HerculesXL phage. Oregon Coast Community College. Chyanna Blackburn, Matthew R Fisher.

**Gene 1**

-Original Glimmer call @bp 1 has strength 9.92

Start 1, Stop 528, Length 528 bp.

It is a gene: GeneMarkS indicates high coding potential; Blastp: E values 7e-62 or less phagedb & NCBI. Gene is in a pham w/ 143 members.

Starterator report: Calls most annotated start, start contains full coding potential and has a relatively good RBS score.

Function: Hypothetical Protein. HHPred: No significant related results. Blastp: many significant results for hypothetical protein on both blastdb and NCBI blast. SOSUI: Soluble protein

**Gene 2**

- Original Glimmer call @bp 525 has strength 14.75

Start 525, Stop 1949, Length 1425 bp.

It is a gene: GeneMarkS indicates high coding potential. Blastp: many e values of zero on phagedb and NCBI blast. In a pham with 681 members.

Starterator report: Does not have most annotated start but there is not another start codon it could call that would make it match with more genes, the start codon is found in 2.3% of genes in the pham. Glimmer and GeneMark agree on start. Start contains full coding potential. 4 bp overlap with gene 1.

Function: terminase. HHPred: Significant probability (90% or higher) for terminase (especially large subunit). Blastp: Terminase

**Gene 3**

- Original Glimmer call @bp 1952 has strength 14.69

Start 1952, Stop 2107, Length 156 bp.

It is a gene: GeneMarkS indicates high coding potential. Blastp: Multiple results with e values around e-14. In a pham with 54 members.

Starterator report: Does not have most annotated start, 29.6% of genes in pham have this start which is called 100% of the time that it is present, no other start could be chosen that would make it match more gene starts. GeneMark and Glimmer agree on start. Start codon contains full coding potential

Function: Hypothetical protein. Blastp: hypothetical protein. HHPred: no significant results (90% or greater probability). SOSUI: Soluble protein

**Gene 4**

-Original Glimmer call @bp 2111 has strength 14.89

Start: 2111, Stop: 3532. Length 1422 bp.

It is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results with an e value of nearly 0. In a pham with 146 members.

Starterator report: Does not have most annotated start, found in 1.4% of phage genes and called 100% of the time it is present, there is another start which is start 32 that is found in 89.7% of the pham members but is only called 0.8% of the time it is present, start 32 is at 2174 and would increase gap between gene 3 and gene 4. GeneMark and Glimmer agree on start. Called start contains full coding potential. RBS scores are the best.

Function: portal protein. HHPred: significant results (90% probability or higher) for portal protein. Blastp: many significant results for portal protein

**Gene 5**

- Original Glimmer call @bp 3532 has strength 6.65

Start 3532, Stop 3741, Length 210 bp.

It is a gene: GeneMarkS indicates high coding potential. Blastp: multiple significant results with e values practically at 0. In a pham with 16 members.

Starterator report: Does not have the most annotated start. Start found in 25% of phage genes. GeneMark and Glimmer agree on start. Called start contains full coding potential. 1 bp overlap with gene 4.

Function: Hypothetical Protein. HHPred: No significant results (90% probability or higher). Blastp: many significant results listed as hypothetical protein. SOSUI: Soluble protein.

**Gene 6**

- Original Glimmer call @bp 3738 has strength 13.34

Start: 3738, Stop: 4013, Length: 276 bp.

It is a gene: GeneMark S indicates high coding potential. Blastp: NCBI e values E-15 or lower. In a pham with 11 members.

Starterator report: called start is the most annotated started. Start contains all coding potential and gives a 4 bp overlap with upstream gene. GeneMark and Glimmer both call the start

Function: Hypothetical Protein. HHPred: No significant results (probability 90% or greater). Blastp: many results indicating a hypothetical protein. SOSUI: Soluble protein.

**Gene 7**

- Original Glimmer call @bp 4017 has strength 10.93

Start 4017, Stop 4871, Length 855.

It is a gene: GeneMark S indicates high coding potential. Blastp: many results with E values practically 0. In a pham with 137 members.

Starterator report: Calls most annotated start, 99.3% of genes have and call this start. Start contains full coding potential. GeneMark and Glimmer agree on start

Function: minor capsid protein. HHPred: MuF-like minor capsid protein or minor capsid protein. Blastp: MuF-like minor capsid protein. But “MuF-like” is not on official SEA PHAGES function list. SOSUI: Soluble protein

**Gene 8**

- Original Glimmer call @bp 4979 has strength 14.38

Start 4979, Stop 5533, Length 555 bp.

It is a gene: GeneMark S indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham with 146 members.

The start is the also possible start in this ORF.

Function: scaffolding protein. HHPred: unknown protein or protein binding. Blastp: Scaffolding protein.

**Gene 9**

- Original Glimmer call @bp 5573 has strength 14.23

Start 5573, Stop: 6568, Length 996 bp.

It is a gene. GeneMark S indicates high coding potential. Blastp: many significant results with E values practically 0. In a pham with 377 members.

Starterator report: Calls most annotated start, found in 93.9% of genes in pham and called 100% when present. GeneMark and Glimmer agree on start. Called start contains full coding potential. RBS score is the best for called start.

Function: major capsid protein. HHPred: Major capsid protein. Blastp: Major capsid protein.

**Gene 10**

- Original Glimmer call @bp 6638 has strength 9.37

Start 6638, Stop 7066, Length 429 bp.

It is a gene: GeneMark S indicates high coding potential. Blastp: many results greater with E values practically 0. In a pham with 146 members.

Starterator report: Calls start that is found in 17.1% of genes in pham and called 92.0% of the time that it is present. GeneMark and Glimmer agree on start. Called start contains full coding potential. RBS score is great at -1.931.

Function: hypothetical protein. HHPred: No significant results (90% probability or greater). Blastp: hypothetical protein. SOSUI: Is a membrane protein with 2 domains. But TMHMM appears to show no transmembrane structure.

**Gene 11**

- Original Glimmer call @bp 7080 has strength 16.29

Start 7038, Stop 7448, Length 411 bp.

It is a gene: GeneMark S indicates high coding potential. Blastp: many significant results with E values practically 0. In a pham with 146 members. greater than 3e-25 , NCBI also indicates many significant results.

Starterator report: original called start was at 7080 found in only 2 of 146 phages, it is never manually annotated, it does not include all coding potential. GeneMark and Glimmer agree on that 7080 start however. Start at 7038 has 34 MA’s and the RBS is more than 100x better than the original. This produces a 29 bp overlap. However, the only other phage is this subcluster has a 23 bp overlap. Changing start to 7038 improves alignment with other genes.

Function: hypothetical protein. HHPred: unknown. Blastp: hypothetical protein. SOSUI: Soluble.

**Gene 12**

- Original Glimmer call @bp 7445 has strength 12.31

Start 7445, Stop 7789, Length 345 bp.

This is a gene: Blastp: Many significant results with E values practically 0. GeneMark S indicates high coding potential. In a pham with 146 members.

Start codon contains full coding potential and gives a 4 bp overlap with upstream gene.

Function: minor capsid protein. HHPred: minor capsid protein. Blastp: Most significant is minor capsid protein, others are hypothetical protein. SOSUI: Soluble.

**Gene 13**

- Original Glimmer call @bp 7792 has strength 10.44; GeneMark calls start at 7786

Start 7786, Stop 8142, Length 357 bp.

It is a gene: GeneMarkS indicates high coding potential. Blastp: significant results with E values practically 0. In a pham with 146 members

GeneMark and Glimmer do not agree: Glimmer call 7792, GeneMark call 7786. Starterator report: Does not have most annotated start. Original start 7792 only found in HerculesXL. Start 7786 found in 17.1% of genes and is called 72.0% of time when present. 7786 gives 4 bp overlap w/ upstream gene. Changing to 7786 gives 1 to 1 alignment with other closely related genes.

Function: Tail terminator. HHPred: 98% probably match with [5A21\_G](http://www.rcsb.org/pdb/explore/explore.do?structureId=5A21), which on the SEA PHAGEs official function list is evidence of tail terminator. Blastp: many results for tail terminator. SOSUI: soluble.

**Gene 14**

- Original Glimmer call @bp 8142 has strength 13.72

Start 8142, Stop 8444, Length 303 bp.

It is a gene: GeneMark S indicates high coding potential. In a pham with 146 members. Blastp: Many significant results with E values practically 0.

GeneMark and Glimmer agree on start. Starterator report: Calls most annotated start, found in 96.6% of genes and called by 100% of genes when it is present and produces 1 bp overlap with upstream gene.

Function: hypothetical protein. HHPred: No significant results (90% probability or higher). Blastp: hypothetical protein. SOSUI: Soluble.

**Gene 15**

-Original Glimmer call @bp 8449 has strength 12.69

Start 8449, Stop 8949, Length 501 bp.

It is a gene: High coding potential according to GeneMarkS. Blastp: Many significant results w/ E values practically 0. In a pham with 146 members.

GeneMark and Glimmer agree on start. Starterator report: Calls most annotated start, found in 65.1% of pham members and called 100% of the time it is present. Start codon contains full coding potential. RBS score is most significant for this start.

Function: Major tail protein. HHPred: Major tail protein. Blastp: Major tail protein.

**Gene 16**

- Original Glimmer call @bp 8985 has strength 13.47

Start 8985, Stop 9557, Length 573 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: many significant related results with E values practically 0. In a pham with 172 phams.

GeneMark and Glimmer agree on start. Starterator report: Calls most annotated start, found in 83.7% of genes and called 100% of the time that it is present. Start codon contains full coding potential. RBS score is most significant for this start

Function: tail assembly chaperone. HHPred: no significant results 90% probability or greater. Blastp: tail assembly chaperone.

**Gene 17:**

- Original Glimmer call @bp 9572 has strength 4.21

Start 8985, Stop 9946, Length 963.

This is a gene: GeneMark S indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham with 103 members.

This is the longer of the two tail assembly chaperone proteins and the -1 frameshift has been noted here at 9518. This is based on the frameshift noted in the other phage in this subcluster (Ixel). Like in Ixel, gene 16 and 17 diverge after the sequence NSGG. The start and stop reported here give it 1:1 alignment with many other genes.

Function: tail assembly chaperone. HHPred: No significant results 90% probability or greater. Blastp: tail assembly chaperone.

**Gene 18**

- Original Glimmer call @bp 9965 has strength 14.22

Start 9965, Stop 12403, Length 2439 bp.

This is a gene: GeneMark S indicates very high coding potential. Blastp: A lot of significant results with e values of 0. Part of pham with 146 members.

GeneMark and Glimmer agree on start. Start contains full coding potential. Starterator report: Does not have most annotated start, has start found in 33.6% of genes and is called 100% of the time it is present. RBS score is the best for called start

Function: Tape measure protein. HHPred: tape measure protein. Blastp: tape measure protein

**Gene 19**

- Original Glimmer call @bp 12396 has strength 10.63

Start 12396, Stop 14645, length 2250 bp.

This is a gene: GeneMark S indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham with 17 members.

GeneMark and Glimmer agree on start. Starterator report: Calls most annotated start, found in 94.1% of genes and called in 100% of the time it is present. Start contains full coding potential. RBS score is most significant and is the first possible start. 8 bp overlap.

Function: minor tail protein. HHPred: Many significant results but I am unsure how to interpret them as they seem related to animal physiology. Blastp: minor tail protein

**Gene 20**

- Original Glimmer call @bp 14645 has strength 15.28

Start 14645, Stop 15814, Length 1170 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Multiple significant results with E values practically 0. In a pham with 18 members

GeneMark and Glimmer agree on start. Starterator report: Does not have most annotated start, has start found in 38.9% of genes and called 100% of time that it is present. Start codon contains full coding potential. 1 bp overlap with upstream gene.

Function: minor tail protein. HHPred: tail protein. Blastp: Minor tail protein

**Gene 21**

- Original Glimmer call @bp 15811 has strength 11.59

Start 15811, Stop 16578, Length 768 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Multiple significant results with E values practically 0. In a pham with 6 members.

GeneMark and Glimmer agree on start. Starterator report: Calls most annotated start, found and called in 100% of pham. Start codon contains full coding potential. 4 bp overlap with upstream gene.

Function: minor tail protein. HHPred: only significant result (90% or greater probability) for unknown function. Blastp: many results minor tail protein, some for tail fiber. This gene is in a region with other tail proteins. Synteny also suggests this to be a tail protein. SOSUI: soluble

**Gene 22**

- Original Glimmer call @bp 16593 has strength 14.39

Start 16593. Stop 16850, Length 258 bp.

This is a gene: GeneMark S indicates high coding potential. Blastp: Multiple significant with E scores practically 0. In a pham with 62 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Start found in 31.7% of genes and called 100% of the time when present. RBS score is most significant for this called start.

Function: hypothetical protein. HHPred: Top result almost to 90% probability show protein of unknown function. Blastp: Majority of significant results show hypothetical protein, but some for lysin A, and endolysin are other significant results. SOSUI: soluble.

**Gene 23**

- Original Glimmer call @bp 16850 has strength 12.85

Start 16850, Stop 17716, Length 867 bp.

This is a gene: GeneMark S indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham with 107 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starerator report: Does not have most annotated start. Has start found in 6 out of 107 genes and called 100% of the time it is present. 1 bp overlap with upstream gene.

Function: endolysin. HHPred: There is a very significant result for endolysin. Blastp: Endolysin or lysin A.

**Gene 24**

- Original Glimmer call @bp 17785 has strength 11.65

Start 17785, Stop 18318, Length 534 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp Many significant results with E values practically 0. In a pham w/ 72 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 54.2% and called 97.4% of time when it is present. RBS score is most significant

Function: membrane protein. HHPred: no significant results (90% probability or greater). Blastp: membrane protein. SOSUI & TMHMM: Membrane protein.

**Gene 25**

- Original Glimmer call @bp 18318 has strength 8.08

Start 18318, Stop 18671, Length 354 bp.

It is a gene: GeneMarkS indicates high coding potential. BlastP: Many significant results with E values practically 0. In a pham w/ 39 members.

GeneMark and Glimmer agree on start. Start codon contains full coding potential. Starterator report: Does not have most annotated start, only found and called in HerculesXL. 1 bp overlap with upstream gene.

Function: membrane protein. HHpred: No significant results. Blastp: hypothetical protein and membrane protein. SOSUI: has 1 domain on SOSUI for membrane protein, same with TMHMM.

**Gene 26**

- Original Glimmer call @bp 18674 has strength 12.73

Start 18674, Stop 19000, 327 bp length.

This is a gene: Coding potential: GeneMark indicated high coding potential. In a pham with 170 members. Blastp: Many significant results with E values practically 0.

GeneMark and Glimmer agree on start. Start codon contains full coding potential. Starterator: doesn’t have most annotated start, has start found in 29.4% of the pham members and called 88% of genes when present. RBS score: most significant

Function: membrane protein. HHpred: Only significant result for unknown function. Blastp: membrane protein. SOSUI: membrane protein with three domains.

**Gene 27**

- Original Glimmer call @bp 19304 has strength 8.02

Start 19304, Stop 19161, Length 144 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: no sign. Results. Orpham.

Glimmer and GeneMark agree on start. Start codon contains full coding potential. This start makes the largest gene. Best RBS score.

Function: hypothetical protein. HHpred: no significant probability. Blastp: n/a. SOSUI: Soluble.

**Gene 28**

- Original Glimmer call @bp 19517 has strength 15.99

Start 19517, Stop 19329, Length: 189 bp.

This is a gene: Coding potential: GeneMarkS indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham w/ 138 members.

GeneMark and Glimmer agree. Start contains full coding potential. Starterator report: Does not have most annotated start. Only found in HerculesXL. No other start would make it match phage genes in pham. RBS score is the best.

Function: hypothetical protein. HHPred: no significant probability. Blastp: hypothetical protein. SOSUI: Soluble.

**Gene 29**

- Original Glimmer call @bp 19680 has strength 14.57

Start 19680, Stop 19528, Length 153 bp

Coding Potential: GeneMark S indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham with 111 members.

This start is the only possible one in this ORF.

Function: hypothetical protein. HHPred: hypothetical but most significant result for d.58.4.0 (A:) automated matches {Burkholderia cepacia [TaxId: 292]}. Blastp: hypothetical protein. SOSUI: Soluble.

**Gene 30**

- Original Glimmer call @bp 19856 has strength 7.68

Start 19856, Stop 19695, Length 162 bp.

This is a gene: GeneMarkS indicates good coding potential. Blastp: several results with significantly low e values. In a pham with 104 members.

GeneMark and Glimmer agree on start. Starterator report: Does not have most annotated start. Found in 1.9% of genes and called 100% when present. Start codon contains nearly all coding potential, but extending out to the next start is a TTG and doesn’t improve RBS score. RBS score is the best for this start and there is a 1 bp overlap with gene 31.

Function: hypothetical protein. HHPred: No significant results. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 31**

- Original Glimmer call @bp 20713 has strength 12.32

Start 20713, Stop 19856, Length 876 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results with E values practically 0. In a pham wtih 136 members.

GeneMark and Glimmer call same start (20713, which has the most MA’s in the pham. BUT, this start doesn’t include all of the coding potential. Start 20731 does has include all coding potential and has a similar RBS score and it reduces the gap between this gene and #32. Starterator report: missing report

Function: hypothetical protein. HHPred: Top result for unknown function, also results for TRYPTOPHAN RNA-BINDING ATTENUATOR PROTEIN-INHIBITORY PROTEIN; TRANSCRIPTION REGULATION, ANTI-TRAP; HET: ZN; 2.8A {BACILL. BOF ; Bacterial OB fold (BOF) protein. Blastp: hypothetical protein. SOSUI: Soluble.

**Gene 32**

- Original Glimmer call @bp 22418 has strength 15.86; GeneMark calls start at 22349

Start 22418, Stop 20745, Length 1674 bp.

This is a gene: GeneMarkS report shows high coding potential. Blastp: A lot of results with e value of 0. In a pham w/ 145 members

GeneMark calls 22349. Glimmer calls 22418. Starterator report: Calls most annotated start found in 100% of phage genes and called 87.6% of the time it is present that start is at 22418 called by Glimmer. Start at 22349 has only 6 MA’s while start 22418 has 118 MA’s. RBS score is best for start 22418 and gives 1:1 alignment with many genes.

Function: RecA-like DNA recombinase. HHPred: Replicative helicase protein repA, RecA-type helicase, and circadian clock protein, Highest result on HHPred is repA helicase. Blastp: RecA-like DNA recombinase.

**Gene 33**

- Original Glimmer call @bp 22687 has strength 3.58

Start 22687, Stop 22394, Length 294 bp.

This is a gene: GeneMarkS indicates high coding potential. In a pham with 202 members. Blastp: A lot of results with e value of 0.

Glimmer and GeneMark agree on start. Starterator report: calls most annotated, found in 73.8% of genes in pham and called 98% of time it is present. Called start contains full coding potential. 1 bp overlap with gene 34.

Function: VRR-Nuc domain protein. HHpred: nuclease/ VRR-NUC/VRR-NUC nuclease. Blastp: nuclease, VRR-NUC domain protein. “nuclease” isn’t on the official SEA PHAGES function list.

**Gene 34**

- Original Glimmer call @bp 23550 has strength 23.27

Start 23550, Stop 22687, Length 864 bp.

This is a gene: GeneMarkS indicates high coding potential. In a pham w/ 198 members. Blastp: A lot of results with e value of 0.

Glimmer and GeneMark agree on start. Starterator report: Calls most annotated start, found in 73.2% of genes and called 97.9% of genes when present. Start codon contains full coding potential within ORF.

Function of gene: hypothetical protein. HHPred: highest result for unknown function, other functions list terminase functions but probability is lower though still significant. Blastp: Hypothetical protein. SOSUI: Soluble.

**Gene 35**

- Original Glimmer call @bp 24250 has strength 13.36

Start 24250, Stop 23582, Length 669 Bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A lot of results with e value of 0. In a pham with 146 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Calls start found in 20.5% of phage genes and called 100% of time when it is present in those genes. RBS score is not the most significant.

Function: AAA-ATPase. HHPred: Most significant result with 99.78 probability is AAA domain, another significant result with 99.14 probability is RecA protein. Blastp: AAA-ATPase (most significant results are for this function). Other significant results are for RecA-like DNA Recombinase.

**Gene 36**

- Original Glimmer call @bp 25428 has strength 12.96; GeneMark calls start at 25299

Start 25428, Stop 24247, Length 1182 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A lot of results with e value of 0. Pham: 198 members.

Glimmer calls start 25428 GeneMark calls 25299, the latter doesn’t contain all coding potential. Starterator report: Start 25428 is most annotated start. Found in 55.6% of genes and called 95.5% of the time when present.

Function: RecB-like exonuclease/helicase. HHPred: Many 99%+ probably results for helicase and nuclease. Blastp: Many results with 0 e values list function as either exonuclease or RecB-like exonuclease/helicase or Cas4 family exonuclease. From official function list If both a helicase and nuclease domain are present, the RecB label should be used. – this appears to be the case.

**Gene 37**

- Original Glimmer call @bp 27313 has strength 12.89

Start 27313, Stop 25415, Length 1899 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results with e values of 0. Pham: 198 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Calls start found in 20.7% of genes and called 100% of time when present. RBS score pretty good start codon at -2.664. 1:1 alignment with several other genes.

Function: DNA polymerase I. HHPred: 100% probability for DNA polymerase/DNA polymerase I. Blastp: DNA polymerase I.

**Gene 38**

- Original Glimmer call @bp 27338 has strength 1.60 \*\* not called by GeneMark

Start 27331, Stop 27483, Length 153 bp.

This gene wasn’t called by either GeneMark or Glimmer. GeneMarkS didn’t call a gene in this region. Glimmer called an orpham gene from 27338-27505 which represents just a tiny portion of a large ORF with multiple regions of high coding potential, which 27338-27505 includes just 1 of those regions. However, GenemarkS shows some moderate coding potential in a different frame, which is what I call here (27331 – 27843) – this gene occupies the entire ORF and therefore seems more deliberate, and blastp indicates some weak similarity to a coiled-coil-helix domain. The gene as called here is also an orpham, but it helps fill the rather large gap between gene 38 and 39. The RBS score of this gene is similar to the gene that Glimmer called (27338-27505).

Function of gene: hypothetical protein. HHPred: No significant probability (90 probability or greater). SOSUI: Soluble protein

**Gene 39**

- Original Glimmer call @bp 27954 has strength 15.86

Start 27954, Stop 27538, Length 417 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results less than 2e-38. Pham: 145 members

GeneMark and Glimmer agree on start. Called start contains full coding potential within ORF. Starterator report: Does not have most annotated start. Has start found in 24.1% of genes and called 97.1% when it is present. RBS score is best for this start

Function: hypothetical protein. HHPred: No significant results, probability equal to or greater than 90. Blastp: hypothetical protein. SOSUI: Soluble

**Gene 40**

- Original Glimmer call @bp 29477 has strength 14.64

Start 29477, Stop 28092, Length 1386 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A lot of significant results with e values of 0. Pham: 198 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 75.8% of genes in pham and is called 74.7% of the time when present. 1 bp overlap with gene 41.

Function: DNA helicase. HHPred: helicase, DNA binding protein, DNA excision repair protein, etc. Blastp: DNA helicase

**Gene 41**

- Original Glimmer call @bp 30550 has strength 14.07

Start 30550, Stop 29477, Length 1074 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results less than 3e-30. Pham: 147 members

GeneMark and Glimmer agree on start. Start codon contains full coding potential. Starterator report: Does not have most annotated start. Found in 1.4% of genes and is called 100% of the time that it is present.

Function: MazG-like nucleotide pyrophosphohydrolase. HHPred: pyrophosphohydrolase/MazG-like nucleoside triphosphate pyrophosphohydrolase. Blastp: MazG-like nucleotide pyrophosphohydrolase

**Gene 42**

- Original Glimmer call @bp 31320 has strength 9.88

Start 31320, Stop 30547, Length 774 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results with e values practically 0. Pham:145 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 89.7% of genes and called 90% of time when present. 23 bp overlap with gene 43.

Function: Hypothetical protein. HHPred: No significant results equal to or greater than 90 probability. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 43**

- Original Glimmer call @bp 31888 has strength 15.69

Start 31888, Stop 31298, Length 591 bp.

This is a gen: GeneMarkS indicates high coding potential. Blastp: Many significant results w/ e values practically 0. Pham: 145 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 98.6% of genes in pham and called 100% of time it is present.

Function: thymidylate kinase. HHPred: Many top significant results for Thymidylate kinase. Blastp: Top results include AAA-ATPase and thymidylate kinase.

**Gene 44**

- Original Glimmer call @bp 32838 has strength 14.32

Start 32838, Stop 31888, Length 951 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results, e values than 2e-88. Pham: 135 members.

Start codon contains full coding potential. GeneMark and Glimmer agree on start. Starterator report: Does not have most annotated start. Calls start found in 11.1% of genes with start called 93.3% of the time. 8 bp overlap with gene 45.

Function: glycosyltransferase. HHPred: significant result for glycosyltransferase. Blastp: glycosyltransferase.

**Gene 45**

- Original Glimmer call @bp 33070 has strength 9.74

Start 33070, Stop 32831, Length 240 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results with low e values. Pham: 119 members

GeneMark and Glimmer agree on start. Start codon contains full coding potential. Starterator report: Calls most annotated start. Found in 88.2% of genes and called 93.3% when it is present.

Function: hypothetical protein. HHPred: No significant results equal to 90 or greater probability. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 46**

- Original Glimmer call @bp 33276 has strength 19.78; GeneMark calls start at 33282

Start 33276, Stop 33136, Length 141 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: several significant results with e values around E-14. Pham: 5 members.

GeneMark calls at 33282, Glimmer calls at 33276. Both starts contain all regions of high coding potential. Starterator report: Does not have most annotated start. Calls start 33276 found in 100.0% of genes in pham and called 40.0% of the time when it is present. Start 33282 is found in 40.0% of the genes (2 genes out of 5) and is called 50.0% of the time that it is present. Start at 33282 causes 4 bp overlap with gene 47. The RBS score is 10X better for 33276 than for 33282.

Function: membrane protein. HHPred: No significant probabilities. Blastp: hypothetical protein. SOSUI: Is a membrane protein. TMHMM: membrane protein.

**Gene 47**

- Original Glimmer call @bp 34103 has strength 17.41; GeneMark calls start at 34154

Start 34103, Stop 33279, Length 825 bp.

This is a gene: GeneMarkS indicates full coding potential. Blastp: Many significant results with e values practically 0. Pham: 145 members.

GeneMark calls start at 34154. Glimmer calls at 34103. Start at 34103 would cause gap of 125 bp. Start at 34154 would cause gap of 74 bp. Both starts contain full coding potential. Starterator report: Does not have most annotated start. Calls start 18 at 34103 that is found in 15.9% of genes in pham and called 100% of time when it is present. Start 34154 which is start 13 doesn’t have any MA’s.

Start at 34103 is -2.443 while start at 34154 is -6.195 for RBS score. 34103 seems to be the best start, this also gives 1:1 alignment with many other genes.

Function: thymidylate synthase. HHPred: thymidylate synthase. Blastp: thymidylate synthase, although one significant result for ThyX-like thymidylate synthase.

**Gene 48**

- Original Glimmer call @bp 34538 has strength 14.96

Start 34538, Stop 34227, Length 312 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant results with e values from E-08 to E-15. Pham: 8 members

GeneMark and Glimmer agree on start. Called start contains all regions of high coding potential. Starterator report: Calls most annotated start. Found in 87.5% of genes and called 100% of time when it is present. RBS score is the best for this start.

Function: hypothetical protein. HHPred: Effector domain of rabphilin-3a Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 49**

- Original Glimmer call @bp 34789 has strength 13.96

Start 34789, Stop 34568, Length 222 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant e values on phagesdb less than 2e-04. None on NCBI. Pham: 10 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Only found in HerculesXL. No other starts would cause it to match other phage members.

Function: hypothetical protein. HHPred: No significant results equal to or greater than 90% probability. Blastp: hypothetical protein or DNA binding protein. SOSUI: Soluble

**Gene 50**

- Original Glimmer call @bp 34968 has strength 15.57

Start 34968, Stop 34792, Start 177 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant e values less than 1e-04. Pham: 108 members

GeneMark and Glimmer agree on start, which is the only start in this ORF.

Function: hypothetical protein. HHPred: Some significant results but none that I see matching the official function list. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 51**

- Original Glimmer call @bp 35234 has strength 12.21

Start 35234, Stop 35022, Length 213 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Numerous e values less than 8e-14. Pham: 13 members.

GeneMark and Glimmer agree on star. Called start contains full coding potential. Starterator report: Does not have most annotated start. Start found in 23.1% of genes and called 100% of the time.

Start gives 4 bp overlap with gene 52.

Function: hypothetical protein. HHPred: No significant results. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 52**

- Original Glimmer call @bp 35965 has strength 16.32

Start 35965, Stop 35231, Length 735 bp.

This is gene: GeneMarkS indicates high coding potential. Blastp: Many significant e values that are practically 0. Pham: 162 members

GeneMark and Glimmer agree on start. Called start contains all coding potential in ORF. Starterator report: Calls most annotated start. Found in 100% of genes in pham and called 98.8% of the time it is present. RBS score is not the best, but other more significant scores would increase gap between gene 53 and 52 by a lot.

Function: hypothetical protein. HHPred: unknown function. Blastp: Hypothetical protein. SOSUI: Soluble protein.

**Gene 53**

- Original Glimmer call @bp 36406 has strength 8.37

Start 36406, Stop 36020, Length 387 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: 1 NCB1 result w/ E value E-16. Pham: 2 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. 4 bp overlap with gene 54.

Function: hypothetical protein. HHPred: No significant results greater than or equal to 90 probability. Blastp: hypothetical protein. SOSUI: Soluble protein

**Gene 54**

- Original Glimmer call @bp 36768 has strength 11.19

Start 36768, Stop 36403, Length 366 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant e values less than e-04. Pham: 10 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Only found in HerculesXL. Starts further upstream would include areas with no coding potential. Start gives 1 bp common overlap with gene 55

Function: hypothetical protein. HHPred: No significant results greater than or equal to 90 probability. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 55**

- Original Glimmer call @bp 37238 has strength 10.05

Start 37238, Stop 36768, Length 471 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Several results with e values practically 0. Pham: 17 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 52.9% of genes in the pham and called 100.0% of the time that it is present. Called start gives 4 bp overlap with gene 56.

Function: hypothetical protein. HHPred: Some significant results but I don’t believe they are included on the official list. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 56**

- Original Glimmer call @bp 37483 has strength 14.86

Start 37483, Stop 37235, Length 249 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant e values less than 2e-18. Pham: 29 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found in 89.7% of genes and called 76.9% of the time that it is present. Gives common 1 bp overlap with gene 57.

Function: Hypothetical protein. HHPred: no significant results. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 57**

- Original Glimmer call @bp 37797 has strength 12.49

Start 37797, Stop 37483, Length 315 bp

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant e values less than e-12. Pham: 3 members.

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Calls most annotated start. Found and called in all three members of pham. Start gives 1 bp overlap with gene 58.

Function: hypothetical protein. HHPred: significant results (92% prob) for “Anti-sigma-F factor Fin”. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 58**

- Original Glimmer call @bp 38054 has strength 15.23

Start 38054, Stop 37797, Length 258 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: Many significant results over 1e-10

Pham: 41 members.

GeneMark and Glimmer agree on start. Called start contains most coding potential. Starterator report: Calls most annotated start. Found in 43.9% of genes in pham and called 55.6% of the time that it is present. RBS score is the best for this start.

Function: hypothetical protein. HHPred: significant result for family of unknown function. Blastp: hypothetical protein. SOSUI: Soluble protein.

**Gene 59**

- Original Glimmer call @bp 38209 has strength 0.39 \*\* not called by GeneMark

Start 38209, Stop 38337, Length 129 bp.

This is probably a gene: Not called by GeneMarkS but it shows moderate coding potential here and helps fill what would otherwise be a large gap (~400 bp) between gene 58 and 60. Blastp: Orpham.

Extending the start would include no additional coding potential, with no substantial improvement in RBS score.

Function: hypothetical protein. HHPred: No significant results greater than or equal to 90 probability. Blastp: no matches. SOSUI: Soluble protein.

**Gene 60**

- Original Glimmer call @bp 38481 has strength 11.64

Start 38481, Stop 38678, Length 198 bp.

This is a gene: GeneMarkS indicates high coding potential in about 2/3 of the gene. Blastp: Many significant results less than e-12. Pham: 54 members

GeneMark and Glimmer agree on start. Called start contains full coding potential. Starterator report: Does not have most annotated start. Calls start found in 16.7% of genes and called 88.9% of the time that it is present. RBS score is the best for this start.

Function: membrane protein. HHPred: No significant results but does show some transmembrane segments. Blastp: Hypothetical protein and membrane protein. SOSUI: Membrane protein with only 1 transmembrane protein domain. TmHmm: membrane protein.

**Gene 61**

- Original Glimmer call @bp 38678 has strength 12.10

Start 38678, Stop 39088, Length 411 bp.

This is a gene: GeneMarkS indicates high coding potential. Blastp: A few significant e values less than e-11. Pham: 27 members

GeneMark and Glimmer agree on start. Called start contain full coding potential. Starterator report: Most annotated start. Found in 51.9% of genes and is called 92.9% of the time that it is present. Called start gives genes 1 bp overlap.

Function: hypothetical protein. HHPred: No significant results equal to or greater than 90% probability. Blastp: Hypothetical protein. SOSUI: Soluble protein