Complete Genome Sequence of *Burkholderia cenocepacia* Phage Mica

James Garcia,[a] Guichun Yao,[a] Maria Guadalupe Vizoso-Pinto,[b] James Clark,[a] Tram Le,[a] Carlos Gonzalez,[a] Jason Gill,[a] Mei Liu[a]

[a]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[b]Infection Biology Lab, Instituto Superior de Investigaciones Biológicas, National Council of Scientific and Technological Research, National University of Tucuman, San Miguel de Tucumán, Argentina

ABSTRACT *Burkholderia cenocepacia* is a multidrug-resistant Gram-negative pathogen known to colonize patients with chronic granulomatous disease and cystic fibrosis. Here, we describe *Burkholderia* phage Mica, which is predicted to be a lysogenic myophage based on the similarity of its structural proteins to *Enterobacteria* phage P2 and *Burkholderia* phage KL3.

Bacteria within the genus *Burkholderia* are known to cause disease in plants. *Burkholderia cenocepacia* is also a multidrug-resistant Gram-negative bacterium within the *B. cepacia* complex (BCC) group of opportunistic pathogens that colonizes patients with cystic fibrosis and chronic granulomatous disease (1, 2). The investigation into *Burkholderia* phage may yield an effective biocontrol agent that can aid in the treatment of afflicted patients as well as protect crop plants.

Phage Mica was isolated from a soil sample collected from Hermann Park in Houston, TX (GPS coordinates, 29.7135373, −95.3910571). The filtered soil extract (soil sample mixed with phosphate-buffered saline [PBS] buffer) was enriched against a *B. cenocepacia* clinical strain (MS1) in tryptic nutrient broth at 37°C overnight. Phage was purified and propagated from the enrichment using the soft-agar overlay method (3) on *B. cenocepacia* strain MS1 grown on tryptic nutrient agar at 37°C. Genomic DNA was purified by a modification of the Promega Wizard DNA kit protocol described by Summer (4). DNA libraries were prepared with 550-bp inserts using an Illumina TruSeq Nano kit. The paired-end 500-bp reads were sequenced on an Illumina MiSeq instrument using v2 500-cycle chemistry. The 307,034 total sequence reads were quality controlled with FastQC v0.11.9 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed with the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The genome sequence was assembled with SPAdes v3.5.0 with default parameters (5). A single contig with 200.1-fold coverage was confirmed to be closed with Sanger sequencing that was PCR amplified by primers (5′-TCGAACGACCATTCGATGC-3′ and 5′-ACATTGCCAGCCTGCAGG-3′). Protein-coding genes were called by GLIMMER v3.0 and MetaGeneAnnotator v1.0 (6, 7). tRNA genes were determined by ARAGORN v2.36, and rho-independent terminators were predicted with TransTermHP v2.09 (8, 9). Gene functions were predicted by conserved domains identified with InterProScan v5.33 and similarities found using BLASTp v2.9.0 with a 0.001 maximum expectation value cutoff against the NCBI nonredundant (nr) and Swiss-Prot databases (10–12). The prediction of transmembrane domains was performed with TMHMM v2.0 (13). The prediction of spanin proteins was performed based on the analysis of lipoprotein signals with LipoP v1.0 (14). Further domain analysis was performed with HHpred v3.2.0 (15). Genome-wide DNA sequence similarity was calculated by progressiveMauve v2.4 (16). All

annotation tools accessed through the CPT Galaxy and Web Apollo interfaces (17–19) were run at default settings unless otherwise stated.

Mica has a genome sequence of 43,707 bp with a coding density of 94% and a G+C content of 62%. A cohesive end site (*cos*) with a 13-bp overhang was predicted by PhageTerm (20). Analysis predicted 69 protein-coding genes, of which 43 were assigned putative functions. Mica was predicted to be a lysogenic myophage based on presence of an integrase and the similarity of its structural proteins to phage P2 (GenBank accession no. NC_001895) and *Burkholderia* phage KL3 (NC_015266). BLASTn indicated that phage Mica shares 47% similarity to myophage phiRSP (MH252365). The lysis cassette genes were found in a cluster and consist of i/o-spanins, a signal anchor release (SAR) endolysin lysozyme, and a class II holin. An ice nucleation protein is predicted through functional analysis, which may act as an accessory phytopathogen virulence factor in the lysogen (21).

Data availability. The genome sequence of phage Mica was deposited under GenBank accession no. MT701586 and BioSample accession no. SAMN14609637. The BioProject accession no. is PRJNA222858, and the SRA accession no. is SRR11558335.

REFERENCES

1. Roszniowski B, McClean S, Drulis-Kawa Z. 2018. Burkholderia cenocepacia prophages—prevalence, chromosome location and major genes involved. Viruses 10:297. https://doi.org/10.3390/v10060297.
2. Vandamme P, Holmes B, Vancanneyt M, Coenye T, Hoste B, Coopman R, Revets H, Lauwers S, Gillis M, Kersters K, Govan JRW. 1997. Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. Int J Syst Bacteriol 47:1188–1200. https://doi.org/10.1099/00207713-47-4-1188.
3. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
13. 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.
14. Juncker AS, Willenbrock H, von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662. https://doi.org/10.1110/ps.0303703.
15. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kubler J, Lozajic M, Gabler F, Soding 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.
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.

20. Garneau JR, Depardieu F, Fortier LC, 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.

21. Pandey R, Usui K, Livingstone RA, Fischer SA, Pfaendtner J, Backus EHG, Nagata Y, Frohlich-Nowoisky J, Schmuser L, Mauri S, Scheel JF, Knopf DA, Poschl U, Bonn M, Weidner T. 2016. Ice-nucleating bacteria control the order and dynamics of interfacial water. Sci Adv 2:e1501630. https://doi.org/10.1126/sciadv.1501630.