**Anekin phage. Oregon Coast Community College. Matthew R Fisher, Leo Chipman, Maggie Browne, Echo Deeg-Davis.**

**Gene 1**

-Original Glimmer call @bp 200 has strength 1.83; 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. Gene is in a pham w/ 25 members.

**Start Codon**: Start codon 200 doesn’t include all coding potential. Start codons 200 and 131 are not called by any other gene in the pham. However, the start at 56 bp is the most manually annotated start in the pham, has the best RBS score, and includes all regions of coding potential.

**Function**: terminase, small subunit. HHPred: 2 results ~95% probability of it being phage terminase. Blastp: significant results from NCBI said minor tail protein/terminase small subunit, phagesdb said terminase. All called functions in this pham say its terminase, small subunit.

**Gene 2**

-Original Glimmer call @bp 597 has strength 8.39

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 1329 members.

**Start codon**: Start is first in the ORF and only one to contain all full coding potential, 11 bp overlap with gene 1. Moving downstream to the next start codon would exclude a large region of high coding potential. This gene does not have the most commonly annotated start in the pham.

**Function**: terminase, large subunit. HHPred: Many 100% probability results for terminase large subunit. Blastp: many significant results for terminase, large subunit on NCBI and phagesdb.

**Gene 3**

- Original Glimmer call @bp 2346 has strength 11.26

Start 2346, stop 3680, Length 1335 bp.

**It is a gene**: GeneMarkS indicates high coding potential throughout ORF. Blastp: Many significant results on phagedb & NCBI. In a pham with 1606 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 2346 is the second-most upstream start. The one before it, at 2334, includes a small amount of extra coding potential but it’s RBS score is -7.4 compared to -3.3 for the 2346 start. Thus, the latter is chosen; Also, of the starts present in this gene, this one has the most MAs.

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

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

Original Glimmer call @bp 4312 has strength 3.05 \*\* not called by GeneMark

DELETE/REMOVE

**It is not gene**: GeneMarkS indicates very poor coding potential. No BLASTp results on NCBI, only 1 significant result on phagedb (Phrank15\_draft, in the same cluster as this phage). It’s problematic that this is a reverse gene tightly tucked in-between 2 forward genes, and there is a 4bp overlap between the end of this gene and the end of gene 3. Also, there is much stronger coding potential in a forward reading frame within this region (see gene 4).

**Gene 4**

- Original Glimmer call @bp 4307 has strength 5.46; GeneMark calls start at 3683

Start 3683, Stop 4900, Length 1218 bp.

**It is a gene**: GeneMarkS indicates high coding potential. Blastp: Many significant results with cluster members having e values of 0 on phagedb and NCBI.

**Start codon**: 4307 start excludes a large region of high coding potential. Using the GeneMark start at 3683 would put this in pham 81256 and start 3683 has the most MA’s. Choosing this start also improves the top BLASTp result from 990 to 1752 and changes the alignment from 1:209 to 1:1.

**Function**: Hypothetical Protein. HHPred: Only two results >90%: Mu\_F protein and minor capsid, although there is very poor coverage (70 aa out of 406). 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. HHPred results don’t support BLAST results, thus function is unknown. Deep TMHMM says not a membrane protein.

**Gene 5**

-Original Glimmer call @bp 5066 has strength 7.85

Start 5066, Stop 5599, 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 is an available upstream start but RBS is much worse and would still leave a 130 bp gap. The start chosen has the most MAs in the pham.

**Function**: scaffolding protein. Blastp: many significant results for scaffolding protein. HHPred: The probability is 98.75% that the function is a scaffolding protein.

 **Gene 6**

- Original Glimmer call @bp 5629 has strength 11.95

Start 5629, Stop 6024, 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: 99.22% probability for 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 6040 has strength 15.51

Start 6040, stop 7080, 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 7092 has strength 15.06

Start 7092, Stop 7391, Length 300 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/ 17 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 and there are no obvious matches to official functions. Most Blastp results are for hypothetical protein/no known function. Deep TMHMM says not a membrane protein.

**Gene 9**

- Original Glimmer call @bp 7408 has strength 4.79

Start 7408, Stop 7764, 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**: Start contains full coding potential, is the most upstream start, and has the highest 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 and database for head-to-tail adaptor.

**Gene 10**

- Original Glimmer call @bp 7761 has strength 8.01

Start 7761, Stop 8,159, 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 the upstream gene. The start is the first in the ORF and the only to contain all coding potential.

**Function**: head-to-tail stopper. HHPred: shows 1 result of 98% probability of being head to tail stopper by the SPP1 structure. 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 8156 has strength 5.81

Start 8,156, Stop 8,539, 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**: 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 8536 has strength 4.42

Start 8,536, Stop 8,937, Length 402 bp.

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

**Start codon**: There is a 4 base pair overlap with the upstream gene. The start contains all coding potential.

**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 8984 has strength 9.77

Start 8,984, Stop 9,169, 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. SOSUI: reports soluble protein. Deep TMHMM says not a membrane protein.

**Gene 14**

-Original Glimmer call @bp 9162 has strength 11.66

Start 9,162, Stop 9,671, Length 510 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. There is an 8 base pair overlap with the upstream gene.

**Function**: major tail protein. HHPred: showed 1 significant result above 90% “tail tube protein” 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 calls @ bp 9771 and has strength 9.95

Start 9771, Stop 10307, 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. It is the first possible start in the ORF. Starterator report: Has most annotated start.

**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 calls @ bp at 10394 and has a strength of 10.01

Start 9771, Stop 10651, Length 882 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 (Phrank15, Isolde, and Persistence).

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

**Gene 17**

-Original Glimmer calls @ bp at 10843 and has a strength of 7.03

Start 10843, Stop 10652, Length, 192

**It is a gene**; GenemarkS indicated high coding potential. BlastP shows a few significant results on Phagesdb and NCBI. In a pham of 2.

**Start Codon**: Starterator report says 100% of genes in the pham use this start point, but there is only one other member in the pham. RBS score is higher than most others, but not the best. This start leaves an appropriate gap before switching back to a forward gene (the next downstream gene). It contains all coding potential.

**Function**: membrane protein. HHpred: showed no significant results over 90%. BlastP shows many matches with significant results as an unknown function protein. SOSUI reports one membrane domain, and Deep TMHMM reports two, therefore it is a membrane protein.

**Gene 18**

-Original Glimmer calls @ bp at 11046 and has a strength of 7.1; GeneMark calls start at 10911

Start 10911, Stop 14747, length: 3837

**It is a gene**: GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI showed many significant results. In a pham of 27.

**Start Codon**: Glimmer and GeneMark do not agree on start, using the Glimmer start point would create a gap and exclude coding potential, but using the GeneMark start is the only start that includes all coding potential and it minimizes the gap with the upstream gene. Other members in this pham use the GeneMark start and have the same length for this gene. Starterator report says 88% of genes in the pham use this start point, 16/19 of them were manual annotation.

**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 calls @ bp at 14747 and has a strength of 7.01

Start: 14747, Stop: 15574, Length: 828

**It is a gene**: GeneMark indicates high coding potential. BalstP on Phagesdb and NCBI showed many significant results. In a pham of 41.

**Start Codon**: Has the best RBS score, Starterator report says there 40 members in this pham and 97.5% use this start and 30/31 were manually annotated. This is the first start in ORF and includes all coding potential. 1 bp overlap with upstream gene.

**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 calls start @bp 15586 and has a strength of 7.87

Start 15586, Stop 16887, 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 members.

**Start codon:** Starterator report says that 69% of genes in this pham have this start. It is called 100% of the time when present. The RBS score for this start is the best. The start is the first in the ORF.

**Function:** minor tail protein. HHpred: Many significant results over 99% that call for this function and BlastP matches the function with many significant results.

**Gene 21**

-Original Glimmer calls start @bp 16896 and has a strength of 4.61.

Start 16896, Stop 17858, Length 963 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 46 members.

**Start Codon:** Starterator report says that 100% of genes in this pham have this start and is called 82.2% of the time when present. 28/35 of the non-draft members manually selected this start. The RBS score on this start was the best score by far. Moving it upstream would create a large overlap with the upstream gene.

**Function:** minor tail protein. HHpred: showed no significant results. Many significant results with BlastP on NCBI and Phagesdb.

**Gene 22**

-Original Glimmer calls start @bp 17858 and has a strength of 7.36.

Start 17858, Stop 18514, Length 657 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 members.

**Start Codon:** Starterator report says that 16 of the non-draft members manually selected this start and 100% of the genes in this pham use this start. Best RBS score by far and first start in the ORF, with a 1 bp overlap with upstream gene.

**Function:** minor tail protein. HHpred: showed no significant results over 90%. BlastP on NCBI and Phagesdb showed many significant results near zero for minor tail protein. Other members of this cluster 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 calls start @bp 18525 and has a strength of 8.47.

Start 18525, Stop 18851, Length 327 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 9 members.

**Start Codon:** Starterator report says this start has 6 of 6 MAs. Best RBS score and it’s the first start in ORF.

**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. SOSUI reports one membrane domain, and Deep TMHMM reports one, therefore it is a membrane protein.

**Gene 24**

Original Glimmer calls start @bp 18826 and has a strength of 3.33.

Start 18826, Stop 19059, Length 234.

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

**Start Codon:** Starterator report says this start has 6 of 6 MAs. Best RBS score and there is a 26 bp overlap with upstream gene. Moving down to next start would create gap, exclude coding potential, and reduce size of gene by ~100 bp.

**Function:** Hypothetical Protein. HHpred: no significant result over 90%. BlastP on NCBI and Phagesdb showed many significant results for unknown function protein. SOSUI reports that the gene is a soluble protein and Deep TMHHM reports 0, therefore it is a Hypothetical Protein.

**Gene 25**

Original Glimmer calls start @bp 19056 and has a strength of 9.52.

Start 19056, Stop 19367, Length 312.

**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 9 members.

**Start Codon:** Starterator report says that 100% of genes in the pham use this start and 5/5 of the non-draft members were manually selected. Best RBS score by far and first start in ORF and there is a 4 bp overlap with upstream gene.

**Function:** membrane protein. HHpred: 2 significant results for unknown function. BlastP shows many matches with significant results as an unknown function protein. SOSUI reports 3 membrane domain, and Deep TMHMM reports 3, therefore it is a membrane protein.

**Gene 26**

-Original Glimmer call @bp 19364 has strength 9.43

Start 19,364, Stop 20,353, Length 990

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

**Start Codon:** Start includes all coding potential. 4 base pair overlap with upstream gene. First start in ORF.

**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 20364 has strength 7.08

Start 20,364, Stop 20,804, Length 441

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

**Start Codon:** First start in ORF. Start includes all coding potential, moving downstream would leave out coding potential. RBS score the best.

**Function:** membrane protein. HHpred: no significant results above 90%. Both database and NCBI report unknown function. All members of pham report unknown function except for a select few that say holin. SOSUI and Deep TMHMM indicate membrane protein by three domains. Blast results also indicate many results for membrane protein.

**Gene 28**

- Original Glimmer call @bp 20866 has strength 10.57

Start 20,866, Stop 21,156, Length 291

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

**Start Codon:** No other start codon in the gene, start codon contains all coding potential.

**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. SOSUI and Deep TMHMM report soluble protein.

**Gene 29**

-Original Glimmer call @bp 21166 has strength 5.26

Start 21,166, Stop 21,600, Length 435

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

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

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

**Gene 30**

-Original Glimmer call @bp 21604 has strength 1.09

Start 21,604, Stop 21,816, Length 213

**It is a gene:** GeneMarkS shows high coding potential. Pham has 17 members. Blastp: Many significant results on phagedb & NCBI.

**Start Codon:** Start codon contains all coding potential. First start in ORF. Moving down would lose coding potential. RBS score is the best.

**Function:** Hypothetical Protein HHpred: no significant results above 90%. Blastp: both NCBI and phagedb indicate hypothetical protein or unknown function. All members of pham report protein as unknown. SOSUI and Deep TMHMM indicate soluble protein.

**Gene 31**

-Original Glimmer call @bp 21817 has strength 12.33

Start 21,817, Stop 23,250, Length 1,434

**It is a gene:** GeneMarkS shows high coding potential. Pham has 4 members. Blastp: many significant results on phagedb and NCBI.

**Start Codon:** Start codon contains all coding potential. Moving down would lose coding potential. First possible start in ORF.

**Function:** hydrolase. HHpred: indicated a result of 99% of being hydrolase. Hit value [3HP4\_A](http://www.rcsb.org/pdb/explore/explore.do?structureId=3HP4) Blastp: NCBI and phagedb indicates many different types of proteins including hydrolase. 2 of 4 pham members report hydrolase.

**Gene 32**

-Original Glimmer call @bp 23894 has strength 11.02

Start 23894, Stop 23247, Length: 648 bp.

**It is a gene**: GeneMarkS indicates high coding potential Blastp: several significant on phagesdb, NCBI shows many significant results, despite being an Orpham.

**Start Codon**: Start contains full coding potential but has the worst RBS among starts. Moving start downstream would result in significant large overlap with downstream gene, whereas moving upstream would lose coding potential.

**Function**: esterase. HHPred: numerous >99% match for esterase. Blastp: many results for esterase and hydrolase family.

*Note: the end of this gene and the end of forward, upstream gene overlap by 4 bp.*

**Gene 33**

-Original Glimmer call @bp 24046 has strength 1.95 \*\* not called by GeneMark.

Start: 24046 Stop: 23918, length 129 bp.

**This is a gene**: DNAMaster indicates Genemark is not calling this gene, however GenemarkS report indicates it does. GenemarkS shows low coding potential. Gene is an Orpham. No blast results on phagesdb or NCBI. When investigated, gene 33 could not be lengthened in place of gene 34 without losing coding potential. Gap of 200 or more bp before downstream gene.

**Start Codon:** Genemark and Glimmer presumably agree on start based on GeneMarkS report. Start contains full coding potential and the best RBS score. This is the first start in the ORF.

**Function**: membrane protein. HHPred shows no significant results. All results under 90%. BlastP: no results. DeepTMHMM: membrane protein.

**Gene 34**

-Original Glimmer call @bp 24275 has strength 12.14

Start: 24,275, Stop: 25195, length 921 bp.

**This is a gene**: GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 194 members.

**Start Codon**: Start contains full coding potential, best RBS score, this start is called 83% of the time when present, 48 MA’s. The gap before this gene leaves room for promoter for reverse gene upstream of it.

**Function**: tyrosine integrase. HHPred: multiple 100% probability outputs for tyrosine recombinase and one 100% for tyrosine integrase. Considering this is a temperate phage, tyrosine integrase was chosen instead of tyrosine homologous recombinase, per the instructions in the official function list. BlastP: tyrosine integrase reported on both phagesdb and NCBI with many significant results.

**Gene 35**

- Original Glimmer call @bp 25319 has strength 4.03; GeneMark calls start at 25205

Start 25,205, Stop: 25,558, length 354 bp.

**This is a gene**: GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Orpham. There is an 11 base pair overlap with downstream, reverse gene (their ends overlap).

**Start Codon**: Genemark and Glimmer do not agree on start. GenemarkS calls start at 25205, reducing the gap between genes and including all regions of high coding potential and this therefore chosen despite having a bad RBS score. The start chosen by Glimmer has a much better RBS score but excludes lots of coding potential.

**Function**: Hypothetical Protein. HHPred shows no significant results for known functions. BlastP: hypothetical protein. SOSUI and Deep TMHMM: not a membrane protein.

**Gene 36**

- Original Glimmer call @bp 25783 has strength 3.12; GeneMark calls start at 25759

Start 25,783, Stop 25,547, length 237 bp.

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

**Start Codon:** The start chosen by Glimmer has 115 MA’s, while the start called by Genemark only has 1. Start called by Genemark does have a better RBS score, however it shortens the gene and would exclude some coding potential. I chose the Glimmer start.

**Function:** helix-turn-helix DNA binding protein. HHPred shows one significant result with a 90.02% probability. BlastP: Many significant results for helix-turn-helix DNA binding protein.

**Gene 37**

- Original Glimmer call @bp 27011 has strength 11.86; GeneMark calls start at 26993

Start: 26,993, Stop: 25,815, length 1179 bp.

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

**Start Codon:** The start chosen by Genemark contains all coding potential, has a slightly better RBS score, and has 248 MA’s. This does shorten the gene by 19 bp, but the Glimmer start only has 2 Mas.

**Function**: tyrosine integrase. HHPred: multiple 100% probabilities for tyrosine integrase, integrase, and recombinase. BlastP: Many significant results for tyrosine integrase.

**Gene 38**

- Original Glimmer call @bp 27545 has strength 5.23

Start: 27,545, Stop: 27,171, length: 375 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 36 members.

**Start Codon:** The start contains all coding potential indicated by Genemark and has the best RBS score. Start has 10 MA’s and is called 100% when present.

**Function:** Hypothetical Protein. HHPred shows no significant results. BlastP: shows many significant results for hypothetical protein. SOSUI & Deep TMHMM: soluble.

*Note: 8 bp overlap with downstream reverse gene.*

**Gene 39**

- Original Glimmer call @bp 27852 has strength 5.38

Start: 27,918, Stop: 27,538, length: 381 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results. Pham has 36 members, same pham as upstream gene.

**Start Codon:** Genemark and Glimmer agree on start, but this excludes a small region of high coding potential. The start at 27918 has almost 10x better RBS and includes all coding potential. None of the starts present in this gene have been MA’ed. It appears to be a duplicate of the upstream gene.

Function: Hypothetical Protein. HHPred shows no significant results. BlastP: many significant results for hypothetical protein. SOSUI and Deep TMHMM: soluble.

**Gene 40**

- Original Glimmer call @bp 28238 has strength 4.45

Start: 28,238, Stop 28,065, length 174 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: A few significant results. Pham has 6 members.

**Start Codon:** Starterator report: only 3 non-draft phages in this pham. This phage doesn’t have the most annotated start. This start has best RBS and includes all coding potential.

**Function:** Hypothetical Protein. HHPred: no significant results. BlastP: some significant results for hypothetical protein. SOSUI and Deep TMHMM: soluble.

**Gene 41**

- Original Glimmer call @bp 28806 has strength 5.66; GeneMark calls start at 28728

Start: 28,728, Stop: 29,228, length 501 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: No significant results. Gene is an Orpham.

**Start Codon:** Glimmer and Genemark do not agree on start. Start called by Genemark reduces large gap between upstream gene, has better RBS and includes all coding potential, which 28806 start does not.

**Function:** Hypothetical Protein. HHPred:no significant results. BlastP: no significant results. SOSUI and Deep TMHMM: soluble.

**Gene 42**

- Original Glimmer call @bp 29225 has strength 8.95

Start: 29,225, Stop: 29,605, length: 381 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 24 members.

**Start Codon:** Start is the first in the ORF and the only to include all possible coding potential. 4 bp overlap with upstream gene and has 2 MAs.

**Function:** Hypothetical Protein. HHPred: no significant results. BlastP: many significant results for hypothetical protein. SOSUI and Deep TMHMM: soluble.

**Gene 43**

- Original Glimmer call @bp 29602 has strength 9.76

Start 29602, Stop 29814, Length 213 bp.

**It is a gene:** GeneMarkS shows high coding potential. Blastp: limited but significant results on phagesdb and NCBI. Pham: 4 members.

**Start Codon:** Calls most annotated start. Start cannot be changed to an upstream start without a large overlap occurring. It contains all coding potential, highest RBS score, and has a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: no significant results. Blastp: Highly significant results for hypothetical protein. SOSUI and Deep TMHMM: soluble.

**Gene 44**

- Original Glimmer call @bp 29811 has strength 7.15

Start: 29811, Stop:30191, Length 381 bp.

**It is a gene:** GeneMarkS shows high coding potential. Blastp: many highly significant results on phagesdb & NCBI. Pham has 5 members.

**Start Codon:** Start called is first available in ORF and has the most MAs. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: no significant results. Blastp: hypothetical protein. SOSUI: Is a soluble protein. SOSUI and Deep TMHMM: soluble.

**Gene 45**

- Original Glimmer call @bp 30188 has strength 7.19

Start 30188, Stop 30352, Length 165 bp.

**It is a gene:** GeneMarkS indicates moderate coding potential. Blastp: several significant values. Pham has 20 members.

**Start Codon:** Does not have the most annotated start in the pham. Start called has one MA, includes all coding potential, and the highest RBS score. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred: No significant results (90% probability or higher). Blastp: hypothetical protein. SOSUI and Deep TMHMM: Is a soluble protein.

**Gene 46**

-Original Glimmer call @bp 30345 has strength 4.47

Start 30,345, Stop 30,494, Length 150

**It is a gene:** GeneMarkS shows high coding potential. Orpham. Blastp: no significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. Moving upstream would cause a huge overlap with upstream gene. 8 bp overlap with upstream gene.

**Function:** Hypothetical Protein HHPred: no significant results above 90%. Blastp: no significant results on NCBI and phagesdb. SOSUI and Deep TMHMM: indicates soluble protein.

**Gene 47 (added gene)**

Not called by Glimmer but predicted by GenemarkS

Start 30,801, Stop 30,547, Length 255 bp.

**It is a gene:** GeneMarkS shows high coding potential. Many BLASTp results on phagesdb and NCBI.

**Start codon:** Several 1:1 alignments in BLASTp results. The start contains all regions of high coding potential that don’t overlap substantially with another gene. RBS score is approximately an average score for this ORF.

**Function:** helix-turn-helix DNA binding domain. HHpred: many significant results for helix-turn-helix, confirmed by identifying the presence of the domain in the HHPred results. BLASTp: helix-turn-helix DNA binding domain protein

**Gene 48**

-Original Glimmer call @bp 31306 has strength 3.54

Start 31,396, Stop 30,938, Length 459

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

**Start codon:** Start 31396 includes more coding potential and is the most manually annotated start in the pham. RBS score has no significant change comparable to the originally called start. Changing to 31396 also reduces the gap to the downstream gene.

**Function:** helix-turn-helix DNA binding domain. HHpred: many significant results for helix-turn-helix, confirmed by identifying the presence of the domain in the HHPred results. Blastp: NCBI and phagesdb indicate either function unknown or helix-turn-helix.

**Gene 49**

-Original Glimmer call @bp 31709 has strength 5.76; GeneMark calls start at 31736

Start 31709, Stop 31930, Length 222

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

**Start codon:** Starterator: start 31709 is the most called start when present in the pham and most manually annotated. Has the best RBS score compared to other starts. Start 31709 includes all coding potential and reduces the gap between genes.

**Function:** helix-turn-helix DNA binding domain. HHPred: many results above 90% calling for helix-turn-helix DNA binding protein. Other top results call for repressor. 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.

**Gene 50**

-Original Glimmer call @bp 31927 has strength 8.87

Start 31927, Stop 32388, Length 462

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

**Start codon:** First start in ORF, only start that contains all coding potential, and there's a 4 pair overlap with upstream gene.

**Function:** HNH endonuclease. HHpred: many significant results above 90% for HNH endonuclease. Blastp: Both NCBI and phagedb show many results for HNH endonuclease. All called functions in the pham are for HNH endonuclease.

**Gene 51**

-Original Glimmer call @bp 32394 has strength 8.55

Start 32394, Stop 32618, Length 225

**It's a gene:** Genemark shows high coding potential. Pham has 17 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is first in ORF and only one that contains all coding potential, moving to the downstream start would lose coding potential.

**Function:** helix-turn-helix DNA binding domain. HHPred: many significant results above 90% saying helix-turn-helix DNA binding protein, HTH confirmed by looking at alignment results. Blastp: Both NCBI and phagedb show many results for helix-turn-helix DNA binding protein or helix-turn-helix DNA binding domain protein. Majority of called function in pham are for helix-turn-helix DNA binding protein or helix-turn-helix DNA binding domain protein.

**Gene 52**

-Original Glimmer call @bp 32704 has strength 7.56

Start 32704, Stop 32967, Length 264

**It's a gene:** GenemarkS shows high coding potential. Pham has 18 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start is first in the ORF and the only that contains all coding potential, moving downstream would lose coding potential. RBS score shows start closest to zero.

**Function:** Hypothetical Protein. HHpred: no significant results above 90%. Blastp: no significant results on NCBI and phagesdb. SOSUI and Deep TMHMM: indicates soluble protein.

**Gene 53**

-Original Glimmer call @bp 33036 has strength 6.50; GeneMark calls start at 33051

Start 32964, Stop 33272, Length 309

**It's a gene:** Genemark shows high coding potential. Blastp: No results on NCBI and very few results on phagedb. Orpham.

**Start codon:** No starterator data. Neither of the called starts include all coding potential and Start 33036 is a TTG start, which is rare. The first start in the ORF (32964) has the best RBS and produces a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: no significant results above 90%. Blastp: no significant results on NCBI (no results) and phagesdb. SOSUI: indicates soluble protein, Deep TMHMM: negative result, soluble protein.

**Gene 54**

-Original Glimmer call @bp 33284 has strength 10.26; GeneMark calls start at 33269

Start 33269, Stop 33547, Length 279

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

**Start codon:** Starterator: Start 33269 is the most annotated start in the pham with 13 manual annotations. Also has the best RBS score. Start 33269 includes more coding potential and reduces the gap with upstream gene, creating a 4 pair overlap.

**Function:** helix-turn-helix DNA binding domain. HHpred: significant results showing HTH domain. Blastp: Both NCBI and phagedb show many results for helix-turn-helix DNA binding protein, helix-turn-helix DNA binding domain protein, or just DNA binding protein. Pham members are either just DNA binding protein or HTH.

**Gene 55**

-Original Glimmer calls start @bp 33544 and has a strength of 7.61.

Start 33544, Stop 33702, Length 159.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. Gene is in a pham with 15 members.

**Start Codon:** Starterator report says that 100% of the genes in this pham use this start and 8/8 non-draft members manually selected this start. This start had the best RBS score and moving upstream would cause a ~ 30 bp overlap.

**Function:** membrane protein. HHpred: showed no significant results over 90%. BlastP shows some matches with significant results as an unknown function protein. SOSUI reports one membrane domain, and Deep TMHMM reports one, therefore it is a membrane protein.

**Gene 56**

-Original Glimmer calls start @bp 33699 and has a strength of 4.86.

Start 33699, Stop 33926, Length 228.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 5 members.

**Start Codon:** Starterator report says that 80% of the genes in this pham use this start. This start had the best RBS score and it has a 4 bp overlap w/ upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed significant results near zero. SOSUI reports soluble protein and no results on Deep TMHMM.

**Gene 57**

-Original Glimmer call @bp 33919 has strength 8.49

Start 33919, Stop 34074, Length 156.

**It is a gene:** GeneMark indicates high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 19 members.

**Start codon:** Starterator report says that 22.2% of the genes in this pham use this start and half of these were manually annotated to this start. Moving the start point would create ~30 bp overlap with the gene before this. This start had the best RBS score and it creates a 8 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed many significant results close to zero. SOSUI and Deep TMHMM show it is a soluble protein and not a membrane protein.

**Gene 58**

-Original Glimmer calls start @bp 34074 and has a strength of 8.44.

Start 34074, Stop 34217, Length 144.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 30 members.

**Start Codon:** Starterator report says that 62.1% of genes in this pham use this start and over half of the non-draft genes were manually annotated for this start. This start has the best RBS score and 1 bp overlap w/ upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed many significant results near zero. SOSUI and Deep TMHMM show it is a soluble protein and not a membrane protein.

**Gene 59**

-Original Glimmer calls start @bp 34214 and has a strength of 5.73.

Start 34214, Stop 34369, Length 156.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 15 members.

**Start Codon**: This start has the best RBS score and 4 bp overlap with upstream gene. First start in the ORF.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed many significant results near zero. SOSUI and Deep TMHMM show it is a soluble protein and not a membrane protein.

**Gene 60**

-Original Glimmer calls start @bp 34455 and GeneMark calls start @bp 34356 and has a strength of 5.69.

Start 34356, Stop 34568, Length 213.

**It is a gene:** GeneMarkS indicates med-high coding potential. BlastP on Phagesdb and NCBI show significant results. Gene is in a pham with 2 members.

**Start Codon:** Starterator report says that 50% of genes in this pham that are not drafts use this start and 1/1 are manually annotated. Glimmer’s called start point would exclude coding potential and would create a large gap between this gene and the gene before it. 34356 start uses all coding potential and has a better RBS score, has a 14 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed many significant results near zero for hypothetical protein. SOSUI and Deep TMHMM show it is a soluble protein and not a membrane protein.

**Gene 61**

-Original Glimmer calls start @bp 34565 and has a strength of 10.90.

Start 34565, Stop 34714, Length 150.

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

**Start Codon:** Starterator report says that 100% of the genes in this pham use this start codon and 15 of them are non-draft genes. This start has the best RBS score. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed many significant results near zero for hypothetical protein. SOSUI and Deep TMHMM show it is a soluble protein and not a membrane protein.

**Gene 62**

Original Glimmer calls start @bp 34711 and has a strength of 6.41.

Start 34711, Stop 34905, Length 195

**It is a gene:** GenemarkS indicates high coding potential. BlastP on NCBI and Phagesdb show just 1 significant result. This gene is in a pham with 2 members.

**Start Codon:** Starterator report says that 100% of the members of this pham use this start. This start has a good RBS score. First start in ORF. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred: showed no significant results over 90%. BlastP showed 1 significant match for unknown function. SOSUI and TMHMM show it is a soluble protein and not a membrane protein.

**Gene 63**

-Original Glimmer calls start @bp 34902 and has a strength of 4.81.

Start 34902, Stop 35039, Length 138.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP only shows no significant results on NCBI and Phagesdb. Gene is in a pham with 2 members.

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

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed no significant results. DeepTMHMM showed no results for membrane protein.

**Gene 64**

-Original Glimmer calls start @bp 35036 and has a strength of 10.93.

Start 35036, Stop 35398, Length 363.

**It is a gene:** GeneMarkS indicates high coding potential. BlastP shows significant results on Phagesdb and NCBI. Gene is in a pham with 5 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.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed few significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 65**

-Original Glimmer calls start @ bp 35395 and has a strength of 8.33.

Start 35395, Stop 35589, Length 195.

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

**Start Codon:** Starterator report says that 66.7% of the members of the pham have this start and 11/18 of them are manually annotated. This start has the best RBS score by far and it contains all coding potential and gives a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed many significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 66**

-Original Glimmer calls start @bp 35586 and has a strength of 11.85.

Start 35586, Stop 35810, Length 225.

**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:** Starterator report says that 65.2% of the genes in this pham have this start and 9/16 are manually annotated. This start had the best RBS score by far and it contains all coding potential and gives a 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed many significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 67**

-Original Glimmer calls start @bp 35803 and has a strength of 3.06.

Start 35803, Stop 36291, Length 489.

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

**Start Codon:** Starterator report says that 23.5% of the genes in this pham have this start and 3/11 are manually annotated. The most annotated start is not present in this gene. This start had the best RBS score by far and moving the start would exclude high coding potential and create a large gap between this gene and the one before it. It is the first start in the ORF and produces a 8 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed many significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 68**

-Original Glimmer calls start @bp 36299 and has a strength of 1.21. GeneMark does not call it.

Start 36284, Stop 36436, Length 153.

**It is a gene:** BlastP shows many significant results despite not being called by GeneMark. It is called by Glimmer and removing the gene would create a large gap. Other members in the cluster call this a gene. Gene is in a pham with 13 members.

**Start Codon:** Starterator report says that 46.2 % of the members of this pham have the start and only 1 manual annotation has been used for it. The most commonly annotated start (36284) is present in this gene and 61.5% of the genes in the pham use that start. Moving the start codon upstream to the first start codon available gives the best RBS score and makes an 8 bp overlap with the gene before it, which is why I chose it.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed many significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 69**

-Original Glimmer calls start @bp 36436 and has a strength of 8.24.

Start 36436, Stop 36609, Length 174.

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

**Start Codon:** Starterator report says that all the genes in this pham have this start and 5 are manually annotated. This start had the best RBS score by far and it’s the first in the ORF and has a 1 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed many significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 70**

-Original Glimmer calls start @bp 36609 and has a strength of 6.97.

Start 36609, Stop 36758, Length 150.

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

**Start Codon:** Starterator report says that all the genes in this pham have this start and 11 of 15 non-drafts are manually annotated. This start has a one bp overlap with the gene before it and has the best RBS score.

**Function:** Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed some significant results near zero for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 71**

- Original Glimmer call @bp 36816 has strength 11.87

Start: 36,816, Stop: 37,665, length: 840 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: Many significant results on phagesdb & NCBI. Pham has 284 members.

**Start Codon:** This gene does not have the most annotated start in the gene. The called start has 6 MAs, is chosen 90% of time when present (found in 10 pham members). This start is the only one this gene has with any MAs. Alternatively, moving the start upstream would result in an RBS score 100x worse and would only include a small amount of additional low coding potential

**Function**: Cas4 exonuclease. HHPred: 99.2% probably and wide coverage match to 4r5q crystal structure of Cas4 family, per the official function list guidelines. BLAST: multiple matches to various types of exonucleases.

**Gene 72**

- Original Glimmer call @bp 37646 has strength 7.72

Start: 37,646, Stop: 38,200, length: 555 bp.

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

**Start Codon:** This start contains all coding potential. Selecting upstream start would results in more than a 50 bp overlap with upstream gene. There is already a 10 bp overlap. Moving downstream would exclude >100 bp of coding potential.

**Function:** Hypothetical Protein. HHPred shows no significant results. BlastP: many significant results for hypothetical protein. DeepTMHMM: not a membrane protein.

 **Gene 73**

- Original Glimmer call @bp 38197 has strength 15.16

Start: 38,197, Stop 39,408, length 1212 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: A few significant results. Pham has 34 members.

**Start Codon:** GenemarkS indicates full inclusion of coding potential, moving the start upstream would result in a much worse RBS score, and result in more than a 50 bp overlap. Start has 6 MA’s. 4 bp overlap.

**Function**: Hypothetical Protein. HHPred: Multiple results for a number of different functions including RecF-like, hydrolase, AAA-ATPase, DNA binding protein, etc. Spoke to Karen Klyczek (SMART team) and she indicated we can’t specify a function currently and to call it hypothetical.

**Gene 74**

- Original Glimmer call @bp 39408 has strength 7.77

Start: 39,408, Stop: 39,929, length 522 bp.

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

**Start Codon:** Called start is first possible, includes all coding potential and has the better RBS score. 1 bp overlap with upstream gene. Anekin doesn’t have the most annotated start. This start is found in 21 pham members and is called 100% of time.

**Function:** SSB protein. HHPred: many highly significant results (over 99%) for single strand DNA binding protein. BlastP: highly significant, every single published phage result reads for ssDNA binding protein.

 **Gene 75**

- Original Glimmer call @bp 40019 has strength 9.92

Start:40,019, Stop:40,330, length: 312 bp.

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

**Start Codon:** Start is first in ORF, contains all coding potential, and has the highest RBS score. This start is found in 23 pham members and is called 83% of time. Anekin doesn’t have the most annotated start.

**Function:** glutaredoxin. HHPred: many highly significant results (over 99%) for glutaredoxin. BlastP: many significant results for glutaredoxin. Nearly all in the large pham call it this way, or NrdH-like glutardoxin.

**Gene 76 (added)**

-Not called by Glimmer or GeneMark

Start 40330, Stop 40560, Length 231

**It is a gene**: GeneMark shows moderate coding potential. Fills a ~200 bp. Has a 1 bp overlap with upstream gene and 4 bp overlap with downstream gene. Many significant BLASTP results.

**Start Codon**: Only start in the ORF! Has a 1 bp overlap with upstream gene.

**Function**: Hypothetical Protein. Hhpred: No significant results. BLASTP results all say hypothetical protein. Deep TMHMM: not a membrane protein.

**Gene 77**

- Original Glimmer call @bp 40557 has strength 14.94

Start 40,557, Stop 43,085, Length 2529 bp.

**It is a gene:** GeneMarkS shows high coding potential throughout the ORF. Blastp: many significant results on phagesdb and NCBI. Pham has 78 members.

**Start Codon:** Called start contains all coding potential, highest RBS score, and is the first possible start in ORF. 14 MAs.

**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 43082 has strength 2.42

Start: 43,082, Stop:43,483, Length 402 bp.

**It is a gene:** GeneMarkS shows high coding potential. Blastp: many highly significant results on phagesdb & NCBI. Pham has 413 members. 4 bp overlaps with upstream and downstream genes.

**Start Codon:** Start called is not the first available, however changing the start codon to the farthest upstream would result in a more than 200 bp overlap with the upstream gene. Start called contains all coding potential, has the best RBS score. This start is found in 19 pham members and is called 100% of the time. Anekin does not have the most annotated start.

**Function:** helix-turn-helix DNA binding protein. HHPred: many highly significant results for helix-turn-helix DNA binding protein (<90%). Blastp: many results for helix-turn-helix DNA binding protein/hypothetical protein.

**Gene 79**

- Original Glimmer call @bp 43480 has strength 14.27

Start 43,480, Stop 44,562, Length 1083 bp.

**It is a gene:** GeneMarkS indicates high coding potential. Blastp: several significant results on phagesdb and NCBI. Pham has 80 members. 4 bp overlaps with upstream and downstream genes.

**Start Codon:** Start called is not the first possible recorded start, however moving the start codon up to the first recorded start would result in a more than 200 bp overlap with half of the upstream gene’s ORF. Called start includes all coding potential, has the highest RBS score, 4 bp w/ upstream gene, and has the most annotated start with 29 MAs.

**Function:** DNA polymerase III sliding clamp (Beta). HHPred: multiple 100% probabilities for DNA polymerase III sliding clamp (beta). Blastp: many significant results for DNA polymerase III sliding clamp (beta) on phagesdb and NCBI.

**Gene 80**

-Original Glimmer call @bp 44559 has strength 9.99

Start 44559, Stop 45014, Length 456

**It’s a gene:** GeneMarkS shows 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.

**Function:** Hypothetical Protein HHpred: top results were for Heme-Oxygenase-like Diiron Oxidase and doesn’t apply to phages. No significant results above 90%. Blastp: no significant results on NCBI and phagesdb, all indicate Hypothetical Protein. All members of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 81**

-Original Glimmer call @bp 45011 has strength 12.63

Start 45011, Stop 46582, Length 1572

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

**Start codon:** Start contains all coding potential. 4 base pair overlap with upstream gene.

**Function:** DNA methyltransferase. HHpred: many results 100% probability for DNA methyltransferase. Blastp: many significant results on NCBI and phagesdb indicating DNA methyltransferase. Majority members of pham call DNA methyltransferase.

**Gene 82**

-Original Glimmer call @bp 46598 has strength 11.31

Start 46598, Stop 47032, Length 435

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

**Start codon:** Start contains all coding potential, first in the ORF. RBS score is the closest to zero. Moving to other start codons in the ORFS would eliminate coding potential and increase gap.

**Function:** Hypothetical Protein HHpred: Only 4 significant results, half indicate unknown function and the other indicate replisome or replication initiator, which aren’t on the official function list. Blastp: many results on phagesdb and NCBI indicate unknown function. Some on NCBI indicate replication initiation protein. Majority of pham indicate unknown function. Deep TMHMM: negative result, soluble protein.

**Gene 83**

-Original Glimmer call @bp 47266 has strength 7.03

Start 47266, Stop 47367, Length 372

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

**Start codon:** Start contains all coding potential, first in the ORF. RBS score is the closest to zero. Moving to other start codons in the ORFS would eliminate coding potential and increase gap.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagedb report only Hypothetical Protein. All members of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 84**

-Original Glimmer call @bp 47634 has strength 7.08

Start 47634, Stop 48191, Length 558

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 2 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. 4 base pair overlap with upstream gene.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagedb report majority Hypothetical Protein. Other member of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 85**

-Original Glimmer call @bp 48188 has strength 13.20

Start 48188, Stop 48622, Length 435

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 164 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. 4 base pair 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 86**

-Original Glimmer call @bp 48622 has strength 7.65

Start 48622, Stop 49053, Length 432

**It’s a gene:** GeneMarkS indicate high coding potential. Pham has 243 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. 1 base pair overlap with upstream gene.

**Function:** RusA-like resolvase. HHpred: many significant results above 90% for RusA-like resolvase. 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 87**

-Original Glimmer call @bp 49046 has strength 3.73

Start 49046, Stop 49711, Length 666

**It’s a gene:** GeneMarkS indicate high coding potential. Pham has 7 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential, first start in ORF, 8 base pair overlap with upstream gene.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagedb report majority being Hypothetical Protein. Other members of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 88**

-Original Glimmer call @bp 49708 has strength 12.14

Start 49708, Stop 49944, Length 237

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 20 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:** Start contains all coding potential. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagedb all report Hypothetical Protein. Other members of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 89**

-Original Glimmer call @bp 49928 has strength 9.17

Start 49928, Stop 50155, Length 228

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 59 members. Blastp: many significant results on NCBI and phagesdb.
**Start codon:** Start contains all coding potential and has best RBS. 16 bp overlap with upstream gene. Starterator: Indicates this start (49928) is the most manually annotated start with 35 MA.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagedb all report Hypothetical Protein. Other members of pham show Hypothetical Protein. Deeep TMHMM: negative result, soluble protein.

**Gene 90**

-Original Glimmer call @bp 50152 has strength 13.13

Start 50152, Stop 50415, Length 264

**It’s a gene:** GeneMarks indicate high coding potential. Pham has 26 members. Blastp: many significant results on NCBI and phagesdb.

**Start codon:**  Start contains all coding potential and is first in the ORF. 4 bp overlap with upstream gene.

**Function:** Hypothetical Protein HHpred: No significant results above 90%. Blastp: Both NCBI and phagesdb all report Hypothetical Protein. Other members of pham show Hypothetical Protein. Deep TMHMM: negative result, soluble protein.

**Gene 91**
-Original Glimmer calls start @ bp 50412 and has a strength of 7.07.
Start 50412, Stop 51011, Length 600.
**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**: This start has the best RBS score by far. It’s the first start in the ORF and has a 4 bp overlap with upstream gene.
**Function**: Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed some significant results, mostly for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 92**
-Original Glimmer calls start @ bp 51689 and has a strength of 12.09.
Start 51650, Stop 51826, Length 177.
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows some significant results on Phagesdb and NCBI. Gene is in a pham with 4 members.
**Start Codon**: Starterator report says that 25% of the members in this pham use the 51650 start, which was not originally called. Using 51650 is that only start that includes all regions of high coding potential, and it would help reduce the large gap between this gene and the one before it. The RBS score is worse though, but keeping the original start would exclude a substantial region of coding potential.

**Function**: Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed 2 significant results for hypothetical protein. DeepTMHMM showed no results for membrane protein.
**Gene 93**
Original Glimmer calls start @ bp 51837 and has a strength of 7.33.
Start 51837, Stop 52076, Length 240.
**It is a gene**: GeneMarkS indicates high coding potential. BlastP shows some significant results on Phagesdb and NCBI. Gene is in a pham with 11 members.
**Start Codon**: Starterator report says that 100% of the genes in this pham use this start and that 8 are manually annotated. This start had the best RBS score. Moving the start to the next one upstream would nearly overlap the entire upstream gene.

**Function**: Hypothetical Protein. HHpred showed no significant results over 90%. BlastP results showed some significant results for hypothetical protein. DeepTMHMM showed no results for membrane protein.

**Gene 94**

- Original Glimmer call @bp 52076 has strength 11.69

Start: 52,076, Stop: 52,237, length: 162 bp.

**This is a gene:** GenemarkS shows high coding potential. Blastp: A few highly significant results on phagesdb & NCBI. Pham has 20 members.

**Start Codon:** Start chosen is the only possible recorded start in the ORF. 1 bp overlap with upstream gene.

**Function:** Hypothetical Protein. HHPred shows no significant results (90% or above). BlastP: shows many significant results for hypothetical protein. DeepTMHMM: not a membrane protein.

 **Gene 95**

- Original Glimmer call @bp 52375 has strength 0.96 \*\* not called by GeneMark

Start: 52,237, Stop: 52,560, length: 324 bp.

**This is a gene:** GeneMarkS essentially zero coding potential, but it’s called by Glimmer and Blastp produces many significant results on phagesdb & NCBI. Pham has 3 members.

**Start Codon:** The first start in the ORF (52237) has the best RBS score, improves the blast results on NCBI and results in a one-to-one alignment, reduces the gap between upstream gene, and produces a 1 bp overlap. Anekin doesn’t have the most annotated start in this small pham.

**Function:** HNH endonuclease. HHPred shows many highly significant results (over 90%). BlastP: Many significant results for HNH endonuclease.

 **Gene 96**

- Original Glimmer call @bp 52648 has strength 5.61; GeneMark calls start at 53041

Start: 52,648, Stop 53,418, length 771 bp.

**This is a gene:** GenemarkS shows several disparate regions of high coding potential. Blastp: Many highly significant results on phagesdb and NCBI. Pham has 5 members.

**Start Codon:** Gimmer and Genemark do not agree on start, however consulting GenemarkS indicates that Start 52,648 is actually called by both Glimmer and GenemarkS. This start has the best RBS by far and has full inclusion of coding potential. Moving to the start upstream would result in a much worse RBS score and would include no additional coding potential. This pham only has drafts, so the starterator is not useful. The start at 53041 would reduce this gene in half and create a large gap.

**Function:** HNH endonuclease. HHPred: Many highly significant results (above 90%) for HNH endonuclease. BlastP: Many highly significant results for HNH endonuclease on phagesdb and NCBI.

----

**tRNA check:**

* nothing found on ARAGORN v1.2.41
* 1 found on tRNAscan-SE v. 2.0: begin 28458, end 28391, PRO. But the infernal score is 20.0 (less than the 35.0 needed for inclusion). Thus, no tRNA is included in the annotation.