match. The best match found is* (A:3-130) with a probability of 16.37 and database score of 21.43.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed; hypothetical protein.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Hypothetical protein is on the SEA-PHAGES function approved list.*
DECISION:	*Hypothetical protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**3**
Stop Coordinate	**2636**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**1176**
Selected Function	**Terminase, large subunit**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 1176 has strength 17.59*
Is there evidence for coding potential?	*Yes.*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Octopus, Nobel, and Kevanna.*
Does the gene violate any major guiding principles?	*This gene does not violate any major guiding principles*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Both call bp 1176*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is –6.887, and the Z score is 0.955 which are both relatively low.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 171, it was called in 155 of the 1231 non-draft genes in the pham.* *Start 110 is 1176 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical terminase in PaoPu, Picnic, and Lola20.*
DECISION:	*@Bp 1176 based on Glimmer/Genemark, coding potential,and 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?	*DNA Master: Terminase (PaoPu)* *E-value: 0.0E0*
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?	*The most informative match is Terminase large subunit with a probability of 100. The database score was 333.08.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, terminase is found after the hypothetical protein and before the portal protein, showing synteny with Yami and Wolfpack.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, terminase, large subunit is found on the SEA-PHAGES approved function list.*
DECISION:	*Terminase, large subunit*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**4**
Stop Coordinate	**3888**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**4**
Selected Start Coordinate	**2761**
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)?	*Glimmer calls bp 2761 and GeneMark calls bp 2830.*
Is there evidence for coding potential?	*somewhat, because the bp values given are pretty somewhat close.*
Is this gene present in other annotated genomes?	*Yes, this gene is present in Clownery and Yubaba.*
Does the gene violate any major guiding principles?	*There is a 3 base pair overlap with the previous gene.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 2761* *GeneMark Start Coordinate bp 2830*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 2761 and bp 2830, the first value has a RBS score of -5.756 and a Z score 1.973. The second value has a RBS score of –6.731 and a Z value of 1.065.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon has a bp length of 1128, which was the longest ORF. The overlap with the previous gene is not excessive. The length of the ORF with the second most likely probability is 1059 bp.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 25, it was called in 111 of the 259 non-draft genes in the pham.* *Site 25 is bp 2761 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical portal protein for PaoPu, Concrete, and McShie.*
DECISION:	*Bp 2761, based on glimmer/genemark, coding potential, starterator, 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* *DNA Master:* *Portal protein [McShie]* *E-value: 0.0E0*
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?	*The most informative match is Portal Protein with 100 probability with the database score of 288.86.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, portal protein, appears in the order after the previous terminase, and before following major capsid and protease fusion protein. Synteny is observed in TomBrady12 and TinyTruffula.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, this is an approved protein on the SEA-PHAGES approved function list.*
DECISION:	*Portal protein.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**5**
Stop Coordinate	**5474**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**3885**
Selected Function	**Major capsid protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Original Glimmer call @bp 3885 has strength 15.73*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in BoomRoasted, JooneeDee, and Nobel.*
Does the gene violate any major guiding principles?	*There is a 3 base pair overlap with the previous gene.*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 3885* *GeneMark Start Coordinate (type NA if not supported):: 3885*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -3.509. The Z score is 3.096, which is the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF but is the best suited and doesn't result in excessive overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called most often in the published annotations is 2, it was called in 128 of 128 non-draft genes in the pham.* *Start 2 is 3885 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Major capsid and protease fusion protein [Microbacterium phage McShie, Namsahir, Biscayne, Belthelas]*
DECISION:	*Bp 3885 based on Glimmer/Genemark, coding potential, phamerator, and starterator.*

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

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*DNA Master: major capsid and protease fusion protean (VitulaEligans)* *E-value: 0.0E0*
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?	*The most informative HHPred match, Major capsid protein Rcc01687, has a probability of 99.93 with a database score of 211.08.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The proposed function, major capsid protein, appears after the portal protein and before the head-to-tail adaptor, displaying synteny with Nobel and Yami.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes.*
DECISION:	*Major capsid protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**6**
Stop Coordinate	**5831**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**5478**
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)?	*Original Glimmer call @bp 5478 has strength 17.87*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Witch, Jahseh, Octopus, and Cheesecake.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 5478* *GeneMark Start Coordinate (type NA if not supported):: 5478*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -4.553, and the Z score is 2.671.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.* *Length is 354*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called most often in the published annotations is 5 @bp 5478, it was called in 140 of 141 non-draft genes in the pham.* *Start 1 is 5478 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Head to tail adaptor [Microbacterium phage Belthelas, Leafy, KillerTomato, KayPaulus,*
DECISION:	*Bp 5478 based off Starterator, Phamerator, Glimmer/Genemaster, and 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?	*DNA Master:* *Head-to-tail-adaptor [Belthelas]* *E-value: 0.0E0*
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?	*According to HHPred, this best match for this protein is Head completion protein with a probability of 99.53 and database score of 82.48. Another plausible match is Head-Tail Connector with a probabilty of 99.36 and database score of 72.26.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yess, the proposed function, head-to-tail adaptor, appears in the order after the previous major capsid and protease fusion protein, and before following tail terminator. Synteny is observed in TomBrady12 and TinyTruffula.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, this protein is apart of the SEA-PHAGES functions approves list.*
DECISION:	*Head-to-tail adaptor*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**7**
Stop Coordinate	**6208**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**5828**
Selected Function	**Minor tail protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Both call bp 5828*
Is there evidence for coding potential?	*Yes, because both genemark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Slizam and LemonPepper.*
Does the gene violate any major guiding principles?	*There is a 3 base pair overlap with the previous gene.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 5828* *GeneMark Start Coordinate bp 5828*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 5828, the assoiated RBS score is –6.591 and the Z value is 1.517, neither are the highest score in their respective fields.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon has a bp length of 381 bp, which is not the longest orf, which has a length of 618, but results in excessive overlap (240 bp).*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 2, it was called in 127 of the 127 non-draft genes in the pham.* *Site 2 is bp 5828 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical minor tail protein for Scamander, PaoPu, and Quaker. Hypothetical tail terminator for TimoTea and Rhysand.*
DECISION:	*Bp 5828 based on glimmer/genemark, coding potential, starterator, 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?	*DNA Master: Minor tail protein (Scamander)* *E-value: 0.0E0*
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?	*The most informative match on HHPred was minor tail protein, with a probability of 98.64 and a database score of 53.47.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The proposed function, minor tail protein, appears after the head-to-tail adaptor and before the major tail protein, displaying synteny with Yami and Wolfpack.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes.*
DECISION:	*minor tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**8**
Stop Coordinate	**6681**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**6247**
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 call bp 6247*
Is there evidence for coding potential?	*Yes, because both genemark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Slizam and LemonPepper.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 6247* *GeneMark Start Coordinate bp 6247*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 6247, The associated RBS score is –5.797 and the Z value is 1.531, neither of which are the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon is 435 bp, which is the longest ORF, and does not result in overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 3, it was called in 126 of the 135 non-draft genes in the pham.* *Site 3 is bp 6247 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical major tail protein for PaoPu, Scamander, and Quaker.*
DECISION:	*Bp 6247 based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: major tail protein [PaoPu]* *E-value: 0.0E0*
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?	*According to HHPred, the best match is Major tail protein with a probability of 99.55 and database score of 92.57.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, major tail protein, appears in the order after tail terminator, and before hypothetical protein. Synteny is observed in TomBrady12 and TinyTruffula.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, this protein is on the approved SEA-PHAGES functions list.*
DECISION:	*Major tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**9**
Stop Coordinate	**7077**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**6694**
Selected Function	**Minor capsid protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, both call bp 6694*
Is there evidence for coding potential?	*Yes, because both genemark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Hulk and Azizam.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 6694* *GeneMark Start Coordinate bp 6694*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 6694, The associated RBS score is – 4.294 and the Z value is 2.250, both of which are the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon is 6694 bp, which is the longest ORF, and does not result in overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 6, it was called in 130 of the 187 non-draft genes in the pham.* *Site 6 is bp 6694 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical protein for PaoPu, VitulaEligans, and PoRanda.*
DECISION:	*Bp 6694 based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: Hypothetical protein (PaoPu)* *E-value: 0.0E0*
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?	*The most informative HHPred match, Minor capsid protein, had a probability of 99.44 and a database score of 73.76.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The proposed gene, Minor capsid protein, appears before the major tail protein and tail assembly chaperone, showing synteny with Yami and Wolfpack.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Minor capsid protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**10**
Stop Coordinate	**7411**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**7091**
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)?	*Original Glimmer call @bp 7091 has strength 15.01*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Jerky, Otwor, and PoRanda*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: bp 7091* *GeneMark Start Coordinate (type NA if not supported)::bp 7091*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -3.826 and the Z value is 2.484.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon has the longest ORF of 321.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Site 1, is bp 7091 in Zahlia. The start number called the most often in the published annotations is 6 @bp 7091, it was called in 127 of the 127 non-draft genes in the pham.* *You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Provide the best BlastP match from NCBI, PhagesDB, and DNA Master with alignment in the format of (Q#:S#), where Q (query) is the sequence you are analyzing and S (subject) is the database match. List the e-value and alignment of the best match for all three BlastP sources.* *Tail assembly chaperone [Microbacterium phage Scamander, Sara, Lola20, Leafy, and Bri160].* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Bp 7091 based on Glimmer/Genemark coding potential, Starterator, and 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?	*DNA Master: tail assembly chaperone [Scamander]* *E-value: 0.0E0*
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?	*According to HHPred, the best match is the Phage tail tube with a probability of 96.89 and database score of 38.85. Another plausible protein is the tail assembly chaperone with a probabilty of 93.61 and database score of 32.31*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, tail assembly chaperone, appears in the order after the hypothetical protein, and is followed by a tail assembly chaperone. Synteny is observed in TomBrady12 and Yubaba.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, the tail assembly chaperone is present on the SEA-PHAGES approved function list.*
DECISION:	*Tail assembly chaperone*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**11**
Stop Coordinate	**7527**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**7438**
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)?	*Original Genemark call @bp 7438.*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*It looks different from Chatham, Naby, and Vanisius, but it has coding potential.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::7438* *GeneMark Start Coordinate (type NA if not supported)::7438*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Z-score is 1.377 and the final score is -6.043.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles?* *It is not the longest ORF, and the longest ORF doesn't have excessive overlap.* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*You will also need to provide the following information from Starterator: does the start match the consensus start site predicted from Starterator? If no, is the consensus start site not found in this ORF? If no, is there a better option for the consensus start site instead of the one predicted by Starterator? If Starterator doesn't reveal a consensus start site, you can record that Starterator was not informative.* *Start 2 is found in 8 of 8 (100%) of genes in pham. Site 2 in Zahlia is bp 7438.* *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.* *Tail assembly chaperone[Bri160]* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
DECISION:	*Bp 7438 based ong Glimmer/Genemark, coding potential, Starterator, and 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?	*PhagesDB: tail assembly chaperone(Bri160)* *E-value: 5.6E-10*
D probability of 90% or greater with appropriate coverage?	*The most informative HHPred match, DUF5347, has an unknown function with a probability of 87.31 and score of 24.87, meaning that it is a hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed. Hypothetical Tail Protein*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes.*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**12**
Stop Coordinate	**9750**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**7645**
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)?	*Original Glimmer call @bp 7645 has strength 17.57*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is present in Charbie, MrGreen, and Slizam.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: bp 7645* *GeneMark Start Coordinate (type NA if not supported):: bp 7645*
Does the start site have an associated Ribosome Binding Site with a high score?	*List the final RBS score and Z-score of the currently predicted start site using the Kibler6/Karlin Medium scoring table. Indicate in your response if this is the best score or not.* *Z-score is 2.754, and the final score is -4.386* *Note: if you are considering more than 1 start site, provide the same information for each proposed start site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Indicate the length of the ORF is with the predicted start and the gap/overlap to the nearest stop codon of the upstream ORF. Does the proposed start site have a gap/overlap with the nearest upstream gene that does not violate the Guiding Principles? It is the longest ORF (2106).*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 17 @bp 7645, it was called in 124 of the 298 non-draft genes in the pham. Start 1 is bp @7645 in Zahlia.*
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.* *Tape measure protein [Microbacterium phage Jannah, Jahseh, Cheesecake, and Hulk]. Tail length tape measure protein [PaoPu]*
DECISION:	*Bp @7645 based on Glimmer/Genemark, Phamerator, coding potential, and Starterator.*

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

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*DNA Master:* *Tape measure protein [Jannah]* *E-value: 0.0E0*
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?	*According to HHPred, Tape measure protein is the best match with a probability of 99.93 and a database score of 190.91*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, tape measure protein, appears in the order after the tail assembly chaperone, and is followed by a minor tail protein assembly chaperone. Synteny is observed in TomBrady12 and Yubaba.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, the proposed tape measure protein is a part of the SEA-PHAGES approved function list.*
DECISION:	*Tape measure protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**13**
Stop Coordinate	**10709**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**3**
Selected Start Coordinate	**9747**
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 call bp 9747*
Is there evidence for coding potential?	*Yes, there is evidence for potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Hulk and Azizam.*
Does the gene violate any major guiding principles?	*This gene does not violate the gene coding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 9747* *GeneMark Start Coordinate bp 9747*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 9747, The associated RBS score is – 4.412 and the Z value is 2.262, both of which are somewhat high.*
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.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 1, it was called in 104 of the 104 non-draft genes in the pham.* *Site 1 is bp 9747 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical minor tail protein for DaftyDuck, Hernandez44, and Hulk.*
DECISION:	*Bp 9747 based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: minor tail protein (DaftyDuck)* *E-value: 0.0E0*
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?	*The most informative HHPred match, Endoglucanase, has a probability of 98.88, with a database score of 88.55. The Endoglucanase is not on the approved function list, so it is a a hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Does not display synteny; hypothetical protein.*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**14**
Stop Coordinate	**12778**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**1**
Selected Start Coordinate	**10709**
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 call bp 10709*
Is there evidence for coding potential?	*Yes, because both genemark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Vanisius, Loca, and Livingwater.*
Does the gene violate any major guiding principles?	*This gene has a 1 base pair overlap with the only tRNA in the DNA sequence.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 10709* *GeneMark Start Coordinate bp 10709*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 10709, The associated RBS score is -5.098 and the Z value is 2.302, neither of which are the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon is 2070 bp, which is the longest ORF, and does not result in overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 1, it was called in 127 of the 127 non-draft genes in the pham.* *Starterator says site 1 is bp 10709 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical minor tail protein for Bradley2, Aztec, and Dongwon.*
DECISION:	*Bp 10709, based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: minor tail protein [Bradley2]* *E-value: 0.0E0*
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?	*According to HHPred, the best match is Receptor-type tyrosine-protein phosphatase F with a probability of 99.71 and database score of 138.58.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function, minor tail protein, appears in the order after the minor tail protein, and is followed by a minor tail protein assembly chaperone. Synteny is observed in TomBrady12 and Yubaba.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, the proposed minor tail protein is present on the SEA-PHAGES approved function list.*
DECISION:	*Minor tail protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**15**
Stop Coordinate	**13328**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**12780**
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)?	*Original Glimmer call @bp 12780 has strength 18.50*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Snuffles.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported)::@bp 12780* *GeneMark Start Coordinate (type NA if not supported)::@bp* *12780*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -2.394, and the Z score is 3.270 which is the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called most often in the published annotations is 3, it was called in 123 of 155 non-draft genes in the pham.* *Start 3 is 12780 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Tail protein [Microbacterium phage PaoPu]* *Minor tail protein [Microbacterium phage DaftyDuck, EdElric, BurtonThePup]*
DECISION:	*@Bp 12780 based on Glimmer/Genemark, coding potential, phamerator, and starterator.*

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

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB: tail protein (PaoPu)* *E-value: 0.0E0*
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?	*The most informative HHPred match, ; S_tail_recep_bd ; Short tail fibre protein receptor-binding domain, has a probability of 66.94 and a database score of 20.63. This means that it is a Hypothetical Protein*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed. Hypothetical Protein*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**16**
Stop Coordinate	**13652**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**13365**
Selected Function	*Hypothetical protein*

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Glimmer call @bp 13365 and GeneMark calls start at 13341*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in ManAs, Minima, and Rhysand.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 13365* *GeneMark Start Coordinate bp 13341*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 13365, The associated RBS score is –3.589 and the Z value is 2.643*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon is 13365 bp, which is the not the longest ORF, but is the best suited.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 5, it was called in 135 of the 136 non-draft genes in the pham.* *Starterator says site 5 is bp 13365 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical protein microbacterium phage Scamander, BurtonThePup, Hernandez44, Yasuo, and Jerky.*
DECISION:	*Bp 13365, based on glimmer/genemark, coding potential, and 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?	*DNA Master: hypothetical protein [Scamander]* *E-value: 3.5E-25*
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?	*There is not quality in HHPred, however the best match is* DUF6776 with a probability of 42.36 and a database score of 23.21
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed; hypothetical protein.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, Hypothetical protein is on the SEA-PHAGES approved function list.*
DECISION:	*Hypothetical protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**17**
Stop Coordinate	**14366**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**13671**
Selected Function	**peptidase**

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 call bp 13671*
Is there evidence for coding potential?	*Yes, because both genemark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Stem and JooneeDee.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 13671* *GeneMark Start Coordinate bp 13671*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 13671, The associated RBS score is –3.978 and the Z value is 2.479, both of which are the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF with the predicted start codon is 696 bp, which is the longest ORF, and does not result in overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 7, it was called in 125 of the 128 non-draft genes in the pham.* *Starterator says site 7 is bp 13671 in Zahlia, but is innacurate due to issues identifying tRNA, which pushed forward the numbers for each gene.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical endolysin for VtuaEligans, and Cheesecake, and hypothetical peptidase for Scamander.*
DECISION:	*Bp 13671, based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: endolysin (VitulaEligans)* *Peptidase (Scamander)* *E-value: 0.0E0*
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?	*The most informative HHPred match, Peptidase M23; membrane protein, enzyme, TRANSPORT PROTEIN; HET: ADP;{Vibrio cholerae}, has a probability of 99.65 with a database score of 119.07.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, the proposed function appears near known functions in that region of the genome in Wolfpack and Nobel, showing synteny.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes.*
DECISION:	*Peptidase*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**18**
Stop Coordinate	**14596**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**Overlap of 21**
Selected Start Coordinate	**14345**
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, the gene was called in two different loacations by Glimmer and GeneMark.*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Octopus, Nobel, and Kevanna.*
Does the gene violate any major guiding principles?	*This gene does not violate any major guiding principles*
DECISION:	*Yes, this is most likely a gene*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Original Glimmer calls bp 14345 has strength 11.69* *GeneMark calls bp 14333*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score for bp 14345 is –5.522 with a Z value of 1.637.* *The final score for bp 14333 is –6.724 with a Z value of 1.587.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Neither of the predicted ORFs are the longest, which results in excessive overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Start 9 was found in 80 of 141 genes in the pham.* *Start number 9 is 14345 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical membrane protein in PaoPU, TimoTea, and BurtonThePup.*
DECISION:	*@Bp 14345 based on Glimmer, coding potential, starterator,and 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?	*E-Value: 0.0E0*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, this protein aligns with* Cell division protein FtsB with the probability score of 96.17.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, endolysin.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, endolysin.*
DECISION:	*Endolysin*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**19**
Stop Coordinate	**14817**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**none**
Selected Start Coordinate	**14593**
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)?	*Original Glimmer call @bp 14593 has strength 14.77*
Is there evidence for coding potential?	*Yes.*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Loca, Cheesecake, and Snuffles.*
Does the gene violate any major guiding principles?	*The gene after this one is a reverse gene.*
DECISION:	*Yes, this is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: @bp 14593* *GeneMark Start Coordinate (type NA if not supported):: @bp 14593*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -3.171, and the Z score is 2.842 which is the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called most often in the published annotations is 5, it was called in 128 of 128 non-draft genes in the pham.* *Start 5 is 12780 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Holin [Microbacterium phage PaoPu, Noelani, Kaijohn]* *Membrane protein [Microbacterium phage Alucard]*
DECISION:	*@Bp 14593 based on Glimmer/Genemark, coding potential, phamerator, and starterator.*

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

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*E-Value: 4.0E-29*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, this protein aligns with* Phage_holin_7_1 with the probability score of 98.24.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, helix-turn-helix DNA binding domain.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, helix-turn-helix DNA binding domain.*
DECISION:	*Helix-turn-helix DNA binding domain*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**20**
Stop Coordinate	**14886**
Direction (For/Rev)	**reverse**
Gap (Overlap) with Previous Gene	**None**
Selected Start Coordinate	**15098**
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)?	*Original Glimmer call @bp 15098 has strength 15.72*
Is there evidence for coding potential?	*Yes, because both GeneMark programs show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Snuffles, Loca, and Cheesecake.*
Does the gene violate any major guiding principles?	*The gene before this is a forward and a reverse gene follows 20.*
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)::@bp 15098* *GeneMark Start Coordinate (type NA if not supported)::@bp 15098*
Does the start site have an associated Ribosome Binding Site with a high score?	*The final score is -3.627, and the Z score is 2.654 which is the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, the predicted start codon is the longest ORF.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called most often in the published annotations is 12, it was called in 125 of 159 non-draft genes in the pham.* *Start 12 is 15098 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Lsr2-like DNA bridging protein [Microbacterium phage PaoPu, Hulk, TeddyBoy, Quaker, HarperAnne]*
DECISION:	*@Bp 15098 based on Glimmer/Genemark, coding potential, phamerator, and starterator.*

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

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*DNA Master: Lsr2-like DNA bridging protein (PaoPu)* *E-value: 2.7E-20*
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?	*The most informative HHPred match, Protein lsr2; anti-parallel beta sheet, dimer, DNA BINDING PROTEIN; 1.728A {Mycobacterium tuberculosis}, has a probability of 99.86 and a database score of 101.74.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The gene is located adjacent to genes of known function, showing high conservation of gene order and synteny, in Yami and Wolfpack*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Lsr2-like DNA bridging protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**21**
Stop Coordinate	**15101**
Direction (For/Rev)	**Reverse**
Gap (Overlap) with Previous Gene	**No Overlap**
Selected Start Coordinate	**15601**
Selected Function	**Helix-turn-helix DNA binding domain (Awaiting Approval)**

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 bp @15595 and GeneMark calls bp @15601*
Is there evidence for coding potential?	*Yes, because both GeneMark programs show coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Sippinontea, Gardevoir, and Vanisius.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp @15595* *GeneMark Start Cooridnate bp @15601*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp @15595, The associated RBS score is -6.398 and the Z value is 1.199, neither of which are the highest.* *Considering bp @15601, The associated RBS score is -7.945 and the Z value is 0.497, neither of which are the highest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The lengths of the ORF with the predicted start codons are 495 (bp @15595) and 501 (bp @15601), and does not result in overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 6, it was called in 87 of the 127 non-draft genes in the pham.* *Starterator says site 7 is bp @15595 in Zahlia*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Helix-turn-Helix DNA Binding Domain Protein for Chatham, Erin, and Gillker. DNA Binding Protein for PaoPu.*
DECISION:	*Bp @15595, based on GeneMark, coding potential, starterator, and 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?	*DNA Master: Helix-turn-helix DNA binding domain protein (Chatham)* *E-value: 0.0E0*
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?	*The most informative match on HHPred was AM32; Repressor, Protein-DNA complex,Type IV secretion system, DNA BINDING PROTEIN; HET: EPE; 1.93A {Enterococcus faeca, with a probability of 99.47 and a database score of 84.54.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The gene is in a region where there is named genes, displaying synteny with Octopus and Horacetta.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes. (Awaiting Dr. Galle's approval)*
DECISION:	*Helix-turn-helix DNA binding domain*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**22**
Stop Coordinate	**15682**
Direction (For/Rev)	**REV**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**15912**
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)?	*Both call bp 15912.*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Owens and Josuke*
Does the gene violate any major guiding principles?	*The previous gene is a reverse gene, however, the gene after this one is forward.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 15912* *GeneMark Start Coordinate bp 15912*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 15912, The associated RBS scores –7.839 and the Z value is 0.933. Neither of which were relatively high.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF is not the longest ORF and doesn't result in any gene overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 5, it was called in 124 of the 127 non-draft genes in the pham.* *Starterator says site 5 is bp 15912 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical transcriptional repressor for Scamander and hypothetical helix-turn-helix DNA binding domain for Godfather and VitulaEligans*
DECISION:	*Bp 15912, based on glimmer/genemark, coding potential, starterator, and 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?	*DNA Master: Transcriptional Repressor (Scamander)* *E-value: 0.0E0*
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?	*The most informative match on HHPred was ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATOR, with a probability of 98.84 and a database score of 40.79.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*The gene is in a region where there is named genes, displaying synteny with Octopus and Horacetta.*
Is the proposed function found on the SEA-PHAGES approved function list?	*No.*
DECISION:	*NKF.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**23**
Stop Coordinate	**16809**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**16423**
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 call @bp 16423 and GeneMark calls start at 16627.*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Josuke, Owens, and Clownery.*
Does the gene violate any major guiding principles?	*The gene before this is a reverse gene and gene 24 is a forward.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 16423* *GeneMark Start Coordinate bp 16627*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 16423 and bp 16627, The associated RBS scores are –5.825 and –5.712 and the Z values are 1.749 and 1.956 respectively.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF of 16423 and 16902 are 387 and 183 respectively, making start 16423 more desirable.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 2, it was called in 95 of the 95 non-draft genes in the pham.* *Starterator says site 2 is bp 16423 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Conserved in hypothetical protein Owens, Goobery, Aztec, TimoTea, and VitulaEligans.*
DECISION:	*Bp 16423, based on glimmer/genemark, coding potential, and 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?	*E-Value: 0.0E)*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No quality match, best match was* DEXHc_TRCF with a probability of 67.7.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No 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?	*No*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**24**
Stop Coordinate	**17111**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**16893**
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)?	*Glimmer call @bp 16893 and GeneMark calls start at 16902.*
Is there evidence for coding potential?	*Yes*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Charbie, JooneeDee, and LaviMo.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 16893* *GeneMark Start Coordinate bp 16902*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 16893 and bp 16902, The associated RBS scores are –4.952 and –6.478 and the Z values are 1.962 and 1.962 respectively.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF of 16893 and 16902 are 219 and 210 respectively, making start 16893 more desirable.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 12, it was called in 103 of the 130 non-draft genes in the pham.* *Starterator says site 12 is bp 16893 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Conserved in helix-turn-helix TimoTea, and MrGreen. Conserved in MerR-like helix-turn-helix DNA binding domain protein Scamander and PaoPu. Conserved in excisionase and transcriptional regulator Azizam.*
DECISION:	*Bp 16893, based on glimmer/genemark, coding potential, and 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?	*E-Value: 5.9E-44*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, this protein aligns with Recombination Directionality Factor RdfS with the probability score of 99.41.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, helix-turn-helix DNA binding domain.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, helix-turn-helix DNA binding domain.*
DECISION:	*Helix-turn-helix DNA binding domain*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**Zahlia**
Gene #	**25**
Stop Coordinate	**17407**
Direction (For/Rev)	**FOR**
Gap (Overlap) with Previous Gene	**N/A**
Selected Start Coordinate	**17108**
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)?	*Both call bp 17108.*
Is there evidence for coding potential?	*Yes, there is evidence for coding potential*
Is this gene present in other annotated genomes?	*Yes, it is conserved in Owens and Josuke.*
Does the gene violate any major guiding principles?	*The gene does not violate any major guiding principles.*
DECISION:	*This is most likely a gene.*

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

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate bp 17108* *GeneMark Start Coordinate bp 17108*
Does the start site have an associated Ribosome Binding Site with a high score?	*Considering bp 17108, The associated RBS score and Z value are not the highest compared to the rest.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*The length of the ORF is not the longest ORF and doesn't result in any gene overlap.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*The start number called the most often in the published annotations is 9, it was called in 127 of the 128 non-draft genes in the pham.* *Starterator says site 9 is bp 17108 in Zahlia.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Hypothetical HNH endonuclease in PaoPu Bri160, and Belthelas.*
DECISION:	*Bp 17108, based on glimmer/genemark, coding potential, starterator, and 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?	*E-Value: 0.0E0*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, this protein aligns with CRISPR with the probability score of 97.75, meaning it may be a protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, HNH endonuclease.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, HNH endonuclease*
DECISION:	*HNH endonuclease*