JVI
Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Genome Sequences of 138 Mycobacteriophages

Graham F. Hatfull,[a] the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science Program,[b] the KwaZulu-Natal Research Institute for Tuberculosis and HIV Mycobacterial Genetics Course Students,[c] and the Phage Hunters Integrating Research and Education Program[d]

Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA[a]; Howard Hughes Medical Institute, Chevy Chase, Maryland, USA[b]; KwaZulu-Natal Research Institute for Tuberculosis and HIV, Durban, South Africa[c]; and Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York, USA[d]

Bacteriophages are the most numerous biological entities in the biosphere, and although their genetic diversity is high, it remains ill defined. Mycobacteriophages—the viruses of mycobacterial hosts—provide insights into this diversity as well as tools for manipulating *Mycobacterium tuberculosis*. We report here the complete genome sequences of 138 new mycobacteriophages, which—together with the 83 mycobacteriophages previously reported—represent the largest collection of phages known to infect a single common host, *Mycobacterium smegmatis* mc^2 155.

The bacteriophage population is large (9), dynamic (8), and old (3). The genetic diversity of the population is sufficiently great that phages of different bacterial hosts typically share little recognizable sequence similarity (2). Comparing the genomes of phages that infect a single common bacterial host—and are therefore likely to be in genetic contact with one another—provides insights into both viral diversity and the mechanisms giving rise to new viruses (4). An emerging central theme is that phage genomes are pervasively mosaic, with segments being readily exchanged between genomes by horizontal exchange and thus present in numerous genomic contexts with different genetic neighbors (2). Eighty-three complete mycobacteriophage genome sequences have been previously reported (5, 6) and are grouped into 22 clusters or subclusters except for 5 singletons that have no close relatives (5, 6).

The isolation, purification, and genomic analysis of new bacteriophages offer an effective platform for providing authentic research experiences to novice scientists (1). Of the 138 phages whose genomes are reported here, 80 were isolated by participants in the Howard Hughes Medical Institute Science Education Alliance's Phage Hunters Advancing Genomics and Evolutionary Science Program (SEA-PHAGES), 10 were isolated in a mycobacterial genetics course at the University of KwaZulu-Natal, and 47 were isolated by teachers, undergraduates, and high school students in the Phage Hunters Integrating Research and Education Program (PHIRE) at the University of Pittsburgh. Each of the new phages was plaque purified, named, sequenced, and annotated by participating students and faculty. In addition, the previously isolated phage DS6A—the only one of the 138 phages that does not infect *M. smegmatis* (7)—was sequenced and annotated.

Genomic DNA isolated from each phage was shotgun sequenced using either Sanger sequencing, 454 pyrosequencing (6), or Illumina sequencing. The qualities of the sequence assemblies were evaluated, and targeted Sanger sequencing was used to resolve weak areas and to determine genome ends. All student-annotated genomes were reviewed and revised as necessary at the University of Pittsburgh prior to GenBank submission.

Of the 138 new genomes, 19 do not belong within an existing cluster and form five new clusters (J, L, M, N, and O), incorporating 3 prior singletons to form new clusters J, L, and O. Six new singletons were identified. Representatives of all existing clusters with the exception of cluster H were identified, and several clusters were further divided to generate subclusters A5 to A9, B5, K4, and K5; cluster L is divided into L1 and L2. The entire collection of 221 genomes forms a total of 36 clusters or subclusters and 8 singletons.

Although the number of mycobacteriophage genomes that have been completely sequenced has grown substantially, the discovery of new clusters, subclusters, and singletons suggests that the population of these phages remains largely undersampled.

Further information on these phage genomes is available at http://www.phagesdb.org. Subsequent reports will describe genomic details and comparative analyses of the phages announced herein.

Nucleotide sequence accession numbers. The sequences of the 138 newly determined mycobacteriophage genomes have been deposited in GenBank under the accession numbers listed in Table 1.

TABLE 1 Newly sequenced mycobacteriophage genomes

Phage	Cluster[a]	GenBank accession no.	Length (bp)	Institution[b]
ABU	B1	JF704091	68,850	[1]Lehigh U
Acadian	B5	JN699007	69,864	[2]U Pitt
Airmid	A5	JN083853	51,241	[1]UT El Paso
Akoma	B3	JN699006	68,711	[2]U Pitt
Alice	C1	JF704092	153,401	[1]U N Texas
Alma	A9	JN699005	53,177	[2]U Pitt
Arbiter	B2	JN618996	67,169	[1]U Mary Wash
Ares	B2	JN699004	67,436	[2]U Pitt
Athena	B3	JN699003	69,409	[2]U Pitt
Avrafan	G	JN699002	41,901	[2]U Pitt
Babsiella	I1	JN699001	48,420	[2]U Pitt
Backyardigan	A4	JF704093	51,308	[1]W Kentucky U
BAKA	J	JF937090	111,688	[2]U Pitt

Continued on following page

Received 1 December 2011 **Accepted** 1 December 2011

Address correspondence to Graham F. Hatfull, gfh@pitt.edu.

doi:10.1128/JVI.06870-11

TABLE 1 (Continued)

Phage	Cluster[a]	GenBank accession no.	Length (bp)	Institution[b]
BarrelRoll	K1	JN643714	59,672	[1]W Kentucky U
Bask21	E	JF937091	74,997	[2]U Pitt
Benedict	A5	JN083852	51,083	[1]Baylor U
BigNuz	Single	JN412591	48,984	[3]UKZN
BillKnuckles	A1	JN699000	51,821	[2]U Pitt
Blue7	A6	JN698999	52,288	[2]U Pitt
Bongo	M	JN699628	80,228	[1]WUSL
BPBiebs31	A1	JF957057	53,171	[1]St Joseph's U
Bruns	A1	JN698998	53,003	[2]U Pitt
Charlie	N	JN256079	43,036	[1]VCU
ChrisnMich	B4	JF704094	70,428	[1]Culver-Stockton
Courthouse	J	JN698997	110,569	[2]IU Southeast
Cuco	A5	JN408459	50,965	[1]UPR Cayey
Daisy	B3	JF704095	68,245	[1]St Joseph's U
Dandelion	C1	JN412588	157,568	[3]UKZN
DaVinci	A6	JF937092	51,547	[2]U Pitt
DeadP	F1	JN698996	56,461	[2]U Pitt
DLane	F1	JF937093	58,899	[2]U Pitt
Doom	A1	JN153085	51,421	[1]Coll William & Mary
Dori	Single	JN698995	64,613	[1]UCSC
DotProduct	F1	JN859129	55,363	[1]UCSD
Drago	F1	JN542517	54,411	[1]JMU
Drazdys	C1	JF704116	156,281	[1]UMBC
Dreamboat	A1	JN660814	51,083	[1]Lehigh U
DS6A	Single	JN698994	60,588	N/A
Elph10	E	JN391441	74,675	[1]UW River Falls
EricB	A6	JN049605	51,702	[1]NC Cen U; [1]QCC
Euphoria	A1	JN153086	53,597	[1]Lehigh U
Eureka	E	JN412590	76,174	[3]UKZN
Faith1	L2	JF744988	75,960	[1]Spelman
Fionnbharth	K4	JN831653	58,076	[1]UCSD
Firecracker	O	JN698993	71,341	[1]UCSC
Gadjet	B3	JN698992	67,949	[2]U Pitt
George	A5	JF704107	51,578	[1]JMU
Ghost	C1	JF704096	155,167	[1]Cabrini Coll
Gladiator	A6	JF704097	52,213	[1]WUSL
GUmbie	F1	JN398368	57,387	[1]Gonzaga
Hammer	A6	JF937094	51,889	[2]U Pitt
Harvey	B1	JF937095	68,193	[2]U Pitt
Hedgerow	B2	JN698991	67,451	[2]U Pitt
HelDan	A3	JF957058	50,364	[1]Loyola M'Mount
Henry	E	JF937096	76,049	[2]U Pitt
Hertubise	B1	JF937097	68,675	[2]U Pitt
Ibhubesi	F1	JF937098	55,600	[3]UKZN
IsaacEli	B1	JN698990	68,839	[2]U Pitt
JacAttac	B1	JN698989	68,311	[2]U Pitt
JAWS	K1	JN185608	59,749	[1]UL Monroe
JC27	A1	JF937099	52,169	[2]U Pitt
Jeffabunny	A6	JN699019	48,963	[1]UCSC
JHC117	A3	JF704098	50,877	[1]U Col Boulder
JoeDirt	L1	JF704108	74,914	[1]CUNY Queens
Kamiyu	B3	JN699018	68,633	[2]U Pitt
Kikipoo	B1	JN699017	68,839	[2]U Pitt
KLucky39	B1	JF704099	68,138	[1]BYU
KSSJEB	A1	JF937110	51,381	[2]U Pitt
Kugel	A1	JN699016	52,379	[2]U Pitt
Larva	K5	JN243855	62,991	[1]Coll William & Mary
Lesedi	A1	JF937100	50,486	[3]UKZN
LHTSCC	A4	JN699015	51,813	[2]U Pitt
Liefie	G	JN412593	41,650	[3]UKZN
Lilac	E	JN382248	76,260	[1]UPR Cayey
LinStu	C1	JN412592	153,882	[3]UKZN
LittleE	J	JF937101	109,086	[2]U Pitt
Marvin	Single	JF704100	65,100	[1]Cabrini Coll
MeeZee	A4	JN243856	51,368	[1]CUNY Queens
Microwolf	A3	JF704101	50,864	[1]NC State
MoMoMixon	C1	JN699626	154,573	[1]Jacksonville St
Morgushi	B1	JN638753	68,307	[1]WUSL
Mozy	F1	JF937102	57,278	[2]U Pitt; [3]UKZN
MrGordo	A1	JN020140	50,988	[1]Purdue U
Murdoc	B1	JN638752	68,600	[1]WUSL
Museum	A1	JF937104	51,426	[2]U Pitt
Mutaforma13	F1	JN020142	57,701	[1]NC State
Nappy	C1	JN699627	156,646	[3]UKZN; [1]UT El Paso
Nova	D	JN699014	65,108	[2]U Pitt
Oline	B1	JN192463	68,720	[1]CMU
Oosterbaan	B1	JF704109	68,735	[1]Calvin Coll
Optimus	J	JF957059	109,270	[1]Hope Coll
OSmaximus	B1	JN006064	69,118	[1]U Alabama Birm
PackMan	A9	JF704110	51,339	[1]UL Monroe
Patience	Single	JN412589	70,506	[3]UKZN
Perseus	A1	JN572689	53,142	[1]U Col Boulder
Phipps	B1	JF704102	68,293	[1]CMU; [1]Georgia St
Pio	C1	JN699013	156,758	[2]Olin Coll Engin
Pleione	C1	JN624850	155,586	[1]WUSL
Rakim	E	JN006062	75,706	[1]UL Monroe
Redi	N	JN624851	42,594	[1]WUSL
Rey	M	JF937105	83,724	[2]U Pitt
RidgeCB	A1	JN398369	50,844	[1]U Montana
Rockstar	A3	JF704111	47,780	[1]UL Monroe
RockyHorror	F1	JF704117	56,719	[1]UCSC
Rumpelstiltskin	L2	JN680858	69,279	[1]JMU
Saintus	A8	JN831654	49,228	[1]Georgia St
Sebata	C1	JN204348	155,286	[3]UKZN; [2]U Pitt
Send513	Single	JF704112	71,547	[1]Georgia St; [1]CMU
Serendipity	B1	JN006063	68,804	[1]Calvin Coll
SG4	F1	JN699012	59,016	[2]U Pitt
Shaka	A4	JF792674	51,369	[1]Culver-Stockton; [1]Loyola M'Mount
Shauna1	F1	JN020141	59,315	[1]BYU
ShiLan	F1	JN020143	59,794	[1]VCU
SirDuracell	E	JF937106	75,793	[2]U Pitt
SirHarley	D	JF937107	64,791	[2]U Pitt
Spartacus	F1	JQ300538	61,164	[2]U Pitt
Stinger	B4	JN699011	69,641	[2]U Pitt
Switzer	A1	JF937108	52,298	[2]U Pitt
TallGrassMM	B1	JN699010	68,133	[2]U Pitt
Thibault	J	JN201525	106,327	[1]Bucknell U
Thora	B1	JF957056	68,839	[1]Calvin Coll
ThreeOh3D2	B1	JN699009	68,992	[2]U Pitt
Timshel	A7	JF957060	53,278	[1]U North Texas
TiroTheta9	A4	JN561150	51,367	[1]W Kentucky U
Toto	E	JN006061	75,933	[1]Cabrini Coll
Trixie	A2	JN408461	53,526	[1]VCU
Turbido	A2	JN408460	53,169	[1]VCU; [1]U Montana
UPIE	L1	JF704113	73,784	[1]Oregon State U
Violet	A1	JN687951	52,481	[1]UCSC
Vista	B1	JN699008	68,494	[2]U Pitt
Vix	A3	JF704114	50,963	[1]Hope Coll
Vortex	B1	JF704103	68,346	[1]WUSL
Wally	C1	JN699625	155,299	[1]VCU
Wee	F1	HQ728524	59,320	[2]U Pitt
Wile	A4	JN243857	51,308	[1]VCU
Yoshand	B1	JF937109	68,719	[2]U Pitt
Yoshi	F2	JF704115	58,714	[1]UL Monroe
Zemanar	B4	JF704104	71,092	[1]U Mary Wash

[a] Six genomes could not be grouped with any cluster and are described as singletons (Single).

[b] Abbreviations: Baylor U, Baylor University; Bucknell U, Bucknell University; BYU, Brigham Young University; Cabrini Coll, Cabrini College; Calvin Coll, Calvin College; CMU, Carnegie Mellon University; Coll William & Mary, College of William and Mary; Culver-Stockton, Culver-Stockton College; CUNY Queens, City University of New York Queens College; Georgia St, Georgia State University; Gonzaga, Gonzaga University; Hope Coll, Hope College; IU Southeast, Indiana University Southeast; Jacksonville St, Jacksonville State University; JMU, James Madison University; Lehigh U, Lehigh University; Loyola M'Mount, Loyola Marymount University; NC Cen U, North Carolina Central University; NC State, North Carolina State University; Olin Coll Engin, Olin College of Engineering; Oregon St U, Oregon State University; Purdue U, Purdue University; QCC, Queensborough Community College; Spelman, Spelman College; St Joseph's U, Saint Joseph's University; U Alabama Birm, University of Alabama at Birmingham; U Col Boulder, University of Colorado—Boulder; UCSC, University of California, Santa Cruz; UCSD, University of California, San Diego; UL Monroe, University of Louisiana at Monroe; UKZN, University of KwaZulu-Natal; U Mary Wash, University of Mary Washington; UMBC, University of Maryland, Baltimore County; U Montana, University of Montana; U N Texas, University of North Texas; U Pitt, University of Pittsburgh; UPR Cayey, University of Puerto Rico at Cayey; UT El Paso, University of Texas at El Paso; UW River Falls, University of Wisconsin—River Falls; VCU, Virginia Commonwealth University; W Kentucky U, Western Kentucky University; WUSL, Washington University in St. Louis; N/A, not applicable (previously isolated). A superscript "1" designates institutions participating in SEA-PHAGES, a superscript "2" the PHIRE program at the University of Pittsburgh, and a superscript "3" the K-RITH mycobacterial genetics course at the University of KwaZulu-Natal.

ACKNOWLEDGMENTS

We recognize the outstanding contributions of the more than 1,600 undergraduate and high school students who participated in this phage discovery and genomics initiative, as well as those of their teachers and instructors. We thank David J. Asai, Lucia P. Barker, William

R. Bishai, Charles Bowman, Kevin W. Bradley, Nell Eisenberg, Roger Hendrix, William R. Jacobs, Jr., Deborah Jacobs-Sera, Paras Jain, Tuajuanda C. Jordan, Victoria Kaspowicz, Razi Khaja, Michelle H. Larsen, Melvina Lewis, Prashini Moodie, Welkin H. Pope, Eric Rubin, Daniel A. Russell, and Willem Sturm for their critical assistance in phage isolation, genome analysis, and program administration. Phage genomes were sequenced at the Genomics and Proteomics Core Laboratories and at the Pittsburgh Bacteriophage Institute at the University of Pittsburgh, the David H. Murdock Research Institute, the Joint Genome Institute, the Nucleic Acids Research Facilities at Virginia Commonwealth University, the University of California Santa Cruz's Genome Sequencing Center, and the Genome Institute at Washington University in St. Louis.

Funding was provided by NIH grants GM093901 and GM51975 and by the Howard Hughes Medical Institute.

REFERENCES

1. **Hanauer DI, et al.** 2006. Inquiry learning: teaching scientific inquiry. Science **314**:1880–1881.
2. **Hatfull GF, Hendrix RW.** 2011. Bacteriophages and their genomes. Curr. Opin. Virol. **1**:298–303.
3. **Hendrix RW.** 2004. Hot new virus, deep connections. Proc. Natl. Acad. Sci. U. S. A. **101**:7495–7496.
4. **Pedulla ML, et al.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113**:171–182.
5. **Pope WH, et al.** 2011. Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. PLoS One **6**:e26750.
6. **Pope WH, et al.** 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One **6**:e16329.
7. **Redmond WB, Cater JC.** 1960. A bacteriophage specific for *Mycobacterium tuberculosis*, varieties *hominis* and *bovis.* Am. Rev. Respir. Dis. **82**:781–786.
8. **Suttle CA.** 2007. Marine viruses–major players in the global ecosystem. Nat. Rev. Microbiol. **5**:801–812.
9. **Wommack KE, Colwell RR.** 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. **64**:69–114.