

Genome Sequences of Cluster A Mycobacteriophages Dorothea and LappelDuVide, Isolated on *Mycobacterium smegmatis* mc²155

Nicole Pasterczyk^{1*}, Claire Wilson^{1*}, Zoe Broker¹, Audrey Chobor¹, DeAndrea Daughtry¹, Odinakachukwu Dibor¹, Annly John¹, Sruthvika Kandru¹, Mokhinur Kodigo¹, Constance Mulligan¹, Chelse Owate¹, Olivia Pistone¹, Sophia Smith¹, Sarah Teitelman¹, Samantha Victor¹, Petra Oganovich¹, Madelyn Hult¹, Rhema Hooper¹, Ellie Novak¹, Sameen Mariaca¹, Lee Graham¹, Stephen Mensah¹, Margaret Kenna^{1*}, Vassie Ware^{1§*}

¹Biological Sciences, Lehigh University, Bethlehem, PA USA

[§]To whom correspondence should be addressed: vcv0@lehigh.edu

*These authors contributed equally.

Abstract

Genome annotation of *Mycobacterium smegmatis* phages Dorothea and LappelDuVide revealed genomic features/organization common to cluster A mycobacteriophages; they are assigned to subclusters A6 and A4, respectively. Both genomes encode a putative immunity repressor, but the mode of prophage inheritance likely differs; Dorothea encodes a ParA/B partitioning system while LappelDuVide encodes a serine integrase. Dorothea encodes 3 tRNAs while none were identified for LappelDuVide.

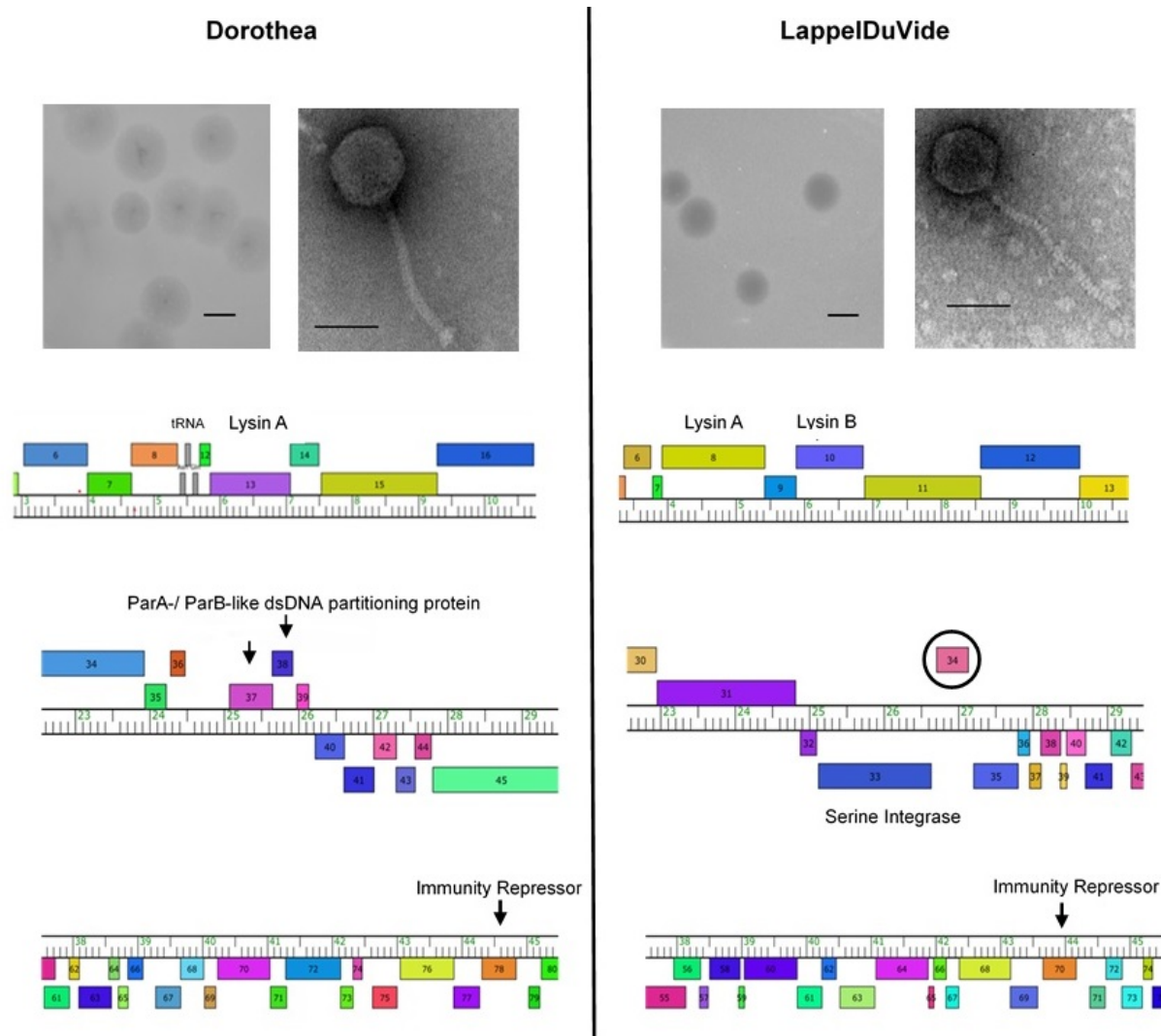


Figure 1. Dorothea and LappelDuVide: Phage Morphologies and Partial Genome Comparisons:

Dorothea, top left; LappelDuVide, top right. Plaque images: scale bar is 5mm. Plaques with clear centers and turbid halos are visible. TEM images: scale bar is 50nm. TEM images show siphoviral morphologies. Phamerator maps: Dorothea, bottom left; LappelDuVide, bottom right. Three distinct regions of each genome are shown: genome coordinates are 3000-

10,500 bp (Region 1); 22,500-29,500 bp (Region 2); and 37,500-45,500 bp (Region 3), respectively. Region 1 shows tRNAs (3 gray boxes) and lysins. Region 2 includes the gene encoding the serine integrase (on the reverse strand) in LappelDuVide and genes encoding ParA (purple box with arrow)/ParB (blue box with arrow) dsDNA partitioning proteins on the forward strand in Dorothea. The gene (34) encircled in black on the forward strand in LappelDuVide disrupts the typical transcriptional direction on the right arm. Region 3 shows the gene encoding the immunity repressor, found distal to machinery involved in lysogeny establishment in both phages. Protein coding genes are depicted with colored boxes.

Description

Given the contributions of bacteriophage research to unravelling fundamentals of molecular biology and developing novel strategies to combat bacterial antibiotic resistance, continued discovery of new bacteriophages and analysis of their unique features is critical to sustain advances in these areas (Haq et al., 2012). Cluster A mycobacteriophages are the most abundant phage group catalogued in the Actinobacteriophage Database (Russell and Hatfull, 2017), sharing many common genetic features and genomic architecture, yet presenting with a high degree of genetic diversity between subclusters (Hatfull, 2012).

Mycobacteriophage LappelDuVide was isolated from direct extraction of soil (Hatfield, PA, USA). Mycobacteriophage Dorothea was isolated from an enriched soil sample (Bethlehem, PA, USA). Briefly for direct isolation, soil samples were suspended in phage buffer with shaking for 1 hour, filtered (0.2 micron pore size), and the filtrate plated in top agar with *Mycobacterium smegmatis* mc²155. Plates were incubated at 37 °C for 2 days. For enriched isolation, the soil and *M. smegmatis* were suspended in 7H9 liquid medium and incubated with shaking for 3 days at 37 °C before the resulting culture was filtered and the filtrate plated (Zorawik et al., 2024). LappelDuVide was purified after 4 rounds of plaque assays; Dorothea was purified after 6 rounds. Plaque morphology characteristics are shown (Table 1, Fig.1). High titer lysates, used in negative-stain transmission electron microscopy (TEM) preparations, revealed siphovirus morphologies for both phages (Table 1, Fig.1).

Table 1: Phage Characterization

	Dorothea (A6)	LappelDuVide (A4)
Plaque Morphology	Clear center, turbid halo	Clear center, turbid halo
Plaque size @ 24hrs (mm)	5 (n=10)	3-5 (n=10)
Phage morphology	Siphovirus	Siphovirus
Capsid width (nm)	50 +/-1 (n=5)	50 +/-1 (n=5)
Tail length (nm)	100 +/-1 (n=5)	125 +/-1 (n=5)
Capsid width : Tail length ratio	1:2	1:2.5
Genome Size (bp)	52,637	51,371
# Raw (150 bp single-end) reads; [shotgun coverage]	319,931 [868]	80,883 [208]
Genome terminus (3' sticky overhang)	5' CGGTCCGGTTA 3'	5' CGGCCGGTAA 3'
GC content (%)	61.6	63.9

LappelDuVide DNA was isolated from lysate by column purification using a Promega Wizard DNA Clean-up Kit. Dorothea DNA was isolated from lysate using a standard phenol/chloroform extraction method (Mülhardt and Beese, 2007). Genomes were prepared for sequencing with an NEB Ultra II FS Library Kit and sequenced on an Illumina MiSeq 1000 (v3 reagents). The resulting 150 base reads were assembled without any trimming or QC. Raw reads were assembled using Newbler version 2.9 (Miller et al., 2010) and Consed version 29 (Gordon and Green, 2013) into a single contig (Table 1). Genomic termini were determined as previously described (Russell, 2018).

Genomes were annotated using DNA Master v5.23.6 (Lawrence, 2007), GeneMark v2.5 (Bessmer and Borodovsky, 2005), Glimmer v3.02 (Delcher et al., 2007), and Starterator v.546 (Pacey, 2016). PECAAN (Rinehart et al., 2016)) was also used for LappelDuVide annotation. BLAST, using the Actinobacteriophage (Russell and Hatfull, 2017) and NCBI non-redundant databases (Altschul et al., 1990) and HHPRED, using the PDB_mmCIF70, Pfam- v.36, NCBI Conserved Domains databases v57v87 (Zimmermann et al., 2018) were used for functional annotation. Phamerator, using Actino_draft database v578 (Cresawn et al., 2011), assessed synteny. tRNAs were identified using tRNAscan v2.0.6 (Lowe and Eddy, 1997) and ARAGORN v1.2.38 (Laslett and Canback, 2004). TMHMM v1.0.24 (Möller et al., 2001) predicted transmembrane domains. Default settings were used for all software.

The Dorothea genome includes 99 predicted protein-encoding genes and three tRNAs: tRNA-Asn, tRNA-Trp, and tRNA-Gln (Fig. 1, Region 1). The LappelDuVide genome has 88 predicted protein-encoding genes and no tRNAs. In general for both phages, a contiguous segment of the genome on the left arm includes the majority of the structural and assembly genes transcribed in one direction from left to right on the forward strand. The remaining genes on the right arm in Dorothea are transcribed in the opposite direction, on the reverse strand. This is also generally true for LappelDuVide, except for one gene (34) positioned just upstream of the serine integrase (located on the reverse strand) (Fig. 1, Region 2). Both genomes encode an immunity repressor in the same pham (identified as a group of proteins with a high degree of amino acid similarity to one another), located in the same relative syntenic position on the reverse strand on the right arm (Cresawn et al., 2011) (Fig. 1, Region 3). Notable differences between the genomes include the presence of lysins A and B in LappelDuVide, but only lysin A in Dorothea. By genome analysis, both phages should be capable of lysogen establishment, but lysogeny would likely occur by different mechanisms. Dorothea has a ParA/ParB partitioning system (Dedrick et al., 2016) on the forward strand and LappelDuVide has a serine integrase (Smith, 2015) (Fig. 1, Region 2).

Nucleotide sequence accession numbers: Dorothea is available at GenBank with Accession No. [PQ184798](#) and Sequence Read Archive (SRA) No. [SRX26413417](#). LappelDuVide is available at GenBank with Accession No. [PP978813](#) and SRA No. [SRX26413423](#).

Acknowledgements: This work was performed through Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) courses at Lehigh University. We thank Daniel Russell and the sequencing team at the Pittsburgh Bacteriophage Institute for genome sequencing, assembly, and input into the Sequence Read Archive. We thank the SEA MINT team for constructive input and review of the manuscript.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410. DOI: [10.1016/S0022-2836\(05\)80360-2](#)
- Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* 33: W451-W454. DOI: [10.1093/nar/gki487](#)
- Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12: 10.1186/1471-2105-12-395. DOI: [10.1186/1471-2105-12-395](#)
- Dedrick RM, Mavrich TN, Ng WL, Cervantes Reyes JC, Olm MR, Rush RE, et al., Hatfull. 2016. Function, expression, specificity, diversity and incompatibility of actinobacteriophage *parABS* systems. *Molecular Microbiology* 101: 625-644. DOI: [10.1111/mmi.13414](#)
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673-679. DOI: [10.1093/bioinformatics/btm009](#)
- Gordon D, Green P. 2013. *Consed*: a graphical editor for next-generation sequencing. *Bioinformatics* 29: 2936-2937. DOI: [10.1093/bioinformatics/btt515](#)
- Haq IU, Chaudhry WN, Akhtar MN, Andleeb S, Qadri I. 2012. Bacteriophages and their implications on future biotechnology: a review. *Virology Journal* 9: 10.1186/1743-422x-9-9. DOI: [10.1186/1743-422X-9-9](#)
- Hatfull GF. 2012. The Secret Lives of Mycobacteriophages. *Advances in Virus Research, Bacteriophages, Part A* : 179-288. DOI: [10.1016/B978-0-12-394621-8.00015-7](#)
- Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32: 11-16. DOI: [10.1093/nar/gkh152](#)
- Lawrence, J. DNAMaster. (2007). Retrieved from <http://cobamide2.bio.pitt.edu/computer.htm>
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research* 25: 955-964. DOI: [10.1093/nar/25.5.955](#)

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95: 315-327. DOI: [10.1016/j.ygeno.2010.03.001](https://doi.org/10.1016/j.ygeno.2010.03.001)

Möller S, Croning MDR, Apweiler R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17: 646-653. DOI: [10.1093/bioinformatics/17.7.646](https://doi.org/10.1093/bioinformatics/17.7.646)

Mülhardt, C., & Beese, E. W. (2007). Molecular biology and genomics. <https://doi.org/10.1016/B978-0-12-088546-6.X5000-3> DOI: [10.1016/B978-0-12-088546-6.X5000-3](https://doi.org/10.1016/B978-0-12-088546-6.X5000-3)

Pacey, M. (2016). Starterator guide. University of Pittsburgh, Pittsburgh, PA.

Reinehart, C., Gaffney, B., Wood, J. D., & Smith, J. (2016). PECAAN, a Phage Evidence Collection And Annotation Network. <https://discover.kbrinsgd.org>

Russell, D. A., & Hatfull, G. F. (2017). PhagesDB: The actinobacteriophage database. *Bioinformatics*, 33(6), 784–786.

Russell DA. 2017. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. *Methods in Molecular Biology, Bacteriophages* : 109-125. DOI: [10.1007/978-1-4939-7343-9_9](https://doi.org/10.1007/978-1-4939-7343-9_9)

Smith MCM. 2015. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiology Spectrum* 3: 10.1128/microbiolspec.mdna3-0059-2014. DOI: [10.1128/microbiolspec.mdna3-0059-2014](https://doi.org/10.1128/microbiolspec.mdna3-0059-2014)

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al., Alva. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* 430: 2237-2243. DOI: [10.1016/j.jmb.2017.12.007](https://doi.org/10.1016/j.jmb.2017.12.007)

Zorawik M, Jacobs-Sera D, Freise AC, SEA-PHAGES, Reddi K. 2024. Isolation of Bacteriophages on Actinobacteria Hosts. *Methods in Molecular Biology, Phage Engineering and Analysis* : 273-298. DOI: [10.1007/978-1-0716-3798-2_17](https://doi.org/10.1007/978-1-0716-3798-2_17)

Funding: Financial support was provided from the Department of Biological Sciences, Lehigh University.

Author Contributions: Nicole Pasterczyk: writing - original draft, visualization. Claire Wilson: writing - original draft, validation, visualization. Zoe Broker: validation. Audrey Chobor: validation. DeAndrea Daughtry: validation. Odinakachukwu Dibor: validation, investigation. Annly John: validation. Sruthvika Kandru: validation. Mokhinur Kodirova: validation. Constance Mulligan: validation. Chelse Owate: validation. Olivia Pistone: validation. Sophia Smith: validation. Sarah Teitelman: validation. Samantha Victor: validation. Petra Oganovich: validation, investigation. Madelyn Hult: validation. Rhema Hooper: validation. Ellie Novak: visualization. Sameen Mariaca: visualization. Lee Graham: resources. Stephen Mensah: supervision, validation, investigation. Margaret Kenna: project administration, supervision, validation, writing - review editing. Vassie Ware: project administration, supervision, validation, writing - review editing.

Reviewed By: Susan Gurney, Anonymous

History: Received October 1, 2025 **Revision Received** November 14, 2025 **Accepted** November 18, 2025 **Published Online** November 18, 2025 **Indexed** December 2, 2025

Copyright: © 2025 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Pasterczyk N, Wilson C, Broker Z, Chobor A, Daughtry D, Dibor O, et al., Ware V. 2025. Genome Sequences of Cluster A Mycobacteriophages Dorothea and LappelDuVide, Isolated on *Mycobacterium smegmatis* mc²155. *microPublication Biology*. [10.17912/micropub.biology.001881](https://doi.org/10.17912/micropub.biology.001881)