**Phrank15 phage. Oregon Coast Community College. Matthew R. Fisher.**

**Gene 1**

-Original Glimmer call @bp 389 has strength 3.48; GeneMark calls start at 131

Start 56, Stop 607, Length 552 bp.

**It is a gene**: GeneMarkS indicates high coding potential. BLASTp: Many significant results (e values smaller than E-4) on phagesdb & NCBI.

**Start codon**: Start codon 389 doesn’t include all coding potential. Changing the start to 56 reduces the gap before the gene, improves the RBS score compared to the other called starts, and means it’s 100% aligned and >99% similar to a gene present in phages Hestia and Persistence (both in the same cluster as Phrank15). Making this change would mean it’s no longer an orpham, but instead in pham 81374 and would have the most common manually- annotated start.

**Function**: terminase, small subunit. HHPred: 2 results ~95% probability of it being phage terminase, small subunit. Blastp: significant results from NCBI said minor tail protein/terminase small subunit, phagesdb said terminase small subunit.

**Gene 2**

-Original Glimmer call @bp 783 has strength 4.20; GeneMark calls start at 690

Start 597, stop 2330, Length 1734 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results on phagedb & NCBI. Gene is in a pham with 1316 members.

**Start codon**: Glimmer-called start at 783 doesn’t include all coding potential. GeneMark-called start at 690 excludes a small region of weak coding potential and leaves a 83 bp gap. Picking start 597 contains all coding potential and creates an 11 bp overlap with gene 1. It also improves the top NCBI BLAST result from score of 2734 to 2903 and makes it 97% identical to gene 2 in cluster-mate Anekin. Starterator: this gene doesn’t have the most annotated start in pham. Start at 783 has no MA’s, the starts at 690 and 597 each have 2 MAs. RBS between 597 and 690 are comparable. 597 is chosen.

**Function**: terminase, large subunit. HHPred: Many >99% probability results for terminase, large subunit. Blastp: many significant results for terminase, large subunit on NCBI and phagesdb.

**Gene 3**

-Original Glimmer call @bp 2347 has strength 6.28

Start 2347, stop 3684, Length 1338 bp.

**It is a gene**: GeneMarkS indicates high coding potential throughout ORF. Blastp: Many significant results on phagedb & NCBI. In a pham with 1628 members.

**Start codon**: Starterator report shows no clear consensus; several different starts are used in the pham. This phage does not have the most annotated start in pham. The start at 2347 is the most annotated start present in this phage, but it excludes some coding potential. The 2 upstream starts have both been manually annotated in other phages, but have much worse RBS scores. Additionally, the most upstream is a TTG start with a 11 bp overlap. The start at 2347 is chosen.

**Function**: portal protein. HHpred: 100% probability it’s a portal protein. Blastp: many significant results for portal protein on NCBI and phagesdb.

**Deleted (auto-annotated as gene 4)**

- Original Glimmer call @bp 4316 has strength 3.14 \*\* not called by GeneMark

DELETE/REMOVE

**It is a not gene**: GeneMarkS indicates very poor coding potential. No BLASTp results on NCBI, only 1 significant result on phagedb (Anekin\_draft, in the same cluster as this phage). This is a reverse gene tightly tucked in-between 2 forward genes. Also, there is much stronger coding potential in a forward reading frame within this region (see gene 4).

**Gene 4**

Original Glimmer call @bp 4311 has strength 5.02; GeneMark calls start at 3687

Start 3687, Stop 4901, Length 1215 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results on phagedb and NCBI.

**Start codon**: 4311 start excludes a large region of high coding potential. Using the GeneMark start at 3687 would put this in pham 81256 and start 3687 has the most MAs. Choosing this start also improves the top BLASTp result from 995 to 1846 and changes the alignment from 1:209 to 1:1.

**Function**: capsid maturation protease. HHPred: Two results >90% for Mu protein and minor capsid.

BLASTp: many significant results for capsid maturation protease on NCBI and phagesdb, including two that are 1:1 aligned and >98% similarity, both are capsid maturation proteases.

**Gene 5**

- Original Glimmer call @bp 5067 has strength 6.92

Start 5067, Stop 5600, Length 534 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results on phagesdb & NCBI. Gene in a pham w/ 16 members.

**Start codon**: Start contains full coding potential and has the best RBS score. There is a 166 bp gap between upstream gene. There are 2 available upstream starts, but RBS are worse and they don’t include any more regions of coding potential. The start chosen has the most MA’s (11 of 11) in the pham.

**Function**: scaffolding protein. Blastp: many significant results for scaffolding protein. HHPred: Top result is for scaffolding protein with high significance. The probability is 99.3% that the function is a scaffolding protein.

**Gene 6**

- Original Glimmer call @bp 5630 has strength 10.55

Start 5630, Stop 6025, Length 396 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results results on phagedb & NCBI. Gene in a pham w/ 144 members.

**Start codon**: Start is the most upstream option and contains full coding potential and best RBS score, multiple 100% alignments with pham members.

**Function**: capsid decoration protein.HHPred: >99% probability for head/capsid decoration protein. Blastp: Many significant results for capsid decoration protein. Blastp has many very high alignments for capsid decoration protein.

**Gene 7**

-Original Glimmer call @bp 6041 has strength 13.76

Start 6041, stop 7081, Length 1041 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results on phagesdb & NCBI. Gene is in a pham w/ 154 other members.

**Start Codon**: Called start contains full coding potential and has the best RBS score, and is the farthest upstream start.

**Function**: major capsid protein**.** HHPred: multiple 100% probability that the function is major capsid protein. Blastp: many significant results for major capsid protein.

**Gene 8**

- Original Glimmer call @bp 7093 has strength 12.41

Start 7093, Stop 7431, Length 339 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant and full alignment results on phagesdb and NCBI. Gene in pham w/ 21 members.

**Start Codon**: Start contains full coding potential, has the best RBS score, and is the farthest upstream start.

**Function**: Hypothetical Protein. HHPred: many significant results but coverage is poor (about 25%) and there are no obvious matches to official functions. The matches include endonuclease, recombination, and HeH. But BLAST results with >95% similarity call it a hypothetical protein. Deep TMHMM says not a membrane protein.

**Gene 9**

- Original Glimmer call @bp 7517 has strength 8.04; GeneMark calls start at 7448

Start 7448, Stop 7804, Length 357 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results on phagesdb and NCBI. Gene in pham w/ 246 members.

**Start codon**: GeneMarkS start has better RBS score (by 2 orders of magnitude) and includes all regions of high coding potential. Start contains full coding potential, is the most upstream start, and has the best RBS score.

**Function**: head-to-tail adaptor. HHPred: Many significant results that support head-to-tail adaptor – matches to SPP1 gp15 and HK97 gp6. Blastp: Many significant results on NCBI database for head-to-tail adaptor.

**Gene 10**

- Original Glimmer call @bp 7801 has strength 7.66

Start 7801, Stop 8199, Length 399 bp.

**It is a gene**: GeneMarkS shows high coding potential. Pham has 26 members. Blastp: Many significant results on phagedb & NCBI.

**Start codon**: There is an overlap of 4 base pairs with upstream gene. The start codon contains all coding potential and has best RBS score.

**Function**: head-to-tail stopper. HHPred: shows 1 result of 98% probability of being head to tail stopper by the SPP1 16 structure, as per instructions on the SEA gene function list. Other top results of HHPred heavily refer to head and tail components. BlastP: many significant results for head-to-tail stopper on NCBI and phagedb. All members of pham say head-to-tail stopper.

**Gene 11**

- Original Glimmer call @bp 8196 has strength 6.47

Start 8196, Stop 8579, Length 384 bp.

**It is a gene**: GeneMarkS shows high coding potential. Pham has 25 members. Blastp: Many significant results on phagedb & NCBI.

**Start codon**: Gene 12 has a 4 base pair overlap with upstream gene. Start contains all coding potential, any other start would exclude coding potential.

**Function**: Hypothetical Protein. HHPred: top result of minor capsid protein and several for tail protein – seemingly contradictory results. Consulted Debbie and she said to call it as hypothetical protein. BlastP: NCBI and database results that are significant show function as function unknown. All members of pham state unknown function. Deep TMHMM says not a membrane protein.

**Gene 12**

- Original Glimmer call @bp 8561 has strength 7.44; GeneMark calls start at 8576

Start 8,576, Stop 8,974, Length 399 bp.

**It is a gene**: GeneMarkS shows high coding potential. Pham has 25 members. Blastp: Many significant results on phagedb & NCBI.

**Start codon**: Glimmer start creates a 19 bp overlap with upstream gene, GeneMark start creates a 4 base pair overlap, but excludes a small amount of med-high coding potential. A 4 bp overlap seems more likely, thus the 8576 start is chosen. Also, the 8576 start has the most manual annotations in the pham.

**Function**: tail terminator. HHPred: second result is tail terminator >99% other top result is “tail-to-head joining” but others are insignificant or unknown. BlastP: many significant results saying tail terminator on both NCBI and phagesdb. Majority members of pham are reporting tail terminator.

**Gene 13**

- Original Glimmer call @bp 9021 has strength 12.30

Start 9021, Stop 9206, Length 186 bp.

**It is a gene:** GenemarkS shows high coding potential. Pham has 22 members. Blastp: Many significant results on phagedb & NCBI.

**Start codon**: Start includes all coding potential, is the first in the ORF. RBS score is the best.

**Function**: Hypothetical Protein. HHPred: No significant results, all below 90% BlastP: Both database and NCBI report unknown function. All members of pham report unknown function. Deep TMHMM says not a membrane protein.

**Gene 14**

- Original Glimmer call @bp 9199 has strength 10.39

Start 9,199, Stop 9,714, Length 516 bp.

**It is a gene**: GeneMarkS shows high coding potential. Pham has 23 members. Blastp: Many significant results on phagedb & NCBI.

**Start codon**: Start includes all coding potential, is the first start in ORF, moving it downstream would exclude coding potential and shorten gene by nearly 100 bp. There is an 8 base pair overlap with the upstream gene.

**Function**: major tail protein. HHPred: 94% result for major tail protein and a 92% result for tail tube protein, which is same as major tail protein in the function list. BlastP: Many significant results saying major tail protein on both NCBI and database. Many pham members report major tail protein.

**Gene 15**

- Original Glimmer call @bp 9804 has strength 8.42

Start 9804, Stop 10340, Length 537 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp shows quite a few significant results on phagesdb and NCBI. Gene is in a pham with 22 other members.

**Start Codon:** Start contains full coding potential and has the best RBS score by far. Is the first possible start in the ORF. Starterator report: this start has 17 out of 17 Mas.

**Function**: tail assembly chaperone. HHpred: many significant results over 90% and multiple significant matches on BlastP that show tail assembly chaperone.

**Gene 16**

- Original Glimmer call @bp 10400 has strength 5.21

Start 9804, Stop 10696, Length 894 bp

**It is a gene**: GenemarkS indicates high coding potential. Blastp shows many significant result on phagesdb and NCBI.

**Start Codon**: There is a -1 translational frameshift for this gene, which is why it has the same start codon as gene 16. The slip occurs at the first T in the sequence CCCTTTTT, where that T is located at base 10259. The slip results in a PF (gene 16) – PF (gene 17) amino acid sequence. This same slippery sequence is found in at least 3 other phages in this AY cluster (Anekin, Isolde, and Persistence).

**Function**: tail assembly chaperone. HHpred and BLAST: many significant results over 90% and multiple significant matches on BlastP close to 0 that show tail assembly chaperone.

**Gene 17**

- Original Glimmer call @bp 10870 has strength 5.66

Start 10870, Stop 10682, Length, 189

**It is a gene**; GenemarkS indicated high coding potential. BlastP shows a few significant results on Phagesdb and NCBI.

**Start Codon**: Starterator report says this start has 7 of 7 MAs. RBS score third best, but the two better scores would shorten gene by ~ 60 bp and would exclude a region of moderate coding potential.

**Function**: membrane protein. HHpred: just one significant results over 90%, for Sigma-M inhibitor. BlastP shows many matches with significant results close to 0 as an unknown function protein. Deep TMHMM reports two membrane domains, therefore it is a membrane protein.

**Gene 18**

- Original Glimmer call @bp 10912 has strength 7.28

Start 10912, Stop 14718, length: 3807

**It is a gene**: GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI showed many significant results.

**Start Codon**: This start includes all regions of coding potential and has the best RBS score by a factor of 10. Starterator says that this start has 17 of 21 MAs in the pham.

**Function**: tape measure protein. HHpred shows many significant results over 90% that match. BlastP shows multiple significant results, all indicating tape measure protein.

**Gene 19**

-Original Glimmer call @bp 14718 has strength 11.92

Start: 14718, Stop: 15545, Length: 828

**It is a gene**: GeneMark indicates high coding potential. BalstP on Phagesdb and NCBI showed many significant results.

**Start Codon**: Has the best RBS score by far, has 1 bp overlap with upstream gene. This is the first start in ORF and includes all coding potential.

**Function**: minor tail protein. HHpred showed no significant results over 90%. BlastP showed many significant results close to zero that matched this function. BlastP on Phagesdb and NCBI agree on function and match other members in the pham. This can be called a minor tail protein based on synteny because it is immediately downstream of the tape measure protein.

**Gene 20**

- Original Glimmer call @bp 15557 has strength 10.75

Start 15557, Stop 16858, Length 1302 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows quite a few significant results on Phagesdb and NCBI. Gene is in a pham with 72 other members.

**Start codon:** The start is the first in the ORF and the only to include all regions of coding potential. Any other start would shorten the gene by >300 bp.

**Function:** minor tail protein. HHpred: Many significant results over 99% for tail protein. BlastP and pham all call it minor tail protein.

**Gene 21**

- Original Glimmer call @bp 16867 has strength 7.00

Start 16867, Stop 17826, Length 960 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows mant significant results on Phagesdb and NCBI. Gene is in a pham with 46 other members.

**Start Codon:** Starterator report says this start has 31 of 39 MAs. First start in the ORF and only one to include all coding potential; the RBS score is the best.

**Function:** minor tail protein. HHpred: showed no significant results over 90%. Many significant results with BlastP on NCBI and Phagesdb for minor tail protein. This can be called a minor tail protein based on synteny because it is a few genes downstream of the tape measure protein.

**Gene 22**

- Original Glimmer call @bp 17826 has strength 7.49

Start 17826, Stop 18485, Length 660 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 22 other members.

**Start Codon:** Best RBS score. 1 bp overlap with upstream gene. Other starts would either create a huge overlap or gap.

**Function:** minor tail protein. HHpred: showed no significant results over 90%. BlastP on NCBI and Phagesdb showed many significant results for minor tail protein. Other members of this pham that share this gene use this as the function. This can be called a minor tail protein based on synteny because it is a few genes downstream of the tape measure protein.

**Gene 23**

-Original Glimmer call @bp 18496 has strength 9.59

Start 18496, Stop 18822, Length 327 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 9 other members.

**Start Codon:** Starterator report says this start has 6 of 6 MAs. Best RBS score. Contains most coding potential. The next upstream start would create a very large overlap with upstream gene.

**Function:** membrane protein. HHpred: showed a few significant results over 90% for unknown function. BlastP shows many matches with significant results as an unknown function protein. Deep TMHMM reports one membrane domain, therefore it is a membrane protein.

**Gene 24**

- Original Glimmer call @bp 18797 has strength 3.71

Start 18797, Stop 19030, Length 234.

**It is a gene:** GeneMark indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 9 other members.

**Start Codon:** Starterator report says this start has 6 of 6 MAs. Best RBS score and GeneMark and Glimmer agree on start. There is a 26 bp overlap with upstream gene. Moving downstream to next start would shorten gene by ~90 bp, creating large gap, and exclude region of coding potential.

**Function:** Hypothetical Protein. HHpred: no significant result over 90%. BlastP on NCBI and Phagesdb showed many significant results for unknown function protein. Deep TMHHM says not a membrane protein. Therefore, it is a Hypothetical Protein.

**Gene 25**

- Original Glimmer call @bp 19027 has strength 8.53

Start 19027, Stop 19338, Length 312.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 9 other members.

**Start Codon:** First start in ORF, 4 bp overlap with upstream gene, it is the only start to include all regions of coding potential.

**Function:** membrane protein. HHpred: Only 1 significant result for unknown function. BlastP shows many matches with significant results as an unknown function protein. Deep TMHMM reports 3 membrane domains, therefore it is a membrane protein.

**Gene 26**

- Original Glimmer call @bp 19335 has strength 14.53

Start 19,335, Stop 20,282, Length 948

**It is a gene:** GeneMarkS shows high coding potential. Pham has 29 members. Blastp: many significant results on phagedb and NCBI.

**Start Codon:** This is the first start in the ORF and only one to include all coding potential. 4 bp overlap with upstream gene.

**Function:** endolysin. HHpred: several significant results for M23 peptidase domain. BlastP: First result of NCBI indicated peptidase but also endolysin. Phagedb show many results of endolysin, unknown, and some peptidase. Official function list says to call it endolysin if it has this domain but doesn’t have a clear lysin A and lysin B gene.

**Gene 27**

- Original Glimmer call @bp 20294 has strength 8.75

Start 20,294, Stop 20,602, Length 309

**It is a gene:** GeneMarkS shows high coding potential. Pham has 94 members. Blastp: many significant results on phagedb and NCBI.

**Start Codon:** First start in ORF and the only one to include all coding potential. RBS score is the best.

**Function:** membrane protein. HHpred: no significant results above 90%. BLAST: results for unknown function and membrane protein. Deep TMHMM indicate membrane protein with 2 domains.

**Gene 28**

- Original Glimmer call @bp 20613 has strength 7.10

Start 20,613, Stop 20,918, Length 306

**It is a gene:** GeneMarkS shows high coding potential. Pham has 22 members. Blastp: many significant results on phagedb and NCBI.

**Start Codon:** First start codon in ORF, only one to contain all coding potential. RBS score is the best.

**Function:** Hypothetical Protein. HHpred: no significant results above 90%. Blastp: both NCBI and phagedb indicate hypothetical protein. All members of pham report protein as unknown. Deep TMHMM: not a membrane protein.

**Gene 29**

- Original Glimmer call @bp 20915 has strength 5.93

Start 20,915, Stop 21,289, Length 375

**It is a gene:** GeneMarkS shows high coding potential. Pham has 20 members. Blastp: many significant results on phagedb and NCBI.

**Start Codon:** RBS score shows is the best. First start in ORF and only one to include all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: no significant results above 90%. Blastp: NCBI lists both tail protein and hypothetical protein. phagedb results show hypothetical protein. All pham members have an unknown protein. Deep TMHMM: not a membrane protein.

**Gene 30**

- Original Glimmer call @bp 21293 has strength 10.50; GeneMark calls start at 21314

Start 21,293, Stop 22,744, Length 1452

**It is a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: Many significant results on phagedb & NCBI.

**Start Codon:** I picked the Glimmer start. This start codon is first in the ORF and only one to contain all coding potential. RBS score is the best by factor of 100x. The GeneMark start would exclude a small region of med-high coding potential

**Function:** Hypothetical Protein. HHpred: many significant results above 90% for “Putative catalytic domain of uncharacterized bacterial proteins from the carbohydrate esterase” and similar results. Blastp: both NCBI and phagedb indicate hypothetical protein/unknown function and minor tail protein. No clear consensus, thus it’s a hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 31**

- Original Glimmer call @bp 22754 has strength 9.42

Start 22,754, Stop 23,343, Length 600

**It is a gene:** GeneMarkS shows high coding potential. Pham has 2 members. Blastp: 2 significant results on NCBI.

**Start Codon:** Start codon is first in the ORF and the only one to contain all coding potential. Has best RBS score.

**Function:** Hypothetical Protein. HHpred: Only 1 significant result, for “lectin like domain”. BLASTp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 32**

- Original Glimmer call @bp 23357 has strength 6.27

Start 23,357, Stop 23,488, Length 132 bp.

**It is a gene**: GeneMarkS indicates high coding potential Blastp: no significant results on phagesdb or NCBI, but in a pham with 4 members

**Start Codon**: This is the only start in the ORF.

**Function**: Hypothetical Protein. HHPred: no significant results > 90%. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 33**

- Original Glimmer call @bp 23639 has strength 6.20

Start: 23,639, Stop: 24,190, length 552 bp.

**This is a gene**: GeneMarkS shows high coding potential. In a pham of 2 members. Many significant BLAST results on NCBI.

**Start Codon:** Start contains full coding potential and the best RBS score by far. This is the first start in the ORF.

**Function**: HNH endonuclease. HHPred: many significant results for endonuclease, including HNH endonuclease. BlastP: multiple significant results for HNH endonuclease.

**Gene 34**

- Original Glimmer call @bp 24383 has strength 7.98; GeneMark calls start at 24389

Start: 24,389, Stop: 24,246, length 144 bp.

**This is a gene**: GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 7 members.

**Start Codon**: GeneMark and Glimmer disagree on start. The GeneMark start is found in all pham members and has 4 of 4 MAs. The Glimmer start is found in only 2 of 7 pham members. The GeneMark start is thus more conserved, but has a slightly worse RBS score but might capture a small region of low coding potential. The GeneMark start was chosen.

**Function**: Hypothetical Protein. HHPred: two results >94% for Bacterial RNA polymerase inhibitor, but that’s not on the official function list. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 35**

- Original Glimmer call @bp 24753 has strength 10.81

Start 24,753, Stop: 24,559, length 195 bp.

**This is a gene**: GenemarkS shows high coding potential. Orpham. Blastp: one significant result on NCBI.

**Start Codon**: This start is the first in the ORF and the only one to include all regions of coding potential. It has the best RBS score. The next available start would cut the gene in half and exclude large regions of coding potential.

**Function**: Hypothetical Protein. HHPred shows no significant results. BlastP: one significant result for hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 36**

- Original Glimmer call @bp 24822 has strength 5.43; GeneMark calls start at 24864

Start 24,822, Stop 25,220, length 399 bp.

**This is a gene**: GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 31 members.

**Start Codon:** Starterator: The Glimmer start only occurs in this phage, and no other phage in the pham. The GeneMark start, however, occurs in >90% of phages and has 20 of 26 MAs. The GeneMark start, however, has a worse RBS score and might exclude a region of low-to-mid coding potential. Regardless, GeneMark start is chosen since it is conserved in the pham. (Besides the GeneMark start, this gene has no other start that has been MA’ed).

**Function:** Hypothetical Protein. HHPred: no significant results. BlastP: Many significant results for hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 37**

- Original Glimmer call @bp 25421 has strength 5.81

Start: 25,445, Stop: 25,209, length 237 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 175 members.

**Start Codon:** The start chosen by Glimmer does not contain all coding potential. Starterator: this gene doesn’t have the most annotated start. But it does have the second most annotated start (26 MAs) at position 25445. This includes all coding potential. For these reasons, it was chosen.

**Function**: helix-turn-helix DNA binding protein. HHPred: multiple significant hits but the results are ambiguous (many different functions listed). BlastP: Many significant results for helix-turn-helix DNA binding protein. All non-draft members of this large pham also call it this.

*Note****:*** *the end of this reverse end and the upstream forward gene overlap by 12 bp.*

**Gene 38**

- Original Glimmer call @bp 25579 has strength 6.26

Start: 25,579, Stop: 26,058, length: 480 bp.

**This is a gene:** GeneMarkS shows high coding potential. Blastp: Many significant results. Pham has 81 members.

**Start Codon:** The start contains all coding potential, except perhaps a small region of weak potential. It has the best RBS score by far. Starterator: it does not have the most common start. The chosen start has 2 MAs, more than any other start that this gene has.

**Function:** HNH endonuclease. HHPred: many significant results for endonuclease, including HNH endonuclease. BlastP: many significant results HNH endonuclease. Nearly all pham members are HNH endonuclease.

*Note****:*** *the end of this forward gene and the end of the downstream reverse overlap by 39 bp.*

**Gene 39**

- Original Glimmer call @bp 27192 has strength 11.13

Start: 27,192, Stop: 26,020, length: 1173 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 727 members.

**Start Codon:** Not the first start in ORF, but contains all coding potential and has a 8 bp overlap with downstream reverse gene. Has the best RBS score. Starterator: has the most annotated start.

**Function**: tyrosine integrase. HHPred: multiple >99% probabilities for tyrosine integrase, integrase, and recombinase. BlastP: many significant results for tyrosine integrase.

**Gene 40**

- Original Glimmer call @bp 27763 has strength 7.18

Start: 27,763, Stop 27,185, length 579 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Numerous significant results. Orpham.

**Start Codon:** First start in ORF and only one to contain all coding potential. 8 bp overlap with downstream reverse gene.

**Function:** Hypothetical Protein. HHPred: one significant result, for unknown function. BlastP: many significant results for hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 41**

-Original GeneMark call @bp 27848

Start: 27,848, Stop: 27,756, length 93 bp.

**This is a gene:** GenemarkS shows a short spike in high coding potential. Not called by Glimmer. Blastp: Just one significant results in NCBI. 2 significant results in phagesdb, all 2 phages within the same cluster as this phage (AY).

**Start Codon:** 4 bp overlap with downstream reverse gene. Has best RBS score. Contains all coding potential.

**Function:** Hypothetical Protein. HHPred:no significant results. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 42**

- Original GeneMark call @bp 27997

Start: 27,997. Stop 27,845, Length: 153 bp.

**This is a gene:** GenemarkS shows high coding potential. Glimmer doesn’t call it. Blastp: One significant result on NCBI. Pham has 2 members.

**Start Codon:** Start is the first in the ORF and the only to include all coding potential. The only other start would reduce gene to only 36 pm. 4 bp overlap with next upstream, reverse ORF.

**Function:** Hypothetical Protein. HHPred: one significant result, for “Queuosine biosynthesis protein”. BlastP: unknown function. Deep TMHMM: not a membrane protein.

**Deleted (auto-annotated as gene 42)**

- Original Glimmer call @bp 27887 has strength 3.45 \*\* not called by GeneMark

**NOT A GENE**. Not called by GeneMark, which shows no coding potential in this region. Also, it would overlap with genes 43 and 45. No BLAST results on NCBI.

**Gene 43**

- Original GeneMark call @bp 28545

Start: 28,545, Stop: 27,994, Length 552 bp.

**It is a gene:** GeneMarkS shows high coding potential. Not called by Glimmer. Blastp: many highly significant results. Pham has 3 members.

**Start Codon:** Start is first available in ORF and only to contain all coding potential. RBS score is slightly worse than the next start, which has a TTG sequence.

**Function:** Hypothetical Protein. HHPred: no significant results. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 44**

- Original Glimmer call @bp 29000 has strength 10.68

Start 29,000, Stop 29209, Length 210 bp.

**It is a gene:** GeneMarkS indicates high coding potential. Blastp: a few significant values. Pham has 5 members.

**Start Codon:** Start is first in the ORF and only one to include all coding potential. RBS is best by far. Every gene in the pham uses this start.

**Function:** Hypothetical Protein. HHPred: only 3 significant results, include “IMMUNITY PROTEIN” and “Nucleoporin”, which don’t match official function list. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 45**

-Original Glimmer call @bp 29206 has strength 8.51

Start 29,206, Stop 29,457, Length 252 bp.

**It is a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: one significant results on NCBI.

**Start codon:** Start contains all coding potential. Moving upstream would cause a 100+bp overlap with upstream gene, moving downstream would exlcude ~40 bp of coding potential. 4 pair overlap with upstream gene.

**Function:** Hypothetical Protein HHPred: no significant results above 90%. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 46**

- Original Glimmer call @bp 29454 has strength 9.98

Start 29,454, Stop 29,732, Length 279 bp.

**It is a gene:** GeneMarkS shows high coding potential. Pham has 24 members. Blastp: many significant results on NCBI.

**Start codon:** Has best RBS score. 4 bp overlap with upstream gene. Contains all coding potential.

**Function:** Hypothetical Protein. HHpred: no significant result. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 47**

- Original Glimmer call @bp 29729 has strength 7.24

Start 29,729, Stop 30,061, Length 333 bp.

**It is a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: many significant results on NCBI.

**Start codon:** First start in ORF, only one to include all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: no results above 90%. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 48**

- Original Glimmer call @bp 30058 has strength 5.81

Start 30,058, Stop 30,402 Length 345 bp.

**It is a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: many significant results on NCBI.

**Start codon:** First start in ORF, only one to include all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: no results above 90%. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 49**

- Original Glimmer call @bp 30399 has strength 7.99

Start 30,399, Stop 30,593, Length 195 bp.

**It's a gene:** Genemark shows high coding potential. Pham has 90 members. Blastp: many significant results.

**Start codon:**  Start is first in ORF and contains all coding potential, gives a 4 bp overlap with upstream gene. Has best RBS score.

**Function:** Hypothetical Protein. HHPred: no significant results above 90%. Blastp: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 50**

- Original Glimmer call @bp 30590 has strength 5.05

Start 30,590, Stop 31,009, Length 420 bp.

**It's a gene:** Genemark shows high coding potential. Pham has 2 members. Blastp: no significant results on NCBI.

**Start codon:**  Start is first in the ORF and the only that contains all coding potential, gives a 4 bp overlap with upstream gene. RBS score is the best.

**Function:** Hypothetical Protein HHpred: no significant results above 90%. Blastp: one significant result on phagesdb: function unknown. Deep TMHMM: not a membrane protein.

**Gene 51**

- Original Glimmer call @bp 31309 has strength 3.77

Start 31,309, Stop 31,055, Length 255 bp.

**It is a gene:** GeneMarkS shows high coding potential. Pham has 15 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:**  This start has 10 of 10 MAs in the pham. This start contains all but a small region of coding potential, but expanding to the next start reduces RBS score and it is a TTG codon.

**Function:** helix-turn-helix DNA binding domain. HHpred: many significant results for “HTH-type transcriptional regulator” and peptidate. Blastp: helix-turn-helix and metallo-protease. All non-draft phages in the pham call it HTH. HTH confirmed by identifying the presence of the domain in the HHPred results.

**Gene 52**

-Original Glimmer call @bp 31814 has strength 1.43

Start 31,904, Stop 31,446, Length 459 bp.

**It is a gene:** GeneMarkS shows high coding potential. Pham has 16 members. Blastp: many significant results.

**Start codon:** The Glimmer-called start doesn’t include all coding potential. However, start 31904 does includes all coding potential and has 10 of 11 MAs in the pham (the start is found in 94% of pham members). RBS score has no significant change comparable to the start called by Glimmer. Changing to 31904 also reduces the gap to the downstream gene.

**Function:** helix-turn-helix DNA binding domain. HHpred: many significant results for helix-turn-helix and many other hits. HTH confirmed by identifying the presence of the domain in the HHPred results. Blastp: NCBI and phagesdb indicate helix-turn-helix. All non-draft phages in the pham call it HTH.

**Gene 53**

-Original Glimmer call @bp 32217 has strength 6.02; GeneMark calls start at 32244

Start 32,217, Stop 32,432, Length 216 bp.

**It is a gene:** GeneMarkS shows high coding potential. Pham has 17 members. Blastp: many significant results.

**Start codon:** The GeneMark start has a much worse RBS and would exclude a small amount of coding potential. The Glimmer start is the only one to include all coding potential, and when present, starterator says it’s called 86% of the time. The most annotated start for this pham is not present in this gene.

**Function:** helix-turn-helix DNA binding domain. HHPred: many results above 90% calling for helix-turn-helix DNA binding protein, among other things. Blastp: Majority of both NCBI and phagesdb indicate helix-turn-helix DNA binding protein, others saying function unknown. Members of pham that have called a function all say helix-turn-helix DNA binding domain. HTH confirmed by identifying the presence of the domain in the HHPred results.

**Gene 54**

- Original Glimmer call @bp 32437 has strength 4.06

Start 32,437, Stop 32,661, Length 225.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI show many significant results. Gene is in a pham with 17 members.

**Start Codon:** Starterator report says that 80% of the genes in this pham use this start and 2 out of the 3 non-draft were manually annotated. This start had the best RBS score and it has a 4 bp overlap w/ upstream gene.

**Function:** helix-turn-helix DNA binding domain. HHPred: many results above 90% calling for helix-turn-helix DNA binding protein, among other things. Blastp: Majority of both NCBI and phagesdb indicate helix-turn-helix DNA binding protein. HTH confirmed by identifying the presence of the domain in the HHPred results.

**Gene 55**

- Original Glimmer call @bp 32743 has strength 8.17

Start 32,743, Stop 32,979, Length 237 bp.

**It is a gene:** GeneMark indicates high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 18 other members.

**Start codon:** Start contains all coding potential and has best RBS score by far. This start is found in 100% of pham members and has 13 of 13 MAs.

**Function:** Hypothetical Protein. HHpred: no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 56**

-Original Glimmer call @bp 32976 has strength 3.11

Start 32,976, Stop 33,476, Length 501 bp.

**It’s a gene:** GeneMarkS shows high coding potential. Pham has 81 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Has the best RBS score for a start that includes all coding potential. 4 bp overlap w/ upstream gene.

**Function:** HNH endonuclease. HHPred: many significant results for endonuclease, including HNH endonuclease. BlastP: many significant results HNH endonuclease. Nearly all pham members are HNH endonuclease. Structure on HHPred meets this condition from official function list: “Must have H-N-H over a 30 aa span”

**Gene 57**

- Original Glimmer call @bp 33473 has strength 10.34

Start 33,473, Stop 33,751, Length 279 bp.

**It’s a gene:** Genemark shows high coding potential. Pham has 27 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Has the best RBS score, includes all coding potential, 4 bp overlap with upstream gene.

**Function:** helix-turn-helix DNA binding domain. HHpred: significant results showing HTH domain, HTH verified by looking at structure. Blastp: Both NCBI and phagedb show many results for helix-turn-helix DNA binding protein, helix-turn-helix DNA binding domain protein, or hypothetical protein.

**Gene 58**

- Original GeneMark call @bp 33748

Start 33,748, Stop 33,906, Length 159 bp.

**It is a gene:** GeneMarkS indicates high coding potential. Not called by Glimmer. BlastP shows one significant result on NCBI. Gene is in a pham with 15 members.

**Start Codon:** Starterator report says that 100% of the genes in this pham have this start and has 9 of 9 MAs. This start has the best RBS score. 4 bp overlap with upstream gene.

**Function:** membrane protein. HHpred: showed no significant results over 90%. BlastP: phagesdb shows mostly function unknown, but 2 hits for oxidoreductase. Deep TMHMM: one membrane domain, therefore it is a membrane protein.

**Gene 59**

-Original Glimmer call @bp 33903 has strength 4.99

Start 33,903, Stop 34,112, Length 210.

**It is a gene:** GeneMark indicates high coding potential. BlastP on NCBI and Phagesdb show significant results. This gene is in a pham with 7 members.

**Start Codon:** First start in ORF, only start to include all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: only one significant results over 90%, for “zinc ribbon”, but it had low coverage. BlastP showed many significant results for hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 60**

-Original Glimmer call @bp 34109 has strength 7.73

Start 34,109, Stop 34,261, Length 153 bp.

**It is a gene:** GenemarkS indicates high coding potential. BlastP: many significant results. This gene is in a pham with 19 members.

**Start Codon:** Has best RBS score. 4 bp overlap with upstream gene. It’s the only start that wouldn’t drastically shorten the gene or create a huge overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 61**

-Original Glimmer call @bp 34258 has strength 7.62

Start 34,258, Stop 34,401, Length 144 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results. Gene is in a pham with 30 members.

**Start Codon:** 4 bp overlap with upstream gene. Contains all coding potential. It’s the only start that wouldn’t drastically shorten the gene or create a big overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 62**

- Original Glimmer call @bp 34398 has strength 9.83

Start 34,398, Stop 34,523, Length 126 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results. Gene is in a pham with 16 members.

**Start Codon:** This is the first start in the ORF and the only that contains all coding potential. Gives a 4bp overlap with upstream gene. Has best RBS score.

**Function:** Hypothetical Protein. HHpred: no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 63**

- Original Glimmer call @bp 34520 has strength 6.83.

Start 34,520, Stop 34,666, Length 147 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant result. Gene is in a pham with 18 members.

**Start Codon:** First in ORF, only one to contain all coding potential, gives a 4 bp overlap with upstream gene. The only other start is downstream and would lessen the size of the gene by >50%.

**Function:** Hypothetical Protein. HHpred: no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 64**

- Original Glimmer call @bp 34663 has strength 4.57

Start 34,663, Stop 34,857, Length 195 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. Gene is in a pham with 27 members.

**Start Codon:** First start in the ORF, 4 bp overlap with upstream gene, it’s the only start that contains all coding potential, best RBS score.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 65**

- Original Glimmer call @bp 34854 has strength 5.01

Start 34,854, Stop 35,078, Length 225 bp.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. Gene is in a pham with 22 members.

**Start Codon:** First start in the ORF, 4 bp overlap with upstream gene, it’s the only start that contains all coding potential, best RBS score.

**Function:** helix-turn-helix DNA binding domain. HHpred: many significant results for transcriptional regulator. Presence of HTH verified by looking at structure. BlastP: several results for both helix-turn-helix DNA binding domain and hypothetical protein.

**Gene 66**

- Original Glimmer call @bp 35071 has strength 2.50

Start 35,071, Stop 35,568, Length 498 bp.

**It is a gene:** GeneMarkS shows high coding potential, many significant BLAST results, in a pham with 18 members.

**Start Codon:** First start in ORF, best RBS score, only one to include all coding potential, 8 bp overlap with upstream gene, removing downstream one start would shorten gene by 80 bp and exclude coding potential.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 67**

- Original Glimmer call @bp 35576 has strength 1.10 \*\* not called by GeneMark

Start 35,561, Stop 35,716, Length 156 bp.

**It is a gene:** Not called be GeneMark but has many significant BlastP results and is in a pham with 13 members.

**Start Codon:** There are only two starts in this ORF. The one not called by Glimmer is upstream at 35,561, improves the RBS score, and produces an 8 bp overlap with upstream gene. Starterator: the start at 35,561 is both the most common in the pham and has the most MAs. Thus, 35,561 is chosen.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 68**

- Original Glimmer call @bp 35713 has strength 4.96

Start 35,713, Stop 35,883, Length 171 bp.

**It is gene:** GeneMarkS indicates high coding potential. BlastP shows significant results on NCBI and Phagesdb. Gene is in a pham with 7 members.

**Start Codon:** Only 2 starts in this ORF. This is the first and the only one to contain all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 69**

- Original Glimmer call @bp 35883 has strength 3.70 \*\* not called by GeneMark

Start: 35,883, Stop: 36,047, length 165 bp.

**This is a gene: Not called by** GenemarkS, but many significant BLASTp results and in a Pham with 18 members.

**Start Codon:** 1 bp overlap with upstream gene. This start is found in 100% of pham and has 10 of 13 MAs.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 70**

- Original Glimmer call @bp 36044 has strength 9.92

Start: 36,044, Stop: 36,862, length: 819 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 46 members.

**Start Codon:** contains all coding potential, 4 bp overlap with the upstream gene. Starterator: 91% of phages in the pham have this start, with 33 of 35 MAs.

**Function:** exonuclease. HHPred: many significant results for exonuclease. BlastP: many significant results for exonuclease and Cas4 family exonuclease.

**Gene 71**

- Original Glimmer call @bp 36862 has strength 2.49

Start: 36,862, Stop 37,323, length 462 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: many significant results. Pham has 2 members.

**Start Codon:** First in the ORF, only one to include all coding potential, best RBS score, 1 bp overlap with the upstream gene.

**Function**: Hypothetical Protein. HHPred: Multiple results for different functions including Tryptophan RNA-binding attenuator protein, Antitermination protein, and Inhibitor of TRAP, none of which are on the official function list. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 72**

- Original Glimmer call @bp 37310 has strength 8.57

Start: 37,310, Stop: 37,858, length 549 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many highly significant results on phagesdb and NCBI. Pham has 48 members.

**Start Codon:** First in the ORF and only one to contain coding potential, has the best RBS score, 14 bp overlap with upstream gene. This start has the most MAs in the pham, 75% of the pham have this start. Moving it down to the next start would shorten gene by ~70 bp and exclude coding potential.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 73**

- Original Glimmer call @bp 37855 has strength 9.29

Start: 37,855, Stop: 38,034, length: 180 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: several significant results. Pham has 7 members.

**Start Codon:** Start is first in ORF, only one that contains all coding potential, has the best RBS score, and 4 bp overlap with upstream gene. There are only 2 starts in the ORF, the other would cut the gene size in half.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 74**

- Original Glimmer call @bp 38031 has strength 14.74

Start: 38,031, Stop: 38,585, length 555 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many highly significant results. Pham has 523 members.

**Start Codon:** includes all coding potential, has the best RBS score, 4 bp overlap with upstream gene.

**Function:** SSB protein. HHPred: many highly significant results (over 99%) for single strand DNA binding protein. BlastP: every single non-draft phage calls it ssDNA binding protein.

**Gene 75**

- Original Glimmer call @bp 38696 has strength 11.15

Start: 38,696, Stop: 39,013length: 318 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 275 members.

**Start Codon:** First in ORF, only one to contain all coding potential, has the second-best RBS score. Starterator: does not have the most annotated start; start is has is called 82.6% of time when present and is the only start in the ORF with any MAs (13).

**Function:** glutaredoxin. HHPred: many highly significant results (over 99%) for glutaredoxin, NrdH-like glutaredoxin, and thioredoxin. BlastP: many significant results for these results, also. Nearly all in the pham use one of these three. I chose glutaredoxin because it’s more specific than thioredoxin, but less specific than NrdH – and I don’t think there is enough evidence to be that specific.

**Gene 76**

- Original Glimmer call @bp 39010 has strength 2.83

Start 39,010, Stop 39,243, Length 234 bp.

**It is a gene:** GeneMarkS indicates high coding potential. Blastp: several significant results. Pham has 21 members.

**Start Codon:** The start contains all coding potential and provides a 4 bp overlap with the upstream gene. The only other start is upstream and would overlap the upstream gene almost entirely.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 77**

- Original Glimmer call @bp 39240 has strength 13.09

Start 39,240, Stop 41,798, Length 2559 bp.

**It’s a gene:** GeneMarkS shows high coding potential. Pham has 78 members. Blastp: many significant results.

**Start codon:** First in the ORF and the only one to contain all coding potential. 4 base pair overlap with upstream gene. This gene does not have the most annotated start in the pham, but this start is called 100% of the time when present (found in 22% of the pham).

**Function:** DNA methyltransferase. HHPred: many highly significant results (97%>), it appears the best results are for methyltransferase which includes a 100% result. There are some results also for helicase, and some in the pham report the function as both helicase and methyltransferase, but it appears that the helicase hits are for proteins the modify chromatin (such as SNF2), which reinforces the function being methyltransferase. Blastp: Highly significant results for DNA helicase/methylase/DNA binding protein.

**Gene 78**

-Original Glimmer call @bp 41870 has strength 6.87; GeneMark calls start at 41795

Start 41,795, Stop 42,016, Length 222 bp.

**It’s a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: no significant results.

**Start codon:** The GeneMark start at 41,795 was chosen because it is the only start to contain all coding potential. Also, it has the best RBS and produces a 4 bp overlap with upstream gene.

**Function:** DNA binding protein. HHpred showed two results over 90%: “Repressor, Protein-DNA complex, Type IV secretion system, DNA BINDING PROTEIN” and “Pyocin activator protein.” BlastP: no results. Deep TMHMM: not a membrane protein.

**Gene 79**

- Original Glimmer call @bp 42147 has strength 3.19

Start 42,147, Stop 42,713, Length 567 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 69 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is the first in the ORF and the only one to contain all coding potential. RBS score is the best. Moving to the next downstream start would shorten the gene in half.

**Function:** HNH endonuclease. HHpred: significant results for endonuclease. Review of its structure shows HNH. Also, phamerator reports a conserved HNH domain. Blastp: HNH endonuclease. All genes in the pham that have a stated function call it as HNH.

**Gene 80**

- Original Glimmer call @bp 42781 has strength 5.24

Start 42,781, Stop 43146, Length 366 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 413 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Excluding a TTG start that would produce a 35 bp overlap, the 42,781 start chosen here is the first in the ORF, contains all coding potential, and has best RBS score. This gene doesn’t have the most annotated start in the pham, but that start chosen in called 100% of the time when present (with 14 MAs).

**Function**: helix-turn-helix DNA binding protein. HHPred: multiple significant hits for various functions – but HTH structure verified. BlastP: Many significant results for helix-turn-helix DNA binding protein. All non-draft members of this large pham also call it this.

**Gene 81**

- Original Glimmer call @bp 43167 has strength 10.53

Start 43,167, Stop 44,249, Length 1083 bp.

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 80 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is first in the ORF, the only one to contain all coding potential, and has the best RBS score. It is the most annotated start in the pham.

**Function:** DNA polymerase III sliding clamp (Beta). HHpred: many significant results above 90% for DNA polymerase III sliding clamp (Beta). Blastp: DNA polymerase III sliding clamp (Beta). Nearly all members of pham call it this.

**Gene 82**

- Original Glimmer call @bp 44246 has strength 9.22

Start 44,246, Stop 44,701, Length 456 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 12 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. 4 base pair overlap with upstream gene. Is the most annotated start in the pham, called 100% of time when present.

**Function:** Hypothetical Protein. HHpred showed two significant results over 90%: oxidoreductase and oxidase. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein. The official gene function list seems to suggest that oxidoreductase should be listed as hypothetical protein.

**Gene 83**

- Original Glimmer call @bp 44698 has strength 12.24

Start 44,698, Stop 46,269, Length 1572 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 133 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is first in the ORF and the only one to contain all coding potential. 4 base pair overlap with upstream gene. Has best RBS score. Is the most annotated start in the pham, called 98% of time when present.

**Function:** DNA methyltransferase. HHpred: 100% probability of methyltransferase. Blastp: methyltransferase.

**Gene 84**

- Original Glimmer call @bp 46287 has strength 9.85

Start 46,287, Stop 46,721, Length 435 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 40 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential, has best RBS score, this start is called 100% when present in the pham (with 11 MAs).

**Function:** Hypothetical Protein. HHpred: a few significant results above 90% for “replication initiator” or equivalent terms, but the only match to the official function list is “RepA-like replication initiator.” Blastp: hypothetical protein, replication initiation protein, DNA binding protein. Most pham members don’t have a function listed. Deep TMHMM: negative result, soluble protein. Considering the ambiguity, I called it a hypothetical protein.

**Gene 85**

-Original Glimmer call @bp 46988 has strength 2.18

Start 46,952, Stop 47,386, Length 435 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 59 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** The 46,988 start called by Glimmer excludes a small region of low coding potential, it is the only one like it in the pham, and has a much worse RBS score than the upstream start at 46,952. I chose the 46,952 start because it includes all coding potential, helps reduce the large gap between the upstream gene and it has 2 MAs, both from phages in the same cluster (AY).

**Function:** HNH endonuclease. HHpred: significant results for endonuclease. Review of its structure shows HNH. Also, phamerator reports a conserved HNH domain. Blastp: HNH endonuclease. All genes in the pham that have a stated function call it as HNH.

**Gene 86**

-Original Glimmer call @bp 47383 has strength 9.25

Start 47,383, Stop 47,754, Length 372 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Pham has 22 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is first in the ORF, the only one to contain all coding potential, has the best RBS score, and has a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 87**

- Original Glimmer call @bp 47751 has strength 12.08

Start 47,751, Stop 47,984, Length 234 bp.

**It’s a gene:** GeneMarkS indicates high coding potential. Orpham. Blastp: no significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential and gives a 4 bp overlap with upstream gene. Doesn’t have the best RBS score. All other starts would either exclude coding potential or create a large overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 88**  
- Original Glimmer call @bp 47981 has strength 6.38  
Start 47,981, Stop 48,337, Length 357 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 10 members.  
**Start Codon**: This start is the first in the ORF, the only one to include all coding potential, and it has a 4 bp overlap with upstream gene.   
**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 89**  
- Original Glimmer call @bp 48334 has strength 7.30  
Start 48,334, Stop 48,513, Length 180 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows several significant results on Phagesdb and NCBI. Gene is in a pham with 5 members.  
**Start Codon**: It contains all coding potential, has the best RBS score, and gives a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed only significant result over 90%, for “unknown function”. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.  
**Gene 90**  
-Original Glimmer call @bp 48510 has strength 7.31  
Start 48,510, Stop 48,791, Length 282 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows a few significant results on Phagesdb and NCBI. Gene is in a pham with 6 members.  
**Start Codon**: Contains all regions of high coding potential, produces 4 bp overlap with upstream gene, has best RBS score. The other two available starts would create large overlaps with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 91**

- Original Glimmer call @bp 48788 has strength 11.34

Start: 48,788, Stop: 49,189, length 402 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 164 members.

**Start Codon:** Start is the first in the ORF, the only one to include all coding potential, and produces a 4 bp overlap with upstream gene.

**Function:** MazG-like nucleotide pyrophosphohydrolase. HHpred: many significant results for MazG-like nucleotide pyrophosphohydrolase above 90%. Blastp: many results in both NCBI and phagedb for this function. Many members of pham indicate this function. “Nucleoside Triphosphate Pyrophosphohydrolase (EC 3.6.1.8) MazG-like domain” identified as conserved domain on phamerator.

**Gene 92**

- Original Glimmer call @bp 49189 has strength 6.39

Start: 49,189, Stop: 49,614, length: 426 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: many significant results on phagesdb & NCBI. Pham has 243 members.

**Start Codon:** First start in the ORF, only one to contain all coding potential, one of the best RBS scores, and produces a 1 bp overlap with upstream gene. This gene does not have the most annotated start in the pham. This start is called 94% when present and has 11 MAs.

**Function:** RusA-like resolvase. HHpred: Two significant results above 90% for RusA-like endonuclease. Blastp: many results in both NCBI and phagedb for this function. Many members of pham indicate this function. “Endodeoxyribonuclease RusA” identified as a conserved domain on phamerator.

**Gene 93**

- Original Glimmer call @bp 49611 has strength 8.08

Start: 49,611, Stop 50,132, length 522 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many highly significant results on phagesdb and NCBI. Orpham.

**Start Codon:** First start in the ORF, only one to contain all coding potential, has the best RBS score, and produces a 1 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: Only 2 significant results (above 90%), both for “PGDYG protein”, which is functionally uncharacterized. BlastP: Two most significant results are for “DnaQ-like DNA polymerase III subunit”, other results are for hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 94**

- Original Glimmer call @bp 50129 has strength 8.94

Start: 50,129, Stop 50,350, length 222 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many highly significant results on phagesdb and NCBI. In a pham with 20 members.

**Start Codon:** Has a bad RBS score but contains all coding potential and produces a 4bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 95**  
- Original Glimmer call @bp 50347 has strength 7.76  
Start 50,347, Stop 50574, Length 228 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. In a pham with 59 members.  
**Start Codon**: First in the ORF, only one to contain all regions of high coding potential, produces 4 bp overlap with upstream gene, has best RBS score.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 96**  
-Original Glimmer call @bp 50577 has strength 8.21; GeneMark calls start at 50571  
Start 50,571, Stop 50,777, Length 207 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows a few significant results on Phagesdb and NCBI. Gene is in a pham with 3 members.  
**Start Codon**: The 50,571 start called by GeneMark has a slightly better RBS score, but more importantly it includes more coding potential and produces a 4 bp overlap with upstream gene. The only MA in this small pham is for the 50,571 start. For these reasons, it was chosen.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 97**  
-Original Glimmer call @bp 50774 has strength 9.06  
Start 50,774, Stop 51,388, Length 615 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows many significant results on Phagesdb and NCBI. Gene is in a pham with 25 members.  
**Start Codon**: First in the ORF (except for a TTG start), contains all regions of coding potential, produces 4 bp overlap with upstream gene, and has best RBS score.

**Function**: helix-turn-helix DNA binding protein. HHPred: multiple significant hits for HTH DNA binding domain, which was verified by looking at the structure. BlastP: Many significant results for helix-turn-helix DNA binding protein and hypothetical protein. All called functions in the pham are for this.

**Gene 98**

* tRNA
* Aragorn 1.2.41: Sequence [51536,51610], tRNA-?(Lys|Gln)(tt), 75 bases, %GC = 62.7
  + Product= tRNA-OTHER
  + Note: tRNA-?(Lys|Gln)(tt)
* tRNAscan-SE v2.0: start 51536, end 51609, infernal score: 42.7. Isotype: Lys. Anticodon: NNN

**Gene 99**  
- Original Glimmer call @bp 51995 has strength 9.59

Start 51,995, Stop 52,150, Length 156 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. Gene is in a pham with 20 members.  
**Start Codon**: First in the ORF, only one to contains all regions of coding potential, has best RBS score. The only other start would produce a gene only 15 bp long.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP: hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 100**  
- Original Glimmer call @bp 52150 has strength 4.84  
Start 52,150, Stop 52,461, Length 312 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. In a pham with 74 members.  
**Start Codon**: Contains all regions of high coding potential, produces 1 bp overlap with upstream gene, the other starts would either create huge overlap with upstream gene or exclude large regions of coding potential.

**Function:** HNH endonuclease. HHpred showed significant results over 90% for HNH endonuclease, and confirmed by looking at structure. BlastP: HNH endonuclease.

**Gene 101**  
- Original Glimmer call @bp 52549 has strength 3.36; GeneMark calls start at 52537  
Start 52,549, Stop 53,265, Length 717 bp.  
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows a few significant results on Phagesdb and NCBI. Gene is in a pham with 5 members.  
**Start Codon**: The start called by Glimmer at 52,549 excludes a small region of coding potential, but it’s RBS score is -2.8 vs -6.6 of the GeneMark-called start at 52,537. The starterator report is not useful because of the small pham (only 3 manually annotated starts, and the start used in those 3 MAs isn’t present in this ORF). Because of the gap with the upstream ORF, this ORF is likely not part of an operon and thus RBS score is important. Thus, I picked the Glimmer start because it has a much better RBS.

**Function:** HNH endonuclease. HHpred showed significant results over 90% for HNH endonuclease, and confirmed by looking at structure. BlastP: HNH endonuclease. Phamerator also shows conserved HNH domains.