Finemlucis

L2 mycobacteriophage

Kansas State University

**NOTE:** Best Annotation list needs phosphoribosyl pyrophosphate synthetase added to it.

10000 - 11000 bps. Several L2 phages have one large forward tail assembly chaperone gene here plus another smaller forward gene completely within the 5’ CDS of the larger one (?). We have 2, non-overlapping forward genes here as there’s no additional start sites 5’ of the downstream gene to stretch it to, but it still identifies as a tail assembly chaperone protein.

In general, >31000 bps. is pretty gappy, if that’s a word, but the gaps are conserved between L2 phages:

* (31326 - 31932 bps.) 606 bps. gap 5’ to integrase conserved in other L2 phages.
* (42271 - 42640 bps.) 369 bps. gap 3’ to Ro RNA binding protein conserved in other L2 phages.
* (53867 - 54025 bps.) 158 bps. gap 5’ to RusA resolvase (endonuclease) conserved in other L2 phages. No upstream starts to stretch to.
* (64869 - 65563 bps.) has 7 tRNA genes (tRNAs 3 - 9)
* (72062 - 71580 bps.) 713 bps. gap seen in Crossroads & Archie, but not Breezona or Loadrie.

gp2 (477 - 866 bps.) is terminase, small subunit by BLASTp (PhagesDB & NCBI) and Phamerator.

gp4 (1069 - 2637 bps.) is terminase, large subunit by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp5 (2638 - 4293 bps.) is portal protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator. Also hits ParA on BLASTp (NCBI) 5e-81.

gp6 (4311- 5369 bps.) is capsid maturation protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp7 (5422 - 6027 bps.) is scaffolding protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp8 (6070 - 7218 bps.) is major capsid protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp9 (7231 - 7695 bps.) is head-to-tail connector by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp10 (7692 - 8030 bps.) is head-to-tail connector by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp11 (8030 - 8461 bps.) is head-to-tail connector by BLASTp (PhagesDB) and Phamerato, but not by BLASTp (NCBI) & HHpred.

gp12 (8471 - 8974 bps.) is head-to-tail connector by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp13 (9002 - 9901 bps.) is major tail protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp14 (9991 - 10290 bps.) is tail assembly chaperone by BLASTp (PhagesDB & NCBI) and Phamerator.

gp15 (9991 - 10814 bps.) is tail assembly chaperone by BLASTp (PhagesDB & NCBI) and Phamerator.

**Frameshift between gp14 (9991 - 10290 bps.) & gp15 (10473 - 10814 bps.) is tail assembly chaperone by BLASTp (PhagesDB & NCBI) and Phamerator. Strange that the frameshift occurs after the 10286 bps. GGGAAA and not a typical slippery GGGAAAA DNA sequence (4th A missing) and it’s a rarer -2 frameshift.**

**MEENKELNDLVGSLFADLIASVRVPEPLQVAPGLVVKNPTKKQANALIKATTEEEAQRIIFGDDYDRAMELFDPQPVQVWNKFMAKYNEHFFGDKDSGKVIYAADIVERYWRAISWDFQTILHVNPLDYFAAPCRCGQCREKYGDDVNARYASRRNWDQFIMYYETLLQIRGSYTQAMYLQDPDVIELQANAKDEDWKASKPPLFGWTAELDALFYIADQVQAGRVQKAEDFKPYPRPELPAEKERKRRKENKVNAGLEAAMARGLETAKWNYLZ**

gp16 (10856 - 15991 bps.) is tape measure protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp17 (16080 - 16994 bps.) is minor tail protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp18 (16995 - 18728 bps.) is minor tail protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp19 (18755 - 19252 bps.) is minor tail protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp20 (19236 - 21032 bps.) is minor tail protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp21 (21038 - 21208 bps.) is NKF, but does hit minor tail protein by BLASTp (PhagesDB & NCBI). Too small to be minor tail protein.

gp22 (21198 - 21473 bps.) is NKF, but does hit minor tail protein by BLASTp (PhagesDB & NCBI). Too small to be minor tail protein.

gp23 (21558 - 21944 bps.) is NKF, but does hit minor tail protein by BLASTp (PhagesDB & NCBI). Too small to be minor tail protein.

gp24 (21945 - 24233 bps.) is minor tail protein by BLASTp (PhagesDB & NCBI) and Phamerator.

gp27 (25356 - 26525 bps.) is Lysin A by BLASTp (PhagesDB & NCBI) and HHpred.

gp28 (26525 - 27553 bps.) is Lysin B by BLASTp (PhagesDB & NCBI) and HHpred.

**gp30 (28028 - 28282 bps.) is NOT holin, as annotated by some. It only has 1 TMH, not the required 2 to be a holin (Welkin).**

gp33 (28698 - 28964 bps.) was stretched from it’s original call and is now NrdH glutaredoxin by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp36 (29656 - 30402 REV bps.) is nucleotidyl transferase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp37 (30489 - 30950 REV bps.) is Lsr2 DNA bridging protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp39 (31932 - 33239 bps.) is integrase (Y-int) by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

**gp40 (33236 - 34015 REV bps.) is NKF even though Winky and Rumpelstiltskin annotated it as HTH DNA binding protein [BLASTp (PhagesDB)], but not by BLASTp (NCBI) or HHpred. May be an annotation error.**

gp41 (33236 - 34015 bps.) is the immunity repressor by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp42 (34743 - 34973 bps.) is CRO (control of repressor’s operator) by BLASTp (PhagesDB & NCBI), HHpred and Phamerator. The gp corresponding to our gp42 has been annotated as HTH DNA binding domain in Archie, but comes up in a different pham in ours and other L2 phages.

gp45 (35525 - 35635 bps.) is NKF and had a huge -97 bps overlap that we shortened to a -19 bps. overlap. Better RBS scores and still NKF. Might want to just delete the whole thing and have a 79 bp. gap.

gp49 (36890 - 37924 bps.) was stretched to 36890 for Q1:T1 BLASTp hits on esterase by by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

**gp52 (38552 - 38680 bps.) was a difficult call because it’s NKF and all of the start sites have poor RBS scores. Some have stretched it for a -33 bps. overlap from the original 8 bps. gap.**

**gp54 (38933 - 39892 bps.) is membrane protein, Band-7-like by BLASTp (PhagesDB & NCBI), HHpred and Phamerator. It has a ton of BLASTp (NCBI) hits to *Mycobactrium* spp. membrane protease subunit genes.**

gp57 (40841 - 42271 bps.) is Ro RNA binding protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator, but Crossroads uses the older Von Willebrand-like domain name.

**gp60 (43089 - 43406 bps.) is a NKF by BLASTp (NCBI & PhagesDB) and HHpred, but Crossroads and another (Loadrie?) name it as a putative secreted protein.**

**gp63 (43817 - 44749 bps.) is NFK by BLASTp (NCBI & PhagesDB) and Phamerator. It does hit as exonuclease on HHpred with 13 hits on exonucleases (99% probab over 64% protein) for the C-terminal half to 2/3rds of the protein. Can’t determine which exonuclease though. Best hit is *E. coli* exonuclease VIII.**

gp65 (45045 - 45800 bps.) is ssDNA binding protein, ERF family by BLASTp (PhagesDB & NCBI), but no good HHpred results.

gp69 (46684 - 47085 bps.) is HNH endonuclease by BLASTp (PhagesDB & NCBI), HHpred and Phamerator. A REV gene called by DNA Master within the CDS of gp69 was deleted.

**gp73 (49120 - 49566 bps.) was added to the original (49166 - 49587 bps.) 418 bps. gap in Finemlucis not seen in other L2 phages. It’s 3’ to DnaB dsDNA helicase conserved in other L2 phages. Endonuclease VII is in this gap in frame +3. It has more starts site 5’ but they have worse RBS scores.**

gp76 (50041 - 50904 bps.) is phosphoribosyl pyrophosphate synthetase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp77 (50901 - 52385 bps.) is nicotinate ribosyltransferase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp78 (52472 - 53398 bps.) is polynucleotide kinase by BLASTp (PhagesDB & NCBI) HHpred and Phamerator.

**gp80 (53631 - 53867 bps.) is scored as NKF, but HHpred does weakly hit GPV baseplate assembly protein V from Enterobacteria phage P2, probability = 92% over only 38% of the protein.**

gp81 (54025 - 54423 bps.) is RusA resolvase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp82 (54386 - 54682 bps.) is WhiB family transcription factor by BLASTp (PhagesDB & NCBI) and Phamerator, but no HHpred hits.

gp83 (54683 - 55087 bps.) is WhiB family transcription factor by BLASTp (PhagesDB & NCBI) and Phamerator, but no HHpred hits.

gp85 (55392 - 55946 bps.) is NKF and was stretched from its original start to maximize CP, LO, and Q1:T1 BLASTp alignments.

**gp86 (55943 - 56167 bps.) is an orpham and gp87 can’t be stretched into that region to account for the CP.**

gp87 (56160 - 56546 bps.) is ssDNA binding protein by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp91 (57428 - 58605 bps.) is RNA ligase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp93 (58780 - 59463 bps.) is ClpP protease by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

**gp96 (59962 - 60183 bps.) is an orpham with no other ORFs in the forward frames with BLASTp hits.**

gp98 (60566 - 60979 bps.) is HNH endonuclease by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp101 (61677 - 62459 bps.) is DnaQ exonulcease (DNA polymerase III subunit) by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp104 (62772 - 63194 bps.) is HNH endonuclease by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

**After gp104 is where there is a lot of variability in the L2 phages.**

gp142 (73836 - 75074 REV bps.) is AAA-ATPase by BLASTp (PhagesDB & NCBI), HHpred and Phamerator.

gp144 (75442 - 75657 REV bps.) is an orpham we left to maximize CP.

Finemlucis has 12 tRNA genes and all were edited for starts & stops by tRNA-scan 2.0. tRNA #12 is REV oriented.

Sequence tRNA Bounds tRNA Anti Intron Bounds Inf HMM 2'Str Isotype Isotype

Name tRNA # Begin End Type Codon Begin End Score Score Score Pseudo CM Score

-------- ------ ---- ------ ---- ----- ----- ---- ------ ----- ----- ------ ------- -------

Mycobacterium 1 63805 63877 Thr CGT 0 0 69.6 43.20 26.40 No Thr 91.6

Mycobacterium 2 64100 64175 Pro TGG 0 0 34.2 0.00 34.20 No Arg 31.9

Mycobacterium 3 64786 64858 Trp CCA 0 0 48.3 25.70 22.60 No Cys 43.2

Mycobacterium 4 64881 64954 Leu TAG 0 0 52.9 25.30 27.60 No Leu 66.1

Mycobacterium 5 64956 65028 His GTG 0 0 33.6 19.40 14.20 No Thr 27.7

Mycobacterium 6 65030 65104 Gln CTG 0 0 51.4 22.30 29.10 No Gln 57.1

Mycobacterium 7 65245 65318 Gly TCC 0 0 49.9 13.30 36.60 No Cys 44.7

Mycobacterium 8 65355 65430 Lys CTT 0 0 84.8 64.30 20.50 No Ile2 94.6

Mycobacterium 9 65432 65520 Ser TGA 0 0 20.7 0.00 20.70 No Ser 28.5

Mycobacterium 10 66615 66685 Cys GCA 0 0 36.9 14.10 22.80 No Cys 38.5

Mycobacterium 11 67011 67080 Asn GTT 0 0 58.0 34.40 23.60 No Asn 72.2

Mycobacterium 12 67349 67277 Lys TTT 0 0 70.6 42.80 27.80 No Lys 86.3