Complete Genome Sequence of *Klebsiella pneumoniae* Jumbo Phage Miami

Daniel Mora,[a] Lauren Lessor,[a] Tram Le,[a] James Clark,[a] (iD) Jason J. Gill,[a] (iD) Mei Liu[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA

ABSTRACT Bacteriophage Miami infects *Klebsiella pneumoniae*, a Gram-negative pathogen that is becoming an increasing threat to public health due to its multidrug resistance. Here, we describe the annotation of the 253,383-bp jumbo phage Miami genome sequence and its similarity to other myophages.

*K*lebsiella pneumoniae is a Gram-negative, encapsulated, nonmotile bacterium that can be found in soil, water, and on human mucosal surfaces (1). Over the last few decades, *K. pneumoniae* has gained a reputation for being an opportunistic nosocomial multidrug-resistant (MDR) pathogen (1). The reputation of *K. pneumoniae* has guided scientific efforts to develop alternative methods to treat this bacterium, largely due to the lack of available antibiotics that effectively treat MDR *K. pneumoniae* infections (2). The reimplementation of bacteriophages as an alternative treatment method in the clinical setting has led to a revival of bacteriophage therapeutics (3). The genome annotation of *K. pneumoniae* myophage Miami is detailed in this report.

Bacteriophage Miami was isolated from a wastewater treatment plant in Houston, TX, against clinical *K. pneumoniae* strain 44809 (GenBank accession no. NDDK01000000), which has sequence type 1502, bioinformatically predicted using the Kaptive database (4, 5). Host bacteria were cultured on tryptic soy broth at 37°C with aeration, and phage isolation and propagation were performed using the soft-agar overlay method (6). DNA was purified following a previously described protocol using the Promega Wizard DNA cleanup system (7). DNA fragmentation and library preparation were done following the manufacturer's protocol to generate inserts with an average size of 300 bp using a TruSeq Nano kit (Illumina) and sequenced on an Illumina iSeq 100 instrument with paired-end 150-bp reads using iSeq 300-bp chemistry. The resulting 802,622 reads were quality controlled with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed with FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). SPAdes v3.5.0 was then used for *de novo* assembly with default settings to generate a raw contig at 37.6-fold coverage (8). Because the sequences of the raw contig ends are often not accurate (either with redundant sequences or missing sequences), the end sequences were corrected by PCR to amplify the end region (with the primers TTTATGCCGTCTCCCCTCT and GAGTAGCCAGCGACACAGAT facing off the ends of the contig) and Sanger sequencing of the product. Identification of genes was conducted using MetaGeneAnnotator v1.0, GLIMMER v3.0, and ARAGORN v2.36 (9–11). Following the gene identification, gene function was predicted using InterProScan v5.33, a BLAST v2.9.0 search against the NCBI nonredundant (nr) and Swiss-Prot databases (12–15), and TMHMM v2.0 with default settings. The genome-wide DNA sequence similarity of Miami was calculated using progressiveMauve v2.4 (16). The entirety of the genome analysis of phage Miami was performed via the CPT Galaxy and Web Apollo interfaces (17–19).

Miami possesses a 253,383-bp genome with a coding density of 94% and 300 predicted protein-coding genes, only 47 of which have a predicted function. The genome G+C content is 43.9%, which is lower than the average G+C content of *K. pneumoniae*

(52.7%). As is common with "jumbo" phages, Miami is phylogenetically distant from other known phages, sharing only 124 similar proteins with *Erwinia* phages Rising Sun and Joad (GenBank accession no. MF459646 and MF459647, respectively), determined by BLASTp (E, <0.001). Miami also shares 105 proteins similar to those in *Vibrio* phages 2 TSL-2019, USC-1, Aphrodite1, and 5 TSL-2019, which are classified in the genus *Aphroditevirus* (MK368614, MK905543, MG720308, and MK358448, respectively). progressiveMauve analyses determined that Miami shares the highest nucleotide identity with *Erwinia* phage Rising Sun at 11.4%.

Data availability. The genome sequence of phage Miami was deposited under GenBank accession no. MT701590 and BioSample accession no. SAMN14609641. The BioProject accession no. is PRJNA222858, and the SRA accession no. is SRR11558348.

REFERENCES

1. Paczosa MK, Mecsas J. 2016. Klebsiella pneumoniae: going on the offense with a strong defense. Microbiol Mol Biol Rev 80:629–661. https://doi.org/10.1128/MMBR.00078-15.
2. Satlin MJ, Chen L, Patel G, Gomez-Simmonds A, Weston G, Kim AC, Seo SK, Rosenthal ME, Sperber SJ, Jenkins SG, Hamula CL, Uhlemann A-C, Levi MH, Fries BC, Tang Y-W, Juretschko S, Rojtman AD, Hong T, Mathema B, Jacobs MR, Walsh TJ, Bonomo RA, Kreiswirth BN. 2017. Multicenter clinical and molecular epidemiological analysis of bacteremia due to carbapenem-resistant Enterobacteriaceae (CRE) in the CRE epicenter of the United States. Antimicrob Agents Chemother 61:e02349-16. https://doi.org/10.1128/AAC.02349-16.
3. Dunavan CP. 2020. Superbug vs. superphage: a review of the perfect predator and an interview with Dr. Chip Schooley. Am J Trop Med Hyg 102:245–248. https://doi.org/10.4269/ajtmh.19-0770.
4. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, Holt KE. 2016. Identification of Klebsiella capsule synthesis loci from whole genome data. Microb Genom 2:e000102. https://doi.org/10.1099/mgen.0.000102.
5. Wick RR, Heinz E, Holt KE, Wyres KL. 2018. Kaptive Web: user-friendly capsule and lipopolysaccharide serotype prediction for Klebsiella genomes. J Clin Microbiol 56:e00197-18. https://doi.org/10.1128/JCM.00197-18.
6. Chen M, Zhang L, Abdelgader SA, Yu L, Xu J, Yao H, Lu C, Zhang W. 2017. Alterations in gp37 expand the host range of a T4-like phage. Appl Environ Microbiol 83:e01576-17. https://doi.org/10.1128/AEM.01576-17.
7. Summer EJ. 2009. Preparation of a phage DNA fragment library for whole genome shotgun sequencing. Methods Mol Biol 502:27–46. https://doi.org/10.1007/978-1-60327-565-1_4.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
9. 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.
10. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396. https://doi.org/10.1093/dnares/dsn027.
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. 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.
13. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
14. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
16. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
17. Ramsey J, Rasche H, Maughmer C, Criscione A, Mijalis E, Liu M, Hu JC, Young R, Gill JJ. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. PLoS Comput Biol 16:e1008214. https://doi.org/10.1371/journal.pcbi.1008214.
18. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. 2019. Apollo: democratizing genome annotation. PLoS Comput Biol 15:e1006790. https://doi.org/10.1371/journal.pcbi.1006790.
19. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, Taylor J, Nekrutenko A. 2020. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res 48:W395–W402. https://doi.org/10.1093/nar/gkaa434.