AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Burkholderia cenocepacia* Phage Paku

Sefanit Rezene,[a] Guichun Yao,[b] Tram Le,[a,b] Ben Burrowes,[c] Carlos Gonzalez,[b] Mei Liu,[a,b] Jason Gill[b]

[a]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA
[b]Center for Phage Technology, Texas A&M University, College Station, Texas, USA
[c]BB Phage Consultancy, LLC, Georgetown, Texas, USA

ABSTRACT *Burkholderia cenocepacia* is able to cause infections in cystic fibrosis patients. *B. cenocepacia* phage Paku has a 42,727-bp genome sharing a phiKMV-like genome arrangement. T7-like tail components were identified in parallel with a tyrosine integrase, suggesting that Paku might exhibit a temperate lifestyle, an atypical feature for an *Autographiviridae* phage.

Burkholderia cenocepacia is an opportunistic pathogen that is found in the environment and is known to cause infections in cystic fibrosis (CF) patients that are difficult to treat, because of its antibiotic resistance and possession of multiple virulence determinants (1). Research has been conducted to better understand *B. cenocepacia* as a pathogen, in hopes of increasing CF patients' life expectancy (2). We are interested in understanding the genomic diversity of *B. cenocepacia* phages in order to develop phage therapy for controlling this bacterium.

Bacteriophage Paku was isolated from a soil sample obtained in 2018 in Lincoln, Nebraska, using *Burkholderia cenocepacia* GIIIa as the host. The soil sample was filtered (0.2-μm pore size), plaque purified three times, and propagated on *B. cenocepacia* using the soft agar overlay method as described previously (3). DNA was purified by the modified Wizard kit protocol described by Summer (4). DNA libraries were prepared with 300-bp inserts with a Swift 2S Turbo kit and sequenced with a MiSeq Nano system using 500-cycle v2 chemistry. The total of 17,386 raw reads were quality controlled using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed with the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit). The genome was assembled with SPAdes v3.5.0 (5), resulting in a single contig with 65.0-fold coverage. The genome was closed bioinformatically based on alignment to another contig with a different opening. Annotation was done on the Center for Phage Technology (CPT) Galaxy-Apollo phage annotation platform (https://cpt.tamu.edu/galaxy-pub) (6–8). This process was broken into two parts, with the structural annotation done by GLIMMER v3 (9) and MetaGeneAnnotator v1.0 (10). tRNAs were detected with ARAGORN v2.36 (11) and tRNAScan-SE v2.0 (12). The functional annotation was completed by use of InterProScan v5.48 (13), BLAST v2.9.0 (14), TMHMM v2.0 (15), HHpred (16), LipoP v1.0 (17), and SignalP v5.0 (18). BLAST searching against the NCBI nonredundant and Swiss-Prot databases (19) was utilized. A genome-wide DNA sequence similarity analysis was performed by progressiveMauve v2.4 (20). All tools were run with default settings unless otherwise specified.

Phage Paku has a 42,727-bp genome with a coding density of 95.6% and a GC content of 61.9%. The precise genome termini could not be determined by PhageTerm (21). In total, 55 protein-coding genes and 1 tRNA gene were predicted. NCBI taxonomy has placed Paku in the subfamily *Okabevirinae*; however, it shares only ~30% nucleotide identity with other members of this subfamily and thus would likely constitute its own genus in this group. Paku shares 31 proteins (BLASTp, E value of $<10^{-5}$) with *Burkholderia* phage Bp-AMP4

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Mei Liu, meiliu@tamu.edu.

The authors declare no conflict of interest.

Received 17 December 2021
Accepted 12 March 2022
Published 28 March 2022

(GenBank accession number HG796221) and 30 proteins with *Burkholderia* phages Bp-AMP1, Bp-AMP2, Bp-AMP3, and AMP1 (GenBank accession numbers HG793132, HG796219, HG796220, and MN191861, respectively) and also shares a phiKMV-like genome arrangement in which the phage RNA polymerase is located in the central portion of the genome. Multiple T7-like tail components were identified in Paku, including homologs of the T7 tail fiber protein gp17 and tail tubular protein gp12. The endolysin of Paku contains an N-terminal signal-arrest-release (SAR) sequence, with the spanin complex directly downstream of the endolysin gene. A tyrosine integrase was identified, raising the possibility that Paku might exhibit a temperate lifestyle, an atypical feature for an *Autographiviridae* phage.

Data availability. Paku's genome was deposited in GenBank with accession number MZ326863. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR14095249, and SAMN18509701, respectively.

ACKNOWLEDGMENTS

Funding was provided by the National Science Foundation (awards EF-0949351 and DBI-1565146), by the Cystic Fibrosis Foundation, by the CPT (https://cpt.tamu.edu), which is an Initial University Multidisciplinary Research Initiative supported by Texas A&M University and Texas AgriLife, and by the Department of Biochemistry and Biophysics.

We are grateful for the advice and support of the CPT members.

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

REFERENCES

1. Drevinek P, Mahenthiralingam E. 2010. *Burkholderia cenocepacia* in cystic fibrosis: epidemiology and molecular mechanisms of virulence. Clin Microbiol Infect 16:821–830. https://doi.org/10.1111/j.1469-0691.2010.03237.x.
2. Scoffone VC, Chiarelli LR, Trespidi G, Mentasti M, Riccardi G, Buroni S. 2017. *Burkholderia cenocepacia* infections in cystic fibrosis patients: drug resistance and therapeutic approaches. Front Microbiol 8:1592. https://doi.org/10.3389/fmicb.2017.01592.
3. Adams M. 1956. Bacteriophages. Interscience Publishers, 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. 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.
7. 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.
8. 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.
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. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
13. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, 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. Krogh A, Larsson B, von Heijne 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.
16. Zimmermann L, Stephens A, Nam SZ, 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.
17. 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.
18. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.
19. UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
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 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.