AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of the Novel *Klebsiella pneumoniae* Phage Marfa

Laith Harb,[a] Justin Boeckman,[a,b] Heather Newkirk,[a] Mei Liu,[a] Jason J. Gill,[a,b] Jolene Ramsey[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[b]Department of Animal Science, Texas A&M University, College Station, Texas, USA

ABSTRACT Here, we describe the complete genome sequence of the T4-like *Klebsiella pneumoniae* myophage Marfa. In its 168,532-bp genome, Marfa has 289 genes, for which 122 gene functions were predicted. Many similar proteins are shared between Marfa and phage T4, as well as its closest phage relatives.

Carbapenemase-producing *Klebsiella pneumoniae* (KPC) is a multidrug-resistant enteric bacterium that has emerged as a global health threat (1). KPC is a causative agent of pneumonia and is one of the most common pathogens in hospital-acquired infections, with an overall mortality rate in infected patients between 22% and 59% (2, 3). Reliable therapeutic drugs are not available; therefore, bacteriophage-based therapy is a potential alternative. Here, we describe the novel *Klebsiella pneumoniae* phage Marfa.

Bacteriophage Marfa was isolated from chloroform-sterilized pooled swine fecal samples from Texas and Michigan. Marfa replicates on a pKpQIL plasmid-cured derivative of *K. pneumoniae* strain 1776c (4) under aerobic conditions in tryptic soy broth or agar (Difco) at 37°C, and phage propagation was done using the soft agar overlay method (5). Genomic DNA was isolated with the Promega Wizard DNA clean-up kit, according to the protocol from Summer (6) and prepared for sequencing with a TruSeq Nano low-throughput kit. Sequencing occurred on an Illumina MiSeq instrument, with 250-bp paired-end reads. From 382,820 total reads in the index, sequence reads were trimmed using FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) after quality control with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). Assembly into a single contig at 119-fold coverage was accomplished with SPAdes v3.5.0 using default parameters (7). The contig was confirmed to be complete by PCR (forward primer, 5′-CCTTGCTGGTCCGTGATTT-3′; reverse primer, 5′-CTGGTGGGTCGTG ATAAGATG-3′) using outward-facing primers and by Sanger sequencing of the product. Protein-coding and tRNA genes were predicted by GLIMMER v3.0, MetaGeneAnnotator v1.0, and ARAGORN v2.36 (8–10). TransTermHP v2.09 was used to annotate rho-independent terminators (11). All of the tools used for analysis and annotation are available on the Center for Phage Technology Galaxy and Web Apollo instances (https://cpt.tamu.edu/galaxy-public/) (12, 13). Functional annotations used evidence from InterProScan v5.22-61 and BLAST v2.2.31 versus the NCBI nonredundant, UniProtKB, Swiss-Prot, and TrEMBL databases, with a cutoff of 0.001 for the E value (14–16). As needed, TMHMM v2.0 and HHpred with ummiclust30_2018_08 for multiple-sequence alignment (MSA) generation and PDB_mmCIF70 for modeling in the HHsuite v3.0 release provided supplementary evidence (17, 18). Marfa was negatively stained with 2% (wt/vol) uranyl acetate and viewed by transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (19).

Marfa is a myophage with a 168,532-bp genome, a coding density of 94.8%, and a G+C content of 41.0%. The G+C content is less than that of *K. pneumoniae*, which has

Citation Harb L, Boeckman J, Newkirk H, Liu M, Gill JJ, Ramsey J. 2019. Complete genome sequence of the novel *Klebsiella pneumoniae* phage Marfa. Microbiol Resour Announc 8:e00748-19. https://doi.org/10.1128/MRA.00748-19.

Editor Simon Roux, DOE Joint Genome Institute

Address correspondence to Jolene Ramsey, jolenerr@tamu.edu.

Received 20 June 2019
Accepted 29 June 2019
Published 18 July 2019

an average G+C content of >50%. Genome analysis revealed 289 protein-coding genes, for which 122 gene functions were predicted. Marfa is T4-like, based on similarity to 111 phage T4 proteins, but it also shares 259 similar proteins with *Klebsiella* myophage vB_Kpn_F48 (GenBank accession number MG746602). Analysis with progressiveMauve demonstrates that Marfa shares 94% sequence identity across 93% of its genome with vB_Kpn_F48 (20). PhageTerm analysis predicts Marfa to have permuted ends and headful packaging, which are expected of T4-like phages (21, 22). Interestingly, the L-shaped tail fiber gene (NCBI accession number QDB71908) has 37% identity and 79% coverage to the L-shaped tail fiber protein of phage T5 (NCBI accession number YP_006961).

Data availability. The genome sequence and associated data for phage Marfa were deposited under GenBank accession number MN044033, BioProject accession number PRJNA222858, SRA accession number SRR8869231, and BioSample accession number SAMN11360393.

ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (awards EF-0949351 and DBI-1565146) and by the National Institutes of Health (NIAID award number AI121689).

We thank Thomas Walsh of the Weill Cornell Medical School and Karen Frank of the National Institutes of Health for the provision of bacterial isolates. Additional support came from the Center for Phage Technology (CPT), an Initial University Multidisciplinary Research Initiative supported by Texas A&M University and Texas AgriLife, and from the Texas A&M University Department of Biochemistry and Biophysics. We are grateful for the advice and support of the CPT staff.

This announcement was prepared in partial fulfillment of the requirements for BICH464 Bacteriophage Genomics, an undergraduate course at Texas A&M University.

REFERENCES

1. Lee GC, Burgess DS. 2012. Treatment of *Klebsiella pneumoniae* carbapenemase (KPC) infections: a review of published case series and case reports. Ann Clin Microbiol Antimicrob 11:32. https://doi.org/10.1186/1476-0711-11-32.
2. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV, Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci U S A 112:E3574–E3581. https://doi.org/10.1073/pnas.1501049112.
3. Chen LF, Anderson DJ, Paterson DL. 2012. Overview of the epidemiology and the threat of *Klebsiella pneumoniae* carbapenemases (KPC) resistance. Infect Drug Resist 5:133–141. https://doi.org/10.2147/IDR.S26613.
4. 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.
5. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol 8:R22. https://doi.org/10.1186/gb-2007-8-2-r22.
12. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
13. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:R93. https://doi.org/10.1186/gb-2013-14-8-r93.
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. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.

17. Krogh A, Larsson B, Heijne von G, Sonnhammer EL. 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.
18. Zimmermann L, Stephens A, Nam S-Z, 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.
19. Valentine RC, Shapiro BM, Stadtman ER. 1968. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. Biochemistry 7:2143–2152. https://doi.org/10.1021/bi00846a017.
20. 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.
21. Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep 7:8292. https://doi.org/10.1038/s41598-017-07910-5.
22. Fokine A, Rossmann MG. 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4:e28281. https://doi.org/10.4161/bact.28281.