AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Genome Sequences of Cluster K Mycobacteriophages DrHayes, Urkel, and SamuelLPlaqson

Kirk R. Anders, Alex M. Murphy,* William F. Ettinger, Douglas Kempthorne, Charles Kittridge, Alex Kures, Sarah Lundgren, Jacob Masters, Rachel Noyes, Christina Winters, Perry Yazzolino, Kasandra Ziebert, Joseph Haydock, Stephen Hayes, Rebecca A. Garlena, Daniel A. Russell, Marianne K. Poxleitner, Ann-Scott H. Ettinger

Department of Biology, Gonzaga University, Spokane, Washington, USA

ABSTRACT Mycobacteriophages DrHayes, Urkel, and SamuelLPlaqson were isolated from soil samples in Spokane, WA, using *Mycobacterium smegmatis* mc²155 grown at room temperature. The three genomes differ by only a few nucleotides, are 60,526 bp long, have 97 predicted protein-coding genes and one tRNA gene, and are members of subcluster K1.

Bacteriophages are the most numerous life form on the planet (1). The genomic sequences of phages isolated on *Mycobacterium smegmatis* mc²155 reveal an astounding genetic variation that provides insight into the diversity and evolution of these viruses (2, 3). We report here the genome sequences of three mycobacteriophages that were isolated and analyzed by undergraduate students in the introductory biology and genetics lab courses as part of the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) research and education program (4–6).

Phages DrHayes, Urkel, and SamuelLPlaqson were isolated from soil samples from different locations in Spokane, WA, following enrichment in liquid cultures of *M. smegmatis* mc²155. The phages were purified and amplified at room temperature (20 to 25°C). DrHayes and SamuelLPlaqson form turbid plaques, whereas Urkel forms clear plaques. Electron microscopy showed that each phage exhibits a siphoviral morphology with an isometric head. Purified DNA was sequenced by Illumina MiSeq using 140-bp single-end reads. Newbler assembled the reads into contigs with average coverages of 114- to 230-fold. The nearly identical genomes are 60,526 bp and have 66.2% G+C content, and each contains 3′ single-strand extensions with the sequence 5′-CTCGTAGGCAT-3′. Urkel carries four 1-bp substitutions compared to DrHayes and SamuelLPlaqson, and SamuelLPlaqson carries an additional substitution compared to the others. The genomes are most closely related to CrimD (93% nucleotide identity spanning 93% of genome lengths) and are members of subcluster K1.

Initial annotations were generated by Glimmer (7), GeneMark (8), Aragorn (9), and tRNAscan-SE (10) and manually refined to identify 97 protein-coding genes and one tRNA ($tRNA^{trp}$) in each genome. Functions were predicted for 41 genes using BLASTP, the Conserved Domain Database at NCBI (11), and HHPred (12). The genome structures match the other K1 genomes closely: genes encoding virion assembly and structure proteins, lysis proteins, and host integration/excision proteins are located on the left arm and, although gene functions on the right arm are mostly unknown, several genes encode proteins related to nucleic acid metabolism, such as exonucleases, DNA primase, RusA resolvase, and RtcB RNA ligase (13). Interestingly, Urkel carries an R55G

Received 14 October 2016 **Accepted** 27 February 2017 **Published** 20 April 2017

Citation Anders KR, Murphy AM, Ettinger WF, Kempthorne D, Kittridge C, Kures A, Lundgren S, Masters J, Noyes R, Winters C, Yazzolino P, Ziebert K, Haydock J, Hayes S, Garlena RA, Russell DA, Poxleitner MK, Ettinger A-SH. 2017. Genome sequences of cluster K mycobacteriophages DrHayes, Urkel, and SamuelLPlaqson. Genome Announc 5:e01388-16. https://doi.org/10.1128/genomeA.01388-16.

Address correspondence to Kirk R. Anders, anders@gonzaga.edu.

* Present address: Alex M. Murphy, Laboratory of Genetics, University of Wisconsin-Madison, Madison, Wisconsin, USA.

missense mutation in the immunity repressor gp44, which correlates with its clear plaque morphology.

The three newly isolated phages differ from CrimD with at least 11 mosaic differences, including gene substitutions, insertions, and deletions. There are three instances (genes *4*, *33*, and *86*) in which genes are conserved among the phages but are more highly diverged (45, 42, and 49% amino acid [aa] identity, respectively) than their flanking genes (73 to 95% aa identity); the functions of these are unknown. Also of note, gene *26*, which is situated within a syntenic block of putative tail protein genes, is unrelated to CrimD gene *26*. In fact, out of all 1,000+ sequenced mycobacteriophages, gene *26* is unique to DrHayes, Urkel, and SamuelLPlaqson. Finally, there are four CrimD genes (*58*, *64*, *69*, and *73*) that are absent from DrHayes, Urkel, and SamuelLPlaqson and thus are likely nonessential for lytic growth.

Accession number(s). The DrHayes, Urkel, and SamuelLPlaqson genomes are available at GenBank under the accession numbers KX657795, KX657796, and KX657794, respectively.

ACKNOWLEDGMENTS

We thank the following people: Amanda Braley for preparation of reagents, lysate archival, and management of the Introductory Biology lab course; Valerie Lynch-Holm at the Franceschi Microscopy and Imaging Center at Washington State University for electron microscopy; David Asai and HHMI for support and encouragement; Sam Chambers for helping isolate DrHayes; the students of the genetics labs at Gonzaga University, spring 2015, for contributions to the genome annotation of SamuelLPlaqson; and Welkin Pope and Graham Hatfull at the University of Pittsburgh for valuable comments on the manuscript.

D.K., A.K., S.L., J.M., and P.Y. isolated the phages and DNA; R.A.G. and D.A.R. sequenced the DNAs; M.K.P. assembled the sequencing reads; K.R.A., A.M.M., W.F.E., C.K., R.N., C.W., K.Z., and A.-S.H.E. annotated, edited annotations of, and analyzed the genomes; J.H., S.H., A.-S.H.E., and K.R.A. supervised the research; and K.R.A. drafted the paper.

This work was supported by an NSF grant (DUE-1245778) to K.R.A., funding from HHMI through the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES), and the Gonzaga University Biology Department.

REFERENCES

1. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812. https://doi.org/10.1038/nrmicro1750.
2. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, 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, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education and Mycobacterial Genetics Course. 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. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, 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.
5. Hatfull GF. 2015. Innovations in undergraduate science education: going viral. J Virol 89:8111–8113. https://doi.org/10.1128/JVI.03003-14.
6. Staub NL, Poxleitner M, Braley A, Smith-Flores H, Pribbenow CM, Jaworski L, Lopatto D, Anders KR. 2016. Scaling up: adapting a phage-hunting course to increase participation of first-year students in research. CBE Life Sci Educ 15:0211. https://doi.org/10.1187/cbe.15-10-0211.
7. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641. https://doi.org/10.1093/nar/27.23.4636.
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, Eddy SR. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.
11. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's Conserved Domain Database. Nucleic Acids Res 43:D222–D226. https://doi.org/10.1093/nar/gku1221.
12. 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.

13. Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, Elgin SC, 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, Jr, 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.