Complete Genome Sequences of 61 Mycobacteriophages

Graham F. Hatfull,*a Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) Program,*b KwaZulu-Natal Research Institute for Tuberculosis and HIV (K-RITH) Mycobacterial Genetics Course,*c University of California–Los Angeles Research Immersion Laboratory in Virology,*d Phage Hunters Integrating Research and Education (PHIRE) Program*e

Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; Howard Hughes Medical Institute, Chevy Chase, Maryland, USA; K-RITH, Durban, South Africa; Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York, USA; Department of Microbiology, Immunology, and Molecular Genetics, University of California–Los Angeles, Los Angeles, California, USA

Mycobacteriophages—viruses of mycobacteria—provide insights into viral diversity and evolution as well as numerous tools for genetic dissection of Mycobacterium tuberculosis. Here we report the complete genome sequences of 61 mycobacteriophages newly isolated from environmental samples using Mycobacterium smegmatis mc²155 that expand our understanding of phage diversity.

Received 25 March 2016 Accepted 23 May 2016 Published 7 July 2016

Bacteriophages are the most numerous biological entities on the planet, with a global population of $10^{31}$ particles. With an estimated $10^{23}$ productive infections per second worldwide, the population is vast, dynamic, and genetically diverse (1–4). As of March 2016, the National Center for Biotechnology Information (NCBI) lists 1,757 Caudovirales genomes, 318 of which infect Mycobacterium hosts. Previous comparative analyses of mycobacteriophages revealed substantial diversity and mosaic architectures resulting from nonhomologous recombination. Integrated research–education programs such as Phage Hunters Integrating Research and Education (PHIRE) (5), Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) (6), the Mycobacterial Genetics Course at the University of Kwazulu-Natal (K-RITH), and the University of California–Los Angeles’s Research Immersion Laboratory in Virology, isolated, sequenced, and annotated the phages reported here (Table 1) using M. smegmatis as a host.

Phages were isolated by direct plating of filtered soil extracts or from enriched cultures, followed by plaque purification. Electron microscopy shows that 60 have siphoviral morphotypes, and HyRo is the sole member of the Myoviridae. Most have isometric capsids, the exceptions being Biper, Bshah, and Zakehe101 with prolacte heads. Genomic DNA was extracted from high titer lysates, sheared, and sequenced at the University of Pittsburgh, University of California–Los Angeles, the DOE Joint Genome Institute, or Virginia Commonwealth University using either Sanger, Illumina, Ion Torrent, or 454 technology. Sequence reads were assembled using Newbler (Roche) and Consed (7) and coverage depths range from 47-fold to 2,308-fold, with an average of 200-fold. Sequence assemblies revealed discrete genome ends for 52 phages, and the 9 with circularly permuted assemblies were bioinformatically linearized such that base one was assigned in accord with other mycobacteriophages. Genomes were annotated using DNA Master (http://cobamide2.bio.pitt.edu), Phamerator (8), Glimmer (9), GeneMark (10), Aragorn (11), and tRNAscanSE (12), and functions were determined using the public databases GenBank, Protein DataBase, pfamA, and phagesdb.org with BLAST (13), and HHPred (14). Genomes were assigned to clusters or subclusters as described previously (15).

Notwithstanding the large extant collection of sequenced mycobacteriophage genomes, these newly sequenced phages considerably expand our understanding of mycobacteriophage diversity. Twenty-two are members of the largest cluster, cluster A, but span 7 of the 15 subclusters. The others are broadly distributed across other clusters, including B, C, E, F, G, I, K, L, M, N, O, and P. Cosmo has substantial nucleotide sequence similarity to the singleton phage Wildcat, forming the new cluster V. The eight cluster N phages, Cedasite (G1), and Brusacoram (P) are notable in that they contain integration-dependent immunity systems in which the phage attachment site (attP) is located within the repressor gene (16).

As is typical of other sequenced phage genomes, functions can be assigned to only ~25% of the predicted genes, primarily those involved in virion capsid and assembly and well-conserved genes associated with DNA metabolism. Two of the cluster A genomes (Eidsmoe, ArcherNM) contain partitioning systems in place of integration cassettes; several genomes (e.g., Phrann, Xeno) encode toxin-antitoxin systems; and three encode Lsr2 homologs (Lolly9, Lumos, and Snenia).

Nucleotide sequence accession numbers. Nucleotide sequence accession numbers for all phages are shown in Table 1.
| Phage name     | Cluster | Genome (bp) | G+C content (%) | GenBank accession no. | Finding/annotating institution |
|---------------|---------|-------------|-----------------|-----------------------|--------------------------------|
| Alvin         | A1      | 49,577      | 63.5            | KP027205              | University of Pittsburgha      |
| ArcherNM      | A2      | 52,561      | 64.2            | KG761559              | Washington State University, University of Floridaa |
| Artemis2UCLA  | A6      | 52,244      | 61.4            | FG603333              | University of California–Los Angelesd |
| Backbacter     | A2      | 52,129      | 63.1            | KU686494              | North Carolina A&T State Universityb |
| Bernardo      | B3      | 68,196      | 67.4            | KG493879              | University of California–Los Angelesd |
| Biper         | Y       | 77,832      | 67.3            | KU1726633             | University of Pittsburgh, Florida Gulf Coast Universityb |
| Bricole       | M1      | 81,128      | 61.6            | KT591491              | Old Dominion Universityb       |
| Bruin         | E       | 74,210      | 63.0            | KG602099              | University of California–Los Angelesd |
| Brusacoram     | P       | 47,618      | 67.0            | KT347311              | College of St. Scholastica b   |
| Carcharodon   | N       | 43,680      | 66.2            | KM885859              | Jacksonville State Universityb |
| Cedasite      | G1      | 41,901      | 66.6            | KT554472              | Morehouse Collegea             |
| Chandler      | B3      | 69,450      | 67.5            | KG272007              | University of Pittsburgha      |
| CloudWang3A   | A6      | 52,873      | 61.4            | KG603332              | University of California–Los Angelesd |
| Conspiracy    | A5      | 50,755      | 60.6            | KG603330              | University of California–Los Angelesd |
| Cosmo         | V       | 78,229      | 56.8            | KG027195              | University of KwaZulu-Natald   |
| Eidsmoe       | A9      | 52,946      | 62.5            | KU176094              | Illinois Wesleyan Universityb  |
| Enkosi        | K1      | 59,052      | 67.2            | KT281789              | University of KwaZulu-Natald   |
| Glass         | B2      | 67,509      | 69.0            | KT880194              | Hope Collegeb                  |
| Graduation    | A1      | 52,823      | 63.5            | KG603331              | University of California–Los Angelesd |
| HanShotFirst  | A1      | 52,390      | 63.8            | KG93880               | University of California–Los Angelesd |
| HufflyPuff     | E       | 76,323      | 63.0            | KG621000              | University of California–Los Angelesd |
| HyRo          | C1      | 153,714     | 64.7            | KT281790              | University of KwaZulu-Natald   |
| Iracema64A    | A4      | 51,637      | 64.0            | KU055616              | La Salle Universityb           |
| JAMal         | B4      | 79,841      | 68.8            | KG493881              | University of California–Los Angelesd |
| JenCasNa      | A3      | 50,877      | 64.0            | KU255188              | Howard Hughes Medical Instituteb |
| Jovo          | A5      | 51,319      | 60.8            | KG493882              | University of California–Los Angelesd |
| Kimberflum    | F1      | 56,826      | 61.4            | KR935214              | Gettyburg Collegeb             |
| LadyBird      | A2      | 53,141      | 63.5            | KT88442               | St. Edward’s Universityb       |
| Lolly9        | L3      | 75,816      | 59.3            | KT281791              | University of KwaZulu-Natalb   |
| Lumos         | L3      | 75,586      | 59.3            | KT372003              | Indian River State Collegeb    |
| MichelleMyBell | N      | 42,240      | 66.0            | KG98246                | Nyack Collegeb                 |
| Mosby         | E       | 74,533      | 63.1            | KG493883              | University of California–Los Angelesd |
| Nala          | E       | 75,894      | 63.1            | KG62101                | University of California–Los Angelesd |
| NaSiaTalie    | A2      | 52,920      | 63.4            | KU297783              | Howard Hughes Medical Instituteb |
| Numberten     | B1      | 68,607      | 66.5            | KJ194583              | University of Pittsburghb      |
| Panchino      | N       | 43,516      | 65.9            | KU935727              | Lincoln Universityb             |
| Phamished     | B1      | 68,515      | 66.5            | KR816508              | Gettyburg Collegeb             |
| PhatBacter    | E       | 76,217      | 63.0            | KG62102                | University of California–Los Angelesd |
| Phatniss      | F1      | 57,293      | 61.3            | KT279576              | Johns Hopkins Universityb       |
| Phrann        | N       | 44,872      | 66.3            | KU935731              | Southern Connecticut State Universityb |
| Pioneer       | A9      | 53,219      | 62.6            | KT285706              | Indian River State Collegeb     |
| PipsqueaksN  | N       | 43,679      | 66.3            | KU935730              | College of Charlestonb          |
| PopTart      | F1      | 55,094      | 61.6            | KT281792              | University of KwaZulu-Natalb    |
| Potter        | B1      | 68,327      | 66.5            | KU867907              | University of Kansasb           |
| Romney       | A4      | 51,370      | 63.9            | KU867906              | Seton Hill Universityb          |
| Sbash         | I2      | 55,832      | 65.6            | KP027201              | University of KwaZulu-Natalb    |
| Seabiscuit    | A1      | 51,781      | 63.7            | KJ194585              | University of Pittsburghb       |
| Seagreen       | F1      | 57,766      | 61.8            | KT281793              | University of KwaZulu-Natalb    |
| SkinnyPete    | N       | 43,478      | 66.4            | KU935729              | Virginia Commonwealth Universityb |
| Snesia        | L3      | 75,626      | 59.3            | KT281794              | University of KwaZulu-Natalb    |
| Sparkdehilly  | F1      | 56,275      | 61.2            | KT895280              | James Madison Universityb       |
| Taspl4        | A1      | 51,409      | 63.9            | KT326768              | Ohio State Universityb          |
| Texage        | A3      | 50,081      | 64.0            | KT326767              | Merrimack Collegeb              |
| TheloniousMonk | A1      | 52,055      | 63.6            | KT363731              | Western Kentucky Universityb     |
| Tres          | B2      | 67,349      | 68.9            | KT365402              | James Madison Universityb       |
| Wooldri       | A3      | 50,797      | 64.0            | KT381277              | Washington State University, University of Floridaa |
| Xeno          | N       | 42,395      | 66.8            | KU935728              | Southern Connecticut State Universityb |
| Xerxes        | N       | 43,698      | 66.3            | KU935726              | University of Floridaa          |
| XFactor       | F1      | 55,617      | 61.7            | KT281795              | University of KwaZulu-Natalb    |
| Zaka          | A6      | 52,122      | 61.5            | KG603534              | University of California–Los Angelesd |
| Zakhe101      | O       | 69,653      | 65.5            | KT281796              | University of KwaZulu-Natalb    |

---

* Phage Hunters Integrating Research and Education (PHIRE) Program, University of Pittsburgh.
* Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES).
* K-RITH Mycobacterial Genetics Course.
* University of California–Los Angeles, Research Immersion Laboratory in Virology.
ACKNOWLEDGMENTS

We thank the many undergraduates and instructors who contributed to the phage discovery and genomics initiative. We especially thank David J. Asai, William Biederman, Charles Bowman, Kevin W. Bradley, Steven Cresawn, Rebecca Garlena, Roger Hendrix, William R. Jacobs, Jr., Deborah Jacobs-Sera, Paras Jain, Victoria Kasporowicz, Michelle H. Larsen, Travis Mavrich, Welkin H. Pope, Eric Rubin, Daniel A. Russell, and Viknesh Sivanathan for their assistance in phage isolation, genome analysis, and program administration.

FUNDING INFORMATION

This work, including the efforts of Graham F. Hatfull, was funded by HHS | National Institutes of Health (NIH) (GM116884). This work, including the efforts of Graham F. Hatfull, was funded by Howard Hughes Medical Institute (HHMI) (54308198).

REFERENCES

1. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812. http://dx.doi.org/10.1038/nrmicro1750.
2. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Jr, Hendrix RW, Hatfull GF. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182. http://dx.doi.org/10.1016/S0092-8674(03)00233-2.
3. Krupovic M, Bamford DH. 2010. Order to the viral universe. J Virol 84:12476–12479. http://dx.doi.org/10.1128/JVI.01489-10.
4. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world’s a phage. Proc Natl Acad Sci USA 96:2192–2197. http://dx.doi.org/10.1073/pnas.96.5.2192.
5. Hatfull GF. 2010. Bacteriophage research: gateway to learning science. Microbe 5:243–250.
6. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, Delong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, Findley AM, Gissendanner CR, Golieblovska GP, 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 I, 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. http://dx.doi.org/10.1128/mBio.01051-13.
7. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. http://dx.doi.org/10.1093/bioinformatics/btt515.
8. 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. http://dx.doi.org/10.1186/1471-2105-12-395.
9. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.
10. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. http://dx.doi.org/10.1093/nar/gki487.
11. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. http://dx.doi.org/10.1093/nar/gkh152.
12. Ackermann KH, Adams N, Adler C, Ahammed Z, Ahmad S, Allgower C, Amsbaugh J, Anderson M, Anderssen E, Arnesen H, Arnold L, Averichev GS, Baldwin A, Balewski J, Barannikova O, Barnby LS, Baudot J, Beddo M, Bekele S, Belaga VV, Bellwied R, Bennett S, Bercovitz J, Berger J, Betts W, Bischel H, Biener F, Bland LC, Bloomer M, Blyth CO, Bochm J, Bonnet BE, Bonnet D, Bossingham R, Botlo M, Boucham A, Bouillo N, Bouvier S, Bradley K, Brady FP, Braythwaite ES, Braythwaite W, Brandin A, Brown RL, Brugalette G, Byrd C, Caines H, Calderon de la Barca Sanchez M, Cardenas A, Carr L. 2001. Elliptic flux in Au+Au collisions at square root(S)NN = 130 GeV. Phys Rev Lett 86:402–407. http://dx.doi.org/10.1103/PhysRevLett.86.402.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
14. Remmert M, Biegert A, Hauser A, Söding J. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173–175. http://dx.doi.org/10.1038/nmeth.1818.
15. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germaine KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O’Brien MB, Vogelberger AM, Hrycikowian 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. http://dx.doi.org/10.1016/j.jmb.2010.01.011.
16. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF. 2013. Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. Mol Cell 49:237–248.