**PHAGE CHARBIE – CLUSTER EE – ANNOTATION NOTES**

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

Original Glimmer call @bp 41 has strength 11.66

SSC: 41 to 379 (Forward)

CP: yes

SCS: both

ST: SS, Start 4 [41], Found in 112 of 113 ( 99.1% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Vanisius\_1, function unknown, query 1-112 subject 1-112, E value = 3e-59, Identity = 100%, Similarity = 100%

GAP: NA

LO: NA

RBS: Kibler7, spacer 13, Z score 2.053, Final -5.036, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Scamander, ID YP\_010050599.1, query 1-112 subject 1-112, E value = 2e-73, Identity = 100%, Similarity = 100%

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

SF-Syn: NKF

**GENE 2 hypothetical protein**

Original Glimmer call @bp 376 has strength 8.76

SSC: 376 to 1059 (Forward)

CP: yes

SCS: both

ST: SS, Start 2 [37], Found in 110 of 110 ( 100.0% ) of genes in pham, Called 96.4% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_2, function unknown, query 1-227 subject 1-227, E value = e-130, Identity = 100%, Similarity = 100%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 13, Z score 2.404, Final -4.334, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium TimoTea, ID QBI97315.1, query 1-227 subject 1-227, E value = 5e-158, Identity = 100%, Similarity = 100%

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

SF-Syn: NKF

**GENE 3 terminase**

Original Glimmer call @bp 1062 has strength 11.58

SSC: 1062 to 2522 (Forward)

CP: yes

SCS: both

ST: NI, Start 89[1062], Found in 115 of 1223 ( 9.4% ) of genes in pham, Called 98.3% of time when present

BLAST-start: aligns with Microbacterium phage Vanisius\_3, terminase, query 1-486 subject 1-486, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: 3bp

LO: NA

RBS: Kibler7, spacer 15, Z score 1.856, Final -5.987, yes

F: terminase

SIF-blast: aligns with terminase from Microbacterium phage Scamander, ID YP\_010050601.1, query 1-486 subject 1-486, E value = 0.0, Identity = 100%, Similarity = 100%

SIF-HHPred: 6Z6D\_A, Terminase large subunit, genome packaging, bacteriophage, ATPase, nuclease, HET: BR; 2.2A, Enterobacteria phage HK97, Probability: 100%, E-value: 3.2e-42,

SF-Syn: This gene is pham 130373 [terminase], upstream is pham 85210, downstream is pham 130481 [portal protein], as seen in Microbacterium phage Luxx

**GENE 4 portal protein**

Original Glimmer call @bp 2758 has strength 12.08; GeneMark calls start at 2716

SSC: 2758 to 3774 (Forward)

CP: yes

SCS: both – GL

ST: NI, Start 16 [2647], Found in 115 of 168 ( 68.5% ) of genes in pham, Called 86.1% of time when present

Start 27 [2758], Found in 114 of 168 ( 67.9% ) of genes in pham, Called 10.5% of time when present

Starterator seems to favor 2647. RBS is close between 2647 and 2758. BLAST has proteins that align with both. I tend to favor 2647 but don’t have enough evidence to change from 2758 officially

BLAST-start: aligns with Microbacterium phage Rhogar\_4, portal protein, query 1-338 subject 1-338, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: 236bp

LO: no

RBS: Kibler7, spacer 11, Z score 1.777, Final -5.300, no

F: portal protein

SIF-blast: aligns with portal protein from Microbacterium phage Azizam, ID QGJ97464.2, query 1-338 subject 1-338, E value = 0.0, Identity = 100%, Similarity = 100%

SIF-HHPred: 8FQL\_L, Portal protein, Prohead I, icosahedral symmetry, HK97, phage, capsid, 3.6A, Escherichia phage HK97, Probability: 100%, E-value: 2.2e-39

SF-Syn: This gene is pham 130481 [portal protein], upstream is pham 130373 [terminase], downstream is pham 1062 [major capsid and protease fusion protein] as seen in Microbacterium phage Luxx

**GENE 5 major capsid and protease fusion protein**

Original Glimmer call @bp 3771 has strength 13.81

SSC: 3771 to 5357 (Forward)

CP: yes

SCS: both

ST: SS, Start 2 [3771], Found in 111 of 113 ( 98.2% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_5, major capsid and protease fusion protein, query 1-528 subject 1-528, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 15, Z score 3.093, Final -3.509, yes

F: major capsid and protease fusion protein

SIF-blast: aligns with major capsid and protease fusion protein from Microbacterium phage Loca, ID URM86189.1, query 1-528 subject 1-529, E value = 0.0, Identity = 98%, Similarity = 99%

SIF-HHPred: 8ECK\_E, Major capsid protein, HK97-fold, T=7, tailed bacteriophage, 2.6A, Gordonia phage Cozz, Probability: 99.91%, E-value: 1.2e-22

SF-Syn: This gene is pham 1062 [major capsid and protease fusion protein], upstream is pham 130481 [portal protein], downstream is pham 1065 [head-to-tail adaptor], as seen in Microbacterium phage Luxx

**GENE 6 head-to-tail adaptor**

Original Glimmer call @bp 5361 has strength 17.90

SSC: 5361 to 5714 (Forward)

CP: yes

SCS: both

ST: SS, Start 2 [5361], Found in 113 of 113 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_6, head-to-tail adaptor, query 1-117 subject 1-117, E value = 3e-60, Identity = 100%, Similarity = 100%

GAP: 4bp

LO: NA

RBS: Kibler7, spacer 16, Z score 2.669, Final -4.553, yes

F: head-to-tail adaptor

SIF-blast: aligns with head-to-tail adaptor from Microbacterium phage Hulk, ID QDF16566.1, query 1-117 subject 1-117, E value = 6e-77, Identity = 97%, Similarity = 99%

SIF-HHPred: 7Z4W\_l, Head completion protein gp15, Bacteriophage, SPP1, Portal Protein, Head completion proteins, Connector Complex, DNA Channel, 2.7A, Bacillus subtilis, Probability: 99.55%, E-value: 1.6e-13

SF-Syn: This gene is pham 1065 [head-to-tail adaptor], upstream is pham 1062 [major capsid and protease fusion protein], downstream is pham 1057 [tail terminator], as seen in Microbacterium phage Luxx

**GENE 7 tail terminator**

Original Glimmer call @bp 5711 has strength 19.94

SSC: 5711 to 6091 (Forward)

CP: yes

SCS: both

ST: SS, Start 2 [5711], Found in 113 of 113 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_7, tail terminator, query 1-126 subject 1-126, E value = 2e-65, Identity = 100%, Similarity = 100%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 7, Z score 1.515, Final -6.591, no [but very similar]

F: tail terminator

SIF-blast: aligns with tail terminator from Microbacterium phage TinyTruffula, ID QKY79162.1, query 1-126 subject 1-126, E value = 2e-81, Identity = 97%, Similarity = 98%

SIF-HHPred: 8GTF\_n, Terminator protein, Marine bacteriophage, Siphophage, Stopper protein, Terminator protein, Head-to-tail interface, 6.6A, Dinoroseobacter phage vB\_DshS-R4C, Probability: 97.95%, E-value: 0.002,

SF-Syn: This gene is pham 1057 [tail terminator], upstream is pham 1065 [head-to-tail adaptor], downstream is pham 84957 [major tail protein], as seen in Microbacterium phage Luxx

**GENE 8 major tail protein**

Original Glimmer call @bp 6130 has strength 16.48

SSC: 6130 to 6564 (Forward)

CP: yes

SCS: both

ST: NI, Start 5 [6130], Found in 113 of 177 ( 63.8% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Yubaba\_8, major tail protein, query 1-144 subject 1-144, E value = 8e-82, Identity = 100%, Similarity = 100%

GAP: 39bp

LO: yes

RBS: Kibler7, spacer 11, Z score 1.530, Final -5.797, yes

F: major tail protein

SIF-blast: aligns with major tail protein from Microbacterium phage PaoPu, ID , query 1-144 subject 1-144, E value = 7e-100, Identity = 100%, Similarity = 100%

SIF-HHPred: 8GTC\_F, Major tail protein, Marine bacteriophage, Siphophage, Baseplate, Megatron protein, Tail fibre protein, Distal tail protein, Hub protein, 4.5A, Dinoroseobacter phage vB\_DshS-R4C, Probability: 99.34%, E-value: 3.2e-11

SF-Syn: This gene is pham 84957 [major tail protein], upstream is pham 1057 [tail terminator], downstream is pham 130571, as seen in Microbacterium phage Luxx

**GENE 9 hypothetical protein**

Original Glimmer call @bp 6577 has strength 14.59

SSC: 6577 to 6960 (Forward)

CP: yes

SCS: both

ST: NI, Start 5 [6577], Found in 115 of 174 ( 66.1% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage YertPhresh\_9, function unknown, query 1-127 subject 1-127, E value = 2e-66, Identity = 100%, Similarity = 100%

GAP: 13bp

LO: yes

RBS: Kibler7, spacer 10, Z score 2.248, Final -4.294, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium PaoPu, ID YP\_009996594.1, query 1-127 subject 1-127, E value = 4e-85, Identity = 100%, Similarity = 100%

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

SF-Syn: NKF

**GENE 10 tail assembly chaperone**

Original Glimmer call @bp 6974 has strength 17.44

SSC: 6974 to 7294 (Forward)

CP: yes

SCS: both

ST: SS, Start 6 [6974], Found in 113 of 113 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_10, tail assembly chaperone, query 1-106 subject 1-106, E value = 2e-53, Identity = 100%, Similarity = 100%

GAP: 14bp

LO: yes

RBS: Kibler7, spacer 10, Z score 2.482, Final -3.826, yes

F: tail assembly chaperone

SIF-blast: aligns with tail assembly chaperone from Microbacterium phage Scamander, ID YP\_010050609.1, query 1-106 subject 1-106, E value = 9e-68, Identity = 100%, Similarity = 100%

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

SF-Syn: this is the first of the tail assembly chaperone proteins the next gene [11] will use gene 10 and a frameshift to form the larger tail assembly chaperone protein

**GENE 11 tail assembly chaperone**

SSC: 7321 to 7410 (Forward) – to be changed due to being part of a frameshift – see below

LaviMo gene 10 (6998 - 7318 )

>LaviMo gp10

METSTKKFDLDQLTLGEVAAIEDLSGVAIGSVSESTPQGKFLAALYMVAKRRDGQPTFTFNAALQASMAEAQSFLGFDAPDAADDADAAESSAEGNGDSSPENAPA

>LaviMo gp11

METSTKKFDLDQLTLGEVAAIEDLSGVAIGSVSESTPQGKFLAALYMVAKRRDGQPTFTFNAALQASMAEAQSFLGFDAPDAADDADAAESSAEGKRRQLARERARLKAQFIVQLGMDPAAYERLTIAERDAIVRELNKRSSRRR

>Charbie\_Draft gp10

METSTKKFDLDQLTLGEVAAIEDLSGVAIGSVSESTPQGKFLAALYMVAKRRDGQPTFTFNAALQASMAEAQSFLGFDAPDAADDADAAESSAEGNGDSSPENAPA

A colorful rectangular shapes with letters

Description automatically generated with medium confidence

gene 10 = in frame 2 - ends at aa G [gga]

frameshift to frame 1 [-1 frameshift] – final a of gga becomes first a in aaa sequence of aa K

section 1 = 6974 to 7258

section 2 = 7258 to 7410

**GENE 12 tape measure protein**

Original Glimmer call @bp 7528 has strength 14.91

SSC: 7528 to 9633 (Forward)

CP: yes

SCS: both

ST: NI, Start 12 [7528], Found in 115 of 213 ( 54.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_12, tape measure protein, query 1-701 subject 1-701, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: 118bp

LO: yes

RBS: Kibler7, spacer 16, Z score 2.752, Final -4.386, yes

F: tape measure protein

SIF-blast: aligns with tape measure protein from Microbacterium phage Luxx, ID QKY78659.1, query 1-701 subject 1-701, E value = 0.0, Identity = 99%, Similarity = 100%

SIF-HHPred: 6V8I\_BF, Tape Measure Protein, gp57, phage tail, tail tip, tape measure protein, VIRAL PROTEIN, 3.7A, Staphylococcus virus 80alpha, Probability: 99.93%, E-value: 2.9e-17

SF-Syn: This gene is pham 130505 [tape measure protein], upstream is tail assembly chaperone, downstream is pham 76903 [minor tail protein], as seen in Microbacterium phage Luxx

SIF-Mem: 1 domain found by DEEPTMHMM, 8 domains found by SOSUI

**GENE 13 minor tail protein**

Original Glimmer call @bp 9630 has strength 10.34

SSC: 9630 to 10655 (Forward)

CP: yes

SCS: both

ST: SS, Start 1 [9630], Found in 93 of 93 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_13, minor tail protein, query 1-341 subject 1-341, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 12, Z score 2.260, Final -4.412, yes

F: minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage Scamander, ID YP\_010050611.1, query 1-341 subject 1-341, E value = 0.0, Identity = 99%, Similarity = 99%

SIF-HHPred: 6VLS\_A Maltose/maltodextrin-binding periplasmic protein,Vip3Aa, Bacillus thuringiensis, Vip3A, biological control, mode of action, glycobiology, TOXIN; HET: PEG; 3.2A, Escherichia coli, Probability: 97.94%, E-value: 0.0018

SF-Syn: This gene is pham 76903 [minor tail protein], upstream is pham 130505 [tape measure protein], downstream is pham 1068 [minor tail protein], as seen in Microbacterium phage Luxx

**GENE 14 minor tail protein**

Original Glimmer call @bp 10655 has strength 10.62

SSC: 10655 to 12724 (Forward)

CP: yes

SCS: both

ST: SS, Start 1 [10655], Found in 113 of 113 ( 100.0% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_14, minor tail protein, query 1-689 subject 1-689, E value = 0.0, Identity = 100%, Similarity = 100%

GAP: TAATG overlap

LO: NA

RBS: Kibler7, spacer 15, Z score 2.300, Final -5.098, yes

F: minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage BurtonThePup, ID AVR56191.1, query 1-689 subject 1-689, E value = 0.0, Identity = 99%, Similarity = 100%

SIF-HHPred: 6TPW\_A, Receptor-type tyrosine-protein phosphatase F, Fibronectin type-III, adhesion protein, CELL ADHESION, HET: SO4; 2.9A, Homo sapiens, Probability: 99.72%, E-value: 2.7e-13

SF-Syn: This gene is pham 1068 [minor tail protein], upstream is pham 76903 [minor tail protein], downstream is pham 85094 [minor tail protein], as seen in Microbacterium phage Luxx

**GENE 15 minor tail protein**

Original Glimmer call @bp 12726 has strength 15.95

SSC: 12726 to 13274 (Forward)

CP: yes

SCS: both

ST: SS, Start 3 [12726], Found in 110 of 138 ( 79.7% ) of genes in pham, Called 99.1% of time when present

BLAST-start: aligns with Microbacterium phage Yubaba\_15, minor tail protein, query 1-182 subject 1-182, E value = e-101, Identity = 100%, Similarity = 100%

GAP: 1bp

LO: NA

RBS: Kibler7, spacer 12, Z score 3.266, Final -2.394, yes

F: minor tail protein

SIF-blast: aligns with minor tail protein from Microbacterium phage PaoPu, ID YP\_009996600.1, query 1-182 subject 1-182, E value = 2e-127, Identity = 100%, Similarity = 100%

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

SF-Syn: This gene is pham 85094 [minor tail protein], upstream is pham 1068 [minor tail protein], downstream is pham 1097, as seen in Microbacterium phage Luxx

**INSERTED GENE**

There is an extra gene at this position present in other EE phage [all but 2]. The first 40 aa’s are shared in Charbie:

MTLIEDVLSEQIAALRASAGLFVRQRADALAAAEAAAVQ

Full DNA and amino acid sequence, and coordinates are in gene 16 below

**GENE 16 hypothetical protein**

DNA:

ATGACACTGATCGAGGACGTGCTGAGCGAGCAGATAGCTGCCCTGCGCGCGTCGGCCGGCCTGTTCGTGCGCCAGCGCGCCGACGCCCTCGCAGCGGCCGAGGCTGCAGCCGTGCAGGCTGGCCGGCGCGGATGCCCGAGCCGCTGCGCTGCAGGCCGAGCTGGACGCCTACCGCGAGCTGCTGCCGGCCCCCGAGATCCCCGAGGACGCGCCGGCGGTCGAGCTGCCGGCCGACGAGACCACGCCCCCGCTCTATCAGCAGCTGCTGACAGAGCGTGA

protein:

MTLIEDVLSEQIAALRASAGLFVRQRADALAAAEAAAVQAGRRGCPSRCAAGRAGRLPRAAAGPRDPRGRAGGRAAGRRDHAPALSAAADRA

SSC: 13311 to 13589 [Forward]

**GENE 17 endolysin**

Original Glimmer call @bp 13618 has strength 8.93

SSC: 13618 to 14313 (Forward)

CP: yes

SCS: both

ST: SS, Start 4 [13618], Found in 113 of 115 ( 98.3% ) of genes in pham, Called 97.3% of time when present

BLAST-start: aligns with Microbacterium phage VitulaEligans\_17, endolysin, query 1-231 subject 1-231, E value = e-136, Identity = 100%, Similarity = 100%

GAP: 29bp [to new inserted gene]

LO: yes

RBS: Kibler7, spacer 12, Z score 2.476, Final -3.978, yes

F: endolysin

SIF-blast: aligns with endolysin from Microbacterium phage Hulk, ID QDF16577.1, query 1-231 subject 1-231, E value = 2e-164, Identity = 99%, Similarity = 99%

SIF-HHPred: 8TZL\_E, Peptidase M23, membrane protein, enzyme, TRANSPORT PROTEIN, HET: ADP, Vibrio cholerae, Probability: 99.77%, E-value: 1.2e-17

SF-Syn: NKF [due to gene insertion above]

**GENE 18 membrane protein**

Original Glimmer call @bp 14280 has strength 10.05

SSC: 14310 to 14543 (Forward)

CP: yes

SCS: both - CS

ST: NI, Start 7 [14280], Found in 111 of 114 ( 97.4% ) of genes in pham, Called 85.6% of time when present

Start 10 [14310], Found in 110 of 114 ( 96.5% ) of genes in pham, Called 12.7% of time when present

Although start 14280 is best in Starterator, both 14280 and 14310 align in BLAST with certain genes. However, 14280 gives an overlap of 31bp, 14310 gives a GTGA overlap and has a much better score in RBS. Therefore, **start changed to 14310**.

BLAST-start: aligns with Microbacterium phage TimoTea\_18, function unknown, query 1-87 subject 1-87, E value = 2e-47, Identity = 100%, Similarity = 100%

GAP: GTGA overlap

LO: NA

RBS: Kibler7, spacer 16, Z score 2.148, Final -5.596, yes [score for 14310]

F: membrane protein

SIF-blast: aligns with membrane protein from Microbacterium phage TimoTea, ID QBI97331.2, query 1-87 subject 1-87, E value = 3e-57, Identity = 100%, Similarity = 100%

SIF-HHPred: 8HHF\_B, Cell division protein FtsB, Bacterial cell division, divisome, FtsB, FtsL, FtsQ, FtsBLQ, FtsQLB, membrane protein complex, heterotrimer, MEMBRANE PROTEIN, 3.04A, Escherichia coli, Probability: 95.84%, E-value: 0.28,

SF-Syn: NKF

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

**GENE 19 hypothetical protein**

Original Glimmer call @bp 14540 has strength 11.64

SSC: 14540 to 14764 (Forward)

CP: yes

SCS: both

ST: SS, Start 5 [14540], Found in 114 of 116 ( 98.3% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Yubaba\_19, function unknown, query 1-74 subject 1-74, E value = Expect = 4e-36, Identity = 100%, Similarity = 100%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 11, Z score 2.840, Final -3.171, yes

F: NKF

SIF-blast: aligns with holin from Microbacterium phage PaoPu, ID YP\_009996604.1, query 1-74 subject 1-74, E value = 2e-43, Identity = 100%, Similarity = 100%

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

SF-Syn: NKF

SIF-Mem: 2 domains found by DEEPTMHMM, 1 domain found by SOSUI

most PhagesDB blasts are hypothetical proteins, most NCBI blasts are holins or membrane proteins

**GENE 20 Lsr2-like DNA bridging protein**

Original Glimmer call @bp 15045 has strength 11.47

SSC: 15045 to 14833 (Reverse)

CP: yes

SCS: both

ST: NI, Start 12 [15045], Found in 113 of 148 ( 76.4% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_20, Lsr2-like DNA bridging protein, query 1-70 subject 1-70, E value = 6e-34, Identity = 100%, Similarity = 100%

GAP: 2bp

LO: NA

RBS: Kibler7, spacer 12, Z score 2.651, Final -3.627, yes

F: Lsr2-like DNA bridging protein

SIF-blast: aligns with Lsr2-like DNA bridging protein from Microbacterium phage TimoTea, ID , query 1-70 subject 1-70, E value = QBI97333.1, Identity = 100%, Similarity = 100%

SIF-HHPred: 4E1P\_A, Protein lsr2; anti-parallel beta sheet, dimer, DNA BINDING PROTEIN; 1.728A, Mycobacterium tuberculosis, Probability: 99.86%, E-value: 6.7e-21

SF-Syn: This gene is pham 85067 [Lsr2-like DNA bridging protein], upstream is pham 1053, downstream is pham 1059 [helix-turn-helix DNA binding domain protein], as seen in Microbacterium phage Luxx

**GENE 21 helix-turn-helix DNA binding domain protein**

Original Glimmer call @bp 15485 has strength 11.53; GeneMark calls start at 15548

SSC: 15548 to 15048 (Reverse)

CP: yes

SCS: both

ST: NI, Start 12 [15485], Found in 112 of 115 ( 97.4% ) of genes in pham, Called 15.2% of time when present

Start 6 [15548], Found in 106 of 115 ( 92.2% ) of genes in pham, Called 74.5% of time when present

start 15548 called more often in Starterator. also aligns 1 to 1 with most genes in blast. therefore, start changed to 15548.

BLAST-start: aligns with Microbacterium phage TimoTea\_21, helix-turn-helix DNA binding domain protein, query 1-145 subject 22-166 [will align when start is changed], E value = 2e-76, Identity = 100%, Similarity = 100%

GAP: 81bp

LO: yes

RBS: Kibler7, spacer 13, Z score 2.992, Final -3.115, yes

F: helix-turn-helix DNA binding domain protein

SIF-blast: aligns with helix-turn-helix DNA binding domain protein from Microbacterium phage Luxx, ID QKY78668.2, query 1-145 subject 22-166, E value = 6e-97, Identity = 99%, Similarity = 99%

SIF-HHPred: 4P0Z\_A AM32; HELIX-TURN-HELIX, DSDNA BINDING, DNA BINDING PROTEIN; HET: MSE; 1.35A, Enterococcus faecalis, Probability: 99.13%, E-value: 9.3e-10

SF-Syn: This gene is pham 1059 [helix-turn-helix DNA binding domain protein], upstream is pham 85067 [Lsr2-like DNA bridging protein], downstream is pham 85194 [helix-turn-helix DNA binding domain protein], as seen in Microbacterium phage Luxx

**GENE 22 helix-turn-helix DNA binding domain protein**

Original Glimmer call @bp 15859 has strength 8.96

SSC: 15859 to 15629 (Reverse)

CP: yes

SCS: both

ST: SS, Start 5 [15859], found in 115 of 115 ( 100.0% ) of genes in pham, Called 99.1% of time when present

BLAST-start: aligns with Microbacterium phage Vanisius\_22, helix-turn-helix DNA binding domain protein, query 1-76 subject 1-76, E value = 5e-37, Identity = 100%, Similarity = 100%

GAP: 715bp

LO: no

RBS: Kibler7, spacer 15, Z score 0.932, Final -7.839, no [RBS shows 15889 being stronger but Starterator and BLAST does not confirm]

F: helix-turn-helix DNA binding domain protein

SIF-blast: aligns with helix-turn-helix DNA binding domain protein from Microbacterium phage VitulaEligans, ID AXC34905.1, query 1-76 subject 1-76, E value = 2e-44, Identity = 100%, Similarity = 100%

SIF-HHPred: 3U3W\_A, Transcriptional activator PlcR protein; ternary complex, PlcR-PAPR7-DNA, HTH DNA-binding domain, HTH\_3 (Helix-turn-helix) domain, TPR\_1 (tetratricopeptide repeats), Pleiotropic regulator; 2.4A, Bacillus thuringiensis, Probability: 98.41%, E-value: 0.000017

SF-Syn: This gene is pham 85194 [helix-turn-helix DNA binding domain protein], upstream is pham 1059 [helix-turn-helix DNA binding domain protein], downstream is pham 3206, as seen in Microbacterium phage Scamander

**GENE 23 hypothetical protein**

Original Glimmer call @bp 16574 has strength 7.96

SSC: 16574 to 16756 (Forward)

CP: yes

SCS: both

ST: SS, Start 3 [16574], Found in 27 of 27 ( 100.0% ) of genes in pham, Called 96.3% of time when present

BLAST-start: aligns with Microbacterium phage Minima\_23, function unknown, query 1-60 subject 1-60, E value = 3e-28, Identity = 100%, Similarity = 100%

GAP: 715bp

LO: NA

RBS: Kibler7, spacer 7, Z score 1.954, Final -5.712, yes

F: NKF

SIF-blast: aligns with hypothetical protein from Microbacterium Nobel, ID QJD50747.2, query 1-60 subject 1-60, E value = 4e-33, Identity = 98%, Similarity = 100%

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

SF-Syn: NKF

**GENE 24 helix-turn-helix DNA binding domain protein**

Original Glimmer call @bp 16840 has strength 8.96; GeneMark calls start at 16849

SSC: 16840 to 17058 (Forward)

CP: yes

SCS: both

ST: SS, Start 12 [16840], Found in 106 of 114 ( 93.0% ) of genes in pham, Called 98.1% of time when present

Start 14 [16849], Found in 114 of 114 ( 100.0% ) of genes in pham, Called 2.6% of time when present

BLAST-start: aligns with Microbacterium phage TimoTea\_24, helix-turn-helix DNA binding domain protein, query 1-72 subject 1-72, E value = 4e-36, Identity = 100%, Similarity = 100%

GAP: 84bp

LO: yes

RBS: Kibler7, spacer 9, Z score 1.960, Final -4.952, yes

F: helix-turn-helix DNA binding domain protein

SIF-blast: aligns with helix-turn-helix DNA binding domain protein from Microbacterium phage MrGreen, ID UVK59396.1, query 1-72 subject 1-72, E value = 9e-43, Identity = 97%, Similarity = 98%

SIF-HHPred: 4LHF\_A, Regulatory protein cox; helix-turn-helix, DNA binding, VIRAL PROTEIN; 2.401A , Enterobacteria phage P2, Probability: 99.11%, E-value: 2.4e-9

SF-Syn: This gene is pham 117549 [helix-turn-helix DNA binding domain protein], upstream is pham 3206, downstream is pham 131857 [HNH endonuclease], as seen in Microbacterium phage Scamander

**GENE 25 HNH endonuclease**

Original Glimmer call @bp 17055 has strength 8.09

SSC: 17055 to 17354 (Forward)

CP: yes

SCS: both

ST: NI, Start 31 [17055], Found in 118 of 172 ( 68.6% ) of genes in pham, Called 100.0% of time when present

BLAST-start: aligns with Microbacterium phage Yubaba\_25, HNH endonuclease, query 1-99 subject 1-99, E value = 3e-56, Identity = 100%, Similarity = 100%

GAP: ATGA overlap

LO: NA

RBS: Kibler7, spacer 9, Z score 1.412, Final -6.051, no [all other results show 17055 as start]

F: HNH endonuclease

SIF-blast: aligns with HNH endonuclease from Microbacterium phage PaoPu, ID YP\_009996610.1, query 1-99 subject 1-99, E value = 8e-68, Identity = 100%, Similarity = 100%

SIF-HHPred: 5H0M\_A HNH endonuclease, Thermophilic bacteriophage, HNH Endonuclease, DNA nicking, HYDROLASE; 1.52A, Geobacillus virus E2, Probability: 97.96%,E-value: 0.000041

SF-Syn: This gene is pham 131857 [HNH endonuclease], upstream is pham 117549 [helix-turn-helix DNA binding domain protein], as seen in Microbacterium phage Scamander