Genome Sequences of Six Cluster N Mycobacteriophages, Kevin1, Nenae, Parmesanjohn, ShrimpFriedEgg, Smurph, and SpongeBob, Isolated on *Mycobacterium smegmatis* mc²155

Russell A. Caratenuto III,[a] Grace O. Ciabattoni,[a] Nicolas J. DesGranges,[a] Cassidy L. Drost,[a] Longhui Gao,[a] Brianna Gipson,[a] Nicholas C. Kahler,[a] Nicole A. Kirven,[a] Julia C. Melehani,[a] Krishna Patel,[a] Alecia B. Rokes,[a] Ryan A. Seth,[a] Matthew C. West,[a] Alexa A. Alhout,[a]* Francis F. Akoto,[a] Nicole Capogna,[a] Netta Cudkevich,[a] Lee H. Graham,[a] Matthew S. Grapel,[a] Maaz M. Haleem,[a] Jamie B. Korenberg,[a] Brooke P. Lichak,[a] Lauren N. McKinley,[a] Kourtney R. Mendello,[a] Caitlin E. Murphy,[a] Lauren M. Pyfer,[a] Wascar A. Ramirez,[a] Julia R. Reisner,[a] Rachel H. Swope,[a] Matthew J. Thoonkuzhy,[a]* Lauren A. Vargas,[a] Croldy A. Veliz,[a] Katherine R. Volpe,[a] Kevin D. Zhang,[a] Dylan Z. Faltine-Gonzalez,[a] Caitlin M. Zuilkoski,[a] Catherine M. Mageeney,[a]* Hamidu T. Mohammed,[a] Margaret A. Kenna,[a] Vassie C. Ware[a]

[a]Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania, USA

ABSTRACT The annotation of six cluster N *Mycobacterium smegmatis* phages (Kevin1, Nenae, Parmesanjohn, ShrimpFriedEgg, Smurph, and SpongeBob) reveals regions of genomic diversity, particularly within the central region of the genome. The genome of Kevin1 includes two orphams (genes with no similarity to other phage genes), with one predicted to encode an AAA-ATPase.

Unique mechanisms of immunity have been discovered for cluster N mycobacteriophages (1). Continued exploration of new cluster N phage genomes may reveal additional novel genes, some of which may be implicated in prophage-mediated defense systems.

Mycobacteriophages were isolated by enrichment from soil on host *Mycobacterium smegmatis* mc²155. Soil samples were placed in 7H9 medium with the host and incubated for 3 days at 37°C, and supernatants were sterile filtered for testing on host lawns. After purification and amplification of the phages, electron microscopy determined a *Siphoviridae* morphology for each of them. Sequencing, assembly, and finishing of the genomes were performed according to Russell (2). Briefly, sequencing libraries were prepared from genomic DNA (isolated by phenol-chloroform extraction) using an Ultra II FS kit (NEB) with dual-indexed barcoding and pooled with others for 48 total libraries run on an Illumina MiSeq instrument, yielding at least 482,600 single-end 150-base reads and at least 201-fold coverage for each genome. Reads were assembled using Newbler version 2.9 (3) with default settings, yielding a single phage contig for each. Consed version 29 (4) was used to check for completeness, accuracy, and genomic termini. The genomic characteristics are summarized in Table 1.

DNA Master (5), embedded with GeneMark 2.5 (6) and GLIMMER 3.0 (7), was used to annotate the open reading frames (ORFs). NCBI BLASTP 2.7, NCBI's conserved domain database (8), and HHpred (9) were used to assign protein functions. Phamerator (10) was used for comparative genomic analysis.

The genomic architecture conforms to that of other cluster N genomes (1). Conserved structural assembly genes are located in the left arm, followed by a central variable region and a diverse right arm that includes DNA maintenance genes (1). The lysis cassette in the left arm encodes lysin A and lysin B in Kevin1; other genomes (reported here) encode only lysin A. One tRNA was predicted in SpongeBob and Parmesanjohn using ARAGORN (11) and tRNAscan (12) but was not annotated, because its location disrupts the leftward transcribed immunity repressor ORF. The central

TABLE 1 Characteristics of cluster N genomes

Phage name	Soil isolation source	Genome size (bp)	Genome terminus (13-base 3′ overhang)	GC content (%)	No. of ORFs
Kevin1	Naugatuck, CT	41,988	5′-CCCGCCGCCCTCG	66.2	69
Nenae	Fairfield, NJ	42,597	5′-CCCGCCGCCTTGG	66.1	70
Parmesanjohn	Bethlehem, PA	43,700[a]	5′-CCCGCCGCAATGG	66.3	72
ShrimpFriedEgg	Bethlehem, PA	42,594	5′-CCCGCCGCCTTGG	66.1	70
Smurph	Manchester, NH	43,700[a]	5′-CCCGCCGCAATGG	66.3	72
SpongeBob	La Cañada Flintridge, CA	41,287	5′-CCCGCCGCCTTGG	66.2	65

[a] Although the Parmesanjohn (isolated in 2017) and Smurph (isolated in 2018) genomes differ by only 2 nucleotides, the geographic location and year of isolation make cross-contamination unlikely.

variable regions of Parmesanjohn, Smurph, and SpongeBob share 97% average nucleotide identity (ANI), while Nenae, ShrimpFriedEgg, and Redi (GenBank accession number JN624851) share 100% ANI. The unique variable region of Kevin1 contains two orphams (*30* and *31*), of which gene *30* is predicted to encode an AAA-ATPase. Functional characterization of these orphams may offer additional insights into cluster N immunity mechanisms.

Data availability. The GenBank and NCBI SRA database (for raw reads) accession numbers, respectively, are MK524500 and SRX5572872 (Kevin1), MK524520 and SRX5572867 (Nenae), MK524515 and SRX5572865 (Parmesanjohn), MK524528 and SRX5572863 (ShrimpFriedEgg), MK524518 and SRX5572864 (Smurph), and MK524509 and SRX5572876 (SpongeBob).

REFERENCES

1. Dedrick RM, Jacobs-Sera D, Bustamante CA, Garlena RA, Mavrich TN, Pope WH, Reyes JC, Russell DA, Adair T, Alvey R, Bonilla JA, Bricker JS, Brown BR, Byrnes D, Cresawn SG, Davis WB, Dickson LA, Edgington NP, Findley AM, Golebiewska U, Grose JH, Hayes CF, Hughes LE, Hutchison KW, Isern S, Johnson AA, Kenna MA, Klyczek KK, Mageeney CM, Michael SF, Molloy SD, Montgomery MT, Neitzel J, Page ST, Pizzorno MC, Poxleitner MK, Rinehart CA, Robinson CJ, Rubin MR, Teyim JN, Vazquez E, Ware VC, Washington J, Hatfull GF. 2017. Prophage-mediated defence against viral attack and viral counter-defence. Nat Microbiol 2:16251. https://doi.org/10.1038/nmicrobiol.2016.251.
2. Russell DA. 2018. Sequencing, assembling, and finishing complete bacteriophage genomes. Methods Mol Biol 1681:109–125. https://doi.org/10.1007/978-1-4939-7343-9_9.
3. Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95:315–327. https://doi.org/10.1016/j.ygeno.2010.03.001.
4. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. https://doi.org/10.1093/bioinformatics/btt515.
5. DNAMaster. http://cobamide2.bio.pitt.edu/computer.htm.
6. 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.
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. 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.
9. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
10. 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.
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. https://doi.org/10.1093/nar/gkh152.
12. 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. https://doi.org/10.1093/nar/25.5.955.