****

**Genome Annotation Notebook**

**MIMG 103BL Fall 2016**

Phage Name (and Host): LilPharaoh (*Mycobacterium smegmatis)*

Team 1 Name: PowerPhagers

Team 1 Member’s Names: Hayley Ennis, Joseph Gaballa, Krishna Govindaraju, Stephanie Orchanian

Team 2 Name: Phage Advice

Team 2 Member’s Names: Alysha Salbato, Justin Miller, Vasileious Ragkousis, Rohan Luhar

**Additional Information that needs to be Included in Lab Notebook:**

* Record the name of the savedfasta.file that represents your Phage’s DNA sequence
* Number of Base Pairs: 56167
* Number of Reads used: 1633
* Phage Cluster: K1
* Character of Genome Ends: 3' sticky overhang
* 3’ Overhang Length (basepairs): 11bp
* Number of Genes: 76
* Number of tRNAs: 0. Confirmed with Aragon, tRNAscan SE.
* Gene numbers assigned to annotate per individual: 8-9

**Supporting Argument of Gene Calls**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 V.R | **Final Location:** | 156-353 (FWD) | | | |
|  | **Autoannotation Location:** | 156-353 | | | |
| **Location Rationale:** | There is coding potential starting at 75 bp. In addition, it had a 1:1 ratio with Enkosi (which is most similar to LilPharaoh). Also the RBS score agrees with a 75bp start site. | | | |
| **Location Visual Evidence:** | Inserting image... | | | |
| **Function:** | Unknown function | | | |
| **Function Rationale:** | Stephanie orchanian:  PhagesDB and NCBI agree that the function is unknown while HHPRED's e-values aren't low enough to make a conclusion. Phamerator had no evince to what the protein function was. | | | |
| **Function Visual Evidence:** | Visual evidence provides no information. | | | |
| **Additional Notes:** | One could consider a start site at position 156 due to starterator evidence and the fact that this is the first gene of the genome suggesting that a promoter could be upstream of the gene. | | | |
| 2 JM | **Final Location:** | 443-643 (FWD) | | | |
|  | **Autoannotation Location:** | 443-643 | | | |
|  | **Location Rationale:** | Starterator, coding potential, RBS Score | | | |
|  | **Location Visual Evidence:** | U,{2f882f40-46c5-4618-835d-3591385a6679}{47},15.166666666666666,5.729166666666667    Image  Self  Image  Smeg:  Image  TB:  Image | | | |
|  | **Function:** | Function unknown | | | |
|  | **Function Rationale:** | No piece of evidence could conclusively establish a putative function.No significant alignment with proteins of known function. HHPRED analysis revealed similar regions to non-viral proteins. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 3 AS | **Final Location:** | 636-863 Forward | | | |
|  | **Autoannotation Location:** | 627-863 Forward | | | |
|  | **Location Rationale:** | More conserved start site in Starterator.  Less gene overlap and still reasonably long ORF with good z-value.  Covers more coding potential in GeneMark. | | | |
|  | **Location Visual Evidence:** | Inserting image...Self  MSmeg  MTb | | | |
|  | **Function:** | terminase, small subunit | | | |
|  | **Function Rationale:** | Sequence similarity in NCBI Blast with 1:1 alignment to terminase small subunits. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 4 RL | **Final Location:** | 847-2271 (FWD) | | | |
|  | **Autoannotation Location:** | 847-2271 (Forward Direction)  1425bp | | | |
|  | **Location Rationale:** | Starterator, Glimmer, Genemark, coding potential  Original Glimmer call @bp 847 has strength 11.70 | | | |
|  | **Location Visual Evidence:** | Inserting image...  Genemark: M. tuberculosis    Genemark: M. smeg  Inserting image... | | | |
|  | **Function:** | Terminase, large subunit | | | |
|  | **Function Rationale:** | PhagesDB blast shows 1:1 alignment with terminase (largely with the large subunit). Phamerator shows conserved domain of phage terminase. NCBI BlastX shows high 1:1 alignment with large subunit. HHPred shows high alignment with terminase subunit. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 5 K.G. | **Final Location:** | 2283-3839 (FWD) | | | |
|  | **Autoannotation Location:** | 2283 –3839 (forward)  1572 bp length | | | |
|  | **Location Rationale:** | Starterator, coding potential (HEURISTIC)  Original Glimmer call @bp 2283 has strength 14.26  gene 5:(pham 2281 conserved) - (phamerator) - conserved domain | | | |
|  | **Location Visual Evidence:** | SELF | | | |
|  | **Function:** | Portal Protein | | | |
|  | **Function Rationale:** | PhagesDB blast provided several results with high probablity and E values of 0, that corresponded to portal proteins. NCBI blast and HHpred showed the same thing. Phamerator also showed conservation of the protein among proteins, as well as specifying that it was a part of the portal protein family. | | | |
|  | **Function Visual Evidence:** | PhagesDB Blast    NCBI Blast | | | |
|  | **Additional Notes:** |  | | | |
| 6 (JG) | **Final Location:** | 3784-6339 (Forward)  Length: 2555 | | | |
|  | **Autoannotation Location:** | 3868 – 6339 (Forward)  Length: 2472 base pairs | | | |
|  | **Location Rationale:** | The autoannotated start site does not seem to be conserved among phams, the GeneMarkS output shows coding potential beyond the auto-annotated parameters, the Z-Score of the final location is higher, and the new start site appears to be conserved among phams. It can be seen in the phamerator output of cluster K phages that pham 17305 is conserved. | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | Methyltransferase - capsid maturation protease | | | |
|  | **Function Rationale:** | Methyl accepting chemotaxis protein [cell motility and secretion/ signal transduction mechanisms] – according to phamerator – domain in this protein is well conserved. PhagesDB and NCBI blast provide several results with high E values corresponding to this function. | | | |
|  | **Function Visual Evidence:** | Inserting Picture...  15 | | | |
|  | **Additional Notes:** |  | | | |
| 7 (H.E.) | **Final Location:** | | 6336 – 6524 (FWD) | |
|  | **Autoannotation Location:** | | Start: 6336  Stop: 6524  Length: 189 | |
|  | **Location Rationale:** | | Glimmer call @bp 6336 has strength 7.46  Starterator shows that the predicted start is conserved across many phage genomes.  Phamerator revealed that gene 7 (middle row) is conserved across several cluster K1 phages.  GeneMark showed that there is coding potential across the full length of the gene for these predicted start sites.  NCBI blast revealed many 1 to 1 alignments.  RBS score has highest z-value. | |
|  | **Location Visual Evidence:** | |  | |
|  | **Function:** | | NKF | |
|  | **Function Rationale:** | | This gene is conserved among mycobacteriophages, however it does not have a known function at this time. | |
|  | **Function Visual Evidence:** | |  | |
|  | **Additional Notes:** | |  | |
| 8 S. O.  F: H.E. | **Final Location:** | 6589 – 7137 (FWD) | | | |
|  | **Autoannotation Location:** | start: 6589, end: 7137, 549 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 6589 has strength 13.90.Starterator’s suggested start site is 6589, which would keep the original start site in DNA Master and Glimmer the same. However, the start site 8 is not conserved. Gene 8 is conserved in every genome. GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Pham 13105 | | | |
|  | **Function:** | Scaffolding protein. | | | |
|  | **Function Rationale:** | Phamerator showed that there is synteny across genomes and that the function of all of the genes is scaffolding protein.  PhagesDB blast showed that many related phages have this function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 9 V.R | **Final Location:** | 7725 - 8145 (FWD) | | | |
|  | **Autoannotation Location:** | 7725 | | | |
|  | **Location Rationale:** | All evidence agrees with auto-annotation call (RBS score, Starterator evidence, NCBI blast and coding potential) | | | |
|  | **Location Visual Evidence:** | Self:        S.O.  I agree with this start site. | | | |
|  | **Function:** | Major capsid protein | | | |
|  | **Function Rationale:** | SO:PhagesDB, HHPRED, and NCBI Blast agree that this is major capsid protein. Also it has a conserved domain. Phamerator had evidence that gene 9 was a major capsid protein. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 10 JM | **Final Location:** | 8266 - 8649 (FWD) | | | |
|  | **Autoannotation Location:** | 8266 | | | |
|  | **Location Rationale:** | Starterator, Coding potential, phamerator, NCBI blast hits | | | |
|  | **Location Visual Evidence:** | Image  Self  ImageImage  Smeg  ImageImage  TB  ImageImage | | | |
|  | **Function:** | Head-to-tail connector protein | | | |
|  | **Function Rationale:** |  | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 11 AS | **Final Location:** | 8651-9004 Forward | | | |
|  | **Autoannotation Location:** | 8651-9004 Forward | | | |
|  | **Location Rationale:** | Glimmer and Genemark agreed.  Glimmer had a score of 13.06.  Starterator agreed with the autoannotated start site and was highly conserved. | | | |
|  | **Location Visual Evidence:** | SelfMSmegMTb | | | |
|  | **Function:** | head-to-tail connector protein | | | |
|  | **Function Rationale:** | 1:1 alignment with NCBI and Phages DB blasts of head-to-tail connectors | | | |
|  | **Function Visual Evidence:** | Image | | | |
|  | **Additional Notes:** |  | | | |
| 12 RL | **Final Location:** | 8985-9266 | | | |
|  | **Autoannotation Location:** | 8985-9266 (forward)  282 length | | | |
|  | **Location Rationale:** | The coding potential for M. tb and M. smeg coding potetnail maps showed homology with the given start site of 8985. Starterator also showed conserved start site on track 5 at start site 4 which corresponds to bp 89885. | | | |
|  | **Location Visual Evidence:** | M. tb    M. smeg | | | |
|  | **Function:** | Portal protein | | | |
|  | **Function Rationale:** | PhagesDB Blast shows good alignment with head-to-tail connector (portal protein). Phamerator shows conservation among cluster K1. NCBI BlastX shows good 1:1 alignment with head-to-tail connector proteins. HHPred was not informative. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 13 K.G. | **Final Location:** | 9263 – 9688 (FWD) | | | |
|  | **Autoannotation Location:** | 9263-9777 (forward)  426 bp length | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 9263 has strength 14.37  Coding potential  gene 13:(pham 4865 conserved) - no conserved domain - however there was synteny in all phages compared in phamerator (aldephagia, ananya, emerson, and SamuelLPlaquson) | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | Portal Protein | | | |
|  | **Function Rationale:** | PhagesDB, NCBI, and HHpred all provided hits corresponding to either "head to tail connector proteins" or portal proteins, with E values of zero. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 14 (JG) | **Final Location:** | 9777-10388 (Forward)  Length: 612 base pairs | | | |
|  | **Autoannotation Location:** | 9777 - 10388 (Forward)  Length: 612 base pairs | | | |
|  | **Location Rationale:** | Starterator suggests that 9777 is the starting site of gp14, and the z-value of start position 9777 is the largest. It can also be seen that pham20131 is conserved among cluster K phages. | | | |
|  | **Location Visual Evidence:** | GeneMarkS  LilPharaoh is in the middle of this phamerator output. Compared to other cluster K phages, it can be seen that pham 2031 is conserved. | | | |
|  | **Function:** | Major tail protein | | | |
|  | **Function Rationale:** | gene 14: 1:1 alignment with Enkosi Major tail protein - PhagesDB blast shows that gp14 is likely also a "Major Tail Protein". PhagesDB and NCBI blast provide several results with high E values corresponding to this function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 15 (H.E.) | **Final Location:** | 10504 – 10920 (FWD) | | | |
|  | **Autoannotation Location:** | Start: 10504  Stop: 10920  Length: 417 | | | |
|  | **Location Rationale:** | Glimmer call @bp 10504 has strength 13.47  Starterator predicted the same start site as Glimmer and GeneMark, however this is only conserved in one other phage.  Phamerator showed Gene 15 is conserved in several related phages.  GeneMark coding potential across the whole region of the gene.  NCBI blast shows 1-to-1 alignement.  RBS score has the highest Z-value. | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | Tail Chaperone Protein | | | |
|  | **Function Rationale:** | 1:1 alignment with Enkosi, syntenny in phamerator. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 16 S.O.  F: H.E. | **Final Location:** | Region 1: Start: 10504 end 10885  Region 2: start 10887 end 11327 | | | |
|  | **Autoannotation Location:** | Start: 10977, end: 11327, 351 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 10977 has strength 7.67. Starterator’s suggested start site is 10977, which would keep the original start site in DNA Master and Glimmer the same. The start site 68 is conserved in most genomes. Gene 16 is conserved in all genomes however genes 15 and 16 have overlapping sequences in all genomes but not in LilPharaoh.GeneMark and RBS scores agree with this start site. Its thought that gene 16 has a frame shift mutation. After checking gene 15 with 16 we decided that gene 16 is the frame shift mutant of 15. | | | |
|  | **Location Visual Evidence:** | Inserting image...    Inserting image... | | | |
|  | **Function:** | Tail assembly chaperone | | | |
|  | **Function Rationale:** | Conserved in Phamerator but no conserved domains.  PhagesDB blast shows it's similar to Enkosi and Amelie and that it is a tail assembly chaperone protein.  NCBI shows many hits with the tail assembly chaperone function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 17 V.R | **Final Location:** | 11327 - 15460 (FWD) | | | |
|  | **Autoannotation Location:** | 11327 | | | |
|  | **Location Rationale:** | All evidence agrees with auto-annotation call (RBS score, Starterator evidence, NCBI blast and coding potential) | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian: I agree with this start site. | | | |
|  | **Function:** | Tape measure protein | | | |
|  | **Function Rationale:** | PhagesDB and NCBI agree that this is a tape measure protein and it has a conserved domain. HHPred has no evidence. Phamerator had evidence that gene 17 is a phage related protein. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 18 JM | **Final Location:** | | | 15554 - 16696 (FWD) |
|  | **Autoannotation Location:** | | | 15554 |
|  | **Location Rationale:** | | | Starterator, Coding potential, phamerator, NCBI blast hits |
|  | **Location Visual Evidence:** | | | Inserting image...        Image  Lilpharoah is track 3 (Biglebops)  Self:  Image  Smeg  ImageTb  Image |
|  | **Function:** | | | Minor tail protein (baseplate) |
|  | **Function Rationale:** | | | This ORF was conserved in its pham (Phamerator) and demonstrated good alignment with proteins of known function (minor tail protein) as determined by HHPRED, NCBI-BLAST, PhagesDB BLAST. HHPRED alignments suggest that this is a baseplate protein. |
|  | **Function Visual Evidence:** | | |  |
|  | **Additional Notes:** | | |  |
| 19 AS | **Final Location:** | 16697-18463 Forward | | | |
|  | **Autoannotation Location:** | 16697-18463 Forward | | | |
|  | **Location Rationale:** | It has the longest ORF and best RBS score.  Good coding potential in GeneMark.  Starterator agreed with autoannotation call. | | | |
|  | **Location Visual Evidence:** | MSmeg | | | |
|  | **Function:** | minor tail protein | | | |
|  | **Function Rationale:** | NCBI and phages DB 1:1 blast hits | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 20 | **Final Location:** | 18463 – 18939 (FWD) | | | |
|  | **Autoannotation Location:** | 18463-18393 Forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 18463 has strength 11.29. Starterator, RBS score, and GeneMark coding potential maps agree with this start site. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | Minor tail protein | | | |
|  | **Function Rationale:** | PhagesDB and NCBI BlastX shows good 1:1 alignment with minor tail proteins. All evidence suggests minor tail protein with good probability. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 21 K.G. | **Final Location:** | 19023 – 20105 (FWD) | | | |
|  | **Autoannotation Location:** | 19023-20105 (forward)  Length = 1083 | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 19023 has strength 14.45  gene 21:(pham 22348 conserved) - no conserved domain - however there was synteny in all phages compared in phamerator (aldephagia, ananya, emerson, and SamuelLPlaquson) | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | Minor Tail Protein | | | |
|  | **Function Rationale:** | Phagesdb, and NCBI blast provided hits corresponding to minor tail proteins with E values of zero. HHpred did not provide any useful information. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 22 (JG) | **Final Location:** | 20111-20413 (Forward)  Length: 303 base pairs | | | |
|  | **Autoannotation Location:** | 20111- 20413 (Forward)  Length: 303 base pairs | | | |
|  | **Location Rationale:** | The GeneMarkS output shows coding potential in this region. It can be seen in the phamerator output of cluster K phages that pham 16195 is conserved. Starterator suggests that 20111 is the start site of the gene, and the z-score of 20111 is the highest among the possible start sites. There is no evidence to suggest that the auto-annotated start site is incorrect. | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | Minor tail protein | | | |
|  | **Function Rationale:** | gene 22: 1:1 alignment with Enkosi, Biglebops, and Amelie Minor tail protein - PhagesDB blast shows that gp22 is likely also a "Minor Tail Protein". PhagesDB and NCBI blast provide several results with high E values corresponding to this function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 23 (H.E.) | **Final Location:** | 20414 – 22900 (FWD) | | | |
|  | **Autoannotation Location:** | Start: 20414  Stop: 22900  Length: 2487 | | | |
|  | **Location Rationale:** | Glimmer call @bp 20414 has strength 12.83  Starterator predicted the same start site, which is highly conserved across many related phage genomes.  Phamerator showed that the gene's size and location are conserved throughout related phages, but that it belongs to a different phamily.  GeneMark coding potential covers the whole region.  NCBI blast shows that there are many 1-to-1 sequences.  RBS score has the highest Z-value. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Minor Tail Protein | | | |
|  | **Function Rationale:** | A PhagesDP BLASTp revealed a 1:1 alignment with phage Enkosi. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 24 S. O. | **Final Location:** | 22900 – 23907 (FWD) , 822 long | | | |
|  | **Autoannotation Location:** | start: 23086, end: 23907, 822 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 23086 has strength 8.81; GeneMark calls start at 22861.Starterator’s suggested start site is 58, 22900, which is different than Glimmer and Genemark. The start site 58 is conserved is most genomes.Gene 24 is conserved location wise and length wise however it is not conserved by its PHAM. RBS scores agree with this start site of 22900. It has a 1 bp overlap but NCBI shows that there are multiple 1 to 1s. | | | |
|  | **Location Visual Evidence:** | Inserting image...    Inserting image... | | | |
|  | **Function:** | Minor tail protein | | | |
|  | **Function Rationale:** | Phamerator shows the gene is conserved and that it has a minor tail protein function.  Phages DB blast shows that it is a minor tail protein as well with hits with an e-value of 0. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 25 V.R | **Final Location:** | 23944 – 24318 ( FWD) | | | |
|  | **Autoannotation Location:** | 23944-24318 F | | | |
|  | **Location Rationale:** | The start site is conserved within its pham and Genemark, Glimmer and coding potential evidence corroborate this. Therefore the start sits 23944 is a sound location call. | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian: I agree with this start site. PhagesDB and NCBI agree that the function is unknown. HHPred has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | SO  PhagesDB and NCBI agree that the function is unknown. HHPred has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function Visual Evidence:** | Visual evidence provides no information. | | | |
|  | **Additional Notes:** |  | | | |
| 26 JM | **Final Location:** | 24344 - 25981 (FWD) | | | |
|  | **Autoannotation Location:** | 24344 | | | |
|  | **Location Rationale:** | All evidence agrees with autoannotation site (coding potential, NCBI blast, phamerator) - 24699 is more conserved according to starterator, but the NCBI blast hits are consistently 1:1 | | | |
|  | **Location Visual Evidence:** | Self:    Smeg:    TB: | | | |
|  | **Function:** | Lysin A | | | |
|  | **Function Rationale:** | Very conserved in its pham and extremely good alignment with known proteins (Lysin A) as determined by HHPRED, NCBI BLAST, PhagesDB BLAST | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 27 AS | **Final Location:** | 25978-26736 Forward | | | |
|  | **Autoannotation Location:** | 25978-26736 Forward | | | |
|  | **Location Rationale:** | Autoannotation start has best RBS score, NCBI shows good 1:1 allignment, start site seems to be conserved on Starterator. | | | |
|  | **Location Visual Evidence:** | MsmegMTbSelf | | | |
|  | **Function:** | lysin B | | | |
|  | **Function Rationale:** | 1:1 blast hits with phages in cluster | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 28 | **Final Location:** | 26747 – 27205 (FWD) | | | |
|  | **Autoannotation Location:** | 26747-27205 Forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 26747 has strength 13.00. GeneMark coding potential maps agree with this call, as does Starterator, and RBS score. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | Holin | | | |
|  | **Function Rationale:** | PhagesDB Blast suggested holin.  NCBI BlastX suggests holin.  HHPred suggests holin.  Holin determined function of gene. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 29 K.G. | **Final Location:** | 27202 – 27549 (FWD) | | | |
|  | **Autoannotation Location:** | (27202-27549) (forward) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 27202 has strength 13.00  gene 29:(pham 3081 conserved) - no conserved domain - however there was synteny in all phages compared in phamerator (aldephagia, ananya, emerson, and SamuelLPlaquson) | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Hypothetical Protein | | | |
|  | **Function Rationale:** | NCBI and Phagesdb Blast showed conservation of the gene among phages, however there was no provided function. HHpred provided hits corresponding to RHO guanine nucleotide transferases, however they had E values of 0.16 and 0.19 and 95% similarity. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 30 (JG) | **Final Location:** | 27546- 27797 (Forward)  Length: 252 | | | |
|  | **Autoannotation Location:** | 27546- 27797 (Forward)  Length: 252 | | | |
|  | **Location Rationale:** | The GeneMarkS output shows coding potential in this region. It can be seen in the phamerator output of cluster K phages that pham 22077 is conserved. Starterator suggests that 27546 is the start site of the gene, and the z-score of 27546 is the highest among the possible start sites. There is no evidence to suggest that the auto-annotated start site is incorrect. | | | |
|  | **Location Visual Evidence:** | Inserting Picture...Inserting Picture... | | | |
|  | **Function:** | UNKNOWN FUNCTION | | | |
|  | **Function Rationale:** | gene 30: 1:1 alignment with Enkosi, Biglebops, and Amelie but function UNKNOWN in all examined phages with pairwise alignment - PhagesDB blast shows that gp30 has UNKNOWN FUNCTION | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 31 (H.E.) | **Final Location:** | 27784 – 28920 (FWD) | | | |
|  | **Autoannotation Location:** | Start: 27784  Stop: 28920  Length: 1137 | | | |
|  | **Location Rationale:** | Glimmer call @bp 27784 has strength 15.40  Starterator agrees with the predicted start site and it is highly conserved across many phage genomes.  Phamerator showed that this gene is conserved across several K1 phages, and that it belongs to the same phamily.  GeneMark coding potential across the whole region.  RBS value has the highest Z-score.  NCBI blast shows many 1-to-1 alignments. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Exonuclease | | | |
|  | **Function Rationale:** | There is a 1:1 query alignment between LilPharaoh and LindNT which has a called function of an exonuclease. This is an approved seaphages function, LilPharaoh gp\_31 is being called as an endonuclease. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 32 S.O.  F: HE | **Final Location:** | 29116 - 29529 (FWD) , 414 long | | | |
|  | **Autoannotation Location:** | start: 29116, end: 29529, 414 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 29116 has strength 9.30. Starterator’s suggested start site is 29116, which would keep the original start site in DNA Master and Glimmer the same. The start site 15 is conserved in most genomes. Gene 32 is conserved in every genome. GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Inserting image... | | | |
|  | **Function:** | Helix-turn-helix DNA binding domain | | | |
|  | **Function Rationale:** | Conserved in phamerator, but the function is unknown for all of the genes.  NCBI showed that it has 1-to-1 coverage for a helix-turn-helix DNA binding domain found in HedwigODU.  HHPRED shows that is has 70% coverage for a helix-turn-helix binding domain. | | | |
|  | **Function Visual Evidence:** | Inserting Picture... | | | |
|  | **Additional Notes:** |  | | | |
| 33 V.R | **Final Location:** | 29526 – 29681 (FWD) | | | |
|  | **Autoannotation Location:** | 29532 | | | |
|  | **Location Rationale:** | starterator and RBS agree with a call site of 29526 and there are multiple 1:1 also the overlap may be because of an operon | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | SO  PhagesDB and NCBI agree that it has an unknown function. HHPRED has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function Visual Evidence:** | Visual evidence provides no information. | | | |
|  | **Additional Notes:** |  | | | |
| 34 JM | **Final Location:** | 29822 - 30511 (FWD) | | | |
|  | **Autoannotation Location:** | 29822 | | | |
|  | **Location Rationale:** | Everything agrees with original start site. | | | |
|  | **Location Visual Evidence:** | Self:    Smeg.:    TB: | | | |
|  | **Function:** | no known function | | | |
|  | **Function Rationale:** | Highly conserved ORF in its pham but no significant alignment with proteins of known function or domain similarity as determined by | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 35 AS | **Final Location:** | 30629-31519 (Forward) | | | |
|  | **Autoannotation Location:** | 30629-31519 Forward | | | |
|  | **Location Rationale:** | There is good coding potential on GeneMark, it has good RBS score and longest ORF. Both Glimmer and GeneMark agree. | | | |
|  | **Location Visual Evidence:** | MSmegMTbSelf | | | |
|  | **Function:** | tRNA nucleotidyltransferase | | | |
|  | **Function Rationale:** | HH-pred found good coverage of a tRNA nucleotidyltransferase as well as Blast hits agreeing with functional call of it being a queuine tRNA-glycosyltransferase | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 36 | **Final Location:** | 31640 – 33070 (FWD) | | | |
|  | **Autoannotation Location:** | 31640-33070 Forward | | | |
|  | **Location Rationale:** | Autoannotation called 31640 as the start site. Original Glimmer call @bp 31640 has strength 12.47, while GeneMark calls start at 31976. The highest RBS score was at 31640 with 2.110, but the RBS for 31976 was still positive at 0.769. However coding potential maps showed a more conserved start site at 31640, and Starterator showed conservation to one other longer genome phage at start site 19, which corresponds to 31640. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | Integrase (y-int) | | | |
|  | **Function Rationale:** | PhagesDB Blast suggests integrase (y-int)  NCBI Blast also suggested integrase.  HHPred suggested integrase, but various forms. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 37 K.G. | **Final Location:** | 34082 – 33180 (REV) | | | |
|  | **Autoannotation Location:** | 34082-33180 (reverse) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 34082 has strength 11.85  gene 37:(pham 7575 NOT conserved) - synteny was observed but in other phages examined in phamerator (aldephagia, ananya, emerson, and SamuelLPlaquson) reverse coding genes that aligned with gene 37in LilPharaoh (were in the same relative region in other K1 phages examined) had reverse coding genes belonging to pham 1438 | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Hypothetical protein | | | |
|  | **Function Rationale:** | Phagesdb and NCBI blast results showed conservation of the protein, however it did not possess a known function. HHpred also did not provide any clear information. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 38 (JG) | **Final Location:** | 34255 – 34097 (Reverse) | | | |
|  | **Autoannotation Location:** | 34097- 34255 (Reverse)  Length: 159 base pairs | | | |
|  | **Location Rationale:** | Given that this gene codes in the reverse direction, that it is conserved within other cluster K phages and Glimmer shows that it has coding potential, there is not sufficient evidence to change the auto-annotated location call. Other putative ORF's result in a gene less than 70 base pairs, or with a very low Z-Scores. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Immunity repressor | | | |
|  | **Function Rationale:** | gene 38: 1:14 alignment with Biglebops, Amelie, Enkosi and Peanam; Functional call for Enkosi and Amelie is Immunity repressor – hypothesized function of LilPharaoh gp38 is Immunity repressor. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 39 (H.E.) | **Final Location:** | 34518 – 34712 (FWD) | | | |
|  | **Autoannotation Location:** | Start: 34518  Stop: 34712  Length: 195 | | | |
|  | **Location Rationale:** | Glimmer call @bp 34518 has strength 5.92. No genemark call.  Starterator called the same start site however it is only conserved in a couple other phages.  Phamerator showed that this gene is not conserved in other K1 phages, and it also appears that there is a large gap.  Not the longest possible ORF, but there is not coding potential for the earlier start, and there is coding potential for the later start.  NCBI blast shows one 1-to-1 alignment.  3rd Highest RBS score. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | helix-turn-helix DNA binding domain | | | |
|  | **Function Rationale:** | This gene has a 1:1 alignment with Enkosi which has a called function of a helix-turn-helix DNA binding protein. HHPred shows that it has homology with another HTH DNA-Binding motif. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 40 S.O.  F: H.E. | **Final Location:** | 34984 - 35289 (FWD) , 306 long | | | |
|  | **Autoannotation Location:** | start: 34984, end: 35289, 306 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 34984 has strength 10.82. Starterator’s suggested start site is 34984, which would keep the original start site in DNA Master and Glimmer the same. However start site 14 is not conserved in most genomes. Gene 40 is conserved in every genome. GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | Phamerator did not show a known function for the gene, but it is conserved.  PhagesDB BLAST all showed unknown function.  NCBI shows 100% alignment with Enkosi, but it doesn't have a function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 41 V.R | **Final Location:** | 35353-35583 | | | |
|  | **Autoannotation Location:** | 35353 | | | |
|  | **Location Rationale:** | The autoannotation call (35353) is supported by evidence generated by all software | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian: I agree with this start site. PhagesDB and NCBI agree the protein has an unknown function. HHPred has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function:** | Unknown Function | | | |
|  | **Function Rationale:** | SO  PhagesDB and NCBI agree the protein has an unknown function. HHPred has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function Visual Evidence:** | Visual Evidence provides no information. | | | |
|  | **Additional Notes:** |  | | | |
| 42 JM | **Final Location:** | 35580-35756 FWD | | | |
|  | **Autoannotation Location:** | 35580 | | | |
|  | **Location Rationale:** | All evidence agreed with Glimmer and GeneMark start sites | | | |
|  | **Location Visual Evidence:** | Self:    Smeg:    TB: | | | |
|  | **Function:** | no known function | | | |
|  | **Function Rationale:** | No software provides conclusive evidence to assign a function to this gene. No alignment to proteins of known function. Poor HHPRED alignments of low coverage and some conservation observed using Phamerator | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 43 AS | **Final Location:** | 35753-35956 Forward | | | |
|  | **Autoannotation Location:** | 35753-35956 Forward | | | |
|  | **Location Rationale:** | There is good coding potential. Glimmer, GeneMark, and Starterator all agree. NCBI doesn't show the greatest allignment but it isn't significant enough to suggest changing the start site. And changing the start site didn't improve allignment. | | | |
|  | **Location Visual Evidence:** | MSmegMTbSelf | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No known function of blast hits and no CDD entries or hhpred results of significance | | | |
|  | **Function Visual Evidence:** | No function listed for blast hits or CDD or HHpred hits of significance or related function | | | |
|  | **Additional Notes:** |  | | | |
| 44 RL | **Final Location:** | 35593-36771 Forward | | | |
|  | **Autoannotation Location:** | 35593-36771 Forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp has strength at 16.25. RBS score, GeneMark coding potential maps, and Starterator agree with the Glimmer call. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | HTH DNA binding protein | | | |
|  | **Function Rationale:** | PhagesDB Blast suggested HTH binding protein.  NCBI BlastX suggested HTH DNA binding protein  HHPred suggested excisionase.  HTH DNA binding protein determined function of gene. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 45 K.G. | **Final Location:** | 36768-37037 Forward | | | |
|  | **Autoannotation Location:** | 36768-37034 (forward) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 36768 has strength 5.50  gene 45:(pham 18003 NOT conserved) - this gene contained a conserved protein domain known as WhiB: WhiB is a transcription factor in Actinobacteria required for differentiation and sporulation  In Adephagia the protein with this domain belongs to the pham 22430, In Ananya the protein with this domain belongs to the pham 22430, In Emerson the protein with this domain belongs to the pham 22430, In SamuelLPlaqson the protein with this domain belongs to the pham 22430. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | WhiB family transcription factor | | | |
|  | **Function Rationale:** | PhagesDB and NCBI blast show hits for the conservation of this gene among related phages corresponding to WhiB-like transcription factor proteins. NCBI blast also shows a conserved domain for this gene. Phamerator also showed the conserved domain. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 46 (JG) | **Final Location:** | 37083-37459 (Forward) | | | |
|  | **Autoannotation Location:** | 37034- 37459 (Forward)  Length: 426 base pairs | | | |
|  | **Location Rationale:** | It can be seen in the Glimmer output that there is coding potential in this region. Syntenny is conserved within the pham as seen in the phamerator output of cluster K1 phages indicating that this is a real gene. In starterator it can be seen that track 8 is conserved within the pham, thus the predicted start site is 37083. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | UNKNOWN FUNCTION | | | |
|  | **Function Rationale:** | 1:1 alignment with Enkosi, Biglebops, and Amelie but function UNKNOWN in all examined phages with pairwise alignment - PhagesDB blast shows that gp46 has UNKNOWN FUNCTION | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 47 (H.E.) | **Final Location:** | Start: 37473  Stop: 37760 | | | |
|  | **Autoannotation Location:** | Start: 37473  Stop: 37760  Length: 288 | | | |
|  | **Location Rationale:** | Glimmer call @bp 37473 has strength 17.22  Starterator called the same start location and it is highly conserved across most phage genomes.  Phamerator showed that this gene is conserved throughout many K1 phages, but is is earlier in the genome than other K1 phages.  GeneMark shows coding potential across the whole region.  Second best RBS score. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | no known function | | | |
|  | **Function Rationale:** | Not enough evidence to make a call. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 48 | **Final Location:** | start: 37775, end: 38326, 552 long | | | |
|  | **Autoannotation Location:** | start: 37775, end: 38326, 552 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 37775 has strength 14.23.Starterator’s suggested start site is 37775, which would keep the original start site in DNA Master and Glimmer the same. However, start site 16 is not conserved in most genomes. Gene 48 is similar to genes in other phages by its conserved sequences and length but it is not conserved by its PHAM.GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Inserting image...Inserting image... | | | |
|  | **Function:** | exonuclease | | | |
|  | **Function Rationale:** | PhagesDB showed phages with e-value of 0 with exonuclease function.  HHPred showed many domains encoding DNA pol III which is involved in exonuclease function.  Phamerator showed it's in the exonuclease pham domain. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 49 V.R | **Final Location:** | 38323-39210 | | | |
|  | **Autoannotation Location:** | 38323 | | | |
|  | **Location Rationale:** | Starterator revealed that 38323 is highly conserved within its pham and thus is the strongest start site candidate. | | | |
|  | **Location Visual Evidence:** | SO  Stephanie Orchanian: I agree with this start site | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | SO  PhagesDB and NCBI agree that it has an unknown function. HHPRED has no evidence. Phamerator had no evidence as to what the gene product is. | | | |
|  | **Function Visual Evidence:** | Visual evidence provides no information | | | |
|  | **Additional Notes:** |  | | | |
| 50 JM | **Final Location:** | 39207-39401 (FWD) | | | |
|  | **Autoannotation Location:** | 39207 | | | |
|  | **Location Rationale:** |  | | | |
|  | **Location Visual Evidence:** | Self:    Smeg:    TB: | | | |
|  | **Function:** | No known function | | | |
|  | **Function Rationale:** | No alignment with proteins of known function, no conservation within its pham, poor HHPRED alignments and generally no evidence to suggest putative function | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 51 AS | **Final Location:** | 39691-39852 Forward | | | |
|  | **Autoannotation Location:** | 39691-39852 Forward | | | |
|  | **Location Rationale:** | Glimmer and GeneMark agreed on the start site.  Starterator also agreed. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No known function of blast hits and no CDD entries or hhpred results of significance | | | |
|  | **Function Visual Evidence:** | * No blast hits of significance or of function | | | |
|  | **Additional Notes:** |  | | | |
| 52 | **Final Location:** | 39849-39968 F | | | |
|  | **Autoannotation Location:** | 39849-39968 Forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp (39849-39968) has strength 8.18. Starterator, RBS and GeneMark Coding Potential maps agreed with the call. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No evidence shows alignment with proteins of known function. | | | |
|  | **Function Visual Evidence:** | Visual evidence is not informative. | | | |
|  | **Additional Notes:** |  | | | |
| 53 K.G. | **Final Location:** | 40052-40573 | | | |
|  | **Autoannotation Location:** | 40052-40574 (forward) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 40052 has strength 14.82  gene 53:(pham 14923 conserved) – synteny is observed, but in all other cluster K phages examined, the genes are much farther downstream (probably due to the fact that LilPharaoh has a much smaller genome when compared to other cluster K phages). | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | No known function | | | |
|  | **Function Rationale:** | NCBI and Phages DB blast showed many high probability hits among other phages, however their functions were unknown. Along with that, HHpred showed no similarity to other phages for this gene. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 54 (JG) | **Final Location:** | 40669- 40914 (forward) | | | |
|  | **Autoannotation Location:** | 40669- 40914 (Forward)  Length: 246 base pairs | | | |
|  | **Location Rationale:** | The Glimmer output shows coding potential in this region which indicates that this is a real gene. It can be seen that the syntenny is gp54 is conserved within the pham. The starterator suggested start site is also the same site called by auto-annotation. At this time there is not sufficient evidence to change the location called by DNA master. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | NrdH-redoxin | | | |
|  | **Function Rationale:** | Phamerator output indicates that gp54 has a conserved domain present in related phages with synteny. PhagesDB and NCBI blast provide several results with high E values corresponding to this function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 55 (H.E.) | **Final Location:** | Start: 40911  Stop: 41282 | | | |
|  | **Autoannotation Location:** | Start: 40911  Stop: 41282  Length: 372 | | | |
|  | **Location Rationale:** | Glimmer call @bp 40911 strength 9.08  Starterator agreed on the start call, but it is only conserved in a few of the phage genomes.  Phamerator showed that this gene is conserved throughout many K1 phages even though it is upstream compared to the other phages.  GeneMark shows that there is coding potential for this whole region, and no coding potential for the earlier possible starts.  Does not have the best RBS score. | | | |
|  | **Location Visual Evidence:** | Inserting Picture... | | | |
|  | **Function:** | HNH endonuclease | | | |
|  | **Function Rationale:** | Phage Enkosi has 1:1 query alignment with this gene and the predicted function is HNH endonuclease. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 56 S.O.  F: H.E. | **Final Location:** | start: 41322, end: 43919, 2598 long | | | |
|  | **Autoannotation Location:** | start: 41322, end: 43919, 2598 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 41322 has strength 15.45. Starterator’s suggested start site is 41322, which would keep the original start site in DNA Master and Glimmer the same. Start site 8 is conserved in most genomes.Gene 48 is similar to genes in other phages by its conserved sequences and length but it is not conserved by its PHAM. GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Inserting image...    Inserting image... | | | |
|  | **Function:** | DNA primase | | | |
|  | **Function Rationale:** | PhagesDB showed three hits with e-values of 0 with DNA primase function.  HHPred also showed a hit with a e-value of 0 with primase activity.  Phamerator also showed a conserved domain with primase activity. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 57 V.R | **Final Location:** | 44353-45063 | | | |
|  | **Autoannotation Location:** | Glimmer: 44314 Genemark: 44353 | | | |
|  | **Location Rationale:** | 44353 is the best candidate since this start site is the LORF, is highly conserved within its pham and demonstrates 1:1 alignment with known genes | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian:  Starterator and RBS say 44353. NCBI has multiple 1:1 ratios. | | | |
|  | **Function:** | RusA resolvase (endonuclease) | | | |
|  | **Function Rationale:** | SO  PhagesDB and NCBI agree that this has an unknown function. HHPRED has no evidence. This protein has conserved domains. Phamerator had evidence that this gene product is a endodeozyribonuclease.When running NCBI many organisms show that this gene product is a RusA resolvase. | | | |
|  | **Function Visual Evidence:** | Inserting image... | | | |
|  | **Additional Notes:** |  | | | |
| 58 JM | **Final Location:** | 45056-45466 (FWD) | | | |
|  | **Autoannotation Location:** | 45056 | | | |
|  | **Location Rationale:** | Start site mostly conserved and 1:1 NCBI blast hits. Coding potential doesn't extend beyond that start site despite there being earlier sites. | | | |
|  | **Location Visual Evidence:** | Self:    Smeg:    TB: | | | |
|  | **Function:** | no known function | | | |
|  | **Function Rationale:** | Some conservation within its pham, no alignment with proteins of known function, poor HHPRED alignments and overall no evidence for putative function | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 59 AS | **Final Location:** | 45463-46644 Forward | | | |
|  | **Autoannotation Location:** | 45463-46644 Forward | | | |
|  | **Location Rationale:** | Glimmer, GeneMark, and Starterator all called 45463. There is good coding potential. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | HTH DNA binding domain | | | |
|  | **Function Rationale:** |  | | | |
|  | **Function Visual Evidence:** | Inserting Picture... | | | |
|  | **Additional Notes:** |  | | | |
| 60 | **Final Location:** | 46654-46887 F | | | |
|  | **Autoannotation Location:** | 46654-46887 forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 46654 has strength 11.21. Starterator, RBS score, and Genemark Coding Potential maps agreed with this call. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No evidence aligned to proteins with known functions. | | | |
|  | **Function Visual Evidence:** | Visual evidence not informative. | | | |
|  | **Additional Notes:** |  | | | |
| 61 K.G | **Final Location:** | 46884-47156(forward) | | | |
|  | **Autoannotation Location:** | 46884-47281 (forward) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 46884 has strength 8.77; GeneMark calls start at 46869  gene 61:(pham 4447 conserved) - synteny is observed, but in all other cluster K phages examined, the genes are much farther downstream (probably due to the fact that LilPharaoh has a much smaller genome when compared to other cluster K phages). | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | No known function | | | |
|  | **Function Rationale:** | Phagesdb, NCBI blast, and HHpred provided no clear confirmation was to what this protein is. However, there was a slight conservation and 100% sequence alignment on NCBI blast. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 62 (JG) | **Final Location:** | 47281-47781 (forward) | | | |
|  | **Autoannotation Location:** | 47281- 47781 (Forward)  Length: 501 base pairs | | | |
|  | **Location Rationale:** | It can be seen in the Glimmer output that there is coding potential in this region indicating that there is infact a gene at this location. The syntenny of gp62 is conserved within the same pham. The starterator output of LilPharaoh was not conclusive as there are not many known genes within pham 6779. At this time there is not sufficient evidence to change the auto-annotated start site of gp62. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | UNKNOWN FUNCTION | | | |
|  | **Function Rationale:** | gene 62: 50% sequence similarity with Enkosi, Biglebops, and Amelie but function UNKNOWN in all examined phages with pairwise alignment - PhagesDB blast shows that gp62 has UNKNOWN FUNCTION | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 63 (H.E.) | **Final Location:** | Start: 47870  Stop: 48046 | | | |
|  | **Autoannotation Location:** | Start: 47870  Stop: 48046  Length: 177 | | | |
|  | **Location Rationale:** | Glimmer call @bp 47870 has strength 14.43  Starterator called the same start site, which is conserved across both other phages.  Phamerator revealed that this gene is not conserved across other K1 phage genomes, although it appears similar to other genes but is in a different phamily.  There is coding potential across this whole region, and not coding potential for the next gene to be extended, suggesting that this is a real gene.  A start site at 47921 has a better RBS score, but there is clearly coding potential for the start prior to this one. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | no known function | | | |
|  | **Function Rationale:** | There is not enough evidence to make a functional call on this gene. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 64 | **Final Location:** | 48043-48912 forward | | | |
|  | **Autoannotation Location:** | start: 48043, end: 48912, 870 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 48043 has strength 13.23. Starterator’s suggested start site is 48043, which would keep the original start site in DNA Master and Glimmer the same. However, start site 14 is not conserved in most genomes. Gene 64 is conserved in every genome.  GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Inserting image... | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | PHamerator shows it's conserved but none of the functions are known.  PhagesDB showed that there are many hits with e-value's of 0 but the function is unknown for all of them.  The significant hits on HHPred and CDD also have unknown function. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 65 V.R | **Final Location:** | 48909-49637 | | | |
|  | **Autoannotation Location:** | 48909 | | | |
|  | **Location Rationale:** | Although starterator does not agree with 48909, there is high coding potential there. There is also 4bp overlap which could be due to an operon. The RBS score also agrees with this. PhagedDB and NCBI agree that this has an unknown function. | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian:  HHPRED has no evidence. This protein has conserved domains. Phamerator had evidence that this gene product is a polynucleotide kinase. | | | |
|  | **Function:** | Polynucleotide kinase | | | |
|  | **Function Rationale:** | PhagedDB and NCBI agree that this has an unknown function. HHPRED has no evidence. This protein has conserved domains. Phamerator had evidence that this gene product is a polynucleotide kinase. | | | |
|  | **Function Visual Evidence:** | Visual evidence provides no information. | | | |
|  | **Additional Notes:** | 48909 and 49188 start sites have similar RBS scores, coding potential and NCBI hits 49188 is highly conserved within pham 22527 and thus Starterator evidence suggests that the latter start site is more appropriate | | | |
| 66 JM | **Final Location:** | 49760-50953 - FWD | | | |
|  | **Autoannotation Location:** | 49760 | | | |
|  | **Location Rationale:** | All evidence agreed with Glimmer and GeneMark start sites | | | |
|  | **Location Visual Evidence:** | (Lilpharaoh on Amelie track (5))  Self:    Smeg:    TB: | | | |
|  | **Function:** | RctB RNA ligase | | | |
|  | **Function Rationale:** | Extremely good alignment with RctB RNA ligases in PhagesDB BLAST and NCBI BLAST as well as HHPRED alignment with the same protein. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 67 AS | **Final Location:** | 50950-51600 Forward | | | |
|  | **Autoannotation Location:** | 50956-51600 Forward | | | |
|  | **Location Rationale:** | Glimmer suggested a start of 5956 and GeneMark suggested 50950. There is better 1:1 allignment on NCBI BlastX. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No known function of blast hits and no CDD entries or hhpred results of significance | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 68 | **Final Location:** | 51593-52153 F | | | |
|  | **Autoannotation Location:** | 51593-52153 Forward | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 51593 has strength 12.63. Starterator, RBS score and GeneMark all show conserved start sites as well as coding potential. | | | |
|  | **Location Visual Evidence:** | M. tb Coding Potential    M. Smeg Coding Potential | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No evidence showed alignment with proteins of known function. | | | |
|  | **Function Visual Evidence:** | Visual evidence was not informative. | | | |
|  | **Additional Notes:** |  | | | |
| 69 K.G. | **Final Location:** | 52146-52304 | | | |
|  | **Autoannotation Location:** | 52146-52537 (159) | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 52146 has strength 13.68  gene 69:(pham 4754 NOT conserved) - synteny is observed between LilPharaoh and Emerson, but in all other cluster K phages examined a protein belonging to this pham is absent. The gene is much farther downstream (probably due to the fact that LilPharaoh has a much smaller genome when compared to other cluster K phages). | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | No known function | | | |
|  | **Function Rationale:** | NCBI blast verified presence of the gene, however with phagesdb and HHpred there was nothing that could confirm the function of the gene. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 70 (JG) | **Final Location:** | 52537-53082 (Forward) | | | |
|  | **Autoannotation Location:** | 52537- 53082 (Forward)  Length: 546 base pairs | | | |
|  | **Location Rationale:** | The Glimmer output shows that there is coding potential in this region indicating that there is a real gene in this location. In the phamerator output of cluster K1 phages, it can be seen that syntenny is conserved within the pham. From the starterator output it can be seen that the start site is also conserved (green track). The auto annotated start site also has the highest z-value and is the same start site called by starterator. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | UNKNOWN FUNCTION | | | |
|  | **Function Rationale:** | gene 70: 50% sequence similarity with Enkosi, Biglebops, and Amelie but function UNKNOWN in all examined phages with pairwise alignment - PhagesDB blast shows that gp70 has UNKNOWN FUNCTION | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 71 (H.E.) | **Final Location:** | Start: 53190  Stop: 53390 | | | |
|  | **Autoannotation Location:** | Start: 53190  Stop: 53390  Length: 201 | | | |
|  | **Location Rationale:** | Glimmer call @bp 53190 has strength 13.98  Starterator called the same start site that is highly conserved across genomes.  Phamerator showed that this gene is conserved in the genome, but not in the exact location.  GeneMark shows that there is coding potential across the whole genome.  This gene does not have the best RBS score, but there is not coding potential for the earlier starts. | | | |
|  | **Location Visual Evidence:** |  | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | There is not enough evidence to make an accurate functional call | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 72 S.O.  F: H.E. | **Final Location:** | start: 53513, end: 53728, 216 long | | | |
|  | **Autoannotation Location:** | start: 53513, end: 53728, 216 long | | | |
|  | **Location Rationale:** | Original Glimmer call @bp 53513 has strength 13.23. Starterator’s suggested start site is 53513, which would keep the original start site in DNA Master and Glimmer the same. Start site 19 is conserved in most genomes. Gene 72 is conserved in every genome. GeneMark and RBS scores agree with this start site. | | | |
|  | **Location Visual Evidence:** | Inserting image... | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | Phamerator, NCBI, PhagesDB, HHPred, and CDD all suggest that the function is unknown. | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 73 V.R | **Final Location:** | 53968-54387 | | | |
|  | **Autoannotation Location:** | 53968 | | | |
|  | **Location Rationale:** | 53968 is the finalized start site since it produces the LORF, it is highly conserved within the its pham and aligns with known genes from other phages | | | |
|  | **Location Visual Evidence:** | Stephanie Orchanian  I agree with this start site. | | | |
|  | **Function:** | Unknown function | | | |
|  | **Function Rationale:** | SO  PhagedDB and NCBI agree that the function is unknown. HHPRED has no evidence. Phamerator has no evidence as to what this gene product is. | | | |
|  | **Function Visual Evidence:** | Visual evidence provides no information. | | | |
|  | **Additional Notes:** |  | | | |
| 74 JM | **Final Location:** | 54497-55198 FWD | | | |
|  | **Autoannotation Location:** | 54497 | | | |
|  | **Location Rationale:** | All evidence agreed with Glimmer and GeneMark start sites | | | |
|  | **Location Visual Evidence:** | Self:    Smeg:    TB: | | | |
|  | **Function:** | DNA binding domain | | | |
|  | **Function Rationale:** | HHPRED data and PhagesDB alignment suggest that gp74 has a DNA binding domain | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 75 AS | **Final Location:** | 55582-55800 Forward | | | |
|  | **Autoannotation Location:** | 55582-55800 Forward | | | |
|  | **Location Rationale:** | Glimmer and GeneMark agreed.  Reasonable RBS and 2nd longest ORF.  Demonstrated good coding potential. | | | |
|  | **Location Visual Evidence:** | MSmeg | | | |
|  | **Function:** | NKF | | | |
|  | **Function Rationale:** | No known function of blast hits and no CDD entries or hhpred results of significance | | | |
|  | **Function Visual Evidence:** |  | | | |
|  | **Additional Notes:** |  | | | |
| 76 | **Final Location:** | 55802-56107 F | | | |
|  | **Autoannotation Location:** | 55802-56107 | | | |
|  | **Location Rationale:** | Glimmer and Genemark originally gave start sites of 55802. Looking at starterator, there wasn't a lot of conservation of this start site, however GeneMark showed the most coding potential for the Glimmer and Genemark called sites. | | | |
|  | **Location Visual Evidence:** | M. Tb Coding Potential    M. smeg Coding Potential | | | |
|  | **Function:** | HNH endonuclease | | | |
|  | **Function Rationale:** | PhagesDB Blast suggested HNH endonuclease. | | | |
|  | **Function Visual Evidence:** | Visual evidence was not informative. | | | |
|  | **Additional Notes:** |  | | | |