Complete Genome Sequence of *Burkholderia gladioli* Phage Maja

Zihao Yu,ᵃ Guichun Yao,ᵃ Maria Guadalupe Vizoso-Pinto,ᵇ Lichang Sun,ᶜ ⓘ Ry Young,ᵃ ⓘ Carlos Gonzalez,ᵃ ⓘ Mei Liuᵃ

ᵃCenter for Phage Technology, Texas A&M University, College Station, Texas, USA
ᵇInfection Biology Lab, Instituto Superior de Investigaciones Biológicas, National Council of Scientific and Technological Research, National University of Tucuman, San Miguel de Tucumán, Argentina
ᶜInstitute of Food Quality and Safety, Jiangsu Academy of Agricultural Sciences, Nanjing, People's Republic of China

ABSTRACT *Burkholderia gladioli* is a Gram-negative bacterium associated with cystic fibrosis infections. Here, we describe the genome sequence of *B. gladioli* phage Maja. Maja is most related to another *Burkholderia* phage, BcepF1, and may be a temperate phage, despite the absence of repressor or integrase homologs in its genome sequence.

Burkholderia gladioli* is a Gram-negative bacterium formerly classified as *Pseudomonas* (1). Although originally regarded essentially as a phytopathogen (2), *B. gladioli* can also colonize the lungs of cystic fibrosis patients (3, 4). Additionally, *B. gladioli* and the closely related species *B. cepacia* are associated with multidrug resistance (5). For the purpose of finding phage for therapeutic applications, *B. gladioli* phage Maja was isolated, and its genome annotation was performed.

Maja was isolated in 2019 from a soil sample 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 overnight against a *B. gladioli* clinical isolate (strain BgPK) in tryptic nutrient broth at 37°C. Phage was purified and propagated from the enrichment using *B. gladioli* strain BgPK as the host on tryptic nutrient broth agar at 37°C by the soft-agar overlay method (6). Phage genomic DNA was extracted from the precipitated phage particles and purified using a Wizard DNA cleanup kit as previously described (7). DNA libraries were prepared using an Illumina TruSeq Nano kit with 300-bp inserts. Genome sequencing was performed on an Illumina MiSeq instrument using v2 300-cycle chemistry. FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and FastX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/download.html) were used for the quality control and trimming of a total of 383,974 sequence reads. *De novo* genome assembly was done by SPAdes v3.5.0 (8). As a result, a single contig was assembled with 270.6-fold sequence coverage. The closure of the genome sequence was done by PCR (forward, 5′-CTTAGAAAGCCGCCCATAGA-3′; reverse, 5′-GGTATCGACATGGCGAAGAA-3′) and confirmed to be complete by Sanger sequencing. The structural annotation of the genome was performed with GLIMMER v3.0 (9) and MetaGeneAnnotator v1.0 (10). TransTermHP v2.09 was used for the prediction of rho-independent termination sites (11), and tRNAs were detected by ARAGORN v2.36 (12). Gene function prediction mainly relied on conserved domain searches with InterProScan v5.22 (13) and sequence similarity search by BLAST v2.2.31 at a 0.001 maximum expectation value (14). Specifically, the sequence similarity search was set at an E value of <0.001 against the NCBI nonredundant and Swiss-Prot/TrEMBL databases (15). Transmembrane domains were found and annotated using TMHMM v2.0 (16). Several structural predictions were performed with HHpred v3.2.0 (17). In addition, progressiveMauve v2.4 calculated genome-wide

DNA sequence similarity to top BLAST nucleotide hits (18). All annotation tools (except HHpred) were used in the Center for Phage Technology Galaxy and Web Apollo platforms hosted at https://cpt.tamu.edu/galaxy-pub (19–21). All software was used with default parameters unless otherwise specified.

Phage Maja has a 68,393-bp genome sequence with 54.5% GC content. A total of 114 protein-coding genes were predicted, with a coding density of 92.4%. Overall, 33 protein coding genes were assigned a function. Comparative genomics and BLASTp analysis revealed that Maja is most related to another *Burkholderia* phage, BcepF1 (GenBank accession number NC_009015), both on the nucleotide level (56.8% nucleotide identity as determined by progressiveMauve) and on the protein level (sharing 90 similar proteins by BLASTp at $E < 0.001$), but Maja's genome sequence contains a 15-kb region where the sequence identity is very low (not possible to align by BLASTn at $E < 0.001$) to all other known phages. Several structural genes were predicted, including the tail fiber, baseplate, and tail sheath proteins. In addition, two free-standing homing endonuclease genes, one GIY-YIG and the other HNH, were predicted in the genome sequence. Four lysis genes were identified, including two spanin genes, two endolysin genes, one glycosyl hydrolase, and one transglycosylase. Finally, many Maja genes exhibit high similarity to bacterial genes, suggesting that Maja may be temperate despite the absence of repressor or integrase homologs, or recently derived from a temperate ancestor, which could explain its low GC content (54.5%) compared to that of its host (~68%) (22).

Data availability. The genome sequence of Maja is available in GenBank under accession number MT708549. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR11558342, and SAMN14609647, respectively.

REFERENCES

1. Stoyanova M, Pavlina I, Moncheva P, Bogatzevska N. 2007. Biodiversity and incidence of Burkholderia species. Biotechnol Biotechnol Equip 21:306–310. https://doi.org/10.1080/13102818.2007.10817465.
2. McCulloch L. 1924. A bacterial blight of gladioli. J Agric Res 27:e0225–e0230.
3. Christenson JC, Welch DF, Mukwaya G, Muszynski MJ, Weaver RE, Brenner DJ. 1989. Recovery of Pseudomonas gladioli from respiratory tract specimens of patients with cystic fibrosis. J Clin Microbiol 27:270–273. https://doi.org/10.1128/JCM.27.2.270-273.1989.
4. Kennedy MP, Coakley RD, Donaldson SH, Aris RM, Hohneker K, Wedd JP, Knowles MR, Gilligan PH, Yankaskas JR. 2007. Burkholderia gladioli: five year experience in a cystic fibrosis and lung transplantation center. J Cyst Fibros 6:267–273. https://doi.org/10.1016/j.jcf.2006.10.007.
5. Simpson IN, Finlay J, Winstanley DJ, Dewhurst N, Nelson JW, Butler SL, Govan JRW. 1994. Multi-resistance isolates possessing characteristics of both Burkholderia (Pseudomonas) cepacia and Burkholderia gladioli from patients with cystic fibrosis. J Antimicrob Chemother 34:353–361. https://doi.org/10.1093/jac/34.3.353.
6. Adams MH. 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.
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. 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. 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.
13. 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.
14. 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.

15. The UniProt Consortium. 2018. UniProt: the universal protein knowledge-base. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.

16. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting trans-membrane 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.

17. 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.

18. 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.

19. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning 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.

20. 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.

21. 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.

22. Ettinger CL, Shehata HR, Johnston-Monje D, Raizada MN, Eisen JA. 2015. Draft genome sequence of Burkholderia gladioli strain UCD-UG_CHAPALOTE (phylum Proteobacteria). Genome Announc 3:e01462-14. https://doi.org/10.1128/genomeA.01462-14.