# **Jamun Annotation**

# **Phage Bioinformatics 2022**

**UNH-Manchester**

**Overall Schema of Gene Orientation in Final Annotated Version:**

**Forward Genes:**  1-24, 35-63

**Reverse Genes:** 25-34

**Presence of tRNA in Jamun**

* ARAGORN v1.2.41 found that no tRNA nor tmRNA genes were found.
* tRNASCAN-SE found one possible tRNA structure, however it was undetermined and its infernal score was less than 35 (31.5). No anticodon found in the program.

**Annotated Genes** take this form below:

**Jamun\_Final\_Gene\_X (stop codon)**

# Jamun\_1 (642): Start site annotated by KP, function annotated by NJ

* **Selected start site: 118**
  + Original Glimmer call bp 118 with a strength of 5.07/ Gene mark did not call the start site for this gene on DNA master
  + Star sites cover all coding potential on Genemark
  + On starterator, Jamun\_1 shares start site 118 with Basilisk
  + Start site 118 found in 2/2 pham members
  + There does not appear to be any coding potential found between Jamun 0bp and 117pb (blasted)
  + Start site 118 covers all of the coding potential in this bp region.
* **Protein Function: (Hypothetical Protein)** annotator: NJ reviewer: AS (03/25/2022)
  + No relatives identified the function on phagesdb
  + According to phagesdb, there are no significant similarities found. The closest related function is a tape measure protein with a e-value of 1.5.
  + According to HHPred, no useful information was found
  + In NCBI Blastp, no significant similarity was found

# Jamun\_2 (1022): Annotated by MSG 3/4/22

* **Selected start site: 714**
  + Original Glimmer call at 714bp has strength 8.33
  + Start site called by Glimmer is 714, which completely encloses all coding potential
  + No close relatives have picked this site.
  + 714 has 1 Manual Annotation from Basalisk\_2
  + 714 has been chosen 100% of the time, when present. (Only present in 2/22 genes in pham)
  + Previous gene Jamun1 stops at 642 and has a 72bp gap between the start of Jamun2, which is 714.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: Terminase small subunit**
  + NCBI BLASTP says it is most likely a ‘terminase small subunit’ with relatives Brynnie, Galaxy, and Orcanus also having it, query cover is 99% with an e-value of 6e-66
  + HHPred says it is most likely terminase small subunit with a 95.3% probability and an e-value of 0.2
  + Both sites seem to be in agreement, so it is most likely a terminase small subunit

# Jamun\_3 (3/4/22): Annotator: RV Reviewer: MB

* **Selected start site: 1027**
  + Start 961 includes most coding potential, 1027 includes all coding potential
  + 961 has 62 bp overlap with Jamun\_2, 1027 has 4 bp gap
  + Original Glimmer call atbp 1027 has strength 14.74
  + Start: at961 has 14 MA's, Start: at1027 has 3 MA's
  + RBS score for 961= -2.219, 1027= -6.014
  + 1027, 1:1 alignment with 3 phages in subcluster on BLAST
* **Protein Function: Terminase large subunit** NJ, reviewed by AS
  + Brynnie identified terminase large subunit, and Abidatro and Galaxy identified terminase
  + According to PhagesDP Blastp, terminase large subunit.
  + According to HHPred, the probability of a terminase large subunit function is 100% and the e-value is 4e-39.
  + In NCBI Blastp, this gene is 99.24% identical to Brynnie, the query covers 100% e-value 0.0.

# **Jamun\_4 (2941):** Annotator: (MJB) reviewer:(RV) date 3/2/22

* **Selected start Site: 2627**
  + Original Glimmer call at 2627/Gene Mark strength 12.12
  + start sites cover all coding potential
  + 12/60 pham members have manually annotated the start site at 2627
  + start 2627 has been called 84.2% of the time, when present.
  + 4 manual annotations for AS1 pham members
  + Previous gene, Jamun3 stops at 2607 and has a 20 bp gap between the start of Jamun4, which is 2627.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: RNA binding protein**
  + Other pham members have called protein a RNA binding protein on PhagesDB
  + According to PhagesDP Blastp, function is RNA binding protein with a e-value of 3e-52
  + No useful information derived from HHPred.
  + In NCBI Blastp, this gene is 100% identical to Brynnie, which has been identified as an RNA binding protein

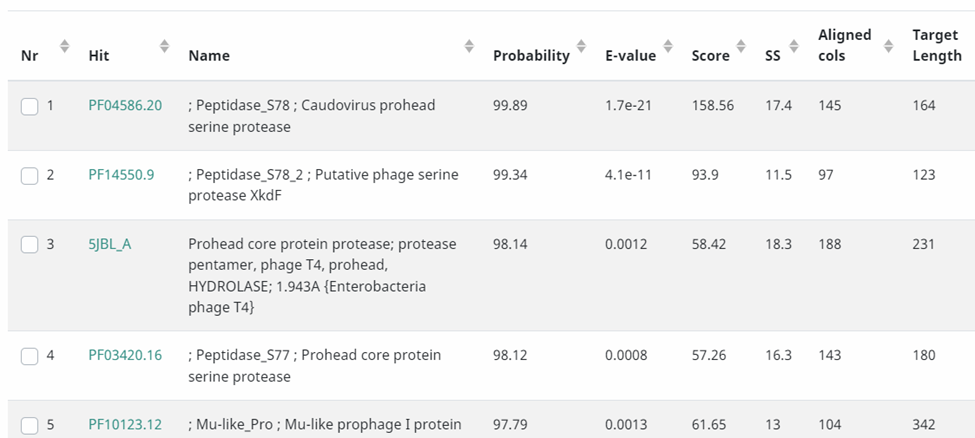
# **Jamun\_5 (4916)** 3/4/22 Annotator: JF Reviewer: LC

* **Selected start site: 2958**
  + Original Glimmer call atbp 2958 has strength 14.92
  + 2958 is the only start site to capture all the coding potential of the gene
  + Close relatives of JamunDraft Abidatro and Brynnie manually annotated start at 2958
  + In Pham, start site 2958 has 142 Manual Annotations out of 193 manually annotated genomes in which 2958was present.
  + 2958 was manually annotated 4 times for cluster AS1
  + There is an 18 bp gap with the upstream gene Jamun4, with no coding potential in the gap
  + RBS Final Score of -3.792, Z-Value of 2.729
  + Blast Query:Subject Ratio 1:1
* **Protein Function: Portal Protein**
  + Phagesdb pham members have annotated portal protein or portal and MuF-like fusion protein with e-values of 0.
  + NCBI showed that Abidatro and Galaxy had called this gene as a portal protein, 94.02% identical to Abidatro
  + HHPred showed that this is similar to a bacteriophage portal protein

# Jamun\_6 (5699): Annotated in class 2/18/22

# **Selected start site: 4941**

* + Original Glimmer call atbp 4941 has strength 12.56; GeneMark calls start at 4968
  + Only the start sites called by Glimmer (4941) and GeneMark (4968) completely encase all coding potential for this gene
  + Close relatives Brynnie and Abidatro have manually annotated the start site at 4941
  + None of the other start sites present in Jamun have been MA, including 4968.
  + 4941 has 7 MA’s (Manual Annotations)
  + 4941 has been chosen 100% of the time, when present.
  + Previous gene Jamun5 (4916) stops and has a 25bp gap between the start of Jamun6, which is 4941.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: Capsid Maturation Protease**
  + NCBI lists all the top functions as capsid maturation proteases, with good e-values.
  + SEAPHAGES database: “sometimes the CMP hits to ClpP proteases. If so, look for a serine-type endopeptidase activity”
  + HHPred results align with the conditions set by the SEAPHAGES database



# Jamun\_7 (7069): annotated in class 2/18/22

# **Selected start site: 5765**

* + Original Glimmer call atbp 5726 has strength 15.01; GeneMark calls start at 5765
  + Start site called by Glimmer (5726) captures all coding potential, while GeneMark (5765) captures the majority of the coding potential.
  + Close relatives Brynnie and Abidatro have manually annotated the start site at 5765.
  + 5765 has 7 Manual Annotations
  + 5765 has been chosen 84.6% of the time, when present.
  + Previous gene Jamun (5699) stops and has a 66bp gap between the start of Jamun (7069).
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: Major capsid Protein**
  + All annotated pham members have a function “major capsid protien”.
  + All fully annotated pham members with a 0.0 E-value claim major capsid protein
  + According to HHPred, the probability of a Major capsid protein is 100% and the e-value is 4.1x10-37.
  + In NCBI Blastp, this gene is 95.85% identical to Brynnie and all search items come up with major capsid protein

# Jamun\_8 (7432)

* **Start Site Evidence: (7154**) annotator: NJ reviewer: AS (03/04/2022)
  + Original Glimmer call 7154.
  + Significant amount of coding potentials covered, but not all.
  + Close relatives Brynnie have manually annotated the same start site.
  + 7154 has 10 manual annotations.
  + 7154 has been called 93.3% of the time, when present.
  + Previous gene Jamun7 stops at 7069bp and has an 85bp gap between the start of Jamun8 at 7432bp.
  + Evaluating the gap, there NCBI blastx showed no significant similarity even though there is coding potential in the gap.
  + RBS: final score is -3.168 and the z score is 2.885
* **Protein Function Hypothetical Protein**
  + Chosen protein: Hypothetical Protein (RV 3/11/22)
  + Hypothetical Proteins for all matches on BLAST, 91.43% similar to Brynnie and 77% similar to Orcanus
  + No useful information on HHPred
  + PhagesDB blastp shows no known functions for annotated phages
  + According to Phagesdb Blastp, the e-value probability of a Hypothetical Protein is 5ex10-30.
  + No additional information on SEA-PHAGES

# Jamun\_9 (8041) Annotated by KP

* **Start site evidence: 7452**
  + Original Glimmer and genemark call bp 7454 with a strength of 9.93
  + There does not appear to be any coding potential outside of the called start site.
  + On starterator, there are 16 manual annotations for this start site
  + This start site is called 95% of the time that it is found present.
  + Start site 7454 covers all of the coding potential on this gene.
  + There is a 22 bp gap where 8 ends and 9 begins.
  + There is no coding potential for another gene
* **Protein Function: Head to tail adapter**
  + NCBI hits show that this is 97.44% identical to Brynnie’s head to tail adapter.
  + HHPred hits show that this is also a possible head to tail adapter for bacteriophages.
  + Phagesdb shows many pham members have called this protein a head-to-tail adapter

# Jamun\_10 (8414): Annotated by MSG 3/4/22

* **Chosen Start Site: 8052**
  + Original Glimmer call atbp 8052 has strength 9.82; GeneMark calls start at 8079
  + Start sites called by Glimmer (8052) and GeneMark (8079) completely enclose all coding potential.
  + Close relatives Brynnie and Abidatro have manually annotated the start site at 8052, as well as Galaxy and Orcanus.
  + 8052 has 16 MA’s (Manual Annotations)
  + 8052 has been chosen 100% of the time, when present.
  + There is an 11 bp gap between the upstream gene, Jamun (8041), and this chosen start site of Jamun (8414).
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: head-to-tail stopper**
  + Abidatro and Brynnie have called this a head-to-tail stopper both in Phamerator and PhagesDB Blastp
  + According to PhagesDP Blastp, this gene is likely to be a head-to-tail stopper.
  + According to HHPred, the probability of a head-to-tail adapter (stopper) is 99.39% with an e-value of
  + In NCBI Blastp, this gene is 98.33% identical to Brynnie, which was identified as head-to-tail stopper, with an e-value of 4e-80.

# Jamun\_11 (8676) (3/4/22) Annotator: RV Reviewer: MB

* **Chosen Start Site: 8419**
  + Original Glimmer call atbp 8419 has strength 9.75
  + Start 8419 captures all coding potential
  + There is a 4 bp gap between this chosen start and with the upstream gene, Jamun (8414)
  + Start site 8419 has 35 Manual annotations
  + RBS final score for 8419is -3.340
  + 8419, 1:1 alignment with 2 phages in subcluster on BLAST
* **Protein Function: Hypothetical Protein** (MJB 3/29/22)
  + All other AS1 phages and pham members have no known function on phagesdb blast, but low e-values.
  + Hypothetical Protein 83.5% identical to Galaxy on NCBI
  + HHPred has no useful results, but 89.93% likely to have an unknown function

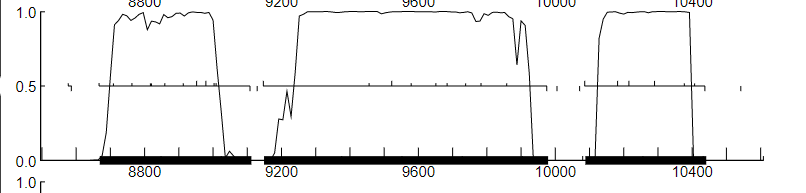
# **Jamun\_12 (9112)** annotator:(MJB) reviewer:(RV) date 3/2/22

* **Chosen Start Site: 8669**
  + Original Glimmer call at bp 8669 has strength 6.61
  + start sites cover all coding potential
  + 20/20 pham members have manually annotated the start site at 8669
  + start 8669 has been called 100% of the time, when present.
  + 4 manual annotations for AS1 pham members
  + Start site 8669 has a 7 bp overlap with the upstream gene Jamun (8676)
  + Evaluating the gap between start site 8669 and the upstream gene Jamun(8676), there is no coding potential for another gene.
* **Protein Function: tail terminator** 
  + All phages in pham annotated function as tail terminator based on Phagesdb Blastp
  + All AS1 annotated phages have determined Tail terminator as gene function.
  + HHPred has 99.1% probability of a tail terminator protein and the e-value is 6.6e-9.
  + In NCBI Blastp, this gene is 98.6% identical to Brynnie, which has been identified as a tail terminating protein.

# Jamun\_13 (9979) Annotator: JF Reviewer: LC 3/4/22

* **Selected start site: 9149**
  + Original Glimmer call at bp 9149 has strength 16.15
  + 9149 Captures all coding potential of the gene, it is the only start site to do so
  + Start 50 @9149 has 200 Manual Annotations, and was called 86.8% of the time when present
  + Close relatives of JamunDraft Abidatro and Brynnie both called start site 50 (9149)
  + Start 50 @9149 was Manually annotated 4 times in cluster AS1
  + There is a 38 bp gap between this chosen start site and the upstream gene Jamun (9112), Genemark shows no coding potential for another gene in this gap
  + RBS Final Score of -3.168, Z-Value of 2.886
  + Blast Query:Subject ratio 1:1
* **Protein Function: major tail protein** annotator: RV reviewer: MB (3/22/2022)
  + Phages Abidatro and Brynnie label gene as major tail protein
  + According to PhagesDP Blastp this gene is a major tail protein
  + According to HHPred, the probability of a major tail protein is 93.3% and the e-value is .000086.
  + In NCBI Blastp, this gene is 98% identical to phage Brynnie major tail protein, query covers 99%, e-value 0.0.

# Jamun\_14 (10441) annotator: (AS) reviewer: (NJ) (3/11/2022)

* **Selected Start Site**: **10088**
  + Original Glimmer call at bp 10088 has strength 12.80
  + Start site covers all coding potential.
    - 
  + Close relatives Abidatro\_14 and Brynnie\_14 have called the start site at 10088.
  + 10088 has 6 manual annotations.
  + 10088 has been called 100.0% of the time, when present.
  + There is a 109 bp gap between this chosen start site and the upstream gene Jamun (9979)
  + Evaluating the gap, there is no coding potential for another gene.
* **Annotating the tail assembly chaperone**
  + GGGGGAA slippage occurs at 10411 and ends at 10417
  + Gene 14 and Gene 15 protein sequences diverge at the slippage site beginning at genome position 10411.
  + The first protein (shown below as blue highlighting) is in reading frame 2 and the predicted second protein sequence for Jamun\_15 (before the translational frameshift was predicted) is shown in reading frame 1 (green), indicating a -1 frameshift. Slippage sequence in frame 2 is shown with blue underlining below.
  + There are identical slippage sequences in Galaxy and Orcanus that both feature a -1 frameshift with the GE to GG amino acid sequence.



# Jamun\_15 (10812)

* This is the tail assembly chaperone programmed frameshift, and shares a start site with Jamun\_14.
* The programmed frameshift in Frame 2 above, produced a Jamun\_15 protein product that finishes in Frame 2 with VG amino acid sequence and continues in Frame 1 with GIGSLCLLLEEY….
* **Protein Function: tail assembly chaperone**
  + Phagesdb BLASTP shows a tail assembly chaperone for relatives Abidatro, Galaxy, Brynnie, and Orcanus. All have low e-values (<9e-58)
  + NCBI BlastP shows tail assembly chaperones for the vast majority of results, including Brynnie, Orcanus, Abidatro and Galaxy. The top result is Brynnie - 90.12% identical with 1e-146 e-value.
  + The only relevant HHPred result is “family of unknown function” with a probability of 97.98% and an e-value of .000038

# Jamun\_16 (13221)

* **Chosen Start Site: 10816** annotator: NJ reviewer: AS (03/04/2022)
  + Original Glimmer call 10816.
  + All coding potential is captured by start 10812
  + Close relatives Brynnie and Abidatro have manually annotated this start site
  + 10816 has 10 manual annotations.
  + 10816 has been called 100% of the time, when present.
  + There is a 4bp gap between this chosen start site and the upstream gene Jamun (10812)
  + Evaluating the gap between this chosen start and the upstream gene Jamun (10812), there is no coding potential for another gene.
  + RBS: final score is -3.515 and the z score is 2.729
* **Protein Function: tape measure protein**
  + Other AS1 and other pham members have called this a tape measure protein on Phagesdb
  + Good blast hits on Phagesdb for tape measure proteins
  + 97.88% similar to Brynnie’s tape measure protein via BLASTP
  + HHPred results say Jamun (13221) is 99.91% likely to be a tape measure protein

# Jamun\_17 (14063) Annotated by KP

* **Chosen Start Site: 13224** 
  + Original Glimmer and genemark call bp 13224 with a strength 14.52
  + There are 23 manual annotations for start site 13224
  + This gene is called 100% of the time that it is found present.
  + Start site 13224 covers all of the coding potential on this gene.
  + There is a 4 base pair gap between Jamun (13221) and this chosen start site
  + There does not appear to be any coding potential in the gap between start 13224 and the upstream gene Jamun (13221)
* **Protein Function: minor tail protein**
  + Phagesdb BlastP shows Minor Tail Protein in Basilisk, Brynnie and Orcanus
  + NCBI Blastp lists the top 6 functions as a minor tail protein, all with good e-values. The top result is Brynnie- 97.2% identical, e-value 5e-176
  + HHPred found no significant results

# **Jamun\_18** (15324): Annotated by MSG 3/4/22, 3/25/22

* **Selected Start Site:14074**
  + Original Glimmer call atbp 14074 has strength 14.42
  + Glimmer (4941) start site completely covers all coding potential.
  + Close relatives Brynnie and Orcanus have manually annotated the start site at 14074.
  + 14074 has 31 MA’s (Manual Annotations)
  + 14074 has been chosen 95.1% of the time, when present.
  + There is a 11bp gap between this start site and the upstream gene, Jamun\_17 (14063)
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: minor tail protein** 
  + All annotated pham members have called minor tail protein as a function
  + According to PhagesDP Blastp all annotated genes have a minor tail protein function with an E-value of 0.0.
  + According to HHPred, the probability of a minor tail protein is 99.91% and the e-value is 6.6e-21.
  + In NCBI Blastp, this gene is 98.56% identical to Brynnie with a minor tail protein

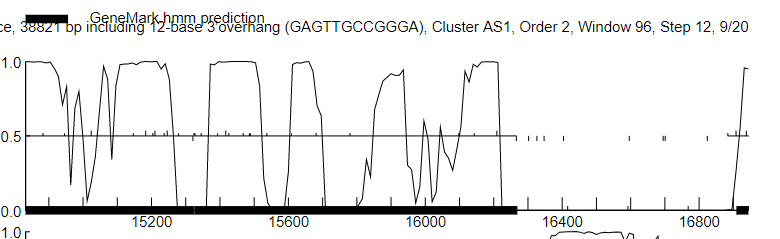
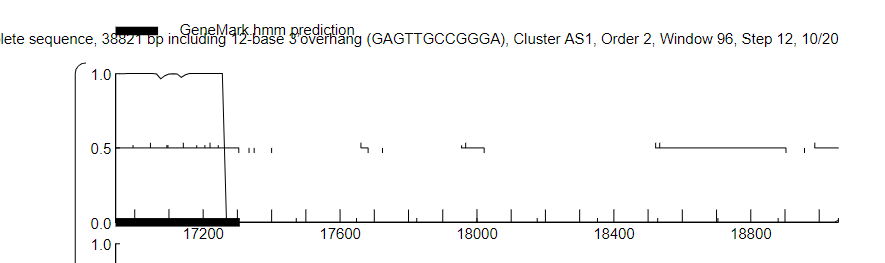
# Jamun\_19 (16269): annotator: NJ reviewer: AS (03/04/2022)

* **Selected Start Site: 15328**
  + Original Glimmer call 15328 and Gene Mark call 15325
  + All coding potentials covered
  + Close relatives Brynnie, Basilisk, and Orcanus have manually annotated the same start site
  + 15328 has 1 manual annotation.
  + 15328 has been called 92.7% of the time, when present.
  + There is 4bp gap between this start site and the upstream gene, Jamun\_18(15324).
  + Evaluating the gap, there is not coding potential for another gene.
  + RBS: final score is -5.368 and the z score is 1.788
* **Selected Function: minor tail protein (RV 3/11/22)**
  + 2 phages from AS1 subcluster show “minor tail protein” with low e values on BLAST
  + No useful information on HHPred
  + SEA-PHAGES notes - “If you have significant hits to either collagen-like or glycine-rich proteins, and are in the syntenic region of minor tail proteins, you can call them minor tail proteins”

# Jamun\_20 (16882): Annotator LC: Reviewer: JF

* **Selected Start Site:** 
  + Original glimmer call at 16319 has strength 9.82; no genemark calls
  + Start site called by glimmer (16319) completely encloses all coding potential for this gene
  + Close relatives Brynnie, Basilisk, and Orcanus all have annotated the start site at 16283
  + No manual annotations for start site called by glimmer (16319)
  + 12 Manual Annotations for start site 16283
  + There is a 14 bp gap between this start site and the upstream gene Jamun (16269).
* **Protein Function: Hypothetical Protein**
  + NCBI showed that there is a relation between this gene and the one found in Brynnie called a Hypothetical Protein
  + HHPred did not show any useful information.

# **Jamun\_21 (17307):** Annotator: AS Reviewer: NJ (03/11/2022)

* **Start Site Evidence: (16885)**
  + Original Glimmer call atbp 16885 has strength 11.34; GeneMark calls start at 16909.
  + Choosing the second starting tick near the 16885 mark because it has less of a gap between before the coding potential yet still covers all of it up to the end codon 17307.
    - 
    - 
  + Start: 11 @16885 has 15 MA's
  + Found in 19 of 22 ( 86.4% ) of genes in pham
  + Called 100.0% of time when present
  + There is a 27bp gap between this start site and the upstream gene, Jamun\_21(17307).
  + No visible overlapping of potential coding upstream of gene.
  + RBS: final score is -4.924 and Z value is 1.937.
* **Protein Function: Hypothetical Protein**
  + Abidatro and Brynnie have not identified this gene in Phamerator or in PhagesDB.
  + According to PhagesDP Blastp, it has an unidentified function.
  + According to HHPred, the lowest similar gene only has an e-value of 110, which means its likelihood is very small.
  + In NCBI Blastp, this gene is 97.14% identical to Brynnie’s, which was only identified as a Hypothetical Protein.

# Jamun\_22 (17621)

* **Selected start site: 17334**
  + Original Glimmer call atbp 17334 has strength 8.51
  + Start at17334 was manually annotated 24/44 times
  + his start was called 100% of the time when present
  + The start site at 17344 was manually annotated 4 times for cluster AS1, including close relatives of JamunDraft, Abidatro and Brynnie
  + There is a 17 bp gap with the upstream gene, Jamun21. There doesn’t seem to be any coding potential in this gap
  + RBS final score of -5.459, Z-value of -4.764
  + Blast Query:Subject ratio 1:1
* **Protein Function: Hypothetical Protein**
  + Relatives Abidatro and Brynnie have not called this gene function
  + NCBI blastP shows the top matches as Hypothetical Protein with low e-values and percentages 96-98%
  + The fourth likely HHPred hits is to “Protein of unknown function” with a probability of 94.8%. However, all of the top hits have poor e-values greater than .06

# **Jamun\_23 (18722)** annotator:(MJB) reviewer:(RV) date 3/10/22

* **Selected start site: (17682)**
  + Original Glimmer call at 17682 /Gene Mark strength 14.01
  + start sites cover all coding potential
  + 17/113 pham members have manually annotated the start site 17682)
  + start 17682 has been called 72.5% of the time, when present.
  + 2 manual annotations for AS1 pham members
  + There is a 61bp gap between this start site and the upstream gene, Jamun (17621).
  + There is no coding potential for another gene in the gap between this start site and the upstream gene Jamun (17621)
* **Protein Function: endolysin** annotator: RV reviewer: MB (3/22/2022)
  + Phages Brynnie and Orcanus label gene as endolysin
  + According to PhagesDP Blastp, vast majority endolysin
  + According to HHPred, the probability of an endolysin is 99.29% and the e-value is 2.4e-10.
  + In NCBI Blastp, this gene is 95% identical to phage Brynnie endolysin, query covers 95%, e-value 0.0.
  + SEA-PHAGES: “some arthrobacter and streptomyces phages have a single endolysin with domains not found in the Mycobacteriophages (like the CHAP domain). use "endolysin" rather than lysin A if the phage does not infect Mycobacterium and no lysin b can be identified.”

# Jamun\_24 (19058) (3/11/22) Annotator: RV Reviewer: MB

# **Selected start site: 18726**

* + Start 18726 includes all coding potential on GeneMark
  + There is a 4bp gap between this start site and the upstream gene, Jamun\_23(18722).
  + Original Glimmer call atbp 18726 has strength 8.06
  + Start: @18726 has 15 MA's
  + Called 71% of the time when present, no other start site has manual annotations
  + -3.210 RBS score for 18726
  + 1:1 alignment with 4 phages in AS1 subcluster on BLAST
* **Protein Function: Hypothetical Protein**
  + Brynnie and Orcanus are some phages who have a similar genetic sequence to this gene.
  + PhagesDB Blast contained no useful information.
  + NCBI contained information saying that this is a possible membrane protein
  + HHPred contained no useful information. Therefore, gene function is unknown.

*Orientation switch to reverse genes through gene 34*

# Jamun\_25 (19188) Reverse: Annotated by MSG 3/7/22

* **Selected start site: 19385**
  + Original Glimmer call atbp 19385 has strength 0.72 \*\* not called by GeneMark
  + Start site called by Glimmer (19385) captures all of the coding potential.
  + Close relative Brynnie has manually annotated the start site at 19385.
  + 19385 has 1 MA’s (Manual Annotations)
  + 19385 has been chosen 66.7% of the time, when present.
  + None of the other start sites of Jamun have been MA’d.
  + There is a 130bp gap between this start site and the upstream gene, Jamun\_24(19058).
  + There doesn’t appear to be any more coding potential for another gene in between in GeneMark.
* **Protein Function: helix-turn-helix DNA binding domain protein**

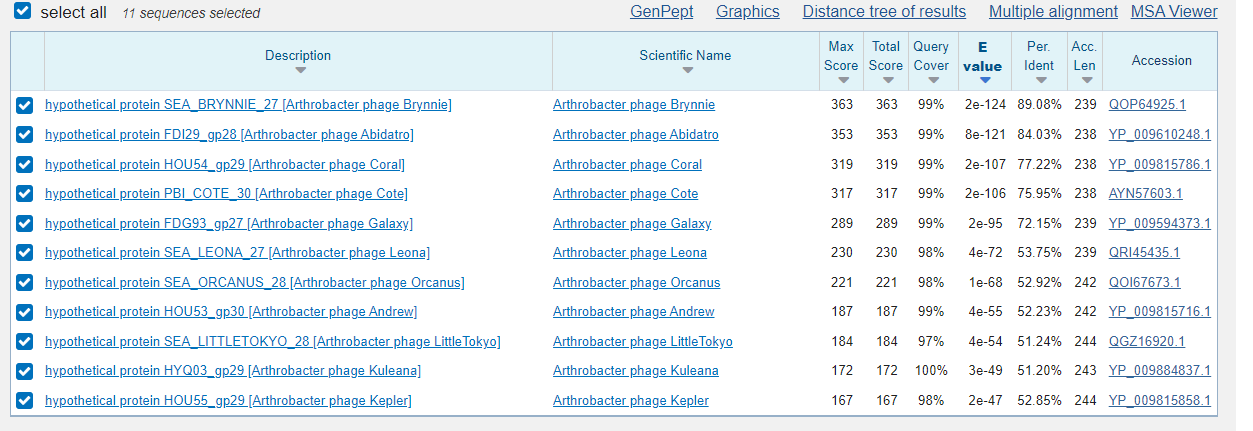
annotator: NJ reviewer: AS (03/11/2022)

* + Brynnie, Galaxy, Abidatro identified a similar gene with the same function
  + According to PhagesDP Blastp, helix-turn-helix DNA binding domain protein
  + According to HHPred, the probability of a helix-turn-helix DNA binding domain protein is 98.33% and the e-value is 0.0000078.
  + In NCBI Blastp, this gene is 94.44% identical to Brynnie, query covers 83% e-value 5e-29.

# Jamun\_26 (19382) Reverse

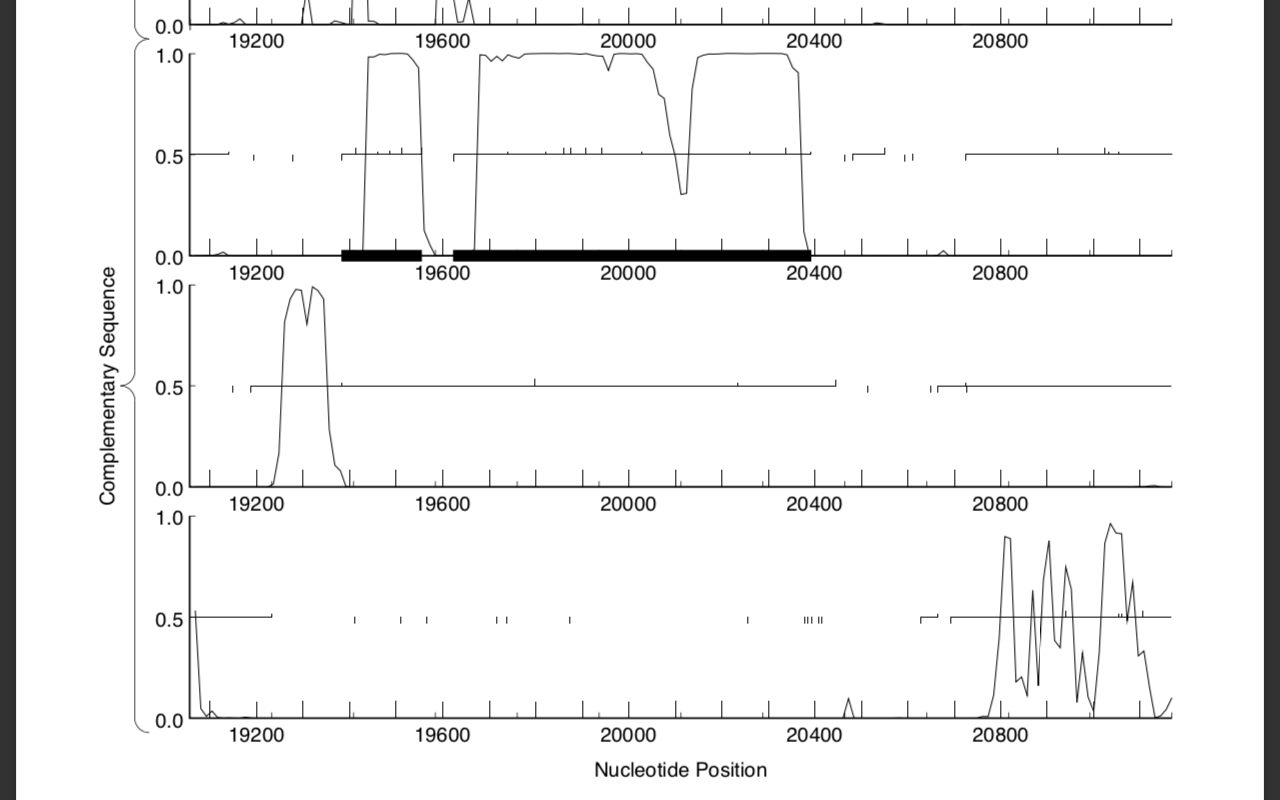
* **Start Site Evidence: 19555** (Annotated by KP)
  + Original Glimmer and gene mark call bp 19555 with a strength of 12.00
  + There is one manual annotation for this start site.
  + This gene is called 100% of the time that it is found present.
  + Start site 19555 does appear to cover all of the coding potential of this gene.
  + There is a 67 base pair gap where Jamun (19622) ends and Jamun (19382) begins.
* **Protein Function: Hypothetical Protein**
  + NCBI BLASTP says it is most likely a Hypothetical Protein with close relative Brynnie also having itm query covers 98% e-value 4e-42
  + HHPred’s highest hit is a protein of unknown function. Probability is 92.94% with an e-value of .22
  + Since both sites seem to be in agreement, i’m going to say it’s a Hypothetical Protein

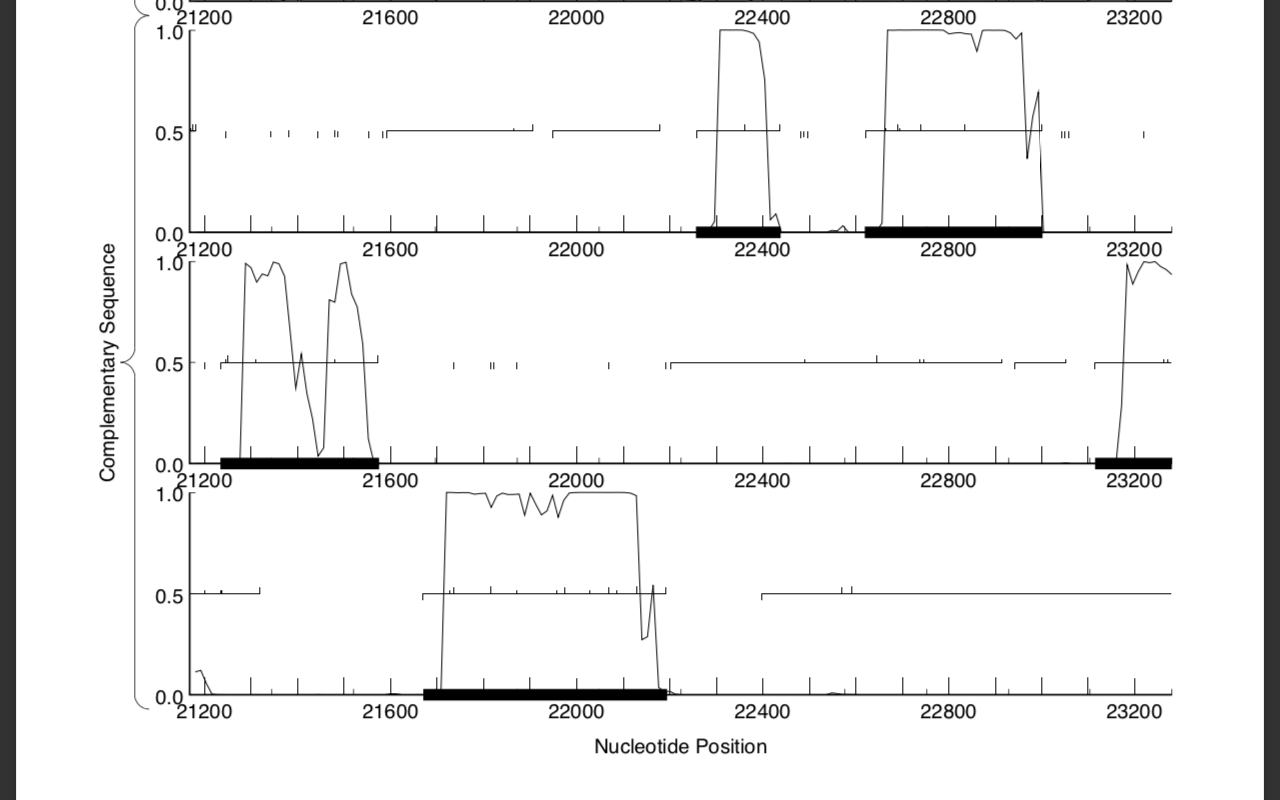
# Jamun\_27 (19622) Reverse

* **Selected start site: 20338** annotator: NJ reviewer: AS (03/11/2022)
  + Original Glimmer call 20338 and Gene Mark call 20392
  + All coding potential covered
  + Close relatives Abidatro, Orcanus, Brynnie and Basilisk have manually annotated the 20338 start site
  + 20338 has 14 manual annotations.
  + 20338 has been called 100% of the time, when present.
  + There is a 353 bp gap between the chosen start (20338) and Jamun (20692).
  + RBS: final score is -2.865 and the z score is 2.989
* **Protein Function (Hypothetical Protein)** Annotator: AS Reviewer: NJ
  + According to HHPred, the probability of an unknown function is 77.04% and the E-value is 41.
  + In NCBI Blastp, this gene is 89.08% identical to a Hypothetical Protein for Arthrocacter phage Brynnie, query covers 99%, E-value 2e-124.
    - 

# Jamun\_28 (20692) Annotator: RV (4/26/22)

* Gene manually added based on coding potential in GeneMark and similarity to other phages in AS1 subcluster (phage Abidatro and phage Brynnie)



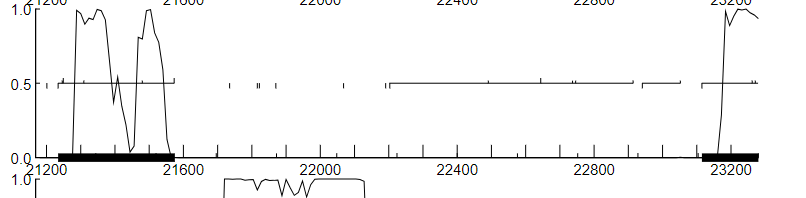


* **Selected start site: 21237**
  + 4 bp overlap (GTGA) with Jamun (21234), highly suggests correct start site
  + 21237 covers all coding potential in GeneMark
  + In BlastX, 1:2 alignment with phage Abidatro and 1:1 alignment with phage Brynnie with start site 21237
  + RBS: final score -4.374, Z value 2.202
* **Protein Function: HNH endonuclease**
  + In Blastx, gene is 91% identical to HNH endonuclease from phage Brynnie e-value 9e-62, 88% identical to HNH endonuclease from phage Abidatro e-value 5e-73
  + In HHPred, “CRISPR-associated endonuclease Cas9/Csn1, HNH domain” with 98.43% probability 5.7e-7 e-value

# Jamun (21132) Forward DELETED

* Small forward gene (orpham) in the middle of reverse genes, violates the Guiding Principles, with no homologs or obvious reasons this is a real gene. Could not find a reason to keep it.

# Jamun\_29 (21234) Reverse

* **Start Site Evidence: (21575)** Annotator: AS Reviewer: NJ (03/11/2022)
  + Original Glimmer call at bp 21575 has strength 4.06
  + Start site covers all of coding potential up to stop codon 21234.
    - 
    - Choosing the first tick at around the 21575 mark → covers all of the coding potential up to stop codon 21234.
  + Shares start site # 106 at 21575 with Abidatro and Brynnie
  + Found in 18 of 403 ( 4.5% ) of genes in the pham
  + Called 55.6% of the time when present
  + Gap of 94 bp between start codon 21575 and Jamun (21670).
  + No visible overlapping of potential coding upstream of the gene.
  + RBS: final score is -2.583 and the Z value is 3.312.
* **Protein Function: helix-turn-helix DNA binding domain protein**
  + Blast on phagesdb, all annotated genes in AS1 pham have called function of helix-turn-helix DNA-binding domain protein
  + HHPred blast 99.05% probability, for HTH protein, 99.2% probability for DNA binding Protein.
  + NCBI 73.39% identical to Orcanus with e-value of 1e-51 for helix turn helix DNA binding protein.

# Jamun\_30 (21670) Reverse Annotator:(JF) reviewer:(RV)

# **Selected start site: 22194**

# Original Glimmer call atbp 22194 has strength 10.53

# Start at 22194 captures all the coding potential of the gene

# Start 3 at bp 22194 was manually annotated by close relative of JamunDraft, Abidatro

# Site was called 100% of the time it was present

# There is a 62 bp gap with the upstream gene Jamun (22256), there doesn’t appear to be significant coding potential in this gap

# RBS final score of -4.557, Z value of 2.404

# Blast Query:Subject ratio 32:4

# **Protein Function: Hypothetical Protein** (annotated by RV 3/11/22)

* + Only Hypothetical Protein matches on BLAST
  + No useful information on HHPred
  + PhagesDB blastp only unknown functions
  + No additional information on SEA-PHAGES

# **Jamun\_31 (22256) Reverse** annotator:(MJB) reviewer:(RV) date 3/10/22

* **Selected start site: (22438)**
  + Original Glimmer call at 22438/Gene Mark strength 10.59
  + start sites cover all coding potential
  + 5/15 pham members have manually annotated the start site 22438
  + 22438 has been called 71.4% of the time, when present.
  + 3 manual annotations for 22438 (Orcanus)
  + Previous gene, Jamun (21670) starts at 22194 and has a 62 bp gap between the stop of Jamun31, which is 22256.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function (Hypothetical Protein)**
  + No phages in cluster have chosen a function
  + According to PhagesDP Blastp, Hypothetical Protein.
  + According to HHPred, the probability of a DNA binding protein is 92.02% and the e-value is 1.4. The high e-values of all the top results show that they should probably be disregarded.
  + In NCBI Blastp, this gene is 90% identical to Orcanus, which has not been identified and is still a Hypothetical Protein.

# Jamun (22399) DELETED (3/11/22) Annotator: RV Reviewer: MB

* **Start Site Evidence**
  + No coding potential in Genemark
  + Would have 40 bp overlap with Jamun\_31 (22256) and 25 bp gap with Jamun\_32 (22619). Deleting the gene produces a gap of 181 bp between Jamun\_31 (22256) and Jamun\_32 (22619). This is similar in size to the gap between AS1 clustermate Orcanus\_32 and \_33, the homologs of Jamun\_31 and \_32. The Orcanus gap is 160 bp. In AS3 relative Andrew, the gap is 144 bp between homologs Andrew\_35 and \_36.
  + Original Glimmer call at bp 22593 has strength 4.48 \*\* not called by GeneMark
  + No starterator data, no Pham members (Orpham)
* **Protein Function**
  + Jamun32 is a orpham and has no other members in its pham.
  + In HHPred, the only possibilities of this gene had very high e-values (20 and 140), meaning they are very unlikely.
  + In NCBI Blastp, this gene is 86.36 identical to Abidatro’s, which was only identified as a Hypothetical Protein.
* **Conclusion: DELETE.** No relatives with obvious function, relatives AS cluster are members of same Pham on either side with a similar-sized gap between.

# **Jamun\_32 (22619) Reverse:** Annotated by MSG 3/7/2022

* **Selected start site: 23002**
  + Original Glimmer call atbp 23002 has strength 10.91
  + Start site called by Glimmer (23002) does capture all coding potential.
  + Close relative Orcanus has manually annotated the start site at 24827.
  + 23002 has 2 MA’s (Manual Annotations)
  + 23002 has been chosen 100% of the time, when present.
  + Jamun (23115) has a 12 bp gap with start 22002.
  + There doesn’t appear to be any more coding potential for another gene in between in GeneMark.
* **Protein Function: Hypothetical Protein**
  + PhagesDB has no published function for other pham members
  + All blast information from phagesdb for AS1 phages shows no known function, other hs have tail assembly chaperone with e-values starting at 0.36
  + According to HHPred, the probability of a MYM-type function is 90.41% and the e-value is 0.46.
  + In NCBI Blastp, this gene is 86.96% identical to Orcanus and 73.9% identical to bacteriophage Andrew, which have not been identified and are still called Hypothetical Proteins.

# Jamun\_33 (23115) Reverse

* **Start Site Evidence: 24299**
  + Original glimmer and genemark call bp 24299 with a strength of 4.08.
  + There are 43 manual annotations for this start site.
  + This gene is called 98% of the time that it is found present.
  + Start site 24299 does appear to cover all of the coding potential for this gene.
  + There is an 8 base pair overlap between Jamun (24292) and start 24299.
* **Protein Function: tyrosine integrase**
  + Brynnie and Galaxy are some phages that have been found to share a similar sequences of this gene
  + PhagesDB Blast contained information stating that this sequence is similar to Brynnie’s gene that has been assigned the function Tyrosine Integrase.
  + NCBI contained information stating that this gene is a possible Tyrosine Integrase found in Brynnie and Galaxy
  + HHPred contained information stating that this is a Tyrosine Integrase

# Jamun\_34 (24292) Reverse

* **Start Site Evidence: 24654** annotator: NJ reviewer: AS (03/11/2022)
  + Original Glimmer call 24441 and Gene Mark call 24654
  + 24654 would allow for all the coding potential to be covered
  + No Starterator information
  + Gene Jamun (23115) has an 8bp overlap with this gene Jamun (24292).
  + RBS: final score is -5.460 and the z score is 1.738
* **Protein Function: immunity repressor**
  + Brynnie and Orcanus have shared similar sequences to this gene.
  + NCBI showed results that it closely resembled both Brynnie and Orcanus which have both assigned this gene the function of immunity repressor
  + HHPred showed no useful information.
  + Phages DB blast showed that gene 35 is similar to Abidatro, Brynnie, and Orcanus all calling the same function.

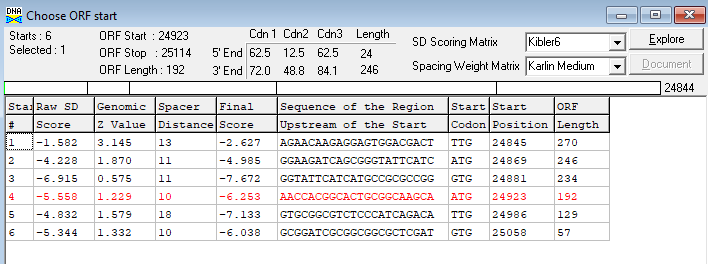
# Jamun (24564) : DELETED 2/25/22 IN CLASS (REACHED CONSENSUS)

* Jamun (24292) is the immunity repressor and the start site was extended substantially to match the GeneMark annotation, per evidence above. Because of this Jamun (24564) would completely overlap immunity repressor Jamun\_34 (24292), violating the Guiding Principles. In addition:
* Gene mark did not call this gene although glimmer did
* No start sites have any coding potential
* No evidence of homologs or ORF matches by BLAST or HHPred
* With this gene deleted the gap between preceding and following genes (see below) is 191 bp. This is slightly smaller than the gap between these genes seen in AS1 Orcanus and AS3 Andrew (~210 bp).
* Eesa and Renna12 have this gene but they are drafts
* **CONCLUSION: DELETE**

*Orientation switch to forward genes for remainder of genome*

# Jamun\_35 (25114): ***(2/25/22)*** A: MSG R: KP

* **Selected start site evidence: 24845**
  + Original Glimmer call at bp 24869 has strength 11.92; GeneMark calls start at 24923
  + Start site called by Glimmer (24869) does not capture all coding potential.
  + Start site called by GeneMark (24923) does not capture all coding potential.
  + Close relatives Brynnie and Abidatro have manually annotated the start site at 24845.
  + 24845 has 5 MA’s (Manual Annotations)
  + 24845 has been chosen 71.4% of the time, when present.
  + Previous gene Jamun (24292) is a reverse gene starting at 24654, so the gap between Jamun\_34 (24292) and this gene Jamun\_35 (25114) is 191 bp.
  + The gaps called by Glimmer and Genemark would result in an even larger gap between genes.
  + There doesn’t appear to be any more coding potential for another gene in between in GeneMark.





* **Protein Function (helix-turn-helix DNA binding domain protein)**
  + Abidatro and Brynnie have deemed this a helix-turn-helix DNA-binding domain protein in Phamerator and PhagesDB
  + According to PhagesDP Blastp, it is also likely to be a helix-turn-helix DNA binding domain protein.
  + According to HHPred, the probability of a helix-turn-helix is 99.16%.
  + In NCBI Blastp, this gene is 97.53% identical to Brynnie’s, which was identified as helix-turn-helix DNA binding domain protein.

# **Jamun\_36 (25355)** Annotator: RV Reviewer: MB (2/25/22)

* **Selected start site: 25116**
  + Start 25116 and stop 25355 encompass all coding potential in GeneMark
  + Start 25116 has 1 bp gap between next gene Jamun (26377)
  + Original Glimmer call atbp 25116 has strength 9.84
  + Start at 25116 has 15 MA's
  + 25116 called 94.1% of the time when present
  + Position 25116 has been called on similar phages in the AS1 subcluster
  + RBS for 25116 is –5.563, high scoring RBS and strong evidence
* **Protein Function: excise**
  + Blast shows helix-turn-helix (HTH) binding protein or excise protein
  + SEA-PHAGES - “Do not call a protein excise unless you can identify the integrase and the immunity repressor in the phage. A more general ‘helix-turn-helix DNA binding protein’ might be more appropriate if you can't distinguish otherwise.”
    - Jamun (23115) is Tyrosine Integrase and Jamun (24292) is Immunity Repressor
  + HHPred predicts binding proteins for this sequence

# Jamun\_37 (26377) annotator:(MJB) reviewer:(RV) date: 2/25/22

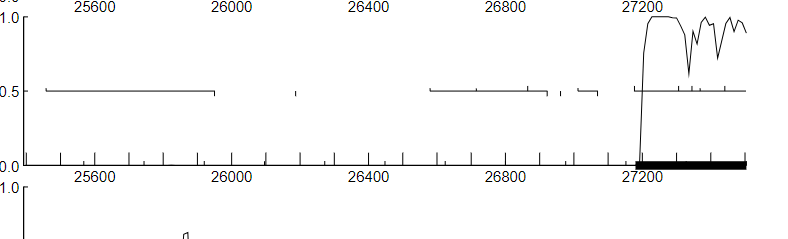
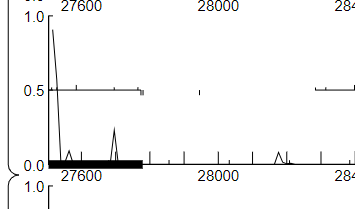
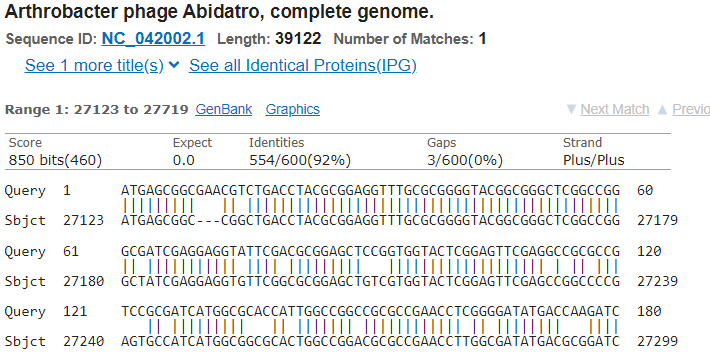
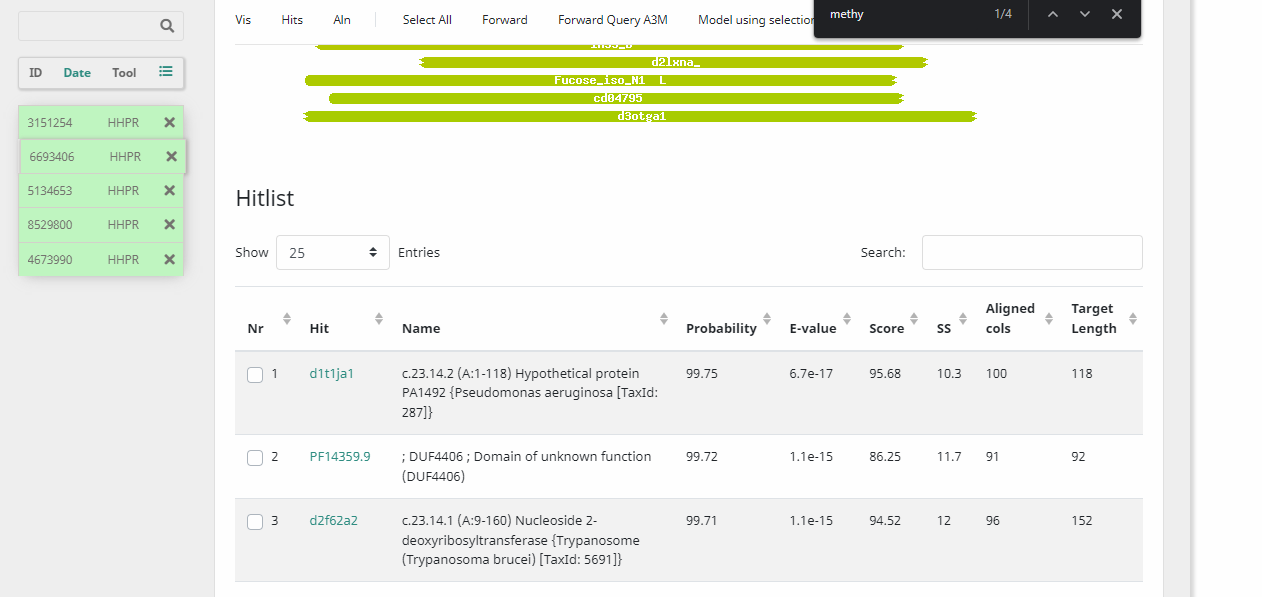
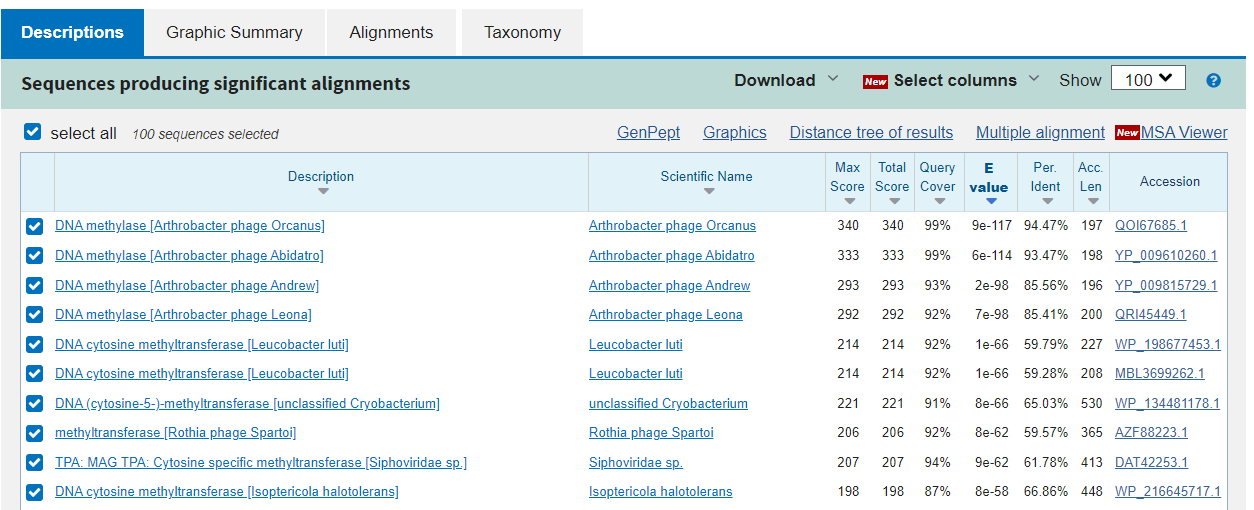
* **Selected start site: (25352)**
  + Original Glimmer call at 25352/Gene Mark strength 11.30
  + start sites cover all coding potential
  + 16/133 pham members have manually annotated the start site at 25352
  + start 25352 has been called 95.2% of the time, when present.
  + 4 manual annotations for AS1 pham members
  + Previous gene, Jamun (25355) stops with ending “ATGA” and has a 4 bp overlap with the start of Jamun (26377), which is at 25352 “ATG”
  + There is no coding potential for another gene.
* **Protein Function: RecE-like exonuclease**
  + Abidatro\_38, Brynnie\_37, Orcanus\_38, and Galaxy\_37 all assigned RecE-like Exonuclease shown on phagesdb
  + The first NCBI blast hit is “Abidatro RecE-like Exonuclease” showing 94.5% similar to Jamun with a 0.0 E-value, other clustermates have similar.

# Jamun\_38 (27183) Annotator: JF Reviewer: LC 2/25/2022

* **Selected start site: 26377**
  + Original Glimmer call atbp 26377 has strength of 14.21
  + The start site 26377 captures all of the coding potential of the gene
  + Start site at 26377 was manually annotated 3 times by other phages in the AS1 cluster- Abidatro, Brynnie and Orcanus
  + This start site was chosen 100% of the time when present in Pham 100427
  + Previous gene Jamun (26377) stops and has a 1 bp overlap between the start of Jamun (27183).
  + Blast Query Subject Ratio 1:1
* **Protein Function: RecT-Like ss DNA binding protein**
  + NCBI BlastP shows the top results as RecT-Like ss DNA binding Protein, and the results have e values of 0, 3e^-180 and 4e^-179
  + HHPred shows a probability of 100% that the function is in the RecT family
  + Both databases match in their chosen listed function.

# Jamun\_39 (27779) Annotator (AS) Reviewer (NJ) (02/25/2022)

# **Selected start site: 27180**

* + Original Glimmer call at 27180 has strength 3.63
  + Chose first upward facing tick as the start site → around the 27180 mark for Jamun (27779) → covers all of coding potential up to 27779.
    - 3rd frame down
    - 
    - 
  + 4 manual annotations of this start codon
  + Found in 7 of 322 of genes in pham → called 100.0% of time when present
  + Start 27180 was manually annotated 2 times for cluster AS1.
  + 4 bp overlap between stop codon 27183 of upstream gene Jamun (27183) and start site 27180 of this gene, Jamun 39.
  + RBS: Final Score = -6.377 and z value = 1.606
  + NCBI Blast
    - Every q:s ratio is 1:1 except for one of them which is 1:4
    - Abidatro genome (the only one with a q:s ratio of 1:4 in it)
      * 
* **Protein Function: (DNA methyltransferase)**
  + According to HHPred, the probability of a Hypothetical Protein is 99.75% and the E-value is 6.7e-17 (poor value).
    - 
  + In NCBI Blastp, this gene is 94.47% identical to DNA methylase for Arthrocacter phage Orcanus, query covers 99%, E-value 9e-117.
    - 
  + Official list says DNA methylase = DNA methyltransferase

# Jamun\_40 (28545):

# **Selected start site: 27823**

* + Original glimmer call atbp 27823 has strength 12.43; no genemark calls
  + In Pham 100599, close relatives Abidatro and Orcanus have manually annotated the start site 27823 in the 3 MA’s, site 27823 has been chosen 100% of the time when present.
  + Only start site called Glimmer (27823) completely encloses all coding potential for this gene
  + It has a 44 bp gap between Jamun (27779) and the start of Jamun (28545).
* **Protein Function: DNA methyltransferase**
  + NCBI BLASTP says it is most likely methyltransferase with relatives Orcanus, Leona, and Andrew all having it, it has a query of 99% and a e-value of 6e-144
  + HHPred also says it is most likely cytosine DNA methyltransferase with a probability of 96.03% and has a e-value of 0.033
  + Due to this agreement between the two, I believe the function is of this gene is for DNA methyltransferase

# Jamun\_41 (28934) annotator: NJ reviewer: AS (02/23/2022)

* **Selected start site: 28557** 
  + Original glimmer call 28557, GeneMark calls start at 286378
  + All coding potential covered
  + Close relatives Abidatro and Brynnie have manually annotated a different start site
  + 28557 has 1 manual annotation.
  + 28557 has been called 100% of the time, when present.
  + Previous gene Jamun (28545) stops and has a 12 bp gap between the start of Jamun (28934).
  + Evaluating the gap, there is not coding potential for another gene.
  + RBS: final score is -2.236 and the z score is 3.312
* **Protein Function: RusA-like resolvase**
  + Abidatro, Brynnie have a similar gene with RusA-like resolvase function
  + According to PhagesDP Blastp, RusA-like resolvase.
  + According to HHPred, the probability of a RusA function is 99.86% and the e-value is 2.7e-20.
  + In NCBI Blastp, this gene is 87% identical to Abidatro, query covers 100% e-value 1e-92.

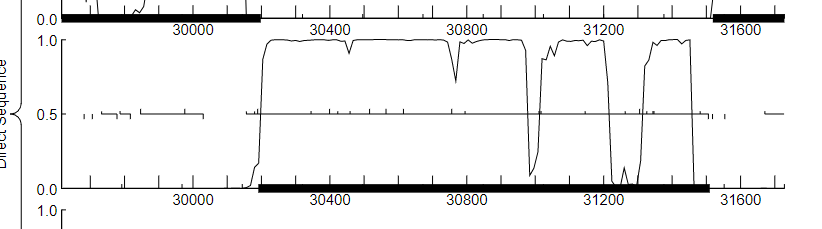
# Jamun\_42 (29401) annotator: NJ reviewer: AS (03/04/2022)

* **Selected start site: (28931)**
  + Original Glimmer call 28931
  + All coding potentials covered
  + Close relatives Brynnie and Abidatro have manually annotated the same start site
  + 28931 has 16 manual annotations.
  + 28931 has been called 100% of the time, when present.
  + Previous gene Jamun (28934) stops and has a 4 bp overlap between the start of Jamun (29401).
  + RBS: final score is -3.116 and the z score is 2.909
* **Protein Function: RusA-like Revolvase**
  + Close Relatives Abidatro and Brynnie both chose the function “RusA like Revolvase” for their equivalent gene.
  + NCBI blastP shows strong hits to RusA and RusA like revolvase, with the top phages in the AS1 cluster
  + The second hit on HHPred shows RusA, with a probability of 99.88% and an e-value of 3.93^-21

# Jamun\_43 (30198) annotator: LC reviewer: JF (3/4/22)

* **Selected start site: (29398)**
  + Original glimmer call atbp 29398 has strength 6.91; Genemark calls start at 29413
  + Close relatives Abitdatro, Basilisk, Brynnie,and Galaxy have manually annotated the start site 29398 in the 16 MA’s, site 29398 has been chosen 100% of the time when present.
  + Only start sites called Glimmer(29398) and Genemark(29413) completely encase all coding potential of this gene
  + Previous gene Jamun (29401) stops and has a 3bp gap between the start of Jamun (30198), which is 29398.
* **Protein Function: helix-turn-helix DNA binding protein**
  + Close relatives Abidatro and Galaxy showed the function Helix turn Helix binding protein on NCBI.
  + HHPred did show a relation to the same Helix turn Helix binding protein.

# **Jamun\_44** **(31510)** Annotator: AS Reviewer: NJ (03/04/2022)

* **Selected start site: (30191)** 
  + Original Glimmer call atbp 30191 has strength 12.88
  + Choosing the first upward-facing tick at around 30191 → covers all of the coding potential up to 31510
    - 
  + Start site # 245 at30191 is called in Abidatro\_46 (one of its most closely related phages).
  + Start: 245 at30191 has 12 MA's → has the most manual annotations
  + Found in 19 of 270 ( 7.0% ) of genes in the pham
  + Overlap of around 7 bp between stop codon of Jamun (30198) and start codon 30191 of Jamun (31510)
  + RBS: final score is -4.825 and the z value is -4.131
    - Not the best final score, but the one it chose with final score -4.314 → Start codon 31484 does not cover all of the possible coding potential
* **Protein Function: DNA methyltransferase**
  + The function of this gene is DNA Methyltransferase
  + NCBI showed that Abidatro and Brynnie both had DNA methyltransferase called for this gene
  + HHPred showed hits towards DNA Methyltransferase.

# Jamun\_45 (31974)

* **Selected start site: (31507)**
  + Original Glimmer call bp 31507 has a strength of 10.38.
  + Gene mark calls start site at 31519
  + 31507 was called 100% when present
  + There is a 4 bp overlap between the stop of Jamun (31510) and the start of Jamun (31974), which is 31507.
  + A 4bp overlap is strong evidence that this is the correct start site.
  + There were 16 manual annotations for 31507
* **Protein Function: helicase loader**
  + The function of this gene is considered to be helicase loader. Based on the blast that was done on NCBI there was a similarity to Hypothetical Protein in Phage Brynnie.
  + In HHPred this protein sequence was found to have been 97% similar to Helicase loader.

# **Jamun\_46 (32600)**: *(2/25/22)* Annotator: MSG Reviewer: KP

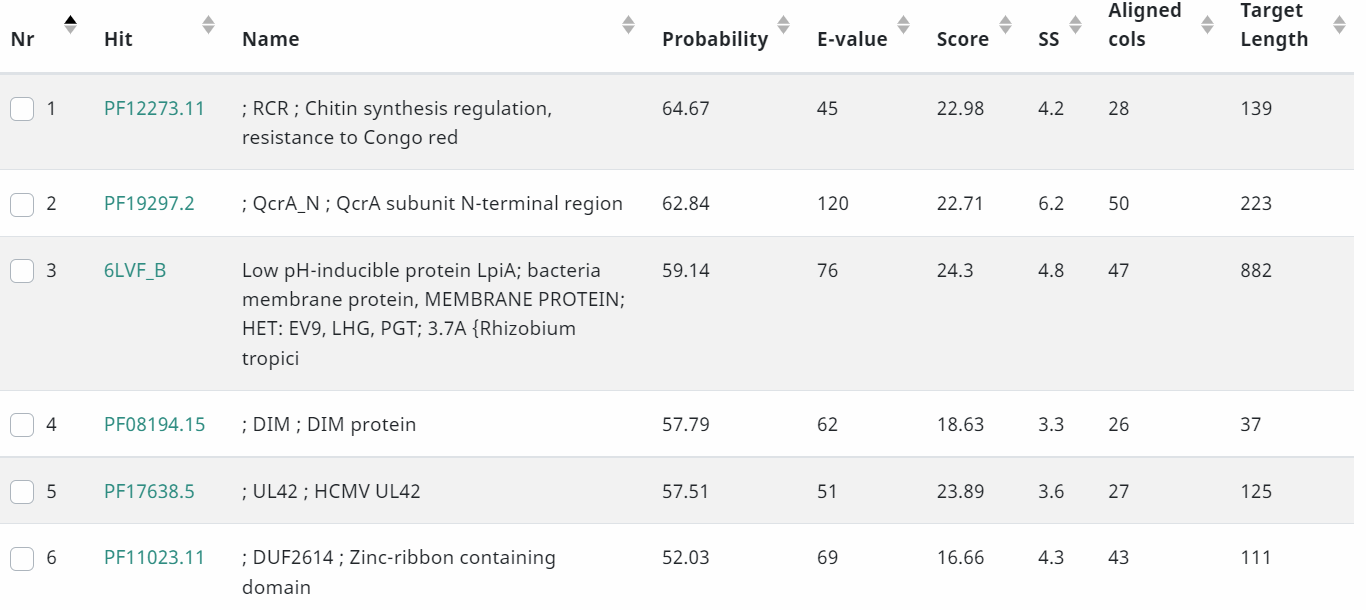
* **Selected start site: 31893**
  + Original Glimmer call atbp 31983 has strength 14.11
  + Start site called by Glimmer (31983) captures almost all coding potential
  + Close relatives Brynnie and Orcanus have manually annotated the start site at 31983.
  + 31983 has 2 manual annotations.
  + 31983 has been called 71.4% of the time, when present.
  + Previous gene Jamun (31974) stops and has a 9 bp gap between the start of Jamun (32600), which is 31983.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: Hypothetical Protein**
  + No other genes in this pham have annotated its function.
  + According to PhagesDP Blastp, its function is unknown.
  + According to HHPred, the probability of a zinc ribbon is 95.06%.
  + In NCBI Blastp, this gene is 86.26% identical to Brynnie’s, which has not been identified and is still a Hypothetical Protein.

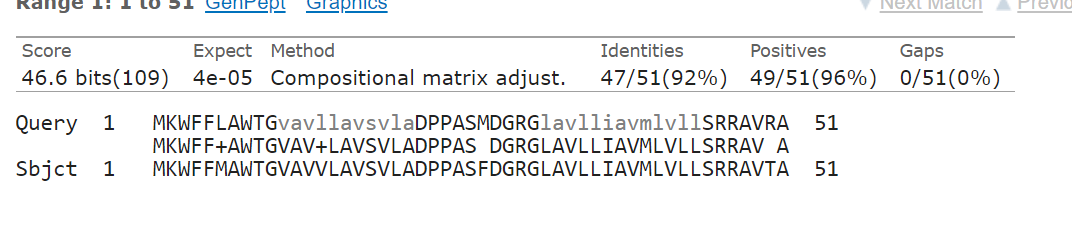
# Jamun\_47 (32964) (2/25/22) (3/4/22) Annotator: RV Reviewer: MB

* **Selected start site: 32737** 
  + Start 32737 and stop 32964 encompass all coding potential in GeneMark
  + Start 32737 has no overlap with Jamun (32600) and a 136 bp gap.
  + Original Glimmer call atbp 32812 has strength 9.01
  + Start: at32737 has 9 MA's
  + 32737 called 60.0% of the time when present
  + Position 32737 has been called on similar phages in the AS1 subcluster
  + RBS for 32737 is -2.231, moderately strong evidence
* **Protein Function: Hypothetical Protein**
  + Blast shows Hypothetical Protein for two phages with high probability
  + HHPred shows DUF6296 (no known function) with 89.01 probability but 1.2 E-value

# **Jamun\_48 (33219)** annotator:(MJB) reviewer:(RV) date:2/25/22

* **Selected start site: (33064)**
  + Original Glimmer call atbp 33064 has strength 9.31
  + Start sites cover all coding potential.
  + Close relatives Abidatro and Brynnie have called the start site at 33064.
  + 33064 has been called 61.9% of the time, when present.
  + 33064 has 9 manual annotations.
  + Previous gene, Jamun (32964) and has a 100 bp gap between the start of Jamun (33219) which is at 33064.
  + Evaluating the gap, there is no coding potential for another gene when run through blast.
* **Protein Function: Membrane Protein**
  + NCBI blast shows an identified membrane protein in Orcanus with a E-value of 4.05e-10 and 92% identities.
  + HHPred’s lowest e-value was 45
  + No genes in pham have identified this as a function published through phagesdb, all has “unknown function listed”
  + HHPred: To be a membrane protein it “must contain at least two (2) transmembrane domains found using TmHmm OR one membrane domain found by two different prediction programs” -Orcanus protein sequence meets TMHMM requirements (2 transmembranes) - Jamun also meets TmHMM requirements.



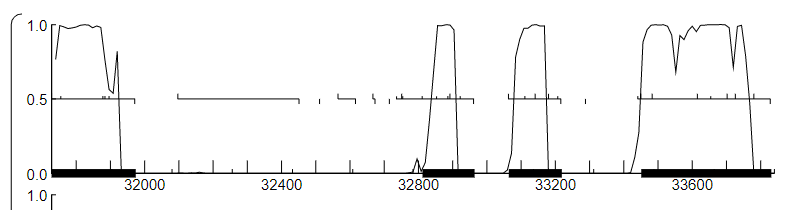
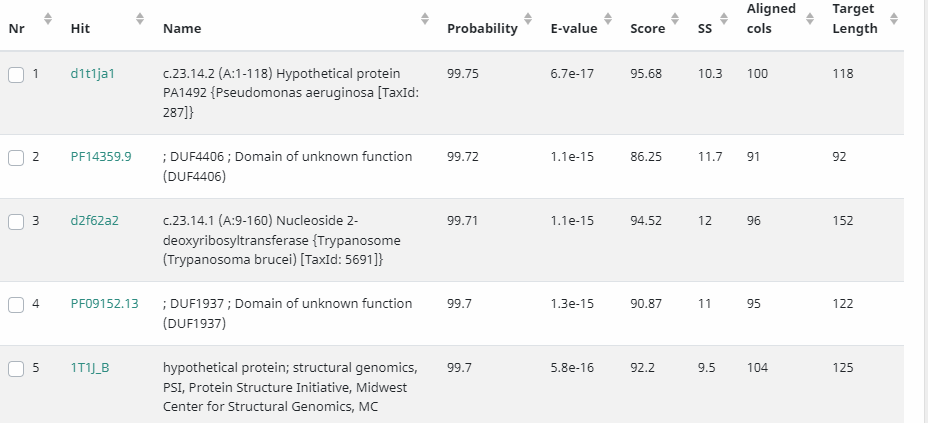
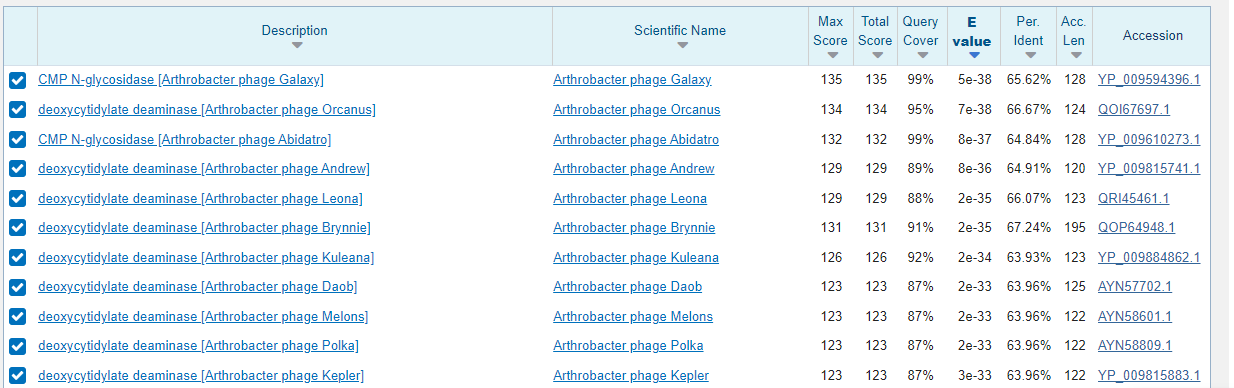


# Jamun\_49 (33458) Annotator: JF Reviewer: LC 2/25/22

# **Start Site Evidence: 33216**

* + Original Glimmer call atbp 33216 has strength of 8.84
  + Start Site captures all coding potential
  + Close relatives of JamunDraft, Brynnie and Abidatro both manually annotated the start at 33216.
  + The start at bp 33216 has been called 100% of the time when present in 6 manual annotations.
  + 4 other phages in cluster AS1 manually annotated the start site at 33216
  + Previous gene Jamun (33219) stops and has a 4)bp gap between the start of Jamun (33458).
  + Has a 4 bp overlap with the upstream gene (Jamun\_50)
  + Blast Query Subject Ratio 1:1
* **Protein Function: Hypothetical Protein**
  + NCBI BlastP shows the top results as Hypothetical Protein, with e values ranging from 4e^-28 to 2e^-21
  + HHPred shows a probability of 86.86% Homo sapiens— therefore I am comfortable stating that this is a Hypothetical Protein with unknown function

# **Jamun\_50 (33831)** Annotator: AS Reviewer NJ (02/25/2022

* **Start Site Evidence: (33451)**
  + Original Glimmer call atbp 33451 has strength 6.23
  + Chose first upward-facing tick as start site → around the 33451 mark for Jamun52 → covers all of coding potential up to 33831.
    - 
  + TaylorSipht\_52 calls this start, which is a relative in AS1, with the same gene number.
  + Start: 50 at 33451 has 23 MA's
  + Start site 33451 is called 77.1% of time when present
  + Gap size of around 7bp between stop codon of upstream gene Jamun (33458) and start codon 33451 of Jamun (33831).
  + Evaluating the gap, there is no coding potential for another gene.
  + RBS: final score is -4.313 and Z value is -3.619.
* **Protein Function (deoxycytidylate deaminase)** Annotator: AS Reviewer NJ
  + According to HHPred, the probability of a Hypothetical Protein is 99.75% and the E-value is 6.7e-17.
    - 
  + In NCBI Blastp, this gene is 66.67% identical to deoxycytidylate deaminase for Arthrocacter phage Orcanus, query covers 95%, E-value 7e-38.
    - 

# Jamun\_51 (34265)

# **Start site evidence: 33828**

* + Original glimmer call atbp 33828 has strength 13.88: no genemark calls
  + In members of this pham, close relatives Abidatro and Galaxy have manually annotated the start site at 33828 in the 2 MA’s, site 33828 has been chosen 75% of the time when present.
  + Only start site called Glimmer(33828) completely encloses all coding potential for this gene
  + Previous gene Jamun (33831) has a 3 bp overlap with the start of Jamun (34265).
* **Protein Function: Hypothetical Protein**
  + NCBI BLASTP says it is most likely a Hypothetical Protein with relatives Abidatro and Galaxy having it, has a query of 99% and a e-value of 7e-40
  + HHPred believes it is peptidoglycan with a 90.35% probability and an e-value of 5.2
  + With NCBI BLASTP having multiple relatives of Jamun having a protein of similar function leads me to believe this gene has the function of a Hypothetical Protein.

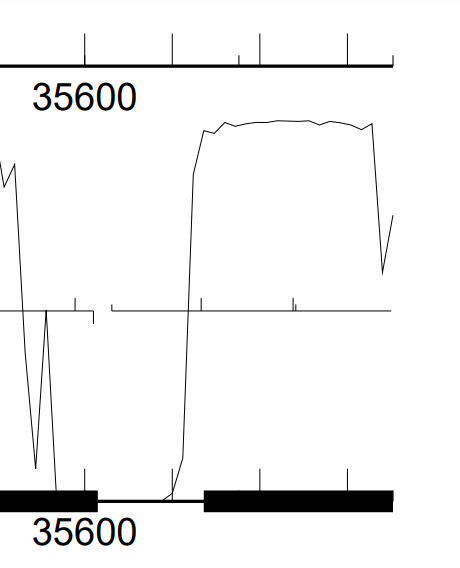
# **Jamun\_52 (34783)** annotator: NJ reviewer: AS (02/25/2022)

* **Start Site Evidence: (34262)**
  + Original glimmer call 34262
  + All coding potentials covered
  + Close relatives Brynnie and Abidatro have manually annotated the same start site
  + 34262 has 143 manual annotations.
  + 34262 has been called 98.9% of the time, when present.
  + Previous gene Jamun (34265) stops and has a 12 bp gap between the start of Jamun (34783) at 34262.
  + Evaluating the gap, there is not coding potential for another gene.
  + RBS: final score is (-5.141) and the z score is (1.946)
* **Protein Function (ssDNA binding protein)** annotator: NJ reviewer: AS (03/04/2022)
  + Brynnie, Galaxy, and Abidatro also annotated similar gene with ssDNA binding protein function
  + According to PhagesDP Blastp, ssDNA binding protein.
  + According to HHPred, the probability of a ss DNA binding protein function is 99.97% and the e-value is 8.2e-29.
  + In NCBI Blastp, this gene is 90.75% identical to Galaxy; the query covers 100% e-value 1e-86.

# **Jamun\_53 (35615)** Annotator: JF Reviewer: LC 3/4/22

* **Selected Start Site:: 34887**
  + Original Glimmer call atbp 34887 has strength 13.64
  + 34887 Captures all the coding potential of the gene, it is the only potential start site to do this
  + Start 9 at 35615 was called 75% of the time when present, and was manually annotated in 11/27 manually annotated genomes that have it.
  + Close relatives of JamunDraft, Brynnie and Abidatro both manually annotated start site 9 (34887)
  + This start was manually annotated 4 times in cluster AS1
  + 105 bp gap with the upstream gene Jamun (34783), but Genemark shows no coding potential in this gap
  + RBS Final score of -2.605
* **Protein Function (Endonuclease) annotator: AS reviewer: NJ (03/23/2022)**
  + According to PhagesDB Blastp, this is an endonuclease.
  + According to HHPred, the probability of a Mitochondrial cardiolipin hydrolase is 99.59% and the e-value is 1.5e-13.
  + In NCBI Blastp, this gene is 97.52% identical to endonuclease Arthrobacter phage Galaxy, query covers 99% e-value (1e-172).

# **Jamun\_54 (36023)** annotator:(MJB) reviewer:(RV) date 3/2/22

* **Selected Start Site: (35634)**
  + Original Glimmer call at 35634/Gene Mark strength 10.36
  + According to DNA master, Gene mark calls start at 35736
  + start sites cover all coding potential
  + There are 2 manual annotations at start site 35634
  + 35634 has been called 80% of the time, when present.
  + 2 manual annotations have been found for start 35634
  + Previous gene, Jamun (35615) stops and has a 8 bp gap between the start of Jamun (36023), which is at 35634.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: (Hypothetical Protein)** annotator:NJ reviewer: AS (03/25/2022)
  + Abidatro, Galaxy and Brynnie said the similar gene has an unknown function
  + According to PhagesDP Blastp, unknown function.
  + According to HHPred, the probability of a terminase small subunit is 99.83% and the e-value is 3.2e-19.
  + In NCBI Blastp, this gene is 97.33% identical to Brynnie, query covers 58% e-value 2e-42.

# Jamun\_55 (36268) (3/4/22) Annotator: RV Reviewer: MB

* **Selected Start site: 36020**
  + Start 36020 includes all coding potential but 36068 does not
  + Start 36020 has 4 bp overlap with Jamun (36023), 36068 has 44 bp gap
  + Original Glimmer call atbp 36020 has strength 6.69; GeneMark calls start at 36068
  + Start: at36020 has 5 MA's
  + RBS score for 36020 -5.227, 36068 -4.846
  + 36020, 1:1 alignment with 3 phages in subcluster on BLAST
* **Protein Function: ssDNA binding protein**
  + Abidatro, Galaxy, and Brynnie said the similar gene was a single stranded DNA binding protein
  + NCBI Blastp says this gene is a single stranded DNA binding protein with 99% query cover and a E-value of 1e-86
  + HHPred also says that this gene is a single stranded DNA binding protein with 99.98% probability and a e-value of 4.7e-29

# Jamun\_56 (36624) Annotated by KP

* **Selected Start Site: 36265**
  + Original Glimmer and genemark call bp 36265 with a strength of 6.92.
  + 36265 is the only site that is present for Jamun that was also manually annotated in other instances
  + 36265 is called 100% of the time that it is found present.
  + 36265 covers all of the coding potential in the region
  + There is a 4bp overlap where Jamun (36268) ends and Jamun (36624) starts.
  + There does not appear to be any coding potential missing between the genes.
  + There are 6 manual annotations of start 36265.
* **Protein Function: Hypothetical Protein**
  + Protein of unknown function.
  + Blast of sequence on NCBI database resulted in a hit with brynnie
  + HHPred resulted in a protein found in bacteriophages of unknown function with strong alignment.

# Jamun\_57 (36801): (2/25/2022) A: MSG R: KP

* **Selected Start Site: 36628**
  + Original Glimmer call atbp 36628 has strength 7.96
  + Start site called by Glimmer (36628) captures all coding potential.
  + Close relatives Brynnie and Orcanus have manually annotated the start site at 36628.
  + 36628 has 13 manual annotations.
  + 36628 has been called 94.1% of the time when present.
  + Previous gene Jamun (36624) stops and has a 4bp gap between the start of Jamun (36801), which is 36628.
  + Evaluating the gap between genes, there was no coding potential found for another gene.
* **Protein Function: Hypothetical Protein**
  + No other genes in this pham have annotated its function.
  + According to PhagesDP Blastp, its function is still unknown.
  + According to HHPred, the probability of a lysine biosynthesis protein is only 23.01%.
  + In NCBI Blastp, this gene is 83.93% identical to Brynnie’s, which was identified as a Hypothetical Protein.

# Jamun\_58 (37456) Annotator: RV Reviewer: MB

* **Selected Start Site: 36794 (2/25/22)** 
  + Start 36794 and 36812 both encompass all coding potential in GeneMark
  + 36794 has a 8 bp overlap with Jamun (36801), 36812 has a 10 bp gap
  + Original Glimmer call atbp 36794 has strength 8.31
  + No other phages called with these starts, no manual annotations
  + RBS score for 36794 -3.293, 36812 -4.646, 36812 better evidence
* **Protein Function: Hypothetical Protein (3/4/22)**
  + Blast shows closest match Hypothetical Protein SEA\_BRYNNIE\_59
  + Nothing of significance found on HHPred
  + No additional information from SEA-PHAGES

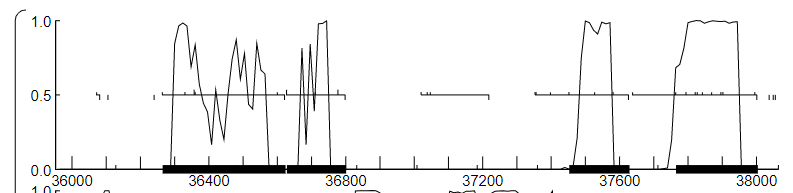
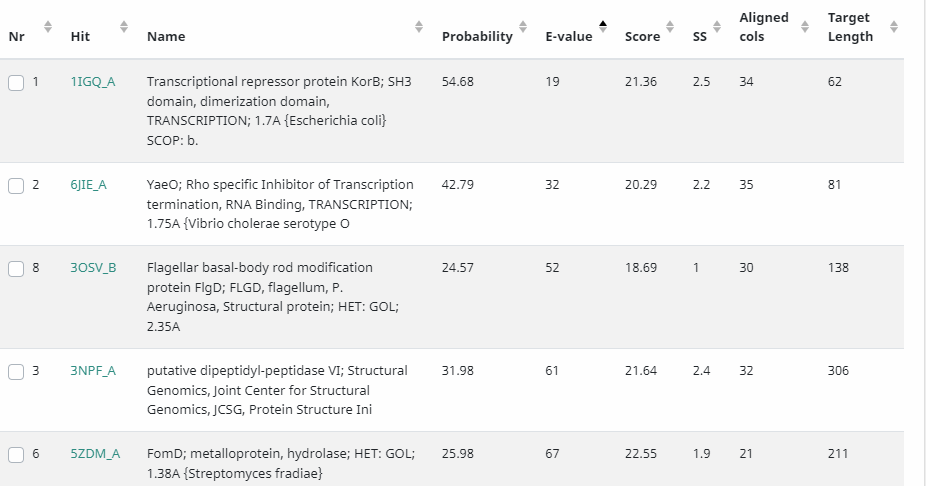
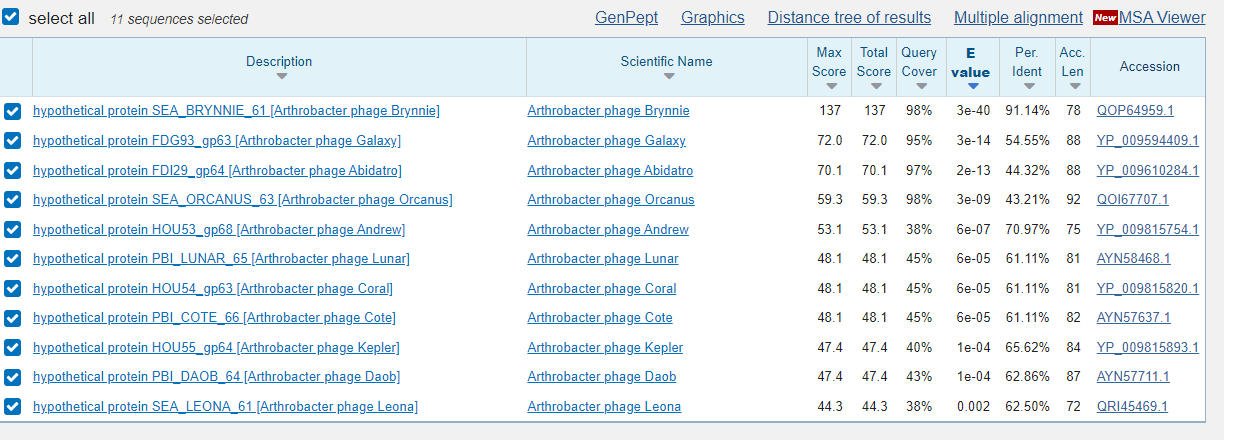
# **Jamun\_59 (37629)** annotator:(MJB) reviewer:(RV) date 3/3/22

* **Selected Start Site: (37453)**
  + Original Glimmer call at 37453/Gene Mark strength 9.21
  + Start Site 37453 covers all coding potential.
  + Close relatives Basilisk, Orcanus, Brynnie have manually annotated the start site at 37453.
  + 37453 has been called 100% of the time, when present.
  + 3 manual annotations for start site 37453
  + Previous gene, Jamun (37456) stops and has a 4 bp overlap between the start of Jamun (37629), which is at 37453.
  + Evaluating the gap, there is no coding potential for another gene.
* **Protein Function: Hypothetical Protein**
  + No pham members have a published function on phagesdb
  + NCBI blast hits were for Hypothetical Proteins in Orcanus and Brynnie.
  + According to HHPred, the probability of a “DNA binding protein HU alpha, Histone-like, DNA binding protein” is 88.93% and the e-value is 2.
  + In NCBI Blastp, this gene is (59)% identical to Brynnie, which is also still identified as a Hypothetical Protein.

# **Jamun\_60 (37775)** Annotator: JF Reviewer: LC (2/25/2022)

* **Selected Start Site: 37626**
  + Original Glimmer call at 37626bp has strength 1.99\*
  + This gene was not called by GeneMark
  + Start site 37626 covers all of the coding potential for this gene
  + Start site at 37626 had 1 Manual Annotation.
  + 4 base pair overlap with the upstream gene.
* **Protein Function: Hypothetical Protein**
  + NCBI Blastp shows the top results as a Hypothetical Protein, and shows relatives Brynnie and Galaxy. The e-values are 4e^-17 and 2e^-4
  + HHPred shows a 90.41% chance of family of unknown function, 89.66% chance of Nudix N-terminal, and 89.53% chance of Hypothetical Protein.

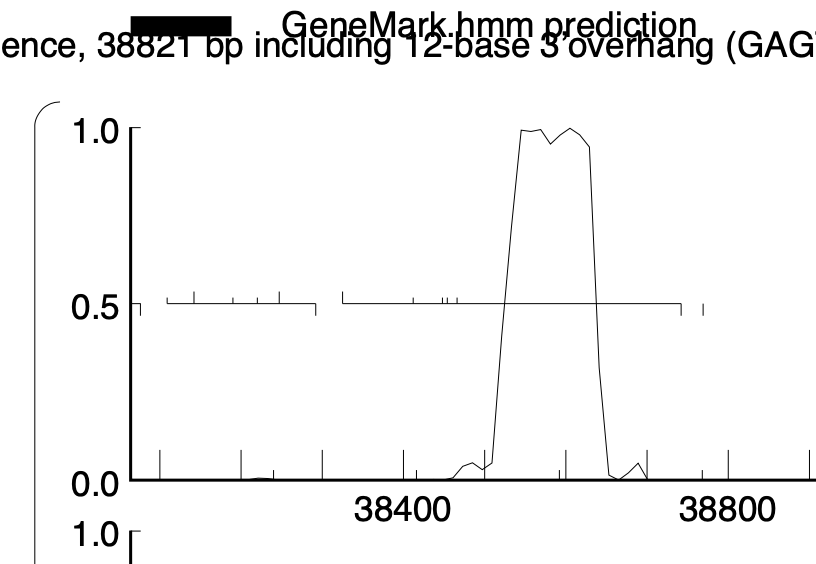
# Jamun\_61 (38004)

* **Selected Start Site (37765)** Annotator: AS Reviewer: NJ (02/25/2022)
  + Original Glimmer call at bp 37765 has strength 11.27
  + 37765, which covers all of coding potential.
    - 
  + Start: 15 at 37765 has 1 MA's → the only possible start codon for Jamun63 that was manually annotated
  + Found in 3 of 15 (20.0%) of genes in pham
  + Called 66.7% of the time when present
  + Gap size of around 10bp between stop codon 37775 of Jamun (37775) Jamun (38004) and start codon 37765 of Jamun63.
  + No visible overlapping of potential coding upstream of the gene.
  + RBS: final score is -3.628 and Z value is 2.494
* **Protein Function: (Hypothetical Protein)**
  + According to HHPred, the probability of a transcriptional repressor protein is 54.68% and the E-value is 19.
    - These are very poor values (ex. too high of an E-value), would disregard this function.
    - 
  + In NCBI Blastp, this gene is 91.14% identical to the Hypothetical Protein for Arthrocacter phage Brynnie, query covers 98%, E-value 3e-40.
    - 

# Jamun\_62 (38213)

* **Selected Start Site: (38028)** annotator:LC reviewer: JF
  + Original glimmer call at 38028bp has strength 2.58 and was not called by genemark
  + Close relative Brynnie has manually annotated the start site 38028, start site 38028 has been chosen 16% of the time when present.
  + Start 38028 has 2 MA, and has been chosen 71.4% of the time when present. Relatives Galaxy, Orcanus, and Basilisk all have this start site along with Jamun.
  + Glimmer (38028) does not cover all of the coding potential,
  + Previous gene Jamun (38004) stops and has a 24 bp gap from the start of Jamun (38028).
  + Evaluating the gap, there is no coding potential for another gene in between.
* **Gene Protein Function: helix-turn-helix binding domain protein, MerR-like**
  + NCBI BLASTP says it is a DNA binding protein, specifically a helix-turn-helix binding domain protein (MerR-like), with relatives Brynnie, Galaxy, and Abidatro all being in the top three for highest likelihood and all sharing this annotation.
  + HHPred’s most likely option is also saying it is a DNA binding protein with a 99.16% chance of being a DNA binding protein
  + With the two different engines seeming to be in agreement, the gene function is for a helix-turn-helix protein
  + “helix-turn-helix binding domain protein, MerR-like” is how it is listed on the Official Functions List although all the other annotations in AS cluster have this listed as “MerR-like helix-turn-helix DNA binding domain protein” Because the former is how it is listed in the OFL, we kept that nomenclature.

# Jamun\_63 (38745)

* **Start Site Evidence: 38467** annotator: NJ reviewer: AS (02/25/2022)
  + Original Glimmer call 38467
  + All coding potential covered. Note that the gene is not marked with a black bar, but it still exists
  + Close relatives Brynnie and Abidatro have manually annotated the same start site.
  + 38467 has 5 manual annotations.
  + 38467 has been called 88.9% of the time, when present.
  + Previous gene Jamun (38213) stops and has a 254 bp gap between the start of Jamun (38745).
  + Evaluating the gap, although big, NCBI blast shows that there is no gene in the gap (no similarity to other known genes/proteins).
  + RBS: final score is -5.040 and the z score is 2.067
* **Protein Function: (HNH endonuclease)** annotator: NJ reviewer: AS (03/04/2022)
  + Brynnie, Galaxy, Abidatro identified a similar gene as HNH endonuclease
  + According to PhagesDP Blastp, HNH endonuclease.
  + According to HHPred, the probability of a CRISPR-associated HNH endonuclease function is 98.72% and the e-value is 3.2e-8, percent of query covered 17%.
  + In NCBI Blastp, this gene is 97.83% identical to Brynnie HNH endonuclease, query covers 100% e-value 2e-59.