**Bachaco (EB) Annotation Notes**

**GENE 1** **hypothetical protein**

Original Glimmer call @bp 397 has strength 15.13; GeneMark calls start at 391

SSC: 397 to 612 (Forward)

CP: yes

SCS: both – GL [all data suggests 397 as the start]

ST: SS, Start 24[397], Found in 38 of 46 ( 82.6% ) of genes in pham, Called 97.4% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_1, function unknown, query 1-71 subject 1-71, E value = 5e-38, Identity = 100%, Similarity = 100%

GAP: NA

LO: NA

RBS: Kibler7, spacer 10, Z score 3.022, Final -2.443, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752263.1, query 1-71 subject 1-71, E value = 2e-44, Identity = 99%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 2 terminase small subunit**

Original Glimmer call @bp 612 has strength 15.63

SSC: 612 to 935 (Forward)

CP: yes

SCS: both

ST: NI, Start 14 [612], Found in 57 of 70 ( 81.4% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_2, terminase small subunit, query 1-107 subject 1-107, E value = 1e-51, Identity = 98%, Similarity = 100%

GAP: TGATG OVERLAP

LO: YES

RBS: Kibler7, spacer 15, Z score 2.016, Final -5.469, yes

F: terminase small subunit

SIF-blast: aligns with terminase small subunit from Microbacterium phage FlameThrower, ID WNO28717.1, query 1-107 subject 1-107, E value = 1e-65, Identity = 98%, Similarity = 99%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: This gene is pham 51743 [terminase small subunit], upstream is pham 85825, downstream is pham 123985 [terminase large subunit], as seen in Microbacterium phage Flame Thrower

**GENE 3 terminase large subunit**

Original Glimmer call @bp 938 has strength 10.57

SSC: 938 to 2503 (Forward)

CP: yes

SCS: both

ST: NI, Start 112 [938], Found in 94 of 1211 ( 7.8% ) of genes in pham, Called 62.8% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_3, terminase large subunit, query 1-521 subject 1-521, E value = 0.0, Identity = 99%, Similarity = 100%

GAP: 3bp

LO: yes

RBS: Kibler7, spacer 12, Z score 2.441, Final -3.808, yes

F: terminase large subunit

SIF-blast: aligns with terminase large subunit from Microbacterium phage FlameThrower, ID WNO28718.1, query 1-521 subject 1-521, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: 6Z6D\_A, Terminase large subunit, genome packaging, bacteriophage, ATPase, nuclease, HET, 2.2A, Enterobacteria phage HK97, Probability: 100%, E-value: 3.1e-38,

SF-Syn: This gene is pham 123985 [terminase large subunit], upstream is pham 51743 [terminase small subunit], downstream is pham 127490 [portal protein], as seen in Microbacterium phage Flame Thrower

**GENE 4 portal protein**

Original Glimmer call @bp 2515 has strength 11.36

SSC: 2515 to 4530 (Forward)

CP: yes

SCS: both

ST: NI, Start 3 [2515], Found in 58 of 176 ( 33.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_4, portal protein, query 1-671 subject 1-671, E value = 0.0, Identity = 99%, Similarity = 99%

GAP: 12bp

LO: yes

RBS: Kibler7, spacer 9, Z score 1.940, Final -4.801, yes

F: portal protein

SIF-blast: aligns with portal protein from Microbacterium phage Celaena, ID YP\_010752266.1, query 1-671 subject 1-671, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: 8FQL\_L Portal protein, Prohead I, icosahedral symmetry, HK97, phage, capsid, 3.6A, Escherichia phage HK97, Probability: 100%, E-value: 1.2e-36

SF-Syn: This gene is pham 127490 [portal protein], upstream is pham 123985 [terminase large subunit], downstream is pham 106150 [capsid maturation protease], as seen in Microbacterium phage Flame Thrower

**GENE 5 capsid maturation protease**

 Original Glimmer call @bp 4538 has strength 14.55

SSC: 4538 to 5275 (Forward)

CP: yes

SCS: both

ST: NI, Start 44 [4538], Found in 190 of 250 ( 76.0% ) of genes in pham, Called 94.7% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_5, capsid maturation protease, query 1-245 subject 1-245, E value = e-134, Identity = 100%, Similarity = 100%

GAP: 8bp

LO: yes

RBS: Kibler7, spacer 8, Z score 2.128, Final -4.853, yes

F: capsid maturation protease

SIF-blast: aligns with capsid maturation protease from Microbacterium phage Celaena, ID YP\_010752267.1, query 1-245 subject 1-245, E value = 3e-176, Identity = 100%, Similarity = 100%

SIF-HHPred: 5JBL\_B, Prohead core protein protease, protease pentamer, phage T4, prohead, HYDROLASE, 1.943A, Enterobacteria phage T4, Probability: 97.78%, E-value: 0.0097

SF-Syn: This gene is pham 106150 [capsid maturation protease], upstream is pham 127490 [portal protein], downstream is pham 104749 [major capsid protein], as seen in Microbacterium phage Flame Thrower

**GENE 6 major capsid protein**

Original Glimmer call @bp 5305 has strength 15.62

SSC: 5305 to 6501 (Forward)

CP: yes

SCS: both

ST: NI, Start 36 [5305], Found in 156 of 265 ( 58.9% ) of genes in pham, Called 94.9% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_6, major capsid protein, query 1-398 subject 1-398, E value = 0.0, Identity = 98%, Similarity = 99%

GAP: 30bp

LO: yes

RBS: Kibler7, spacer 9, Z score 3.255, Final -2.034, yes

F: major capsid protein

SIF-blast: aligns with major capsid protein from Microbacterium phage Celaena, ID YP\_010752268.1, query 1-398 subject 1-398, E value = 0.0, Identity = 98%, Similarity = 99%

SIF-HHPred: 8FXR\_n1, Major capsid protein, gp9, Myophage, 4.5A, Agrobacterium phage Milano, Probability: 99.98%, E-value: 8.9e-30,

SF-Syn: This gene is pham 104749 [major capsid protein], upstream is pham 106150 [capsid maturation protease], downstream is pham 85888, as seen in Microbacterium phage Celaena

**GENE 7 hypothetical protein**

Original Glimmer call @bp 6574 has strength 13.41

SSC: 6574 to 6873 (Forward)

CP: yes

SCS: both

ST: SS, Start 4 [6574], Found in 44 of 44 ( 100.0% ) of genes in pham, Called 88.6% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_7, function unknown, query 1-99 subject 1-99, E value = 2e-52, Identity = 100%, Similarity = 100%

GAP: 73bp

LO: yes

RBS: Kibler7, spacer 9, Z score 2.570, Final -3.474, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Katzastrophic, ID YP\_010752333.1, query 1-99 subject 1-99, E value = 6e-63, Identity = 99%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 8 head-to-tail adaptor**

Original Glimmer call @bp 6906 has strength 11.12

SSC: 6906 to 7457 (Forward)

CP: yes

SCS: both

ST: NI, Start 22 [6906], Found in 58 of 369 ( 15.7% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_8, head-to-tail adaptor, query 1-183 subject 1-183, E value = e-101, Identity = 100%, Similarity = 100%

GAP: 33bp

LO: yes

RBS: Kibler7, spacer 12, Z score 3.022, Final -2.584, yes

F: head-to-tail adaptor

SIF-blast: aligns with head-to-tail adaptor from Microbacterium phage Celaena, ID YP\_010752270.1, query 1-183 subject 1-183, E value = 2e-128, Identity = 99%, Similarity = 99%

SIF-HHPred: 8HQO\_S, Head completion protein, Neck, Portal, T5, VIRUS, VIRAL PROTEIN, 3.2A, Escherichia phage DT57C, Probability: 99.95%, E-value: 1e-25

SF-Syn: This gene is pham 127388 [head-to-tail adaptor], upstream is pham 85888, downstream is pham 127435 [head-to-tail stopper], as seen in Microbacterium phage Celaena

**GENE 9 head-to-tail stopper**

Original Glimmer call @bp 7457 has strength 16.35

SSC: 7457 to 7786 (Forward)

CP: yes

SCS: both

ST: SS, Start 37 [7457], Found in 196 of 241 ( 81.3% ) of genes in pham, Called 99.5% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_9, head-to-tail stopper, query 1-109 subject 1-109, E value = 3e-60, Identity = 100%, Similarity = 100%

GAP: TGATG overlap

LO: yes

RBS: Kibler7, spacer 9, Z score 3.022, Final -3.493, yes

F: head-to-tail stopper

SIF-blast: aligns with head-to-tail stopper from Microbacterium phage Celaena, ID YP\_010752271.1, query 1-109 subject 1-109, E value = 8e-74, Identity = 100%, Similarity = 100%

SIF-HHPred: 7Z4W\_5, Head completion protein gp16, Bacteriophage, SPP1, Portal Protein, Head completion proteins, Connector Complex, DNA Channel, VIRAL PROTEIN, 2.7A, Bacillus subtilis, Probability: 99.48%, E-value: 8.5e-13

SF-Syn: This gene is pham 127435 [head-to-tail stopper], upstream is pham 127388 [head-to-tail adaptor], downstream is pham 127444, as seen in Microbacterium phage Celaena

**GENE 10 hypothetical protein**

Original Glimmer call @bp 7813 has strength 10.05; GeneMark calls start at 7786

SSC: 7786 to 8049 (Forward)

CP: yes

SCS: both – GM [Starterator and RBS suggest 7786; 7786 allows for better alignment in BLAST]

ST: NI: Start 27 [7786], Found in 191 of 227 ( 84.1% ) of genes in pham, Called 97.4% of time when present

BLAST-start: aligns with Microbacterium phage FlameThrower\_10, function unknown, query 1-78 subject 10-87, E value = 3e-37, Identity = 98%, Similarity = 100%

GAP: TGATG overlap

LO: NA

RBS: Kibler7, spacer 10, Z score 1.983, Final -4.630, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Katzastrophic, ID YP\_010752336.1, query 1-78 subject 10-87, E value = 9e-47, Identity = 99%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 11 Tail terminator**

Original Glimmer call @bp 8051 has strength 13.66

SSC: 8051 to 8449 (Forward)

CP: yes

SCS: both

ST: NI, Start 9 [8051], Found in 13 of 58 ( 22.4% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_11, Tail terminator, query 1-132 subject 1-132, E value = 1e-67, Identity = 100%, Similarity = 100%

GAP: 2bp

LO: NA

RBS: Kibler7, spacer 12, Z score 2.769, Final -3.116, yes

F: Tail terminator

SIF-blast: aligns with tail terminator from Microbacterium phage Celaena, ID QDH92390.1, query 1-132 subject 1-132, E value = 5e-88, Identity = 100%, Similarity = 100%

SIF-HHPred: 6TE9\_F, Tail terminator protein, Rcc01690, 3.58A, Rhodobacter capsulatus, Probability: 99.48%, E-value: 2e-12

SF-Syn: Tail terminator, upstream gene is head to tail stopper (pham 6861), downstream gene is major tail protein (pham 68574), as seen in phage Katzastrophic

**GENE 12 Major tail protein**

Original Glimmer call @bp 8464 has strength 20.68

SSC: 8464 to 9054 (Forward)

CP: yes

SCS: both

ST: NI, Start 45 [8464], Found in 96 of 509 ( 18.9% ) of genes in pham, Called 87.5% of time when present

BLAST-start: aligns with Microbacterium phage Quenya\_13, major tail protein, query 1-196 subject 1-196, E value = e-107, Identity = 97%, Similarity = 99%

GAP: 15bp

LO: YES

RBS: Kibler7, spacer 12, Z score 3.255, Final -2.095, yes

F: Major tail protein

SIF-blast: aligns with major tail protein from Microbacterium phage Quenya, ID QOP64248.1, query 1-196 subject 1-196, E value = 1e-136, Identity = 98%, Similarity = 99%

SIF-HHPred: 6XGR\_M, Major tail protein, YSD1\_22, Bacteriophage tail, helical assembly, VIRAL PROTEIN; 3.5A, Bacteriophage sp., Probability 98.64%, E-value 0.0000024

SF-Syn: Major tail protein, upstream gene is tail terminator (pham 68679), downstream gene is tail assembly chaperone (pham 64535), as seen in phage Quenya (EB)

**GENE 13 tail assembly chaperone**

Original Glimmer call @bp 9138 has strength 14.71

SSC: 9138 to 9446 (Forward)

CP: yes

SCS: both

ST: SS, Start 7 [9138], Found in 76 of 90 ( 84.4% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_13, tail assembly chaperone, query 1-102 subject 1-102, E value = 2e-51, Identity = 100%, Similarity = 100%

GAP: 84bp

LO: yes

RBS: Kibler7, spacer 10, Z score 1.949, Final -4.701, no

F: tail assembly chaperone

SIF-blast: aligns with tail assembly chaperone from Microbacterium phage Celaena, ID YP\_010752276.1 , query 1-102 subject 1-102, E value = 3e-67, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: This gene is pham 126716 [tail assembly chaperone], upstream is pham 126528 [major tail protein], downstream is pham 2167 [tape measure protein], as seen in Microbacterium phage Celaena

**GENE 14 tail assembly chaperone**

frameshift in conjunction with gene 13 above

Katzastrophic Gene 13

MTFKVPESKRSIRQNQFEFQVPGDRKTYRIPKAKYLSIGQVEALASKGDEVQITDILEILGQGEAREAVRTLDQEQLMALMEAWQDDSGITVGESSASTETS

Katzastrophic Gene 14

MTFKVPESKRSIRQNQFEFQVPGDRKTYRIPKAKYLSIGQVEALASKGDEVQITDILEILGQGEAREAVRTLDQEQLMALMEAWQDDSGITVGGILGLHRDLLSKGPARQALQYDLLVRGLSLEDLGEAFTWYDLLAFSKHVQQETNSALARELHGPAWSVEGQLMAIVADHLAVANWQRAGRKSAPKPQRIPRPWEKPKTTVLGKGAIPLAAFADWWDSRKSKRSRKRAKKPPTD

Bachaco\_Draft gene 13 (9138 - 9446 ) – frame 3

MTFKVPESKRSIRQNQFEFQVPGDRKTYRIPKAKYLSIGQVEALASKGDEVQITDILEILGQGEAREAVRTLDQEQLMALMEAWQDDSGITVGESSASTETS

CGGTATCACCGTGGGG = 9416

Bachaco\_Draft gene 14 (9416 - 9847 )

MTFKVPESKRSIRQNQFEFQVPGDRKTYRIPKAKYLSIGQVEALASKGDEVQITDILEILGQGEAREAVRTLDQEQLMALMEAWQDDSGITVG

GILGLHRDLLSKGPARQALQYDLLVRGLSLEDLGEAFTWYDLLAFSKHVQQETNSALARELHGPAWSVEGQLMAIVADHLAVANWQRAGRKSAPKPQRIPRPWEKPKTTVLGKGAIPLAAFADWWDSRKSKRSRKRAKKPPTD

G at 9416 used for GGG [G] in frame 3 and then for GGA [G] in frame 2

section 1 = 9138 to 9416

section 2 – 9416 to 9847

**GENE 15 Tape measure protein**

Original Glimmer call @bp 9893 has strength 13.52

SSC: 9893 to 12670 (Forward)

CP: yes

SCS: both

ST: SS, Start 3 [9893], Found in 48 of 48 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_15, tape measure protein, query 1-925 subject 1-925, E value = 0.0, Identity = 99%, Similarity = 99%

GAP: 46bp

LO: yes

RBS: Kibler7, spacer 12, Z score 2.990, Final -2.812, yes

F: Tape measure protein

SIF-blast: aligns with tape measure protein from Microbacterium phage FlameThrower, ID WNO28730.1, query 1-925 subject 1-925, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: Tape measure protein, gp57, VIRAL PROTEIN; 3.7A, Staphylococcus virus 80 alpha, 6V8I\_CF, probability 99.57%, E-value 4.3e-8

SF-Syn: Tape measure protein, upstream gene is tail assembly chaperone (pham 68675), downstream gene is minor tail protein (pham 68631), as seen in phage Celaena

SIF-Mem: 2 domains found by TMHMM; 5 domains found by SOSUI.

**GENE 16 Minor tail protein**

Original Glimmer call @bp 12667 has strength 15.08

SSC: 12667 to 13521 (Forward)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 34 – nt12667 – Found in 82 of 273 (30.0%) – of genes in pham – Called 67.1% of the time when present.

BLAST-start: aligns with Microbacterium phage Celaena\_16, Minor tail protein, query 1-284 subject 1-284, E value = e-165, Identity = 99%, Similarity = 99%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 1.815, Final -5.305, no

F: Minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage Celaena, ID QDH92395.1, query 1-284 subject 1-284, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: Distal tail protein, gp58; phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A, {Staphylococcus virus 80alpha} 6V8I\_BD, probability 99.95%, E-value 6.1e-25

SF-Syn: Minor tail protein, upstream gene is tape measure protein (pham 2167), downstram gene is minor tail protein (pham 69307), just like in phage Celaena (EB)

**GENE 17 Minor tail protein**

Original Glimmer call @bp 13521 has strength 14.37

SSC: 13521 to 16133 (Forward)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 4 – nt13521 – Found in 4 of 12 (33.3%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Celaena\_17, minor tail protein, query 1-870 subject 1-872, E value = 0.0, Identity = 81%, Similarity = 86%

GAP: TGATG Overlap

LO: NA

RBS: Kibler7, spacer 16, Z score 1.991, Final -5.108, no

F: Minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage Celaena, ID QDH92396.1, query 1-870 subject 1-872, E value = 0.0, Identity = 81%, Similarity = 86%

SIF-HHPred: Tail protein, 43 kDa; PSI, MCSG, 3D37\_A, probability 98.93%, E-value 0.0000019

SF-Syn: Minor tail protein, upstream gene is minor tail protein (pham 68631), downstream gene is NKF (pham 65724), as seen in phage Celaena (EB)

**GENE 18 minor tail protein**

Original Glimmer call @bp 16153 has strength 14.23

SSC: 16153 to 16986 (Forward)

CP: yes

SCS: both

ST: SS, The start site called the most often in the published annotations is 13 – nt16153 – Found in 115 of 119 (96%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_18, minor tail protein, query 1-276 subject 1-276, E value = e-144, Identity = 89%, Similarity = 94%

GAP: 20 bp

LO: yes

RBS: Kibler7, spacer 11, Z score 3.061, Final -2.615, yes

F: minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage BAjuniper, ID: WIC89443.1, query 1-276 subject 1-276, E value = 7e-163, Identity = 84%, Similarity = 90%

SIF-HHPred: 4L9B\_A Receptor Binding Protein, beta sandwich domain, phage receptor binding protein, Lactococcus lactis pellicle cell wall polyphosphosaccharide, VIRAL PROTEIN; 1.75A, Lactococcus phage 1358, Probability: 99.72%, E-value: 1.1e-16,

SF-Syn: This gene is pham 124234 [minor tail protein], upstream is pham 128329 [minor tail protein], downstream is pham 126638 [minor tail protein], as seen in Microbacterium phage Katzastrophic

**GENE 19 Hypothetical protein**

Original Glimmer call @bp 16998 has strength 10.56

SSC: 16998 to 17840 (Forward)

CP: yes

SCS: both

ST: SS, Start 7 [16998], Found in 83 of 92 ( 90.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena, function unknown, query 1-280 subject 1-280, E value = e-170, Identity = 97%, Similarity = 99%

GAP: 12bp

LO: yes

RBS: Kibler7, spacer 11, Z score 3.266, Final -1.993, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Sea\_Celaena\_19, ID QDH92398.1, query 1-280 subject 1-280, E value = 0.0, Identity = 98%, Similarity = 99%

SIF-HHPred: NKF, there are partial matches with above 90% probability but they are homo sapien adhesion or receptor proteins

SF-Syn: NKF

**GENE 20 Endolysin**

Original Glimmer call @bp 17854 has strength 12.51

SSC: 17854 to 19551 (Forward)

CP: yes

SCS: both

ST: SS, The start site called the most often in the published annotations is 7 – nt17854 – Found in 18 of 18 (100%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic, Endolysin, query 1-565 subject 1-565, E value = 0.0, Identity = 98%, Similarity = 98%

GAP: 14bp

LO: yes

RBS: Kibler7, spacer 13, Z score 3.169, Final -2.485, yes

F: Endolysin

SIF-blast: aligns with Endolysin from Microbacterium phage FlameThrower, ID WNO28735.1, query 1-563 subject 1-563, E value = 0.0, Identity = 98%, Similarity = 98%

SIF-HHPred: 6AKV\_A LysB4, endolysin, LAS type enzyme, L-Alanoyl D-Glutamate endopeptidase, HYDROLASE, 2.4A, Bacillus phage B4, Probability: 99.17%, E-value: 4.2e-10

SF-Syn: This gene is pham 86883 [endolysin], upstream is pham 126638, downstream is pham 2497, as seen in Microbacterium phage Katzastrophic

**GENE 21 membrane protein**

Original Glimmer call @bp 19551 has strength 9.44

SSC: 19551 to 20033 (Forward)

CP: yes

SCS: both

ST: SS, Start 10 [19551], Found in 43 of 43 ( 100.0% ) of genes in pham, Called 69.8% of time when present

BLAST-start: aligns with Microbacterium phage Celaena, function unknown, query 1-160 subject 1-160, E value = 6e-85, Identity = 96%, Similarity = 98%

GAP: TGATG Overlap

LO: NA

RBS: Kibler7, spacer 6, Z score 2.586, Final -3.925, yes

F: membrane protein

SIF-blast: aligns with membrane protein from Microbacterium phage Katzastrophic, ID YP\_010752347.1, query 1-159 subject 1-159, E value = 1e-105, Identity = 97%, Similarity = 99%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

SIF-Mem: 4 domains found by DEEPTMHMM; 4 domains found by SOSUI

**GENE 22 membrane protein**

Original Glimmer call @bp 20044 has strength 8.40

SSC: 20044 to 20370 (Forward)

CP: yes

SCS: both

ST: NI, Start 27 [20044], Found in 43 of 145 ( 29.7% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena, function unknown, query 1-108 subject 1-108, E value = 5e-51, Identity = 99%, Similarity = 99%

GAP: 11bp

LO: yes

RBS: Kibler7, spacer 9, Z score 3.022, Final -2.523, yes

F: membrane protein

SIF-blast: aligns with membrane protein from Microbacterium phage Katzastrophic, ID YP\_010752348.1, query 1-108 subject 1-108, E value = 1e-61, Identity = 97%, Similarity = 98%

SIF-HHPred: 6FKF\_p, ATP synthase subunit b', chloroplastic, ATP synthase, membrane protein complex, molecular motor, MEMBRANE PROTEIN, HET, ATP, ADP, 3.1A Spinacia oleracea, Probability: 91.66%, E-value: 5.1

SF-Syn: NKF

SIF-Mem: 1 domain found by both DEEPTMHMM and SOSUI

**GENE 23 membrane protein**

Original Glimmer call @bp 20374 has strength 15.40

SSC: 20374 to 20691 (Forward)

CP: yes

SCS: both

ST: NI, Start 24 [20374], Found in 18 of 92 ( 19.6% ) of genes in pham, Called 94.4% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_23, function unknown, query 1-105 subject 1-105, E value = 2e-50, Identity = 98%, Similarity = 99%

GAP: 3bp

LO: NA

RBS: Kibler7, spacer 10, Z score 3.022, Final -2.443, yes

F: NKF

SIF-blast: aligns with membrane protein from Microbacterium phage Katzastrophic, ID YP\_010752349.1, query 1-105 subject 1-105, E value = 4e-63, Identity = 97%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

SIF-Mem: 2 domains found by DEEPTMHMM; 2 domains found by SOSUI

**GENE 24 hypothetical protein**

Original Glimmer call @bp 20900 has strength 11.76

SSC: 20900 to 21460 (Forward)

CP: yes

SCS: both

ST: SS, Start 6 [20900], found in 47 of 47 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_24, function unknown, query 1-186 subject 1-186, E value = e-102, Identity = 95%, Similarity = 98%

GAP: 209bp

LO: no

RBS: Kibler7, spacer 6, Z score 2.202, Final -5.220, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage FLAMETHROWER\_24, ID WNO28739.1, query 2-186 subject 3-187, E value = 9e-126, Identity = 95%, Similarity = 96%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 25 hypothetical protein**

Original Glimmer call @bp 21471 has strength 12.66

SSC: 21471 to 22142 (Forward)

CP: yes

SCS: both

ST: SS, Start 15 [21471], Found in 43 of 48 ( 89.6% ) of genes in pham, Called 97.7% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_25, function unknown, query 1-223 subject 1-223, E value = e-116, Identity = 93%, Similarity = 95%

GAP: 11bp

LO: yes

RBS: Kibler7, spacer 10, Z score 2.786, Final -2.940, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Quenya, ID YP\_010752220.1, query 1-223 subject 1-223, E value = 4e-123, Identity = 79%, Similarity = 86%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 26 Cas4 exonuclease**

Original Glimmer call @bp 22486 has strength 13.89

SSC: 22486 to 23376 (Forward)

CP: yes

SCS: both

ST: NI, Start 33 [22486], Found in 48 of 154 ( 31.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage FlameThrower\_26, Cas4 exonuclease, query 1-296 subject 1-296, E value = e-169, Identity = 97%, Similarity = 99%

GAP: 344bp

LO: no

RBS: Kibler7, spacer 9, Z score 1.732, Final -5.238, no

F: Cas4 exonuclease

SIF-blast: aligns with Cas4 family exonuclease from Microbacterium phage Katzastrophic, ID YP\_010752352.1, query 1-296 subject 1-296, E value = 0.0, Identity = 97%, Similarity = 99%

SIF-HHPred: 3H4R\_A, Exodeoxyribonuclease 8, Exonuclease, Recombination, Hydrolase, Nuclease, 2.8A, Escherichia coli, Probability= 98.29%, E-value= 0.0000083

SF-Syn: NKF

**GENE 27 hypothetical protein**

Original Glimmer call @bp 23369 has strength 12.72

SSC: 23369 to 23752 (Forward)

CP: yes

SCS: both

ST: NI, Start 19 [23369], Found in 11 of 58 ( 19.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_27, function unknown, query 1-127 subject 1-127, E value = 7e-72, Identity = 100%, Similarity = 100%

GAP: GTGAGTAA overlap

LO: NA

RBS: Kibler7, spacer 11, Z score 2.016, Final -4.624, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Katzastrophic, ID YP\_010752353.1, query 1-127 subject 1-127, E value = 3e-86, Identity = 98%, Similarity = 99%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 28 HicA-like toxin**

Original Glimmer call @bp 23946 has strength 13.96

SSC: 23946 to 23749 (Reverse)

CP: yes

SCS: both

ST: NI, Start 97 [23946], Found in 28 of 177 ( 15.8% ) of genes in pham, Called 71.4% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_28, HicA-like toxin, query 1-65 subject 1-65, E value = 2e-31, Identity = 100%, Similarity = 100%

GAP: 105bp

LO: yes

RBS: Kibler7, spacer 10, Z score 1.829, Final -4.954, yes

F: HicA-like toxin

SIF-blast: aligns with HicA-like toxin from Microbacterium phage Celaena, ID YP\_010752290.1, query 1-65 subject 1-65, E value = 6e-40, Identity = 100%, Similarity = 100%

SIF-HHPred: 6G26\_H, HicA, N-terminal domain of the antitoxin HicB which acts as an inhibitor to HicA, 2.49A, Burkholderia pseudomallei K96243, Probability = 99.65%, E-value= 2.1e-15,

SF-Syn: This gene is pham 127489 [HicA-like toxin], upstream is pham 1723, downstream is pham 120481 [deoxyuridine triphosphatase], as seen in Microbacterium phage Katzastrophic

**GENE 29 deoxyuridine triphosphatase**

Original Glimmer call @bp 24051 has strength 13.42

SSC: 24051 to 24608 (Forward)

CP: yes

SCS: both

ST: NI, Start 11 [24051], Found in 40 of 42 ( 95.2% ) of genes in pham, Called 72.5% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_29, deoxyuridine triphosphatase, query 1-185 subject 1-186, E value = 2e-98, Identity = 93%, Similarity = 94%

GAP: 105bp

LO: yes

RBS: Kibler7, spacer 8, Z score 2.738, Final -3.568, no

F: deoxyuridine triphosphatase

SIF-blast: aligns with deoxyuridine triphosphatase from Microbacterium phage FlameThrower, ID WNO28744.1, query 1-185 subject 1-186, E value = 2e-123, Identity = 92%, Similarity = 94%

SIF-HHPred: 5MYF\_A, dUTPase from DI S. aureus phage, Staphylococcus aureus, pathogenicity island, SaPI, dUTPases, signalling, gene transfer, mobile genetic elements, Hydrolase, 1.85A, Staphylococcus aureus, Probability= 99.53%, E-value= 2.1e-13

SF-Syn: This gene is pham 120481 [deoxyuridine triphosphatase], upstream is pham 127489 [HicA-like toxin], downstream is pham 1756 [thymidylate kinase], as seen in Microbacterium phage Katzastrophic

**GENE 30 thymidylate kinase**

Original Glimmer call @bp 24605 has strength 15.48

SSC: 24605 to 25111 (Forward)

CP: yes

SCS: both

ST: SS, Start 21 [24605], Found in 47 of 58 ( 81.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_30, thymidylate kinase, query 1-168 subject 1-168, E value = 2e-91, Identity = 99%, Similarity = 99%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 15, Z score 2.947, Final -3.509, yes

F: thymidylate kinase

SIF-blast: aligns with thymidylate kinase from Microbacterium phage FlameThrower, ID WNO28745.1, query 1-168 subject 1-168, E value = 8e-117, Identity = 99%, Similarity = 98%

SIF-HHPred: 5ZB0\_B, Thymidylate kinase, NMP kinase, TRANSFERASE, 1.19A, Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579), Probability= 99.72%, E-value= 3.2e-15

SF-Syn: This gene is pham 1756 [thymidylate kinase], upstream is pham 120481 [deoxyuridine triphosphatase], downstream is pham 848 [recombination directionality factor], as seen in Microbacterium phage Katzastrophic

**GENE 31 recombination directionality factor**

Original Glimmer call @bp 25123 has strength 16.05

SSC: 25123 to 25872 (Forward)

CP: yes

SCS: both

ST: NI, Start 37 [25123], Found in 58 of 146 ( 39.7% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage FlameThrower\_31, recombination directionality factor, query 1-249 subject 1-249, E value = e-142, Identity = 100%, Similarity = 100%

GAP: 12bp

LO: yes

RBS: Kibler7, spacer 12, Z score 2.128, Final -4.467, yes

F: recombination directionality factor

SIF-blast: aligns with recombination directionality factor from Microbacterium phage Celaena, ID YP\_010752293.1, query 1-249 subject 1-249, E value = 4e-179, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: This gene is pham 848 [recombination directionality factor], upstream is pham 1756 [thymidylate kinase], downstream is pham 6212, as seen in Microbacterium phage FlameThrower

**GENE 32 hypothetical protein**

Original Glimmer call @bp 26232 has strength 9.28

SSC: 26232 to 26113 (Reverse)

CP: yes

SCS: both

ST: SS, Start 5 [26232], Found in 9 of 10 ( 90.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_32, function unknown, query 1-39 subject 1-39, E value = 2e-16, Identity = 97%, Similarity = 97%

GAP:

LO: TGATG overlap

RBS: Kibler7, spacer 11, Z score 1.579, Final -5.543, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage FlameThrower, ID WNO28747.1, query 1-39 subject 1-39, E value = 1e-15, Identity = 97%, Similarity = 97%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

SIF-Mem: 1 domain found by DEEPTMHMM, 1 domain found by SOSUI

**GENE 33 hypothetical protein**

Original Glimmer call @bp 26540 has strength 13.19; GeneMark calls start at 26486

SSC: 26540 to 26232 (Reverse)

CP: yes

SCS: both - GL

ST: SS, Start 10 [26540], Found in 19 of 22 ( 86.4% ) of genes in pham, Called 94.7% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_34, function unknown, query 1-102 subject 1-102, E value = 6e-52, Identity = 98%, Similarity = 100%

GAP: 74bp

LO: NA

RBS: Kibler7, spacer 11, Z score 1.459, Final -5.796, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage FlameThrower, ID WNO28748.1, query 1-102 subject 1-102, E value = 2e-65, Identity = 98%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 34 NrdH-like glutaredoxin**

original gene 34 and 36 [forward] overlapped with gene 35 [reverse] – gene 35 was removed – gene 34 start overlapped with gene 33 – coordinates of gene 34 changed to 26614 – 26842 to mimic amino acid sequence of NrdH-like glutaredoxin in other EB phage

Original Glimmer call

SSC: 26614 to 26842 [forward]

CP: yes

SCS: both

ST: original coordinates suggest orpham, therefore, no data

BLAST-start:

GAP: 74bp

LO: yes

RBS: Kibler7, spacer , Z score , Final -, yes

F: NrdH-like glutaredoxin

SIF-blast: aligns with NrdH-like glutaredoxin from Microbacterium phage Celaena, ID YP\_010752296.1, query 1-76 subject 1-76, E value = 8e-47, Identity = 100%, Similarity = 100%

SIF-HHPred: 1R7H\_A, NrdH-redoxin, NRDH, THIOREDOXIN, GLUTAREDOXIN, REDOX PROTEIN, DOMAIN SWAPPING, ELECTRON TRANSPORT, 2.69A, Corynebacterium ammoniagenes, Probability= 98.9%, E-value= 5.7e-8,

SF-Syn: NKF

genes from here are +1 in dna master e.g. next is gene 36

**GENE 35 holliday junction resolvase**

Original GeneMark call @bp 26844

SSC: 26844 to 27173 (Forward)

CP: yes

SCS: both

ST: SS, Start 46 [26844], Found in 58 of 147 ( 39.5% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_36, holliday junction resolvase, query 1-109 subject 1-109, E value = 3e-52, Identity = 94%, Similarity = 95%

GAP: TGATG overlap

LO: NA

RBS: Kibler7, spacer 9, Z score 1.149, Final -6.466, yes

F: holliday junction resolvase

SIF-blast: aligns with holliday junction resolvase from Microbacterium phage Celaena, ID YP\_010752297.1, query 1-109 subject 1-109, E value = 4e-65, Identity = 94%, Similarity = 95%

SIF-HHPred: 7BGS\_A, Holliday junction resolvase, archeal holliday junction resolvase helicase DNA binding enzyme phage 15-6 thermus thermophilus, 2.5A, Thermus thermophilus phage 15-6, Probability= 99.58%, E-value= 7.1e-14

SF-Syn: changes made to previous gene means there is no current synteny

**GENE 36 hypothetical protein**

Original Glimmer call @bp 27201 has strength 14.57

SSC: 27201 to 27506 (Forward)

CP: yes

SCS: both

ST: SS, Start 4 [27201], Found in 58 of 58 ( 100.0% ) of genes in pham, Called 98.3% of time when present

BLAST-start: aligns with Microbacterium phage Eden\_36, function unknown, query 1-99 subject 13-110, E value = 2e-31, Identity = 64%, Similarity = 80%

GAP: 28bp

LO: yes

RBS: Kibler7, spacer 14, Z score 1.794, Final -5.680, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Franklin22, ID YP\_010752161.1, query 1-100 subject 1-98, E value = 1e-25, Identity = 51%, Similarity = 70%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 37 hypothetical protein**

Original Glimmer call @bp 27591 has strength 16.00

SSC:

CP: yes

SCS: both

ST: NI, Start 18 [27591], Found in 10 of 45 ( 22.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_37, function unknown, query 1-86 subject 1-86, E value = 3e-45, Identity = 100%, Similarity = 100%

GAP: 85bp

LO: yes

RBS: Kibler7, spacer 19, Z score 3.022, Final -2.443, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Katzastrophic, ID YP\_010752364.1, query 1-86 subject 1-86, E value = 3e-55, Identity = 99%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 38 hypothetical protein**

Original Glimmer call @bp 27844 has strength 9.23

SSC: 27844 to 28032 (Forward)

CP: yes

SCS: both

ST: NI, Start 15 [27844], Found in 6 of 47 ( 12.8% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_39, function unknown, query 1-62 subject 1-62, E value = 1e-28, Identity = 100%, Similarity = 100%

GAP: ATGCCTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 1.922, Final -4.899, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752300.1, query 1-62 subject 1-62, E value = 5e-34, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 39 DNA primase/helicase**

Original Glimmer call @bp 28143 has strength 12.80

SSC: 28143 to 30599 (Forward)

CP: yes

SCS: both

ST: NI, Start 35 [28143], Found in 59 of 144 ( 41.0% ) of genes in pham, Called 96.6% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_39, DNA primase/helicase, query 1-818 subject 1-818, E value = 0.0, Identity = 99%, Similarity = 99%

GAP: 111bp

LO: no

RBS: Kibler7, spacer 8, Z score 2.516, Final -4.036, yes

F: DNA primase/helicase

SIF-blast: aligns with DNA primase/helicase from Microbacterium phage FlameThrower, ID WNO28754.1, query 1-818 subject 1-818, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: 8IQI\_C, Putative primase C962R, polymerase, primase, PrimPol, Helicase, DNA BINDING PROTEIN, HET: ANP, African swine fever virus BA71V, Probability= 100%, E-value= 1.7e-33

SF-Syn: This gene is pham 85082 [DNA primase/helicase], upstream is pham 2276, downstream is pham 127340 [DNA polymerase I], as seen in Microbacterium phage FlameThrower

**GENE 40 DNA polymerase I**

Original Glimmer call @bp 30596 has strength 12.73

SSC: 30596 to 32467 (Forward)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 224 – nt 30596 – Found in 1185 of 1600 (74.1%) – of genes in pham, Called 71.4% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_41, DNA polymerase I, query 1-623 subject 1-623, E value = 0.0, Identity = 98%, Similarity = 99%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 2.638, Final -3.558, yes

F: DNA polymerase I

SIF-blast: aligns with DNA polymerase I from Microbacterium Celaena, ID QDH92419.1, query 1-623 subject 1-623, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: DNA polymerase I; mycobacteria, DNA polymerase, apoenzyme, TRANSFERASE-DNA complex; HET: DCT; 1.9A {Mycolicibacterium smegmatis} 6VDD\_A, probability 100%, E-value 4.8e-66

SF-Syn: DNA polymerase I (pham 68543), upstream gene is DNA primase/helicase (pham 880), downstream gene is NKF (pham 1750), just like in phage Celaena (EB)

**GENE 41 Hypothetical protein**

Original Glimmer call @bp 32467 has strength 9.49

SSC: 32467 to 32664 (Forward)

CP: yes

SCS: both

ST: SS, The start site called the most often in the published annotations is 9 – nt 32467 – Found in 56 of 56 (100%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_42, function unknown, query 1-65 subject 1-65, E value = 5e-35, Identity = 100%, Similarity = 100%

GAP: A overlap

LO: NA

RBS: Kibler7, spacer 15, Z score 1.912, Final -5.209, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Celaena\_41, ID QDH92420.1, query 1-65 subject 1-65, E value = 4e-41, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 42 Hypothetical protein**

Original Glimmer call

SSC: 32661 to 32963 (Forward)

CP: yes

SCS: both

ST: SS, The start site called the most often in the published annotations is 37 – nt 32661 – Found in 114 of 121 (94.2%) – of genes in pham, Called 99.1% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic, function unknown, query 1-100 subject 1-100, E value = 1e-51, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 3.061, Final -2.661, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Celaena\_42, ID QDH92421.1, query 1-100 subject 1-100, E value = 7e-66, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 43 DNA binding protein**

Original Glimmer call @bp 32963 has strength 9.14

SSC: 32963 to 33685 (Forward)

CP: yes

SCS: both

ST: NA, This gene does not have the start site called the most often in the published annotations. The start site called is 37 – nt 32963 – Found in 44 of 123 (35.8%) – of genes in pham, Called 97.7% of the time when present.

BLAST-start: aligns with Microbacterium phage Celaena\_43, DNA binding protein, query 1-240 subject 1-240, E value = e-118, Identity = 90%, Similarity = 94%

GAP: A overlap

LO: NA

RBS: Kibler7, spacer 6, Z score 2.119, Final -4.916, no

F: DNA binding protein

SIF-blast: aligns with DNA binding protein from Microbacterium Celaena, ID QDH92422.1, query 1-240 subject 1-240, E value = 4e-144, Identity = 90%, Similarity = 94%

SIF-HHPred: RNA polymerase sigma factor RpoS; transcription initiation, Pseudomonas aeruginosa, RNA polymerase, sigmaS, SutA, 7XL3\_F, probably 99.94%, E-value 1e-23.

SF-Syn: DNA binding protein (pham 63728), upstream gene is NKF (pham 965), downstream is pham 4319, as seen in phage Vitas

**GENE 44 Hypothetical protein**

Original Glimmer call @bp 33819 has strength 13.11

SSC: 33819 to 33682 (Reverse)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 17 – nt 33819 – Found in 19 of 19 (100%) – of genes in pham, Called 84.2% of the time when present.

BLAST-start: aligns with Microbacterium phage Quenya\_46, function unknown, query 1-57 subject 1-57, E value = 1e-21, Identity = 100%, Similarity = 100%

GAP: 19 bp

LO: yes

RBS: Kibler7, spacer 6, Z score 1.633, Final -5.945, no

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Quenya\_46, ID QOP64281.1, query 1-57 subject 1-57, E value = 2e-25, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 45 SprT-like protease**

Original Glimmer call @bp 33838 has strength 8.95; GeneMark calls start at 33928

SSC: 33928 to 34509 (Forward)

CP: yes

SCS: Both-GM [Starterator, BLAST and RBS favor 33928]

ST: NA, This gene does not have the start site called the most often in the published annotations. Start site 49 – nt 33928 – Found in 22 of 97 (22.7%) – of genes in pham, Called 22.7% of the time when present. This matches the call made by GeneMark and has a better RBS score than what is called both by Glimmer and Starterator. However, the start site called is 29 – nt 33838 – Found in 1 of 103 (1.0%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_46, SprT-like protease, query 31-223 subject 1-193, E value = e-108, Identity = 99%, Similarity = 99%

GAP: 19 bp

LO: yes

RBS: Kibler7, spacer 6, Z score 1.639, Final -5.933, no

F: SprT-like protease

SIF-blast: aligns with SprT-like protease from Microbacterium Katzastrophic, ID UKH48483.1, query 31-223 subject 1-193, E value = 1e-135, Identity = 99%, Similarity = 99%

SIF-HHPred: SprT-domain-containing protein Spartan: DPC repair protease, DNA BINDING PROTEIN; HET: ADP, MLZ, FLC; 1.5A {Homo sapiens} 6MDW\_A, probability 99.62%, E-value 9.9e-16

SF-Syn: SprT-like protease (pham 1210), downstream gene is ParB-like nuclease domain protein (pham 2226), just like in phage Katzastrophic (EB).

**GENE 46 ParB-like nuclease domain protein**

Original Glimmer call @bp 34511 has strength 9.10

SSC: 34511 to 34969 (Forward)

CP: yes

SCS: both

ST: NA, This gene does not have the start site called the most often in the published annotations. The start site called is 18 – nt 34511 – Found in 6 of 43 (14.0%) – of genes in pham, Called 100% of the time when present.

BLAST-start: aligns with Microbacterium phage Kieran\_46, ParB-like nuclease domain protein, query 1-152 subject 1-152, E value = 9e-75, Identity = 84%, Similarity = 90%

GAP: 2 bp

LO: NA

RBS: Kibler7, spacer 11, Z score 2.918, Final -2.918, yes

F: ParB-like nuclease domain protein

SIF-blast: aligns with ParB-like nuclease domain protein from Microbacterium Kieran, ID AVR57207.1, query 1-152 subject 1-152, E value = 7e-89, Identity = 85%, Similarity = 90%

SIF-HHPred: ParB-like nuclease domain protein; ParABS, cytoskeleton, bactofilin, CTP, CELL CYCLE; HET: CTP; 1.7A {Myxococcus xanthus DK 1622} 6RYK\_B, probability 98.97%, E-value 5.2e-9

SF-Syn: ParB-like nuclease domain protein (pham 2226), upstream gene is SprT-like protease (pham 1210), just like in phage Kieran (EB). However, downstream gene does not match.

**GENE 47 Hypothetical protein**

Original Glimmer call @bp 34966 has strength 7.79

SSC: 34966 to 35400 (Forward)

CP: yes

SCS: both

ST: NA, This gene does not have the start site called the most often in the published annotations. The start site called is 21 – nt 34966 – Found in 44 of 94 (46.8%) – of genes in pham, Called 88.6% of the time when present.

BLAST-start: aligns with Microbacterium phage Celaena\_47, function unknown, query 1-144 subject 1-144, E value = 2e-84, Identity = 97%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 14, Z score 1.749, Final -5.495, no

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Celaena\_47, ID QDH92426.1, query 1-144 subject 1-144, E value = 8e-97, Identity = 98%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 48 Hypothetical protein**

Original Glimmer call @bp 35405 has strength 16.19

SSC: 35405 to 35857 (Forward)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 30 – nt 35405 – Found in 32 of 37 (86.5%) – of genes in pham, Called 81.2% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrohpic\_49, function unknown, query 1-150 subject 1-150, E value = 2e-82, Identity = 96%, Similarity = 100%

GAP: 5 bp

LO: NA

RBS: Kibler7, spacer 6, Z score 2.133, Final -4.886, no

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Katzastrophic\_49, ID UKH48486.1, query 1-150 subject 1-150, E value = 4e-106, Identity = 96%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 49 Hypothetical protein**

Original Glimmer call @bp 35861 has strength 11.69

SSC: 35861 to 36301 (Forward)

CP: yes

SCS: both

ST: NA, This gene does not have the start site called the most often in the published annotations. The start site called is 20 – nt 35861 – Found in 17 of 461 (3.7%) – of genes in pham, Called 94.1% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_50, function unknown, query 1-146 subject 1-146, E value = 5e-79, Identity = 97%, Similarity = 99%

GAP: 4 bp

LO: NA

RBS: Kibler7, spacer 13, Z score 3.061, Final -2.686, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Katzastrophic\_50, ID UKH48487.1, query 1-146 subject 1-146, E value = 2e-99, Identity = 97%, Similarity = 99%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 50 HNH endonuclease**

Original Glimmer call @bp 36378 has strength 8.43

SSC: 36378 to 36770 (Forward)

CP: yes

SCS: both

ST: SS, The start site called the most often in the published annotations is 40 – nt 36378 – Found in 46 of 56 (82.1%) – of genes in pham, Called 89.1% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_51, HNH endonuclease, query 1-130 subject 1-130, E value = 4e-74, Identity = 100%, Similarity = 100%

GAP: 77 bp

LO: no

RBS: Kibler7, spacer 10, Z score 2.660, Final -3.425, no

F: HNH endonuclease

SIF-blast: aligns with HNH endonuclease from Microbacterium Katzastrophic, ID UKH48488.1, query 1-130 subject 1-130, E value = 1e-91, Identity = 100%, Similarity = 100%

SIF-HHPred: HNH endonuclease domain protein; CRISPR-Cas, Cas9, HNH, RuvC, RNA-guided DNA endonuclease, cytoplasmic, Hydrolase; HET: 4OGE\_A, probability 98.34%, E-value 0.0000011

SF-Syn: HNH endonuclease (pham 1748), upstream gene is NKF (pham 70112), downstream gene is NKF (pham 1738), as seen in phage Katzastrophic (EB)

**GENE 51 Hypothetical protein**

Original Glimmer call @bp 36917 has strength 13.66

SSC: 36917 to 37228 (Forward)

CP: yes

SCS: both

ST: NA, The start site called the most often in the published annotations is 10 – nt 36917 – Found in 45 of 55 (81.8%) – of genes in pham, Called 86.7% of the time when present.

BLAST-start: aligns with Microbacterium phage Katzastrophic\_52, function unknown, query 1-103 subject 1-103, E value = 9e-56, Identity = 100%, Similarity = 100%

GAP: 147 bp

LO: no

RBS: Kibler7, spacer 12, Z score 1.985, Final -4.944, no

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Sea\_Katzastrophic\_52, ID UKH48489.1, query 1-103 subject 1-103, E value = 9e-69, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 52 hypothetical protein**

Original Glimmer call @bp 37225 has strength 6.57

SSC: 37225 to 37389 (Forward)

CP: yes

SCS: both

ST: SS, Start 3 [37255], Found in 56 of 56 ( 100.0% ) of genes in pham, Called 98.2% of time when present

BLAST-start: aligns with Microbacterium phage FlameThrower\_51, function unknown, query 1-54 subject 1-54, E value = 2e-23, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 9, Z score 2.439, Final -3.761, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Katzastrophic, ID YP\_010752379.1, query 1-54 subject 1-54, E value = 3e-29, Identity = 98%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 53 hypothetical protein**

Original Glimmer call @bp 37440 has strength 15.00

SSC: 37440 to 37922 (Forward)

CP: yes

SCS: both

ST: SS, Start 7 [37440], Found in 30 of 58 ( 51.7% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_53, function unknown, query 1-160 subject 1-160, E value = 1e-85, Identity = 100%, Similarity = 100%

GAP: 51bp

LO: yes

RBS: Kibler7, spacer 12, Z score 1.697, Final -5.372, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage FlameThrower, ID WNO28767.1, query 1-160 subject 1-160, E value = 2e-110, Identity = 99%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 54 hypothetical protein**

Original Glimmer call @bp 37983 has strength 7.04

SSC: 37983 to 38204 (Forward)

CP: yes

SCS: both

ST: NI, Start 10 [37983], Found in 11 of 26 ( 42.3% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_55, function unknown, query 1-73 subject 1-73, E value = 2e-40, Identity = 100%, Similarity = 100%

GAP: 61bp

LO: yes

RBS: Kibler7, spacer 10, Z score 1.829, Final -4.954, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage FlameThrower, ID WNO28768.1, query 1-73 subject 1-73, E value = 9e-46, Identity = 99%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 55 dihydrofolate reductase**

Original Glimmer call @bp 38194 has strength 15.23

SSC: 38194 to 38778 (Forward)

CP: yes

SCS: both

ST: SS, Start 14[38194], Found in 58 of 61 ( 95.1% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_55, dihydrofolate reductase, query 1-194 subject 1-194, E value = e-111, Identity = 99%, Similarity = 100%

GAP: 10bp overlap

LO: NA

RBS: Kibler7, spacer 10, Z score 1.811, Final -4.991, yes

F: dihydrofolate reductase

SIF-blast: aligns with dihydrofolate reductase from Microbacterium phage FlameThrower, ID WNO28769.1, query 1-194 subject 1-194, E value = 1e-137, Identity = 98%, Similarity = 99%

SIF-HHPred: 3CSE\_B, Dihydrofolate reductase, protein-ligand complex, reductase, OXIDOREDUCTASE, HET: NDP, N22, 1.6A, Candida glabrata, SCOP: c.71.1.1, l.1.1.1, Probability= 99.9%, E-value= 5.4e-22

SF-Syn: This gene is pham 85568 [dihydrofolate reductase], upstream is pham 115270, downstream is pham 127696, as seen in Microbacterium phage FlameThrower

**GENE 56 hypothetical protein**

Original Glimmer call @bp 38775 has strength 14.95

SSC: 38775 to 39110 (Forward)

CP: yes

SCS: both

ST: NI, Start 31 [38775], Found in 21 of 71 ( 29.6% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_57, function unknown, query 1-111 subject 1-111, E value = 2e-59, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 13, Z score 2.947, Final -2.953, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752318.1, query 1-111 subject 1-111, E value = 1e-75, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 57 thymidylate synthase**

Original Glimmer call @bp 39140 has strength 11.55; GeneMark calls start at 39113

SSC: 39107 to 40042 (Forward)

CP: yes

SCS: both – CS [39107 favored by Starterator, BLAST and RBS]

ST: NI, Start 40 [39107], Found in 16 of 98 ( 16.3% ) of genes in pham, Called 87.5% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_58, thymidylate synthase, query 1-300 subject 12-311, E value = e-173, Identity = 99%, Similarity = 99%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 2.439, Final -3.812, yes

F: thymidylate synthase

SIF-blast: aligns with thymidylate synthase from Microbacterium phage FlameThrower, ID WNO28771.1, query 1-300 subject 12-311, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: 7DP4\_A, Thymidylate synthase, THYMIDYLATE SYNTHASE, NUCLEOTIDE SYNTHASE, TRANSFERASE, METHYLTRANSFERASE, HET: MTX, BME, UMP, 1.5A, Brugia malayi, Probability= 100%, E-value= 1.9e-54

SF-Syn: This gene is pham 127616 [thymidylate synthase], upstream is pham 127696, downstream is pham 99537, as seen in Microbacterium phage FlameThrower

**GENE 58 membrane protein**

Original Glimmer call @bp 40052 has strength 13.80

SSC: 40052 to 40240 (Forward)

CP: yes

SCS: both

ST: SS, Start 7 [40052], found in 21 of 58 ( 36.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_58, function unknown, query 1-62 subject 1-62, E value = 4e-27, Identity = 98%, Similarity = 98%

GAP: 10bp

LO: NA

RBS: Kibler7, spacer 9, Z score 1.565, Final -5.589, yes

F: NKF

SIF-blast: aligns with membrane protein from Microbacterium phage Katzastrophic, ID YP\_010752385.1, query 1-62 subject 1-62, E value = 2e-29, Identity = 97%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

SIF-Mem: 2 domains found by DEEPTMHMM, 2 domains found by SOSUI

**GENE 59 hypothetical protein**

Original Glimmer call @bp 40387 has strength 19.06

SSC: 40387 to 40521 (Forward)

CP: yes

SCS: both

ST: SS, Start 10 [40387], Found in 49 of 49 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Celaena\_59, function unknown, query 1-44 subject 1-44, E value = 1e-17, Identity = 100%, Similarity = 100%

GAP: 147bp

LO: no

RBS: Kibler7, spacer 12, Z score 3.022, Final -2.584, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752321.1, query 1-44 subject 1-44, E value = 9e-22, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 60 hypothetical protein**

Original Glimmer call @bp 40521 has strength 17.20

SSC: 40521 to 40727 (Forward)

CP: yes

SCS: both

ST: SS, Start 17 [40521], Found in 19 of 44 ( 43.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_61, function unknown, query 1-68 subject 1-68, E value = 1e-31, Identity = 100%, Similarity = 100%

GAP: TAGTG overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 3.022, Final -2.584, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752322.1, query 1-68 subject 1-68, E value = 1e-39, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 61 hypothetical protein**

Original Glimmer call @bp 40737 has strength 13.12

SSC: 40737 to 41015 (Forward)

CP: yes

SCS: both

ST: SS, Start 1 [40737], Found in 5 of 5 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_62, function unknown, query 1-92 subject 1-92, E value = 7e-48, Identity = 100%, Similarity = 100%

GAP: 10bp

LO: NA

RBS: Kibler7, spacer 9, Z score 2.737, Final -3.119, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752323.1, query 1-92 subject 1-92, E value = 1e-59, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 62 hypothetical protein**

Original Glimmer call @bp 41012 has strength 13.24

SSC: 41012 to 41230 (Forward)

CP: yes

SCS: both

ST: NI, Start 10 [41012], Found in 6 of 58 ( 10.3% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_63, function unknown, query 1-72 subject 1-72, E value = 3e-31, Identity = 98%, Similarity = 98%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 3.022, Final -2.584, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752324.1, query 1-72 subject 1-72, E value = 7e-39, Identity = 99%, Similarity = 98%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 63 hypothetical protein**

Original Glimmer call @bp 41227 has strength 12.56

SSC: 41227 to 41607 (Forward)

CP: yes

SCS: both

ST: SS, Start 6 [41227], Found in 58 of 58 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_64, function unknown, query 1-126 subject 1-126, E value = 1e-69, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 1.910, Final -4.924, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium phage Celaena, ID YP\_010752325.1, query 1-126 subject 1-126, E value = 2e-85, Identity = 100%, Similarity = 100%

SIF-HHPred: NKF, no matches with a probability above 90%

SF-Syn: NKF

**GENE 66** **tRNA**

ARAGORN

 tRNA-Asn(gtt)

 73 bases, %GC = 57.5

 Sequence [41842,41914]

tRNAscan-SE

41842, 41913

tRNA – Asn (gtt)

Infernal Score = 57.9

tRNA-Asn(gtt) detected by both Aragorn and tRNAscan-SE, infernal score 57.9, ends have been trimmed to match the Aragorn web-version

**GENE 67 tRNA**

ARAGORN

tRNA-Gln(ctg)

 72 bases, %GC = 62.5

 Sequence [41969,42040]

tRNAscan-SE

41969, 42039

tRNA – Gln (ctg)

Infernal score = 57.4

tRNA-Gln(ctg) detected by both Aragorn and tRNAscan-SE, infernal score 53.0, ends have been trimmed to match the Aragorn web-version

**GENE 66 HNH endonuclease**

Original Glimmer call @bp 42043 has strength 3.07; GeneMark calls start at 41983

SSC: 42043 to 42318 (Forward)

CP: yes

SCS: both – GL [42043 favored by Starterator, BLAST and RBS]

ST: NI, Start 128 [42043], Found in 62 of 543 ( 11.4% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Katzastrophic\_67, HNH endonuclease, query 1-91 subject 1-91, E value = 4e-52, Identity = 100%, Similarity = 100%

GAP: 436bp

LO: YES

RBS: Kibler7, spacer 10, Z score 2.527, Final -3.486, yes

F: HNH endonuclease

SIF-blast: aligns with HNH endonuclease from Microbacterium phage Celaena, ID YP\_010752326.1, query 1-91 subject 1-91, E value = 4e-60, Identity = 99%, Similarity = 98%

SIF-HHPred: 6M0X\_A, CRISPR-associated endonuclease Cas9 1, Cas9, CRISPR, DNA BINDING PROTEIN, HET: BA; 2.561A , Streptococcus thermophilus LMD-9, Probability= 97.73%, E-value= 0.00013

SF-Syn: NKF