DUNAMIS – CLUSTER FD FULL ANNOTATION NOTES

Gene 1: Both Glimmer and GeneMark Call the start site 749. There is coding potential for this start and stop site. Tillums and Anjali both have this gene which are similar phages in this pham with low e values of 1e-111 and 2 e-90. There is no significant overlap with other genes (gap: -4). It is long enough to be considered a gene. The final score is -3.368 and the z score is 2.597. This is not the longest ORF, however, the longest has excessive overlap. The start site has 3 MAs, Anjali and Swim also have this start with 3 MAs each. There is also 1:1 alignment with Swim (e value: e-112), Tillums (e value: e-111), and Anjali (e value: 2e-90). Therefore, the start site is 749.

FUNCTION: DB and NCBI BLAST calls the function a DNA Methyltransferase with high identity and low evalues when compared to Phage Tillums ,Anjali and Swim. HHRED evidence: DNA cytosine methyl transferase with 99% probability, 77% coverage and e value of 3.4 e-17

Gene 2 : This is called a gene by Glimmer and GeneMark. There is coding potential for the proposed start and stop sites. Anjali, Swim, and Tillums all call this gene. The e values are low at 7e-23, 2e-22 and 5e-19 respectively. The length of this gene is 167 which is sufficient to be called a gene with no significant overlap with other genes. The genes before and after this one are both in the reverse direction. Glimmer and GeneMark both call the start site 913 for this gene. There is a z score of 1.586 and a final score of -5.534 for this start site.. The scores are not the best proposed ones, however, the better scores have significant overlap. The length of the ORF is 168 bp, the gap is -4 which is the best proposed gap. Anjali calls this start site with 3 MAs, Dunamis has the start site in starterator with 3 MAs and it is also found in 100% of genes in this pham . It is called 100% of the time when present. Phages DB BLAST shows 1:1 alignment with Anjali, and Swim with 7e-23 and 2e-22 respectively. The start site for this gene is 913. Phages DB BLAST gives 1:1 alignment with Anjali, and Swim with e values of 7e-23 and 2e-22 respectively. These functions are unknown for these phages.

Function: For NCBI Anjali is listed as a hypothetical protein with 98.15 % alignment and 96.36% coverage with an e value of 3.71e-28. The accession is YP\_009842150. Additionally, Tillums has the function listed as a hypothetical protein with 65% alignment, 90.9% coverage and an e value of 1.64e-23. The accession is WNT44574. HHpred does not have any listed phage genes with good probability. The e values are also all above 0.5.- Hypothetical protein.

Gene 3: Only Glimmer calls this a gene. There is evidence for coding potential with a start at approximately 1044 and a stop at 910. Glimmer calls a start site of 1044 where GeneMark does not call a start site at all. The final score for this is -3.976, the z score is 2.44. These are the best scores. The length of the open reading frame is 135 for this start site which is long enough and has no significant overlap with genes up or downstream. This is present in other annotated genomes such as Tillums (e value: 2e-16), Anjali (e value: 5e-15), Mendel (e value: 4e-15), and Swim (e value: 1e-13) all have this gene annotated. There is no significant overlap with other genes (gap: -4), and the length is sufficient to be considered a gene (135 bp). This is a gene. Start site 1044 has 4 MAs for Dunamis. Phages DB says that this gene aligns 1:1 with Tillums (e value: 2e-16), Mendel (e value: 4e-15), Anjali (e value: 6e-15), and Swim (e value 1e-13).

Function call: Unknown function. From NCBI Blast, Tillums calls this a hypothetical protein. HH pred has no matches for bacteriophages that match this protein. No quality hits. These all call this a hypothetical protein.

Gene 4: Only Glimmer calls this a gene at 1154. There is coding potential for this gene. This gene is an orpham. Starterator report was not informative for this gene. The phages DB blast for the selected gene of 1154 has no 1:1 alignment with any phages. There is a 1:18 alignement with Tillums with an e value of 2e-17. HH pred has no matches with any phages. NCBI Blast has 68% alignment 100% coverage, and 1:18 alignment with an e value of 4.62e-18 for Tillums.

NCBI, Phagesdb and HHPRED gave no significant hits. In Tillums, this protein has an unknown function and an e value of 2e-28 for phages DB. In NCBI it is called a hypothetical protein with 100% alignment, 100% coverage, and e value of 9.49e-31 with 1:1 alignment. There are no matches in HH pred

Gene 5: Only Glimmer called this gene. There is evidence of coding potential. This gene was only called in Tillums (e-value 5e-16). There is not significant overlap. The gene after it is in the other direction but the gene before it is on the same strand. The gene is only 123 base pairs. Glimmer called 1320 as the start site. The Z-score is 2.237 and final score is -5.172. This is not the longest ORF. The gap is large, 179 base pairs. The start site called is a GTG. This start site has 1 manual annotation in Tillums. There is 1:1 alignment with Tillums (e-value 5e-16). NCBI called this gene a hypothetical protein with no known function, only found in Tillums (100% identity, 100% alignment, 100% coverage).

Gene 6: Both GeneMark and Glimmer called this gene. There is coding potential on the GeneMark. This gene is present in Tillums, Mendel, Swim and Anjali (e-values 3e-82, 6e-81, 6e-77, 2e-76). Glimmer and GeneMark disagree with on the start site. Glimmer calls 1455 which is a GTG and GeneMark calls 1500 which is an ATG. Start site 1455 has a Z-score of 1.181 and final score of -6.396. Start site 1500 has a Z-score of 2.758 and final score of -3.314. Start site 1455 is the longest ORF and has a gap of 134 base pairs. Start site 1500 is the smaller ORF and has a gap of 179 base pairs. There is no significant overlap with other genes at either startsite. Start site 1500 has 4 manual annotations on Starterator and is called 71.4% of the time when present. Start site 1455 is never called and has 0 manual annotations. Start site 1455 has 16:1 alignment with Tillums, Mendel, Swim, and Anjali. Start site 1500 has 1:1 alignment with Tillums, Mendel, Swim, and Anjali (evalues 3e-82, 6e-81, 7e-77, 2e-75). Start site 1500 has the most evidence and is preferred by Starterator. Phages Db Blast and NCBI blast both call this protein a SSB protein, 1:1 alignment with Tillums. HHPred also calls this a SSB with 100% probability and e-value 2e-35.

Gene 7: This gene was called by both Glimmer and GeneMark. The Gene Mark shows coding potential between the correct regions. This gene is present in Tillums, Mendel and Anjali (e-values 1e-151, 1e-147,1e-143). Both Glimmer and GeneMark called site 2057 which has a Z-score of 1.606 and Final score of -6.717 and is the longest ORF. There is no significant overlap (gap 89). Start 2057 is the most annotated site with 5 manual annotations. Site 2057 has 1:1 alignment with Tillums, Mendel, Swim, and Anjali (e-values 1e-151, 1e-147, 1e-146, 1e-143). Phages DB blast suggest this is a DNA terminal protein, as well as NCBI blast suggests DNA Terminal Protein (100% Identity, 100% Aligned, 100% Coverage). HHPred also suggests "Primer terminal protein, DNA Terminal Protein" (probability 100%, e-value 2.9e-41). This gene is located in a region of high conservation, preceded by a single stranded DNA binding protein and followed by DNA Polymerase.

Gene 8: Both GeneMark and Glimmer call this gene and there is coding potential within the correct region. This gene is present in Tillims, Mendel, Swim, and Anjali (e-value 0, 0, 0, 0). Glimmer calls 2923 as the start site and GeneMark calls 2857, however Startertor supports an alternate start site of 2878 with 5 manual annotations. Start site 2878 has the better Z score (2.502 compared to 2.029 and 1.586) and final score (-5.161 compared to -4.513 and -5.534). Start sites 2923 and 2857 have no MAs. Start site 2878 has 1:1 alignment with Tillums, Anjali, Swim, and Mendel (e-value 0,0,0,0), where the other starts have 1:16 and 1:8 alignment. Start site 2878 is the most likely start. PhagesDB suggests DNA polymerase as the function with 1:1 alignment with Tillums DNA Polymerase and NCBI also supports DNA Polymerase (100% Identity, 100% Alignment, 100% Coverage). HHPred suggests DNA polymerase with 100% probability, E-value 3.2e-55). This gene is in a region of high conservation, preceded by DNA terminal protein and followed by ribbon-helix-helix DNA.

Gene 9: This gene was called by both Glimmer and GeneMark and there is coding potential in the correct region. This gene is present in Tillums, Mendelm, Swim, and Anjali (e-value 7e-41, 2e-28, 3e-27, 2e-28). Glimmer calls 4796 as the start site and GeneMark calls 4805. 4805 has the better Z score (2.948) and Final score (-2.700), however is not the longest ORF, but is the second longest. 4805 is an ATG whereas 4796 is a TTG, and ATG is more preferred. 4805 has 4 manual annotations whereas 4796 has none. Site 4805 has 1:1 alignment with Tillums, Mendel, Anjali and Swim (e-values 8e-41, 3e-28, 3e-28, 4e-27). Phages DB blast calls this gene a ribbon-helix-helix DNA binding domain (e-value 8e-42, 1:1 alignment) and NCBI also called it a ribbon-helix-helix DNA binding domain protein (100% identity, 100% alignment, 100% coverage). HHPred calls it a Ribbon-helix-helix Homodimer (98.9% probability, e-vlaue 7.2e-8). This gene is in a region of high conservation, preceded by DNA polymerase and followed by a gene of unknown function.

Gene 10: Glimmer and GeneMark both call the start site at 5226. There is clear coding potential for this start and stop site. The gene has no significant overlap with the previous gene (gap: 187), and is long enough to be considered a valid ORF. Although not the most called start in Starterator, the selected start site is supported by the Pham Starterator report with 1 manual annotation. PhagesDB BLAST shows alignment with Swim and Tillums with e-values of 1e-17 and 5e-19, respectively, indicating a strong 1:1 alignment. NCBI confirms this with similarly strong alignments: Swim (e-value: 2e-21) and Tillums (e-value: 3e-23). The RBS score is -2.794 and the z-score is 3.003, indicating a reasonable ribosome binding strength. This gene does not have a known function. PhagesDB and NCBI both return hypothetical proteins with no functional assignments and strong alignment (PhagesDB e-value: 5e-19; NCBI e-value: 2e-21). HHpred does not provide a match with ≥90% probability only a low confidence match to a possible ribosomal protein (57.37% probability).

Gene 11: Both Glimmer and GeneMark call the start site at 5366, and the GeneMark-host trained map shows clear coding potential for this gene. Starterator report with 3 manual annotations, although it is not the most commonly called start in published annotations. The RBS score is -2.559 with a z-score of 2.986, indicating a strong ribosome binding site. This start site is not the longest ORF, but the best possible ORF. Conservation with other phages is strong: PhagesDB shows 1:1 alignment with Swim, Mendel, Tillums, and Anjali (e-value: 1e-54), and NCBI confirms alignment with Swim and Tillums (e-value: 1e-58). Therefore, the selected start site is 5366.PhagesDB and NCBI both identify this as a hypothetical protein with no known function (PhagesDB e-value: 1e-54; NCBI e-value: 1e-58). HHpred does not return any matches above 90% probability, the closest is a possible scaffolding protein at 71%, which is insufficient for confident function prediction. Therefore, this gene has no known function (NKF).

Gene 12: Both Glimmer and GeneMark call this gene, with Glimmer suggesting the start site at 5690 and GeneMark at 5696. The selected start site, 5690, is supported by Pham Starterator with 1 manual annotation, even though it was not the most commonly called start site across all annotations (called ~57% of the time). This start site also corresponds to the longest ORF, avoiding excessive overlap and ensuring full coverage of coding potential. The RBS score is -4.039, and the z-score is 2.252, suggesting a reasonable ribosome binding strength. Coding potential is clearly evident throughout the ORF. This start site is conserved across multiple phages in the FD cluster with 1:1 alignment: PhagesDB shows perfect hits to Tillums, Anjali, Mendel, and Swim (e-value: 0), and NCBI confirms these alignments as well (e-value: 0 for all). These high-quality, no e-value matches strongly support this start site and gene presence. BLAST in both PhagesDB and NCBI call this gene a Major Capsid Protein, with very strong alignment (e-value: 0). HHpred confirms this function with a match to a major capsid protein from Bacillus phage Phi29 (Probability: 100%, e-value: 0, coverage: 82.4%). This gene is also located in a region highly conserved for capsid structure genes among FD cluster phages, supporting functional conservation., and the functional assignment is on the SEA-PHAGES approved function list. Therefore, this gene is a Major Capsid Protein.

Gene 13; This gene is called only by Glimmer, not by GeneMark, but there is coding potential visible in the GeneMark-host trained map, supporting its presence. The start site is 7421, is the longest ORF GeneMark did not predict a start site, Glimmer's suggested start at 7421 is well-supported. Starterator confirms this as the most commonly called start site in published annotations, chosen 100% of the time, with 2 manual annotations reinforcing this selection. The RBS score is -6.323, with a z-score of 1.207, which is on the lower side but still acceptable given the strong conservation. This gene aligns well with phages in the FD cluster, showing 1:1 alignment with Swim and Tillums. BLAST results show strong support: PhagesDB: e-value 7e-23,NCBI: e-value 1e-28.Despite the strong alignment, this gene does not have a known function. BLAST results in both PhagesDB and NCBI return hypothetical proteins. HHpred gives no quality hits (best match only has 45.5% probability to an SsrA-binding protein), which is not high enough to confidently assign a function.The gene is not a transmembrane protein and lies in a region with conserved gene order among phages in the FD cluster, though neighboring genes also lack known functions. Therefore, this gene has no known function (NKF).

Gene 14: Both Glimmer and GeneMark call this gene and agree on the start site at 7551. There is clear coding potential shown in the GeneMark-host trained maps. The gene overlaps the previous gene by 29 bp, which is within acceptable limits and does not violate guiding principles. The selected start site is the longest ORF. The RBS score is -2.505 with a z-score of 3.003, indicating a strong ribosome binding site. According to Starterator, this start site is among the most frequently called in published annotations and is supported by 2 manual annotations, further validating this decision. There is strong 1:1 alignment with related phages in the FD cluster:PhagesDB: Swim and Tillums (e-value: 1e-30) NCBI: Swim and Tillums (e-value: 3e-38), The function of this gene is unknown. Both PhagesDB and NCBI return this as a hypothetical protein. HHpred suggests a possible structural binding protein with 77.93% probability, which falls short of the 90% threshold for a confident call. Therefore, this gene has no known function (NKF).

Gene 15: Glimmer and GeneMark both call this gene with a start of 7745. There is coding potential. This gene is called in other annotated phages and is long enough without significant overlap. The start site has a RBS score of -3.000 which is the best and a z-score of 2.807. The predicted start codon is the longest ORF. Supported by Phamerator with a start at 7745 with 4MA's. Has a 1:1 alignment with Swim, Tillums, and Anjali on NCBI blast and Phages DB. Has alignment with major tail protein on NCBI blast and Phages DB. On HHPred, the best alignment is: Major tail protein; phage tail, VIRUS;{Staphylococcus phage Andhra} Probability: 100, e-value: 1.3e-107. Compared to other phages in the same cluster (Swim and Tillums), this gene is located in a similar region. The proposed function of major tail protein is found on the approved function list.

Gene 16: Glimmer and GeneMark both call this gene with a start of 9940. There is coding potential. This gene is called in other annotated phages and is long enough without significant overlap. The start site has a RBS score of -3.298 which is the best and a z-score of 2.908. The predicted start codon is the longest ORF. Supported by Phamerator with a start at 9940 with 4MA's. Has a 1:1 alignment with Swim, Mendel, and Anjali on NCBI blast and Phages DB. Has alignment with portal protein on NCBI blast and Phages DB. On HHPred, the best alignment is: Head-tail connector (Portal protein); Bacteriophage, Portal protein, cryoEM, VIRAL PROTEIN; 3.2A {Bacillus phage GA-1} Probability:100, e-value: 3.1e-61. Compared to other phages in the same cluster (Swim and Tillums) this gene is located in a similar region. The proposed function of portal protein is found on the approved function list.

\*\*\*Gene 17: Glimmer and GeneMark both call this gene with a start of 10920. There is coding potential. This gene is called in other annotated phages and is long enough without significant overlap. The start site has a RBS score of -3.212 which is the best and a z-score of 2.707. The predicted start codon is the longest ORF. Supported by Phamerator with a start at 10920 with 14MA's. Has a 1:1 alignment with Mendel, Swim, and Tillums on NCBI blast and Phages DB. Has alignment with lower collar protein on NCBI blast and Phages DB. On HHPred, the best alignment is: Lower collar protein; phage tail, VIRUS;{Staphylococcus phage Andhra}Probability: 100, e-value: 9.6e-39. Compared to other phages in the same cluster (Swim and Tillums), this gene is located in a similar region. The proposed function of lower collar protein is not found on the approved function list so listed it as Hypothetical Protein but this should be changed

Gene 18: Glimmer and GeneMark both call this gene with a start of 11575. There is coding potential. The gene is called in other annotated phages and has an overlap of 35bp. The start site has a RBS score of -3.311 which is the best and a z-score of 2.594. The predicted start codon is not the longest ORF. Supported by Starterator at 11575 with 4MA's. Has a 1:1 alignment with Swim, Mendel, Tillums, and Anjali on NCBI blast and Phages DB. Has alignment with minor tail protein on NCBI blast and Phages DB. On HHPred, there is no quality match, but the closest alignment is: Tail Fiber of phage P68; structural protein, receptor binding protein, beta-propeller, SUGAR BINDING PROTEIN; 2.0A {Staphylococcus aureus phage P68} Probability: 66.43, e-value: 53. Compared to other phages in the same cluster (Swim and Mendel), this gene is located in a similar region. The proposed function of minor tail protein is found on the approved function list.

\*\*Gene 19: Glimmer and GeneMark both call this gene with a start of 13118. There is coding potential. The gene is called in other annotated phages and has a gap of 235bp with the previous gene. There is no coding potential within that gap region. The start codon has a RBS score of -7.894 and a z-score of 0.865. The predicted start codon is not the longest ORF. Supported by Phamerator with a start at 13118 with 1MA's. Has a 1:1 alignment with Swim on NCBI blast and Phages DB. Has high alignment with endolysin, protease M23 domain on NCBI blast and Phages DB. On HHPred, the best alignment is: Peptidase M23; membrane protein, enzyme, TRANSPORT PROTEIN; HET: ADP;{Vibrio cholerae} Probability: 99.14, e-value: 5.2e-9. Compared to other phages in the same cluster (Swim and Tillums), this gene is located in a similar region. The gene is not a transmembrane protein. The proposed function of endolysin (protease M23 domain) is not found on the approved function list but needs to be added

Gene 20; Both GeneMark and Glimmer call this gene.Coding potential confirmed between start site 13779 and stop site 14846. Present in other FD cluster phages (Mendel, Anjali) with Pham 11023. Suggested start site: 13779 by both Glimmer and GeneMark. Strong RBS score and Z-score. Longest ORF without excessive overlap. Supported by Starterator and BLAST analyses (E value = 0).Decision: Start site 13779.Function of Putative Protein: Aligns with terminase function in PhagesDB and NCBI (alignment 1:1, E value = 0). High probability in HHPred (99.76% for terminase).Located in conserved genomic region with functional genes. Decision: Terminase.

Gene 21:Gene Identification: Both GeneMark and Glimmer call this gene.Coding potential confirmed by GeneMark-host trained map. Present in other FD cluster phages (Mendel, Anjali) with Pham 205777.Suggested start site: 14864 by both Glimmer and GeneMark. Strong RBS score and Z-score. Longest ORF without excessive overlap. Supported by Starterator and BLAST analyses (E value = 3e-32). Decision: Start site 14864. Function of Putative Protein: Aligns with holin function in PhagesDB (alignment 1:1, E value = 4e-05) and "membrane protein" in NCBI (alignment 1:1, E value = 5e-39). No high-probability match in HHPred (highest probability 88.6%). Possible transmembrane protein according to DeepTMHMM so labelled it as such.

Gene 22: Both GeneMark and Glimmer call this gene. Coding potential confirmed by GeneMark-host trained map. Present in other FD cluster phages (Mendel, Swim) with Pham 3204. Suggested start sites: 15112 by Glimmer and 15070 by GeneMark. Start site 15070 has an RBS score of -3.873 and Z-score of 2.33 has better final score and smaller gap. - Start site 15070 aligns with related phages (Mendel, Swim) at 1:1; 15112 lacks 1:1 alignment. Decision: Start site 15070, supported by 4 manual annotations and evidence from BLAST analyses.

Function of Putative Protein:Aligns with a protein of unknown function in PhagesDB (E value = 8e-76) and membrane protein in NCBI (E value = 4e-92). No high-probability match in HHPred (highest probability 38.41%). Located in conserved genomic region with functional genes. 4Possible transmembrane protein according to DeepTMHMM. Decision: Membrane Protein.

Gene 23: GeneMark and Glimmer call this gene. Coding potential confirmed. Present in other FD cluster phages (Tillums, Anjali) with Pham 176016. Suggested start site: 15491 by both Glimmer and GeneMark. RBS score of -6.081 and Z-score of 1.291 for start site 15491 andaligns with related phages (Mendel, Anjali) at 1:1. Decision: Start site 15497, supported by 31 manual annotations and evidence from BLAST analyses. Function of Putative Protein: Unknown function in PhagesDB (E value = 2e-49) and hypothetical protein in NCBI (E value = 3e-62. High-probability match in HHPred (PspB; Probability Score: 97.63% but e values are not great) Possible transmembrane protein according to DeepTMHMM. Decision: membrane protein

\*\*Gene 24; Gilmmer and GeneMark agree for start site of 15879. This has gene has coding potential, supported by Phamerator with a start site of 15879 with 4 MA's. In PhagesDB the start site 15879 has 1:1 alignment with Swim (1:1) (e-value e-125), Seldom (1:1)(e-value e-125), Anjali (1:1)(e-value e-122), and Tillums (1:1) (e-value e-121). In NCBI, start site 15831 has a 1:1 alignment with Swim (2e-value-156), Anjali (1e-152), and Tillums (4e-152). In phages DB Blast, phage Swim aligns the protein sequence and lists it as an endolysin N-acetylmuramyl-L-alanine amidase domain (FD cluster, e-value e-125). Additionally, phage Tillums aligns with protein sequence for endolysin N-acetylmuramyl-L-alanine amidase domain (FD cluster, e-value -121). In HHPred, the protein sequence aligns with protein N-acetylmuramyl-L-alanine amidase (99.39 % Probability, e-value of 1.3e-10). In HHPred, the most informative match is N-acetylmuramoyl-L-alanine amidase protein with a probability of 99.28 and e-value of 7.5e-10. As there is no Lysin B we cannot call it Lysin A, N-acetylmuramyl-L-alanine amidase doesn’t have a lysin B, therefore it is an endolysin. This gene has a function of endolysin, N-acetylmuramoyl\_ L alanine amidase domain and has to be added.

Gene 25: Both Gilmmer and GeneMark agree at a start site of 16523. Start site 16523 has longest ORF. This gene has coding potential, supported by Phamerator with a start at 16523 with 2MA's. Has a 1:1 alignment with phages Mendel, Swim, Tillums, Anjali on NCBI blast and Phages DB databases. Phages DB and NCBI blast state protein sequence aligns with minor tail protein (e-value 0). In HHPred, the most informative match is Collagen-like protein 6 (98.03% Probability, e-value of .000034). Functional assignments list, states the minor tail protein has significant hits in HHPred as a collagen-like protein and therefore can be called a minor-tail protein.

Gene 26; Both Gilmmer and GeneMark agree with a start site of 18676, this gene has coding potential and longest ORF. Supported by Phamerator with a start at 18676 with 3MA's. The best PhagesDB match with 1:1 alignment is with phages Anjali (e-value of 2e-44), Tillums (e-value 3e-44), GoodGraces ( e-value 3e-44), and Swim (e-value 3e-43). NCBI blast didn’t show results for alignment with start site 18676. In phages DB Blast, phage Anjali aligns the protein sequence and lists it as no known function (FD cluster, e-value 2e-44). In HHPred, the protein sequence aligns best with protein ORF 4 (72.97% Probability & e-value of 24), however due to the low probability and e-value, gene 27 likely doesn't encode this protein. Therefore, protein has no known function.

Gene 27; Both Gilmmer and GeneMark agree with start site of 18918, has coding potential. Start site 18918 supported by Phamerator with 5MA's. In PhagesDB the start site 18918 has 1:1 alignment with phage Tillums (1:1) (e-value 2e-15) and phage Good graces (1:1) (e-value 6e-15). In NCBI, start site 18918 has a 1:1 alignment with Tillums (e-value of 4e-17), Swim (1e-14), and Anjali (1e-11). In phages DB Blast, the best match is with phage Tillums aligns 1:1 with protein sequence (e-value 2e-15). In HHPred, the protein sequence aligns best with small protein subunit found in plants (Probability of 70.54%, e-value of 7.7). However, since the protein sequence alignment has low probability and a high e-value, this is not the best match. Therefore, this gene has no known function.

Gene 28: Gilmmer and GeneMark agrees at a start of 19043. Has coding potential. Has no 1:1 alignment with any other phages on NBI blast and Phages DB has a no known function. Is an orpham

Gene 29: Gilmmer and GeneMark do not agree on a start site. Has coding potential. Supported by Phamerator with GeneMark at a start of 9338 with 1MA's. Has no 1:1 alignment with other phages on NCBI blast and Phages DB has no known function.

Gene 30: There is coding potential in this areaGilmmer and GeneMark does not agree at a start.. Staterator is not informative. Has coding potential. Chose 19558 as it has less of an overlap and Has 1:1 alignment with Swim on NCBI blast and Phages DB has no known function.

Gene 31: Glimmer and GeneMark agrees at a start of 19719. There is coding potential. Has no 1:1 alignment with other phages on NBI blast and Phages DB has a function of no known function. No information on starterator. Is an orpham