**Gene 1**

**1a. FINAL Gene Co-ordinates:** Start**:**1, Stop: 567

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Adenylate kinase

**1d. Final Summary: START**:Longest ORF, called in GeneMark, no overlap or gap with previous gene (final score: -6.387, z-value: 1.715, spacer distance: 6). Clear coding potential in first forward reading frame from ~1-540, 1:1 target:query for top 20 gene hits in BLAST, most manual start called in Starterator with 261 MAs and called 95.9% of the time when present. Synteny conserved. **FUNCTION**: Function supported by HHPred (high coverage and low e-value), adenylate kinase called in 9 of top 10 hits in BLAST, called by 81% (320 out of 395) of phages in the same pham (84713), called by three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 2**

**1a. FINAL Gene Co-ordinates:** Start: 564, Stop: 2351

**1b. Is it a protein-coding gene?** Yes

**1c. What is its function?** Terminase

**1d.** **Final Summary: START**: Not longest ORF but same call in Glimmer and GeneMark, overlap of 4 with previous gene (final score: -3.723, z-value: 2.967, spacer distance: 15). Clear coding potential in third forward reading frame from ~564-2351, 1:1 target:query for top 10 genes, most manual start called in Starterator with 267 MAs and called 100% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHPred (high coverage and low e-value), terminase called in top 20 hits in BLAST, called by 86% (498 out of 582) of phages in the same pham (220995) and called by three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 3**

**1a. FINAL Gene Co-ordinates:** Start: 2436, Stop: 2651

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical Protein

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 84 with previous gene compared to all other candidate starts (final score: -2.666, z-value: 3.261, spacer distance: 13). Clear coding potential in third forward reading frame from ~2436-2651, 1:1 target:query for 9 of top 10 gene hits, most manual start called in Starterator with 267 MAs and called 100% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called in top 10 hits in BLAST, called by 100% (296 out of 296) of phages in the same pham (216), called in three closest whole-genome relatives. Function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 4**

**1a. FINAL Gene Co-ordinates:** Start: 2648, Stop: 3013

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START**: Longest ORF, same call in Glimmer and GeneMark, overlap of 4 (final score: -4.161, z-value: 2.623, spacer distance: 14). Clear coding potential within the second forward reading frame from ~2648-3013, 1:1 target:query for 6 of top 10 gene hits in BLAST, most manual start called in Starterator with 260 MAs and called 96.3% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called in top 20 hits in BLAST, called by 100% (301 out of 301) of phages in the same pham (84759), called in three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 5**

**1a. FINAL Gene Co-ordinates:** Start:3081, Stop: 3218

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** Longest ORF**,** same call in Glimmer and GeneMark, smallest gap of 67 with previous gene compared to all other candidate starts (final score: -4.991, z-value: 1.887, spacer distance: 10). Clear coding potential in the third forward reading frame from ~3050-3218, 1:1 target:query for 6 of 8 total gene hits in BLAST, most manual start called in Starterator with 264 MAs and called 99.3% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein called in all 8 total hits in BLAST, called by 100% (289 out of 289) of phages in the same pham (84776), called in three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 6**

**1a. FINAL Gene Co-ordinates:** Start: 3919, Stop: 3365 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** RuvC-like resolvase

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 46 with previous gene compared to all other candidate starts (final score: -5.989, z value: 1.715, spacer distance: 14). Coding potential in first reverse reading frame from ~3919-3365, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 324 MAs and called 90.2% of the time when present. Synteny conserved. **FUNCTION**: Function supported by HHPred score (high coverage), RuvC-like resolvase called in 100% of top hits in BLAST, called by 83% (707 out of 847) of phages in same pham (228150), called by three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 7**

**1a. FINAL Gene Co-ordinates:** Start: 3966, Stop: 4412

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical Protein

**1d. Final Summary: START:** Not longest ORF but same call in Glimmer and GeneMark, smallest gap of 46 with previous gene compared to all other candidate starts (final score: -5.157, z value: 2.128, spacer distance: 14). Clear coding potential within second forward reading frame from ~3966-4412, 1:1 target:query for 8 of top 10 gene hits, most manual start called in Starterator with 259 MAs and called 95.6% of the time when present. Synteny conserved. **FUNCTION**: Function supported by HHPred (high coverage, low e-value), hypothetical protein called in 8 out of top ten hits in BLAST, called by 100% (583 out of 583) of phages in the same pham (229800), called by three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 8**

**1a. FINAL Gene Co-ordinates:** Start:4442, Stop: 6361

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Portal protein.

**1d. Final Summary: START:** Longest ORF with the gap of 29, most manual start called in Starterator with 405 MAs and called 90.4% of the time when present, best RBS scores among start candidates (final score: -5.138, z value: 2.361, spacer distance: 16). Strong, clear coding potential in the second forward reading frame from ~4442-6361, 1:1 target:query for top 20 genes in BLAST, synteny conserved. Not called by Glimmer and Genemark (start @4553) which has worse RBS values and is not most called start in Starterator. **FUNCTION**: Function supported by HHPred score (high probabilities and low e-values), portal protein called in top ten gene hits in BLAST, called by 82% (594 out of 726) of phages in the same pham (228153), called by three closest whole-genome relatives**.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 9**

**1a. FINAL Gene Co-ordinates:** Start: 6361, Stop: 8934

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Capsid maturation protease

**1d. Final Summary: START:** Not longest ORF and GeneMark and Glimmer disagree with this start; however, overlap of 1 with previous gene compared to all other candidate starts, and some of the best RBS scores (final score: -4.416, z value: 2.212, spacer distance: 9). Strong, clear coding potential in the first forward reading frame from ~6700-8934, 1:1 target:query for top 20 gene hits, most manual start called in Starterator with 409 MAs and called 95.1% of the time when present. Synteny conserved. **FUNCTION:** Capsid maturation protease called in top 10 gene hits in BLAST, called by 77% (383 out of 497) of phages in same pham (90), called in three closest whole-genome relatives. Function not supported by HHPred (low coverage).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DELETED THIS AUTO-CALLED GENE**

**1a. FINAL Gene Co-ordinates: Start: 6768, Stop: 6583
1b. Is it a protein-coding gene? No, this is not a gene.**

**1c. What is its function? n/a**

**1d: Final summary:** This auto-called gene was determined to be deleted as there was only 1 start candidate, 1 BLAST hit, minimal coding potential in the -3 frame, and there were only 7 other phages in the pham (11126). None of the three closest whole-genome relatives call this gene. Additionally, this is a reverse gene surrounded by forward genes both upstream and downstream.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**ADDED GENE, NOT AUTO-CALLED**

**Gene 10**

**1a. FINAL Gene Co-ordinates:** Start: 8944, Stop: 9378

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** HNH endonuclease

**1d. Final Summary: START:** Longest ORF, gap of 9 with previous gene compared to all other candidate starts (final score: -6.354, z value: 1.396, spacer distance: 13). Although minimal coding potential in the third forward reading frame, close whole-genome relatives call this gene. 1:1 target:query for 6 of top 10 gene hits. Synteny conserved in top closest relative. **FUNCTION:** Function supported by HHPred (high coverage and low e-value), HNH endonuclease called in 9 out of top ten gene hits in BLAST, called by 80% (322 out of 400) of phages in same pham (229829).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 11**

**1a. FINAL Gene Co-ordinates:** Start: 9507, Stop: 9644

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 128 with previous gene compared to all other candidate starts (final score: -5.627, z value: 2.118, spacer distance: 13). Clear coding potential in the third forward reading frame from ~9500-9644, 1:1 target:query for 3 of top 5 gene hits, most manual start called in Starterator with 239 MAs and called 100% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called 100% of the time in top hits in BLAST, called by 100% (262 out of 262) of phages in the same pham (84814), function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 12**

**1a. FINAL Gene Co-ordinates:** Start: 9758, Stop: 11503

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Major capsid hexamer protein

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 113 with previous gene compared to all other candidate starts (final score: -2.794, z value: 3.152, spacer distance: 13). Clear coding potential in the second forward reading frame from ~9758-11503, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 268 MAs and called 100% of the time when present. Synteny conserved. **FUNCTION:** Major capsid hexamer protein function called in 7 out of top 10 hits in BLAST, called by 82% (418 out of 507) of phages in the same pham (228170), and called by closest whole-genome relative.Function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 13**

**1a. FINAL Gene Co-ordinates:** Start: 11603, Stop: 12406

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Major capsid pentamer protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer, GeneMark, smallest gap of 99 with previous gene compared to all other candidate starts (final score: -3.347, z value: 2.734, spacer distance: 11). Strong, clear coding potential from ~11603-12406 in third forward reading frame, 1:1 target: query for top ten gene hits, most manual start called in Starterator with 343 MAs and called 99.8% of the time when present. Synteny conserved. **FUNCTION:** Major capsid pentamer protein function called in 6 out of top ten hits in BLAST, called by 84% (424 out of 506) of phages in the same pham (227117), and called by closest whole-genome relative. Function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 14**

**1a. FINAL Gene Co-ordinates:** Start: 12418, Stop: 13266

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF but same call in Glimmer and GeneMark, smallest gap of 11 with previous gene compared to all other candidate starts (final score: -4.467, z value: 2.217, spacer distance: 12). Clear coding potential from ~12418-13266 in the first forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 395 MAs and called 100% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHPred (high probability and low e-value), hypothetical protein function called by all top ten hits in BLAST, called by 100% (365 out of 365) of phages in the same pham (229839).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 15**

**1a. FINAL Gene Co-ordinates:** Start: 13266, Stop: 13655

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, overlap of 1 (final score: -3.535, z value: 2.680, spacer distance: 12). Strong, clear coding potential from start to stop site, 1:1 target:query for 9 of top 10 gene hits, most manual start called in Starterator with 300 MAs and called 98.6% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called by 7 of top 10 hits in BLAST, called by 90.3% (290 out of 321) of phages in the same pham (229852), function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 16**

**1a. FINAL Gene Co-ordinates:** Start: 13714, Stop: 14013

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START**: Not longest ORF but same call in Glimmer and GeneMark (final score: -6.174, z-value 1.623). Clear coding potential from start to stop site, 1:1 target:query for 4 of top 10 gene hits, most manual start called in Starterator with 72 MAs and called 100% of the time when present, conserved synteny. **FUNCTION:** 100% of phages similar to closest relative from phamerator analysis call a hypothetical protein function, hypothetical protein function called by 7 out of top ten hits in BLAST, called by 100% (95 out of 95) of phages in the same pham (224471), function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 17**

**1a. FINAL Gene Co-ordinates:** Start: 14010, Stop: 14624

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Differing calls in Glimmer and GeneMark but supported by GeneMark, longest ORF, overlap of 4 with previous gene (final score: -4.523, z value: 2.15, spacer distance: 11). Clear coding potential from ~14019-14624, 1:1 target:query for top 20 genes, most manual start called in Starterator with 338 MAs and called 97.7% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHPred (high coverage), hypothetical protein function called by 8 out of top ten hits in BLAST, called by 100% (452 out of 452) of phages in the same pham (227151), and PhagesDB predicts the function of a hypothetical protein for 100% of similar phages to this gene.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 18**

**1a. FINAL Gene Co-ordinates:** Start: 14712, Stop: 15512

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Major tail protein.

**1d. Final Summary: START:** Longest ORF,same call in Glimmer, GeneMark, smallest gap of 87 with previous gene compared to all other candidate calls (final score: -3.836, z value: 2.491, spacer distance: 11). Clear coding potential from start to stop site in the third forward reading frame, 1:1 target: query for top ten gene hits, most manual start called in Starterator with 288 MAs and called 97.7% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHPred with tail tube protein called in top two hits with high probability, major tail protein called by top 20 hits in BLAST, called by 84% (640 out of 765) of phages in the same pham (225750), called by closest whole-genome relative.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 19**

**1a. FINAL Gene Co-ordinates:** Start: 15800, Stop: 15570 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Longest ORF**,** same call in Glimmer, GeneMark, smallest gap of 77 with previous gene compared to all other candidate starts (final score: -3.307, z value: 3.135, spacer distance: 7). Clear coding potential from start to stop site in the first reverse reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 450 MAs and called 97.5% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called by top ten hits in BLAST, called by 100% (537 out of 537) of phages in the same pham (76668), 100% of similar phages to the closest relative from phamerator analysis called a hypothetical protein function, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 20**

**1a. FINAL Gene Co-ordinates:** Start: 16654, Stop: 15878 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Queuine tRNA-ribosyltransferase.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 55 with previous gene compared to all other candidate calls (final score: -3.514, z value: 2.651, spacer distance: 11). Clear coding potential from start to stop site in the first reverse reading frame, 1:1 target:query for 7 of top 10 gene hits, most manual start called in Starterator with 328 MAs and called 98.9% of the time when present. Synteny conserved**. FUNCTION:** Function supported by HHpred (high coverage), queuine tRNA-ribosyltransferase function called by top fifteen hits in BLAST, called by 78% (289 out of 369) of phages in the same pham (208288), called by closest whole-genome relative.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 21**

**1a. FINAL Gene Co-ordinates:** Start: 16973, Stop: 16710 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF but same call in Glimmer and GeneMark, strongest RBS scores compared to all other candidate calls (final score: -3.307, z value: 3.135, spacer distance: 7). Clear coding potential from start to stop in the third reverse reading frame, 1:1 target:query for 4 of top ten gene hits, most manual start called in Starterator with 167 MAs and called 65.8% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called by top 20 hits in BLAST, called by 100% (284 out of 284) of phages in the same pham (252), 100% of similar phages to the closest relative from phamerator analysis called a hypothetical protein function, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 22**

**1a. FINAL Gene Co-ordinates:** Start: 17083, Stop: 17820

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Head-to-tail adaptor.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 109 with previous gene compared to all other candidate calls (final score: -5.680, z value: 1.869, spacer distance: 14). Clear coding potential from start to stop site in the first forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 337 MAs and called 86.0% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHpred (high coverage), head-to-tail adaptor function called by top 50 or hits in BLAST, called by 80% (639 out of 796) of phages in the same pham (217440), called by closest whole-genome relative.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 23**

**1a. FINAL Gene Co-ordinates:** Start: 17820, Stop: 18338

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF and differing calls in Glimmer and Genemark but call supported by Glimmer, overlap of 1 with previous gene, strongest RBS scores (final score: -4.583, z value: 2.129, spacer distance: 9). Clear coding potential from start to stop site in the third forward reading frame, 1:1 target:query for top 9 gene hits, most manual start called in Starterator with 364 MAs and called 97.9% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called by top ten hits in BLAST, called by 100% (497 out of 497) of phages in the same pham (224150), PhagesDB predicts the function of hypothetical with a 100% frequency in all similar phages to ElvisPhasley for this gene, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 24**

**1a. FINAL Gene Co-ordinates:** Start: 18354, Stop: 18704

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF but same call in Glimmer and GeneMark, strongest RBS scores compared to all other candidate calls (final score: -5.680, z value: 1.869, spacer distance: 14). Clear coding potential between the start and stop site in the third forward reading frame, 1:1 target:query for second top gene hit, most manual start called in Starterator with 330 MAs and called 77.2% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHpred (high probability, coverage, and low e-value), hypothetical protein function called by top twenty hits in BLAST, called by 99.5% (569 out of 572) of phages in the same pham (225763), called by PhagesDB predicts the function of hypothetical with a 97% frequency in all similar phages to ElvisPhasley for this gene.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 25**

**1a. FINAL Gene Co-ordinates:** Start: 18708, Stop: 19130

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Tail assembly chaperone

**1d. Final Summary: START:** Differing calls in Glimmer and GeneMark but supported by GeneMark, longest ORF, strongest RBS scores compared to all other candidate calls (final score: -5.633, z value: 1.599, spacer distance: 11). Clear coding potential from the start to stop in the third forward reading frame, most manual start called in Starterator with 55 MAs and called 94.6% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHpred (high probability, coverage, and low E value), tail assembly chaperone function called by top five hits in BLAST, called by 54% (382 out of 708) phages in the same pham (228154), called by closest whole-genome relative.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 26**

**1a. FINAL Gene Co-ordinates:** Start: 19181, Stop: 19744

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, but same call in Glimmer and GeneMark, smallest gap of 50 with previous gene compared to all other candidate starts (final score: -3.030, z value: 2.892, spacer distance: 11). Clear coding potential from the start to stop site starting at ~19175 in the second forward reading frame, 1:1 target:query for top ten gene hits in BLAST, most manual start called in Starterator with 266 MAs and called 100% of the time when present. Synteny conserved**. FUNCTION:** Hypothetical protein function called by top five hits in BLAST, called by 95.3% (446 out of 468) of phages in the same pham (225780), called by three closest whole-genome relatives, function not supported by HHPred (low coverage).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 27**

**1a. FINAL Gene Co-ordinates:** Start: 20150, Stop: 19809 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, but same call in Glimmer and GeneMark, strongest RBS scores compared to all other candidate calls (final score: -3.030, z value: 2.892, spacer distance: 11). Clear coding potential from the start to stop site in the second reverse reading frame, 1:1 target: query for 6 of the top 10 gene hits, most manual start called in Starterator with 214 MAs and called 81.9% of the time when present. Synteny conserved**. FUNCTION:** Hypothetical protein called by top ten hits in BLAST, called by 100% (293 out of 293) of phages in the same pham (84764), function not supported by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 28**

**1a. FINAL Gene Co-ordinates:** Start: 20270, Stop: 26248

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Tape measure protein.

**1d. Final Summary: START:** Longest ORF,same call in Glimmer and GeneMark, smallest gap of 119 with previous gene compared to all other candidate starts (final score: -7.561, z value: 0.934, spacer distance: 14). Clear coding potential from start to stop site in the second forward reading frame, 1:1 target:query for top 20 gene hits, most manual start called in Starterator with 407 MAs and called 99.8% of the time when present. Synteny conserved**. FUNCTION:** Function supported by HHpred (high probability and low E value), tape measure protein function called by top 15 hits in BLAST, called by 87% (412 out of 472) of phages in the same pham (100), called by closest whole-genome relative.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 29**

**1a. FINAL Gene Co-ordinates:** Start: 26258, Stop: 27691

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Minor tail protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 9 compared to all other candidate starts (final score: -3.814, z value: 2.471, spacer distance: 10). Clear coding potential from start to stop site in the second forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 313 MAs and called 99.4% of the time when present. Synteny conserved. **FUNCTION:** Minor tail protein called by top 30 hits in BLAST, called by 86% (427 out of 498) of phages in the same pham (219275), called by top three closest whole-genome relatives, function called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 30**

**1a. FINAL Gene Co-ordinates:** Start: 27688, Stop: 28800

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Minor tail protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, overlap of 4 with previous gene (final score: -4.367, z value: 2.267, spacer distance: 12). Clear coding potential from start to stop site, starting at ~27720 in the first forward reading frame and levels off at end of 28800; 1:1 target:query for top ten gene hits, most manual called start in Starterator with 259 MAs and called 96.9% of the time when present. Synteny conserved**. FUNCTION:** Function supported by HHpred (high coverage), minor tail protein function called by top 30 hits in BLAST, called by 77% (390 out of 507) of phages in the same pham (187775), called by top three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 31**

**1a. FINAL Gene Co-ordinates:** Start:28797, Start: 31055

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Minor tail protein.

**1d. Final Summary: START:** Longest ORF,same call in Glimmer and GeneMark, overlap of 4 with previous gene (final score: -4.742, z value: 2.042, spacer distance: 11). Clear coding potential from start to stop site, starting at ~28840 in the third forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 300 MAs and called 99.7% of the time when present. Synteny conserved. **FUNCTION:** Minor tail protein function called by top 20 hits in BLAST, called by 76% (302 out of 400) of phages in the same pham (219305), called by top three closest whole-genome relatives, function supported by HHpred (high probability and low e-value, but low coverage).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 32**

**1a. FINAL Gene Co-ordinates:** Start:31059, Stop: 32405

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Minor tail protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, gap of 3 (final score: -3.628, z value: 2.564, spacer distance: 10). Clear coding potential between start and stop site in the third forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 346 MAs and called 100% of the time when present, synteny conserved**. FUNCTION:** Function supported by HHpred (high probability and low E-value), minor tail protein function called by approximately top 50 hits in BLAST, called by 87% (346 out of 400) of phages in the same pham (141), called by top three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 33**

**1a. FINAL Gene Co-ordinates:** Start:32409, Stop: 33569

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Minor tail protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 3 with previous gene compared to all other candidate starts (final score: -3.671, z value: 2.543, spacer distance: 10). Clear coding potential between start and stop site, starting at ~32450 in the third forward reading frame, 1:1 target:query for top 20 gene hits, most manual start called in Starterator with 296 MAs and called 90.3% of the time when present. Synteny conserved. **FUNCTION:** Function supported by HHpred (high probability and low E value), minor tail protein function called by top 10 hits in BLAST, called by 64% (77 out of 121) of phages in the same pham (230018), called by 2 of 3 closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 34**

**1a. FINAL Gene Co-ordinates:** Start:33600, Stop: 33989

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 30 with previous gene compared to all other candidate starts, best RBS scores (final score: -6.520, z value: 1.578, spacer distance: 15). Clear coding potential from start to stop site in the third forward reading frame, 1:1 target:query for top ten gene hits, most manual start called in Starterator with 507 MAs and called 95.6% of the time when present. Synteny conserved. **FUNCTION:** Hypothetical protein function called by 9 of top 10 hits in BLAST, called by 100% (994 out of 994) of phages in the same pham (227106), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 35**

**1a. FINAL Gene Co-ordinates:** Start:33998, Stop:34384

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, smallest gap of 8 with previous gene compared to all other candidate starts (final score: -3.524, z value: 2.877, spacer distance: 8). Clear coding potential between start and stop site, 1:1 target:query for top 20 gene hits, most manual start called in Starterator with 307 MAs and called 98.5% of the time when present. Synteny conserved. **FUNCTION**: Hypothetical protein function called by top 15 hits in BLAST, called by 98.3% (466 out of 474) of phages in the same pham (228173), called in three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 36**

**1a. FINAL Gene Co-ordinates:** Start: 34381, Stop: 35031

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, overlap of 4 bp with previous gene, most annotated start in Starterator with 532 MA’s and called 74.8% of time when present, second best RBS scores (final score: -5.979, z-value: 1.918, spacer distance: 6). Strong coding potential for the whole gene in the first forward reading frame, 1:1 target:query for 6 of top 10 gene hits by BLAST, conserved synteny. **FUNCTION:** Hypothetical protein function called by top 15 hits in BLAST, called by 100% (994 out of 994) of phages in the same pham (227106), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 37**

**1a. FINAL Gene Co-ordinates:** Start: 35028, Stop: 35627

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical Protein.

**1d. Final Summary: START:** Not longest ORF, but same call in Glimmer and GeneMark, overlap of 1 with previous gene (final score: -5.009, z value: 1.918, spacer distance: 9). Strong coding potential for most of the gene in the third forward reading frame, 1:1 target:query for top 10 genes, most manual start called in Starterator with 532 MAs and called 74.8% of the time when present, conserved synteny. **FUNCTION:** Hypothetical protein function called by top ten hits in BLAST, called by 100% (994 out of 994) of phages in the same pham (227106), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 38**

**1a. FINAL Gene Co-ordinates:** Start: 35624, Stop: 36247

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical Protein

**1d. Final Summary: START:** Not longest ORF, but same call in Glimmer and GeneMark, overlap of 4 with previous gene (final score: -4.853, z value: 2.217, spacer distance: 8). Strong coding potential in the second forward reading frame within gene range, 1:1 target:query for top 20 gene hits, most manual start called in Starterator with 296 MAs and called 97.6% of the time when present. Conserved synteny. **FUNCTION:** Hypothetical protein function called by top 20 hits in BLAST, called by 100% (384 out of 384) of phages in the same pham (172562), called by three closest whole-genome relatives, function not called by HHPred (low coverage and high e-values)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 39**

**1a. FINAL Gene Co-ordinates:** Start: 36281, Stop: 36649

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, but same call in GeneMark and Glimmer, second smallest gap of 33 bp with previous gene, most manual start called in Starterator with 295 MAs and called 95.1% of the time when present, better RBS scores than LORF start (final score: -4.292, z value: 2.265, spacer distance: 11). Strong coding potential in the second forward reading frame, 1:1 target:query for 6 of top 10 gene hits, conserved synteny. **FUNCTION:** Function supported by HHPred score (high probability, relatively high coverage of 67%, and E-value of 0.47), hypothetical protein called by top ten hits in BLAST, called by 100% (491 out of 491) of phages in the same pham (87), called by three closest whole-genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 40**

**1a. FINAL Gene Co-ordinates:** Start: 36649, Stop: 36960

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, but same call in GeneMark and Glimmer, overlap of 1 (final score: -7.729, z value: 0.851, spacer distance: 14). Strong coding potential in the first forward reading frame, 1:1 target:query for 7 of top 10 gene hits, most manual start called in Starterator with 248 MAs and called 68.0% of the time when present, conserved synteny. **FUNCTION:** Hypothetical protein function called by top ten hits in BLAST, called by 100% (477 out of 477) of phages in the same pham (202564), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 41**

**1a. FINAL Gene Co-ordinates:** Start: 37000, Stop: 37407

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical Protein.

**1d. Final Summary: START:** Longest ORF, same call in GeneMark and Glimmer, smallest gap of 39 with previous gene (final score: -7.581, z value: 0.862, spacer distance: 8). Strong coding potential in the first forward reading frame, 1:1 target:query for top 10 genes, most manual start called in Starterator with 370 MAs and called 93.7% of the time when present, conserved synteny. **FUNCTION:** Hypothetical protein function called by 9 of top 10 hits in BLAST, called by 100% (497 out of 497) of phages in the same pham (81), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 42**

**1a. FINAL Gene Co-ordinates:** Start: 37407, Stop: 38528

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical Protein.

**1d. Final Summary: START:** Longest ORF, same call in GeneMark and Glimmer, overlap of 1 with previous (final score: -5.832, z value: 2.267, spacer distance: 18). Strong coding potential for the entirety of the gene in the third forward reading frame, 1:1 target:query for top 13 gene hits, most manual start called in Starterator with 305 MAs and called 99.7% of the time when present, conserved synteny. **FUNCTION:** Hypothetical protein function called by 8 of top 10 hits in BLAST, called by 98.6% (362 out of 367) of phages in the same pham (162), called by three closest whole-genome relatives, function not called by HHPred.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DELETED THIS AUTO-CALLED GENE**

**1a. FINAL Gene Co-ordinates:** Start: 38727, Stop: 39227

**1b. Is it a protein-coding gene***?* Not a gene**.**

**1c. What is its function?** N/A

**1d. Final Summary:** This auto-called gene was determined to be deleted due to minimal coding potential in the forward reading frames in the range of this gene, there being only 2 BLAST hits for this gene and 5 other members in the called pham for this gene (9340), and the fact that the gene is not called in any of ElvisPhasley’s 3 closest whole-genome relatives. This gene also overlaps 2 reverse genes that are called in the closest relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 43**

**1a. FINAL Gene Co-ordinates:** Start: 38728, Stop: 38510 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, butsame call in Glimmer and GeneMark, gap of 1 with previous gene (final score: -4.649, z value: 2.127, spacer distance: 12). Clear coding potential from start to stop site in the first reverse reading frame, 1:1 target:query for 7 of top 10 gene hits, most manual start called in Starterator with 259 MAs and called 98.3% of the time when present, synteny conserved. **FUNCTION:** Hypothetical protein function called by top 10 hits in BLAST, called by 100% (291 out of 291) of phages in the same pham (84773), called by three closest whole-genome relatives, function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 44**

**1a. FINAL Gene Co-ordinates:** Start: 39023, Stop: 38730 Reverse

**1b. Is it a protein-coding gene*?***Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF but most manual start called in Starterator with 223 MAs and called 80.6% of the time when present (LORF is uncalled in Starterator), overlap of 4 with previous gene (final score: -4.513, z value: 2.124, spacer distance: 5), clear coding potential in the second reverse reading frame, 1:1 target:query for 8 of top 10 genes, conserved synteny. No start called by Glimmer, Genemark start is not the most annotated in Starterator and has poorer RBS scores. **FUNCTION:** Hypothetical protein called by 9 of top ten hits in BLAST, called by 99% (285 out of 288) of phages in the same pham (240), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 45**

**1a. FINAL Gene Co-ordinates:** Start: 39235, Stop: 39020 Reverse

**1b. Is it a protein-coding gene*?*** *Yes.*

**1c. What is its function?** Helix-turn-helix DNA binding domain

**1d. Final Summary: START:** Longest ORF, same call by Genemark overlap of 4 with previous (final score: -4.299, z value: 2.23, spacer distance: 10). Strong coding potential from start to stop site in the first reverse reading frame, 1:1 target: query for 3 of top 10 gene hits, most manual start called in Starterator with 291 MAs and called 100% of the time when present, conserved synteny. No start called by Glimmer. **FUNCTION:** Helix-turn-helix DNA binding domain function called by 7 of top ten hits in BLAST**,** called by 82% (276 out of 338) of phages in the same pham (222813), called by three closest whole-genome relatives, called by HHPred (high coverage and low e-value).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 46**

**1a. FINAL Gene Co-ordinates:** Start: 39657, Stop: 39232 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Helix-turn-helix DNA binding domain

**1d. Final Summary: START:** Not the longest ORF but same call in Glimmer and GeneMark, overlap of 4 with previous gene, strongest RBS scores (final score: -2.794, z value: 3.152, spacer distance: 13). Strong coding potential in first reverse frame, 1:1 target:query for 4 of top 10 gene hits, most manual start called in Starterator with 51 MAs and called 76.8% of the time when present, conserved synteny. **FUNCTION:** Helix-turn-helix DNA binding domain called by top 10 BLAST hits, called by 70% (113 out of 161) of phages in the same pham (228341), called by top two closest whole-genome relatives, function not called by HHPred (low coverage).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 47**

**1a. FINAL Gene Co-ordinates:** Start: 39890, Stop: 39654 Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark, gap of 4 with previous gene (final score: -4.635, z value: 2.103, spacer distance: 9). Strong coding potential from start to stop in the second reverse reading frame, 1:1 target:query for 4 of 6 total BLAST hits, most manual start called in Starterator with 42 MAs and called 93.5% of the time when present. Conserved synteny. **FUNCTION:** Hypothetical protein called in 5 of 6 total BLAST hits, called by 100% (46 out of 46) of phages in the same pham (1947), called by three closest whole-genome relatives, function not called by HHPred (low coverage).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 48**

**1a. FINAL Gene Co-ordinates:** Start: 40455, Stop: 39895Reverse

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Hypothetical protein.

**1d. Final Summary: START:** Not longest ORF, but same call in Glimmer and GeneMark, gap of 107 (this occurs between a reverse-reading gene and a forward-reading gene), best RBS scores compared to all other candidate calls (final score: -3.212, z value: 2.84, spacer distance: 12). Strong coding potential from start to stop site in the third reverse reading frame, 1:1 target:query for 5 of top 10 gene hits, most manual start called in Starterator with 206 MAs and called 75.3% of the time when present, conserved synteny. **FUNCTION:** Hypothetical protein function called in top 20 hits in BLAST, called by 98.7% (516 out of 523) of phages in the same pham (229809), called by three closest whole-genome relatives, function not called by HHPred (low coverage).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**DELETED THIS AUTO-CALLED GENE**

**1a. FINAL Gene Co-ordinates:** Start: 40594, Stop: 40505 (Reverse)

**1b. Is it a protein-coding gene?** No, this is not a gene.

**1c. What is its function?** n/a

**1d. Final Summary:** This gene this was determined to be deleted because a) this gene is not called in any of ElvisPhasley’s three closest relatives and b) there is no coding potential in any of the reverse frames in the predicted coordinates for this gene.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 49**

**1a. FINAL Gene Co-ordinates:** Start: 40554, Stop: 41870

**1b. Is it a protein-coding gene*?*** Yes.

**1c. What is its function?** Lysin A.

**1d. Final Summary: START:** Not the longest ORF but most manual start called in Starterator with 144 MAs and called 50.9% of the time when present, 2 closest whole-genome relatives also call the most called start in Starterator, 659 bp gap (this occurs between a reverse-reading gene and a forward-reading gene). Strong coding potential from start to stop site in first forward reading frame, 1:1 target:query for 7 of top 10 genes in BLAST, conserved synteny. Both Glimmer call (start @40563) and Genemark call (start @40575) are not the most annotated start in Starterator and would create a larger gap. **FUNCTION:** Lysin A called by top 10 hits in BLAST, called by 91% (348 out of 381) of phages in the same pham (229834), called by three closest whole-genome relatives, function supported by HHPred (high coverage, low e-value).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 50**

**1a. FINAL Gene Co-ordinates:** Start: 41880, Stop: 43235

**1b. Is it a protein-coding gene?** Yes.

**1c. What is its function?** Lysin B.

**1d. Final Summary: START:** Longest ORF, same call in Glimmer and GeneMark (final score: -4.326, z value: 2.217, spacer distance: 10), 10 bp gap with previous gene, most manual start called in Starterator with 267 MAs and called 100% of the time when present. Strong coding potential in the third forward reading frame, 1:1 target: query for top 20 gene hits in BLAST, conserved synteny.

**FUNCTION:** Lysin B called by top 20 gene hits in BLAST, called by 92% (460 out of 501) of phages in the same pham (229811), called by three closest whole-genome relatives, function supported by HHPred (low e-values but low coverage ~38%).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 51**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 43957**, Stop:** 43337 (Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical Protein
 **(d) Final Summary: START**: Longest ORF, same call in Glimmer and Genemark, 57bp gap with previous gene (final score: -4.173, z-value: 2.84 spacer distance: 16). Strong coding potential in –1 frame from ~43350 – 43950, 1:1 target:query for top 12 genes, most manual called start in Starterator (called 100% of the time when present), conserved synteny. **Function**: Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (493 out of 493) of phages in the same pham (221004), called by 3 closest whole genome relatives. HHpred hits show low coverage and high e value, not supportive.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 52**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 45418**, Stop:** 44015 (Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Exonuclease
 **(d) Final Summary: START**: Not longest ORF – longest ORF has 163bp overlap, same call in Glimmer and Genemark, 2bp gap with previous gene (final score: -2.305, z-value: 3.95 spacer distance: 13). Strong coding potential in –1 frame from ~44050– 45400, 1:1 target: query for top 16 genes, most manual called start in Starterator (called 84.3% of the time when present), conserved synteny. **Function**: Function supported by high HHPRED score for exonuclease, Function supported by BLAST with exonuclease called top 5 genes, called by 90/493 of phages in the same pham (222760), not called by 3 closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 53**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 47043**, Stop:** 45421(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** DNA Helicase
 **(d) Final Summary: START**: Not longest ORF – longest ORF has 152bp overlap and other smaller ORFs have similarly unacceptable overlaps, called by Glimmer, 22bp gap with previous gene (final score: -4.702, z-value: 2.1, spacer distance: 12). Strong coding potential in –3 frame from ~45450– 47080, highly manually called start in Starterator (called 74.7% of the time when present), 1:1 target:query with 1 of top 10 gene hits in BLAST, conserved synteny. **Function**: Function supported by high HHPRED score for top 250 hits for helicase, Function supported by BLAST with DNA helicase called top genes, called by 463/542 of phages in the same pham (229807), called by 3 closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DELETED THIS AUTO-CALLED GENE**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 47066**, Stop:** 47194 (Forward)
 **(b) Is it a protein-coding gene?** No, not a gene.
 **(c) What is its function?** n/a
 **(d) Final Summary:** This auto-called gene was determined to be deleted because a) it is a forward gene flanked by many reverse genes both up and downstream, b) there is extremely minimal to no coding potential predicted in the forward frames (+1 specifically) within the predicted coordinates of this genes, c) none of ElvisPhasley’s three closest whole-genome relatives call this gene and as a result keeping this gene would cause synteny to not be conserved.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 54**

**Final Gene Information**

1. **FINAL Gene Co-ordinates: Start:** 47367**, Stop:** 47266 (Reverse)
2. **Is it a protein-coding gene?** Yes.
3. **What is its function?** Hypothetical protein
4. **Final Summary:** Not longest ORF—longest ORF overlaps previous gene by 152 bp—called ORF is the first ORF without significant overlap (7 bp gap), same call by Genemark (final score: -5.440, z-value: 1.838, spacer distance: 13). Strong coding potential in the -3 frame from ~47300-47400 bp, 1:1 target:query for zero hits in BLAST, conserved synteny. Glimmer called start @47559 which has zero manual annotations in Starterator and creates a smaller ORF. **Function:** Hypothetical protein called by 9 of 9 total gene hits in BLAST, called by 100% (44 out of 44) of phages in same pham (223237), called by two closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 55**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 47860**, Stop:** 47645(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein
 **(d) Final Summary: START**: Not longest ORF – longest ORF has 83bp overlap, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 2.377, z-value: 3.216, spacer distance: 11). Strong coding potential in –1 frame from ~47650– 47850, most manually called start in Starterator (called 100% of the time when present) 250MA, 1:1 target:query for top 16 genes, conserved synteny with 8 of most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top genes, called by 100% of phages in the same pham (84757), called by 3 closest whole genome relatives. HHpred hits show low coverage and high e value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 56**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 48537**, Stop:** 47857(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein
 **(d) Final Summary: START**: Not longest ORF – longest ORF has 256bp overlap and other smaller ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 4.760, z-value: 2.071, spacer distance: 12). Strong coding potential in –3 frame from ~47900– 48520, most manually called start in Starterator (called 96% of the time when present) 136MA, 1:1 target: query for top 20 genes, conserved synteny with 8 of most similar phage. **Function**: Function supported by high HHPRED score for top hit of family of unknown function, Function supported by BLAST with hypothetical protein called for top genes, called by 100% of phages in the same pham (229924), called by 8 of closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 57**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 48905**, Stop:** 48534(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein
 **(d) Final Summary: START**: Not longest ORF – longest ORF has 208bp overlap and other smaller ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 7bp overlap with previous gene (final score: - 7.021, z-value: 1.527, spacer distance: 17). Strong coding potential in –3 frame from ~48550– 48900, highly manually called start in Starterator (called 59.8% of the time when present) 157MA, 1:1 target: query for top 32 genes, conserved synteny with 8 of most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top 3 genes and many of top 100, called by 772/775 of phages in the same pham (229924), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 58**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 49189**, Stop:** 48899(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, but same call by Glimmer and Genemark, 75bp gap with previous gene (final score: - 2.016, z-value: 3.395, spacer distance: 11). Strong coding potential in –1 frame from ~48940– 49200, highly manually called start in Starterator (called 77.4 of the time when present) 189MA, 1:1 target: query for top 15 genes, conserved synteny with 10 most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top 3 genes and many of top 100, called by 864/871of phages in the same pham (163226), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 59**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 52012**, Stop:** 49265(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** DNA primase/helicase

 **(d) Final Summary: START**: Not longest ORF, longest ORF has 103bp overlap, same call by Glimmer and Genemark, 119 bp gap with previous gene (final score: - 3.201, z-value: 2.816, spacer distance: 9). Strong coding potential in –1 frame from ~49300– 52000, most manually called start in Starterator (called 74.3% of the time when present) 304MA, 1:1 target: query for many of top 100 genes, conserved synteny with 10 most similar phage. **Function**: Function supported by HHPRED with high scores for top 4 hits as helicase, Function supported by BLAST with variations of DNA primase/helicase called for 100, called by 401/495 of phages in the same pham (84), called by 10 closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DELETED THIS AUTO-CALLED GENE**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates:** Start: 52015, Stop: 52134 (Forward)
 **(b) Is it a protein-coding gene?** No, not a gene.
 **(c) What is its function?** n/a

1. **Final Summary:** This gene was determined to be deleted because a) it is a singular forward gene flanked by many reverse genes both upstream and downstream, b) there is very minimal coding potential in the +1 frame in the auto-called region, c) none of ElvisPhasley’s 3 closest whole-genome relatives call this gene.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 60**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 52358**, Stop:** 52131(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 8bp overlap with previous gene (final score: - 4.624, z-value: 2.1, spacer distance: 11). Strong coding potential in -3 frame from ~52150– 52350, most manually called start in Starterator (called 100% of the time when present) 51MA, 1:1 target: query for top 13 genes, conserved synteny with 10 most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top genes, called by 57/57 of phages in the same pham (85674), called by 3 of closest whole genome relatives. HHpred hits show low coverage and high e value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 61**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 52920**, Stop:** 52351(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 70 bp overlap and other smaller ORFS have similarly unacceptable overlaps, called by Genemark, 4bp overlap with previous gene (final score: - 5.585, z-value: 1.592, spacer distance: 10). Strong coding potential in -3 frame from ~52370– 53850, most manually called start in Starterator (called 80.5%% of the time when present) 209MA, 1:1 target: query for top 9 genes, conserved synteny with 3 of most similar phage. **Function**: Function supported by high HHPRED score for top hit of domain of unknown function, Function supported by BLAST with hypothetical protein called for top hits, called by 290/290 of phages in the same pham (228236), called by 10 closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 62**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 54776**, Stop:** 52917(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** DNA polymerase I

 **(d) Final Summary: START**: Not longest ORF, LORF has 28bp overlap, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 4.888, z-value: 1.969, spacer distance: 11). Strong coding potential in -2 frame from ~52960– 54760, most manually called start in Starterator (called 97.9%% of the time when present) 305MA, 1:1 target: query for top 96 genes, conserved synteny with 10 most similar phage. **Function**: Function supported by high HHPRED score for top 10 hits of DNA Polymerase, Function supported by BLAST with DNA polymerase I called for top hits, called by 1721/1972 of phages in the same pham (229768), called by 10 closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 63**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 55231**, Stop:** 54773(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 92bp gap with previous gene (final score: - 6.053, z-value: 1.39, spacer distance: 11). Strong coding potential in -1 frame from ~54810– 55220, most manually called start in Starterator (called 95.4% of the time when present) 303MA, 1:1 target: query for top 40 genes, conserved synteny with 7 of the most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top hits, called by 417/417 of phages in the same pham (228187), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 64**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 55476**, Stop:** 55324(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 184bp overlap other longer ORFS have similarly unacceptable, same call by Glimmer and Genemark, 1bp overlap with previous gene (final score: - 4.977, z-value: 1.964, spacer distance: 12). Strong coding potential in -3 frame from ~55360– 55470, most manually called start in Starterator (called 81.9%% of the time when present) 200MA, 1:1 target: query for top 96 genes, conserved synteny with 7 of most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top hits, called by 271/271 of phages in the same pham (274), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 65**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 55832**, Stop:** 55476(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 176bp overlap, longer ORF with 31bp gap not called by similar phage, same call by Glimmer and Genemark, 112bp gap with previous gene (final score: - 4.161, z-value: 2.623, spacer distance: 14). Strong coding potential in -3 frame from ~55500– 55850, most manually called start in Starterator (called 81.6%% of the time when present) 197MA, 1:1 target: query for many of top genes, conserved synteny with 10 most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top hits, called by 289/289 of phages in the same pham (224230), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 66**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 56304**, Stop:** 55945(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has smaller gap not called by relatives, same call by Glimmer and Genemark, 229bp gap with previous gene (final score: - 7.414, z-value: 1.332, spacer distance: 5). Strong coding potential in -2 frame from ~55990– 56310, most manually called start in Starterator (called 83.2%% of the time when present) 271MA, 1:1 target:query for many of top genes, conserved synteny with 7 of most similar phage. **Function**: Function supported by BLAST with hypothetical protein called for top hits, called by 456/456 of phages in the same pham (206463), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DELETED THIS AUTO-CALLED GENE**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 56534**, Stop:** 56653(Forward)
 **(b) Is it a protein-coding gene?** No, not a gene.
 **(c) What is its function?** n/a

1. **Final Summary:** This auto-called gene was determined to be deleted. This gene is an orpham, having no other similar genes in any other phages, including ElvisPhasley’s 3 closest whole-genome relatives. This gene is not called in the top close relative (OSmaximus). Additionally, there is no coding potential where this gene was predicted to be. While this leaves a considerable gap between gene stop @55945 and gene stop @57811, this sizable gap is seen in OSmaximus and makes sense due to the flipping from reverse to forward here.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 67**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 57008**, Stop:** 57811(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF is not commonly called, not called by Glimmer or Genemark, 764bp gap with previous gene – occurs at change in direction (final score: - 5.590, z-value: 1.589, spacer distance: 10). Strong coding potential in +2 frame from ~57100– 57770, most manually called start in Starterator (called 55.6%% of the time when present) 144MA, 1:1 target:query for 0 hits in BLAST, conserved synteny with 4 of most similar phage. Three closest phage relatives also call the manually called start in Starterator. **Function**: Function supported by BLAST with hypothetical protein called for top hits, called by 474/475 of phages in the same pham (229818), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 68**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 57946**, Stop:** 58125(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Helix-turn-helix DNA binding domain

 **(d) Final Summary: START**: Not longest ORF, LORF is not called by similar phage, same call by Glimmer and Genemark, 134bp gap with previous gene (final score: - 3.810, z-value: 2.473, spacer distance: 10). Strong coding potential in +3 frame from ~57940– 58070, most manually called start in Starterator (called 88.9%% of the time when present) 32MA, 1:1 target:query for 2 of top genes, conserved synteny with 1 of most similar phage. **Function**: HHPRED supports function of ribbon-ribon helix transcriptional repressor with high scores, Function supported by BLAST with calls of hypothetical protein and ribbon-ribbon helix DNA binding domain called for top hits, called by 34/78 of phages in the same pham (230102), ribbon-ribbon helix DNA binding domain called by 1 of closest whole genome relatives.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 69**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 58122**, Stop:** 58304(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 175bp overlap, other longer ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 4.874, z-value: 2.207, spacer distance: 8). Strong coding potential in +1 frame from ~58120– 58270, most manually called start in Starterator (called 79.7%% of the time when present) 207MA, 1:1 target:query for 3 top genes and many of top hits, no synteny with 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, called by 277/277 of phages in the same pham (84796), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 70**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 58297**, Stop:** 58512(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 17bp overlap, same call by Glimmer and Genemark, 8bp overlap with previous gene (final score: - 4.874, z-value: 2.207, spacer distance: 8). Strong coding potential in +1 frame from ~58300– 58450, most manually called start in Starterator (called 87.1% of the time when present) 57MA, 1:1 target:query for top gene hits, synteny with 4 of most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, called by 131/132 of phages in the same pham (219516), called by 4 of closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 71**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 58509**, Stop:** 59207(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 4.426, z-value: 2.491, spacer distance: 14). Strong coding potential in +3 frame from ~58500– 5900 and unstable coding potential till 59200, most manually called start in Starterator (called 100% of the time when present) 267MA, 1:1 target:query for top 20 gene hits, synteny with 4 of most similar phage. **Function**: Function supported by BLAST with calls of RNASE E, hypothetical protein, and DNA binding protein for top hits, RNASE E called by 18/482 of phages in the same pham (229816), variations of DNA binding domain called by 179/482 of phages in the same pham (229816), remainder called as hypothetical protein, RNASE E called by 1, DNA binding domain called by 1, hypothetical protein called by 8 of closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 72**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 59200**, Stop:** 59784(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 233bp overlap, other longer ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 8bp overlap with previous gene (final score: - 6.182, z-value: 1.557, spacer distance: 8). Strong coding potential in +3 frame from ~59200– 59750, most manually called start in Starterator (called 77.9% of the time when present) 208MA, 1:1 target:query for top gene hits, synteny with 9 of most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 293/294 of phages in the same pham (229863), called by 10 of closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 73**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 59781**, Stop:** 60179(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 3.915 z-value: 2.491, spacer distance: 12). Strong coding potential in +1 frame from ~59790– 60160, most manually called start in Starterator (called 100% of the time when present) 265MA, 1:1 target:query for top 5 gene hits, no synteny with most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 288/288 of phages in the same pham (84781), called by 9 of closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 74**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 60220**, Stop:** 60348(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 116bp overlap, other longer ORFs have similarly unacceptable overlaps, called by Glimmer, 40bp gap with previous gene (final score: - 6.780 z-value: 1.26, spacer distance: 8). Strong coding potential in +3 frame from ~60200– 60300, most manually called start in Starterator (called 100% of the time when present) 47MA, 1:1 target:query for 3 of top gene hits, no synteny with most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 66/66 of phages in the same pham (85754), not called by of closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 75**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 60383**, Stop:** 60727(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 53bp overlap, called by Glimmer, 34bp gap with previous gene (final score: - 7.668, z-value: 0.627, spacer distance: 12). Strong coding potential in +2 frame from ~60380– 60700, most manually called start in Starterator (called 87.9% of the time when present) 234MA, 1:1 target:query for top 12 gene hits, no synteny with most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 293/293 of phages in the same pham (223), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 76**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 60729**, Stop:** 61331(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 1bp gap with previous gene (final score: -6.893, z-value: 1.489, spacer distance: 16). Strong coding potential in +1 frame from ~60750– 61200, dip in coding potential at 60830, most manually called start in Starterator (called 100% of the time when present) 267MA, 1:1 target:query for top 18 gene hits, synteny with 8 of most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for 12 top hits, hypothetical protein called by 383/491 of phages in the same pham (224151), called by 8 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 77**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 61390**, Stop:** 61626(Forward)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 35bp overlap, same call by Glimmer and Genemark, 58bp gap with previous gene (final score: - 6.150, z-value: 1.573, spacer distance: 8). Strong coding potential in +3 frame from ~61400– 61600, most manually called start in Starterator (called 99.6% of the time when present) 254MA, 1:1 target:query for top 10 gene hits, synteny with 8 of most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 278/278 of phages in the same pham (225810), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 78**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 61910**, Stop:** 61701(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, called by Glimmer, 58bp gap with previous gene (final score: - 4.731, z-value: 2.047, spacer distance: 11). Strong coding potential in -2 frame from ~61700– 61900, most manually called start in Starterator (called 84.8% of the time when present) 225MA, 1:1 target:query for top 9 gene hits and many of top hits, synteny with 8 of most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 289/289 of phages in the same pham (235), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 79**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 62410**, Stop:** 61991(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, called by Glimmer, 35bp gap with previous gene (final score: - 4.845, z-value: 1.99, spacer distance: 11). Strong coding potential in -1 frame from ~62030– 62370, most manually called start in Starterator (called 85.8% of the time when present) 210MA, 1:1 target:query for top 23 gene hits and many of top hits, synteny with 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 291/291 of phages in the same pham (226), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 80**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 62589**, Stop:** 62446(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** hypothetical protein

 **(d) Final Summary: START**: LORF (only start), same call by Glimmer and Genemark, 54bp gap with previous gene (final score: - 4.669, z-value: 2.047, spacer distance: 10). Strong coding potential in –3 frame from ~62440– 62600, most manually called start in Starterator (called 100% of the time when present) 267MA, 1:1 target:query for top 7 gene hits, synteny with 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 353/353 of phages in the same pham (84733), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 81**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 62907**, Stop:** 62644 (Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: not longest ORF, LORF has 424bp overlap, other longer ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 6.181, z-value: 1.335, spacer distance: 9). Strong coding potential in –3 frame from ~62750– 62880, most manually called start in Starterator (called 100% of the time when present) 260MA, 1:1 target:query for top 12 gene hits and many of top hits, synteny with 2 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 289/289 of phages in the same pham (84777), called by 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 82**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 63239**, Stop:** 62904(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: not longest ORF, LORF has 169bp overlap, other longer ORFs have similarly unacceptable overlaps, same call by Glimmer and Genemark, 1bp overlap with previous gene (final score: - 4.905, z-value: 1.969, spacer distance: 9). Strong coding potential in –2 frame from ~62920– 63220, most manually called start in Starterator (called 100% of the time when present) 42MA, 1:1 target:query for top 3 gene hits and many of top hits, synteny with 9 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 46/46 of phages in the same pham (2153), called by 2 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 83**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 63469**, Stop:** 63239(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 8bp overlap with previous gene (final score: - 4.198, z-value: 2.543, spacer distance: 8). Strong coding potential in –1 frame from ~63270– 63450, most manually called start in Starterator (called 98.6% of the time when present) 262MA, 1:1 target:query for top gene hits, synteny with 2 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 289/289 of phages in the same pham (233), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 84**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 63848**, Stop:** 63462(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: Not longest ORF, LORF has 113bp overlap, other longer ORFs are not generally called, same call by Glimmer and Genemark, 43bp gap with previous gene (final score: - 4.846, z-value: 2.506, spacer distance: 16). Strong coding potential in –2 frame from ~63500– 63850, most manually called start in Starterator (called 85.5% of the time when present) 218MA, 1:1 target:query for many of top gene hits, synteny with 7 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 285/285 of phages in the same pham (227200), called by 7 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 85**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 64194**, Stop:** 63892(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** HNH endonuclease

 **(d) Final Summary: START**: Not longest ORF, LORF has 85bp overlap, same call by Glimmer and Genemark, 4bp overlap with previous gene (final score: - 7.384, z-value: 0.729, spacer distance: 11). Strong coding potential in –3 frame from ~63940– 64160, most manually called start in Starterator (called 87.2% of the time when present) 230MA, 1:1 target:query for many of top gene hits, synteny with 3 of 10 most similar phage. **Function**: Function supported by BLAST with call of HNH endonuclease for top hits, HNH endonuclease called by 315/413 of phages in the same pham (228188), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value, top hits for endonuclease.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 86**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:**64301**, Stop:** 64191(Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF (only start) , called by Genemark, 4bp overlap with previous gene (final score: - 4.696, z-value: 2.103, spacer distance: 12), minimal coding potential in -2 frame, most manually called start in Starterator (called 100% of the time when present) 208MA, 1:1 target:query for many of top gene hits, synteny with 6 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 225/225 of phages in the same pham (84865), called by 6 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 87**

**Final Gene Information:**
 **(a) FINAL Gene Co-ordinates: Start:** 64477 **Stop:** 64298 (Reverse)
 **(b) Is it a protein-coding gene?** Yes.
 **(c) What is its function?** Hypothetical protein

 **(d) Final Summary: START**: LORF, same call by Glimmer and Genemark, 12bp gap with previous gene (final score: - 3.258, z-value: 2.818, spacer distance: 12), strong coding potential in –1 frame from ~64350-64450 , most manually called start in Starterator (called 100% of the time when present) 266MA, 1:1 target:query for 8 of top gene hits, synteny with 4 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 290/290 of phages in the same pham (192283), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Gene 88**

**1a. FINAL Gene Co-ordinates:** Start: 64750, Stop: 64490 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Not longest ORF—longest ORF overlaps previous gene over 200 bp—called ORF is the first ORF without significant overlap (4 bp overlap), same call in Glimmer and Genemark (final score: -4.257, z-value: 2.251, spacer distance: 10). Strong coding potential in -1 frame from ∼64550–64730 bp, 1:1 target:query for 2 of top 10 genes, most manual called start in Starterator (called 84.4% of the time when present), conserved synteny. **FUNCTION:** Hypothetical protein function called by top 20 genes in BLAST, called by 99.6% (256 out of 257) of phages in same pham (225864), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 89**

**1a. FINAL Gene Co-ordinates:** Start: 64944, Stop: 64747 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Longest ORF, same call by Glimmer (final score: -5.605, z-value: 2.103, spacer distance: 6), 4 bp overlap, most manual called start in Starterator (called 100% of time when present), 3 closest whole-genome relative also call most called start in Starterator for this gene. Moderate coding potential in -3 frame from ~64850-64900 bp, 1:1 target:query for 4 of 4 total genes in BLAST, conserved synteny. No start called by Genemark. **Function:** Hypothetical protein called by 4 total gene hits in BLAST (>5 phages per hit), called by 100% (255 out of 255) of phages in same pham (302), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 90**

**1a. FINAL Gene Co-ordinates:** Start: 65165, Stop: 64941 (Reverse)
**1b. Is it a protein-coding gene*?*** Yes
**1c. What is its function?** Hypothetical protein
**1d. Final Summary: Start:** Not longest ORF—longest ORF overlaps previous gene by 34 bp—called ORF is the first ORF without significant overlap (7 bp overlap), same call in Glimmer and Genemark (final score: -5.352, z-value: 1.969, spacer distance: 8). Strong coding potential in the -2 frame from ~64970-65160 bp, 1:1 target:query for 9 of top 10 gene hits in BLAST, most manual called start in Starterator (called 62.8% of time when present), conserved synteny. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (286 out of 286) of phages in same pham (84783), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 91**

**1a. FINAL Gene Co-ordinates:** Start: 65344, Stop: 65159 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Not longest ORF—longest ORF overlaps previous gene by 67 bp—called ORF is the first ORF without significant overlap (4 bp overlap), same call in Glimmer and Genemark (final score: -5.526, z-value: 1.691, spacer distance: 12). Strong coding potential in the -1 frame from ~65200-65350 bp, 1:1 target:query for top 10 gene hits in BLAST, most manual called start in Starterator (called 99.7% of time when present), conserved synteny. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (290 out of 290) of phages in same pham (84771), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 92**

**1a. FINAL Gene Co-ordinates:** Start: 65562, Stop: 65341 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Longest ORF, same call in Glimmer and Genemark (final score: -4.101, z-value: 2.36, spacer distance: 11). Strong coding potential in the -3 frame from ~65370-65550, 1:1 target:query for top 15 gene hits in BLAST, most manual called start in Starterator (called 96.3% of time when present), conserved synteny. 120 bp gap with previous gene, ORF present within the gap in the -3 frame however there is no coding potential within the gap and none of 3 closest phage relatives call a gene in the gap. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (297 out of 297) of phages in same pham (196245), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 93**

**1a. FINAL Gene Co-ordinates:** Start: 65985, Stop: 65683 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Not longest ORF—longest ORF overlaps previous gene by over 200 bp—called ORF is the first ORF without significant overlap (29 bp gap), same call by Glimmer and Genemark (final score: -4.137. z-value: 2.573, spacer distance: 8). Strong coding potential in the -3 frame from ~65700-65950 bp, 1:1 target:query for top 20 gene hits in BLAST, conserved synteny. ElvisPhasley does not have the most annotated start in Starterator for this gene, chosen start was called 100% of time when present and 2 of 3 closest whole-genome relatives also did not have the most called start and made same call. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (353 out of 363) of phages in same pham (84730), called by three closest whole-genome relatives. Function not called by HHPred (high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 94**

**1a. FINAL Gene Co-ordinates:** Start: 66212, Stop: 66015 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Longest ORF, same call by Glimmer and Genemark (final score: -3.110, z-value: 3.083, spacer distance: 8), most manual called start in Starterator (called 100% of time when present). Strong coding potential in the -2 frame from ~66030-66190 bp, 1:1 target:query for top 20 gene hits in BLAST, conserved synteny. 75 bp gap with previous gene, no ORFs with in gap and no close relatives call a gene within the gap. **Function:** Hypothetical protein called by top 20 genes in BLAST, called by 100% (339 out of 339) of phages in same pham (197781), called by three closest whole-genome relatives. Function not called by HHPred (high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 95**

**1a. FINAL Gene Co-ordinates:** Start: 66521, Stop: 66288 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** Longest ORF, same call by Glimmer and Genemark (final score: -3.331, z-value: 2.973, spacer distance: 8), most manual called start in Starterator (called 100% of time when present). Strong coding potential in -2 frame from ~66360-66500 bp, 1:1 target:query for 8 of 8 total gene hits in BLAST, conserved synteny. 36 bp gap, no ORFs or coding potential within gap and closest whole-genome relatives do not call a gene within the gap. **Function:** Hypothetical protein called by all 8 total gene hits in BLAST, called by 100% (26 out of 26) of phages in same pham (3656), called by two closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 96**

**1a. FINAL Gene Co-ordinates:** Start: 66914, Stop: 66558 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Not longest ORF, but is same call by Glimmer and Genemark (final score: -2.297, z-value: 3.295, spacer distance: 12), most manual called start in Starterator (called 73.4% of time when present). Strong coding potential in -2 frame from ~66570-66900 bp, 1:1 target:query for 6 of top 10 gene hits in BLAST, conserved synteny. 162 bp gap with previous gene, ORF present in -1 frame however no coding potential and none of three closest whole-genome relatives call a gene within their similarly sized gap. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (290 out of 290) of phages in same pham (230), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 97**

**1a. FINAL Gene Co-ordinates:** Start: 67655, Stop: 67077 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** Longest ORF, same call by Glimmer (final score: -8.958, z-value: -0.053, spacer distance: 11), 43 bp gap, ElvisPhasley does not have the most annotated start in Starterator for this gene (three closest whole-genome relatives do have the most annotated start). Strong coding potential in the -2 frame from ~67130-67650 bp, 1:1 target:query for 2 of top 10 gene hits in BLAST (Genemark called start @67616 which has zero 1:1 target:query), conserved synteny. **Function:** Hypothetical protein called by top 20 gene hits in BLAST, called by 100% (290 out of 290) of phages in same pham (229867), called by three closest whole-genome relatives. Function not called by HHpred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 98**

**1a. FINAL Gene Co-ordinates:** Start: 67878, Stop: 67699 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** Longest ORF, same call by Glimmer and Genemark (final score: -5.972, z-value: 1.429, spacer distance: 9), most manual called start in Starterator (called 100% of time when present). Strong coding potential in the -3 frame from ~67750-67860 bp, 1:1 target:query for 8 of top 10 gene hits in BLAST, conserved synteny. **Function:** Hypothetical protein called by 7 of top 10 gene hits in BLAST, called by 100% (289 out of 289) of phages in same pham (224206), called by three closest whole-genome relatives. Function not called by HHPred (low coverage and high e-values).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 99**

**1a. FINAL Gene Co-ordinates:** Start: 68111, Stop: 67875 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: START:** not LORF, LORF is not called by closest relatives, same call by Glimmer and Genemark, 53bp gap with previous gene (final score: - 3.222, z-value: 3.216, spacer distance: 15), strong coding potential in –2 frame from ~67920-68110, most manually called start in Starterator (called 85.9% of the time when present) 210MA, 1:1 target:query for many of top gene hits, synteny with 10 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 325/325 of phages in the same pham (195), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 100**

**1a. FINAL Gene Co-ordinates:** Start: 68566, Stop: 68165 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** not LORF, LORF has 34bp overlap, same call by Glimmer and Genemark, 0bp gap with previous gene (final score: - 3.296, z-value: 2.799, spacer distance: 12), strong coding potential in –1 frame from ~68210-68560, most manually called start in Starterator (called 99.7% of the time when present) 281MA, 1:1 target:query for 30 of top gene hits, synteny with 10 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 360/360 of phages in the same pham (170), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 101**

**1a. FINAL Gene Co-ordinates:** Start: 68901, Stop: 68566 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** not LORF, LORF has 23bp overlap, longer ORF available with 19bp gap is not called by similar relatvies, same call by Glimmer and Genemark, 70bp gap with previous gene (final score: - 3.331, z-value: 2.973, spacer distance: 8), strong coding potential in –3 frame from ~68600-68940, most manually called start in Starterator (called 90% of the time when present) 239MA, 1:1 target:query for many of top gene hits, synteny with 2 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 294/294 of phages in the same pham (229864), called by 10 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Gene 102**

**1a. FINAL Gene Co-ordinates:** Start: 69394, Stop: 68972 (Reverse)

**1b. Is it a protein-coding gene*?*** Yes

**1c. What is its function?** Hypothetical protein

**1d. Final Summary: Start:** not LORF, same call by Glimmer and Genemark, 0bp gap with previous gene (final score: - 3.307, z-value: 3.135, spacer distance: 7), strong coding potential in –1 frame from ~68900-69400 , most manually called start in Starterator (called 79.3% of the time when present) 41MA, 1:1 target:query for 4 of top gene hits, synteny with 10 of 10 most similar phage. **Function**: Function supported by BLAST with call of hypothetical protein for top hits, hypothetical protein called by 58/58 of phages in the same pham (2046), called by 2 of 10 closest whole genome relatives. HHpred hits show low coverage and high e-value.