AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of Cluster K Mycobacteriophages Deby, LaterM, LilPharaoh, Paola, SgtBeansprout, and Sulley

Joseph M. Gaballa,[a] Keeyon Dabrian,[a] Rachel Desai,[a] Ryan Ngo,[a] Diane Park,[a] Erin Sakaji,[a] Yiwei Sun,[a] Boon Tan,[a] Marcia Brinck,[a] Olivia Brobst,[a] Rebecca Fernando,[a] Hannah Kim,[b] Siobhan McCarthy,[a] Michael Murphy,[a] Alexandra Sarkis,[a] Parker Sevier,[a] Amitoj Singh,[a] Darwin Wu,[a] Min-Ying Wu,[a] Hayley A. Ennis,[a] Rohan Luhar,[a] Justin E. Miller,[a] Stephanie B. Orchanian,[a] Alysha N. Salbato,[a] Sai Alam,[a] Lauren Brenner,[a] Zilla Kailani,[a] Joel Laskow,[a] Xinyu Ma,[a] Aika Miikeda,[a] Paola Nol-Bernardino,[a] Alisa Sukhina,[a] Nikolina Walas,[a] Wenyuan Wei,[a] Nam Phuong Do,[a] Christina T. Fournier,[a] Christy J. Kim,[a] Samantha F. Mosier,[a] Carly Pierson,[a] Ivonne G. Romero,[a] Mikael Sanchez,[a] Oyinlola Sawyerr,[a] Joyce Wang,[a] Rina Watanabe,[a] Samuel Wu,[a] Annie Chen,[a] Katelynn Kazane,[a] Yousif Kettoola,[a] Emma C. Goodwin,[a] Andrew J. Lund,[a,b] William Villella,[a] Drake Williams,[a] Amanda Freise,[a] Jordan Moberg Parker[a]

[a]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, California, USA
[b]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California, USA

ABSTRACT Mycobacteriophages Deby, LaterM, LilPharaoh, Paola, SgtBeansprout, and Sulley were isolated from soil using *Mycobacterium smegmatis* mc²155. Genomic analysis indicated that they belong to subclusters K1 and K5. Their genomic architectures are typical of cluster K mycobacteriophages, with most variability occurring on the right end of the genome sequence.

Mycobacteriophages, viruses that selectively infect mycobacteria, such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*, are of translational interest because of their potential to counter the rising threat of antibiotic resistance among bacterial pathogens (1). The vast genetic diversity among bacteriophages allows for them to be categorized into clusters and subclusters based on nucleotide similarity (2, 3). Cluster K, which contained 119 members as of October 2018 (http://phagesdb.org/clusters/K/), is one of the only known clusters of mycobacteriophages able to infect *M. tuberculosis*, the pathogen responsible for tuberculosis (4). Here, we report the novel genome sequences of the five cluster K1 phages Deby, LaterM, LilPharaoh, SgtBeansprout, and Sulley and of Paola, a cluster K5 phage.

The mycobacteriophages were all isolated by enrichment cultures of *Mycobacterium smegmatis* mc²155 from soil samples collected in Los Angeles, CA, by students in the SEA-PHAGES program (5). Viral DNA was extracted with the Promega Wizard DNA clean-up kit (product number A7280), and sequencing libraries were prepared using an NEB Ultra II kit with dual-indexed barcoding or a Titanium emulsion PCR (emPCR) kit. Libraries were then pooled and run on an Illumina MiSeq platform or a GS FLX system, yielding at least 30,000 single-end reads and at least 80-fold coverage for each genome (Table 1). These reads were then assembled with Newbler version 2.9 with default settings and in each case yielded a single-phage contig, which was checked for completeness, accuracy, and phage genomic termini with Consed version 29, as previously described (6). The phage genomes were linear, double-stranded DNA with 11 nucleotide 3′ sticky overhangs. Genome sequence sizes ranged from 56,167 to 61,535 bp, and GC content varied between 65.0 and 67.1%, with an average of 66.4%, which is slightly less than the average of cluster K phages (66.9%) (Table 1).

Genome annotation was performed with DNA Master (http://cobamide2.bio.pitt.edu/) and PECAAN (https://pecaan.kbrinsgd.org/), which integrate both Glimmer (7) and GeneMark (8) to predict potential open reading frames. ARAGORN (9) and tRNAscan-SE (10) were used to detect the presence of tRNAs. Gene locations were

Citation Gaballa JM, Dabrian K, Desai R, Ngo R, Park D, Sakaji E, Sun Y, Tan B, Brinck M, Brobst O, Fernando R, Kim H, McCarthy S, Murphy M, Sarkis A, Sevier P, Singh A, Wu D, Wu M-Y, Ennis HA, Luhar R, Miller JE, Orchanian SB, Salbato AN, Alam S, Brenner L, Kailani Z, Laskow J, Ma X, Miikeda A, Nol-Bernardino P, Sukhina A, Walas N, Wei W, Do NP, Fournier CT, Kim CJ, Mosier SF, Pierson C, Romero IG, Sanchez M, Sawyerr O, Wang J, Watanabe R, Wu S, Chen A, Kazane K, Kettoola Y, Goodwin EC, Lund AJ, Villella W, Williams D, Freise A, Moberg Parker J. 2019. Genome sequences of cluster K mycobacteriophages Deby, LaterM, LilPharaoh, Paola, SgtBeansprout, and Sulley. Microbiol Resour Announc 8:e01481-18. https://doi.org/10.1128/MRA.01481-18.

Editor Irene L. G. Newton, Indiana University, Bloomington

Address correspondence to Jordan Moberg Parker, jparker@microbio.ucla.edu.

Received 7 November 2018
Accepted 27 November 2018
Published 10 January 2019

TABLE 1 Novel cluster K mycobacteriophages

Phage name	Cluster	GenBank genome sequence accession no.	Sequence Read Archive raw read accession no.	Sequencing coverage (×)	Genome length (bp)	GC content (%)	No. of coding genes	No. of tRNAs
Deby	K1	MG962364	SRX5036139	4,562[a]	60,463	66.5	95	1
LaterM	K1	MG962371	SRX5036140	4,028[a]	60,143	66.5	95	1
LilPharaoh	K1	MF919518	SRX5036137	1,633[b]	56,167	67.1	78	0
SgtBeansprout	K1	MH020245	SRX5036138	1,052[b]	56,439	67.1	78	0
Sulley	K1	MF919532	SRX5036141	80[c]	59,873	66.4	94	1
Paola	K5	MG962374	SRX5036142	1,263[b]	61,535	65.0	92	1

[a] Sequencing performed at the NC State Genomic Sciences Laboratory (Illumina).
[b] Sequencing performed at the Pittsburgh Bacteriophage Institute (Illumina).
[c] Sequencing performed at the UCLA Genotyping and Sequencing Core (454 GS FLX pyrosequencing).

curated with Phamerator, which compares phage genes and genomes, and Starterator, which identifies conserved start sites (11). Gene functions were predicted with BLASTp (12) against the PhagesDB (13) and NCBI databases (https://www.ncbi.nlm.nih.gov/protein), as well as HHPred (14) and TMHMM (15).

Annotation revealed that the genome sequences of phages Deby, LaterM, Sulley, and Paola contained 92 to 95 coding genes and one $tRNA^{trp}$, and LilPharaoh and SgtBeansprout had 78 coding genes and no tRNAs. In line with other cluster K mycobacteriophages, the left end of each genome was highly conserved among the phages, whereas the right end of each genome was much more variable (4) and contained the majority of genes without any known functions. The majority of the structural and the assembly genes, such as the tape measure protein and major capsid proteins, dominated the first 25 kb of the genome sequences. All phages contained lysis cassettes with lysin A, lysin B, and holin genes, and a conserved integrase was found downstream of the lysis cassettes in all phages, which suggests that they are all temperate phages potentially capable of undergoing the lysogenic life cycle.

Data availability. GenBank accession numbers for the six mycobacteriophages discussed in this paper are provided in Table 1.

ACKNOWLEDGMENTS

Krishna P. Govindaraju, Kyerra Jones, Emily Lundberg, Vasileios Ragkousis, Carlos Rojas, Jinhua Shen, and Ximena Ibarra contributed to the isolation, annotation, and genome analysis of the novel phages. We thank Krisanavane Reddi for preparation of materials, lysate archiving, and management of the instructional laboratories; Hong Zhou of the UCLA Electron Imaging Center for NanoMachines for electron microscopy support; the UCLA Genotyping and Sequencing Core, the NC State Genomic Sciences Laboratory, and Rebecca A. Garlena and Daniel A. Russell at the Pittsburgh Bacteriophage Institute for phage sequencing and assistance with genome assembly; and Debbie Jacobs-Sera, Welkin Pope, Graham Hatfull, and the SEA-PHAGES community for programmatic support.

J.M.G. drafted the paper and contributed to the isolation, annotation, and genome analysis of the novel phages; J.M.P. revised the paper and performed quality control on the annotations; A.F. and J.M.P. supervised the research; and all other authors contributed to the isolation, annotation, and genome analysis of the novel phages.

This project was funded by the Dean of Life Sciences Division at UCLA, with additional support for sequencing from the HHMI Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program.

REFERENCES

1. Gondil VS, Chhibber S. 2018. Exploring potential of phage therapy for tuberculosis using model organism. Biomed Biotechnol Res J 2:9–15. https://doi.org/10.4103/bbrj.bbrj_93_17.
2. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko C-C, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J Mol Biol 397:119–143. https://doi.org/10.1016/j.jmb.2010.01.011.
3. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG,

Jacobs WR Jr, Hendrix RW, Lawrence JG, Hatfull GF. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416. https://doi.org/10.7554/eLife.06416.

4. Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, Elgin SCR, Guilfoile FR, Forsyth MH, Harris AD, Harvey SE, Hughes LE, Hynes PM, Jackson AS, Jalal MD, MacMurray EA, Manley CM, McDonough MJ, Mosier JL, Osterbann LJ, Rabinowitz HS, Rhyan CN, Russell DA, Saha MS, Shaffer CD, Simon SE, Sims EF, Tovar IG, Weisser EG, Wertz JT, Weston-Hafer KA, Williamson KE, Zhang B, Cresawn SG, Jain P, Piuri M, Jacobs WR, Hendrix RW, Hatfull GF. 2011. Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. PLoS One 6:e26750. https://doi.org/10.1371/journal.pone.0026750.
5. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5:e01051-13. https://doi.org/10.1128/mBio.01051-13.
6. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes, p 109–125. *In* Clokie MRJ, Kropinski AM, Lavigne R (ed), Bacteriophages: methods and protocols, vol. 3. Springer, New York, NY.
7. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009.
8. Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki487.
9. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
10. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
11. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395. https://doi.org/10.1186/1471-2105-12-395.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
13. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
14. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
15. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.