AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Stenotrophomonas maltophilia* Myophage Marzo

Janki Patel,[a] Brenda Godoy,[a] James Clark,[a,b] Ben Burrowes,[a,b] Ry Young,[a,b] Mei Liu[a,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

ABSTRACT *Stenotrophomonas maltophilia* is a Gram-negative opportunistic bacterium that is increasingly being associated with infections. Here, we report the complete genome of the *S. maltophilia* myophage Marzo, with a 159,384-bp genome encoding 268 proteins, 23 tRNAs, and 1 transfer-messenger RNA. Marzo is closely related to *S. maltophilia* phages IME-SM1 and Mendera.

Stenotrophomonas maltophilia is found in aqueous habitats, including plant rhizospheres and animals, and is an opportunistic Gram-negative bacterium that can cause infections in tissues ranging from the skin to the heart in immunocompromised individuals (1). We are interested in studying *S. maltophilia* phage genomes in the interest of exploring potential therapeutic treatment options.

Phage Marzo was isolated from an activated sludge sample collected from the Texas A&M wastewater treatment plant in September 2019, using the soft agar overlay method (2) with *S. maltophilia* (ATCC 17807) as the propagation host grown aerobically at 30°C in nutrient broth or agar (BD). Marzo DNA was purified from ~8 mL phage lysate using the Promega Wizard DNA cleanup system, as described previously (3). Sequencing libraries were prepared as 300-bp inserts using a Swift 2S Turbo kit and sequenced on an Illumina MiSeq system with paired-end 150-bp reads using 300-cycle v2 chemistry. The 106,506 total sequence reads were quality controlled using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc), trimmed with FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit), and assembled using SPAdes v3.5.0 (4). A raw contig of 159,439 bp was obtained, and its end sequences were manually corrected with Sanger sequencing of a PCR product amplified from the contig ends (forward primer, TGAACTTCTCCAGCCCGAAC; reverse primer, TGTAGCGAGCC CTGATCTCT). PhageTerm was used to predict phage termini from the raw sequencing reads (5). Phage Marzo was annotated using the Center for Phage Technology (CPT) Galaxy-Apollo phage annotation platform (https://cpt.tamu.edu/galaxy-pub) (6–8). Gene calling included GLIMMER v3.0 (9) and MetaGeneAnnotator v1.0 (10). tRNA and transfer-messenger RNA (tmRNA) genes were detected using ARAGORN v2.36 (11) and tRNAscan-SE v2.0 (12). Gene function was predicted using InterProScan v5.48 (13), BLAST v2.9.0 (14) with the NCBI nonredundant and Swiss-Prot databases (15), TMHMM v2.0 (16) for transmembrane domains, HHPred (17), LipoP v1.0 (18) for lipoproteins, and SignalP v5.0 (19). Genome-wide DNA sequence similarity to top BLAST nucleotide hits (from the NCBI nucleotide database) was calculated by progressiveMauve v2.4 (20). All tools were run with default settings unless otherwise specified.

Phage Marzo was determined to be a myophage via negative staining with 2% (wt/vol) uranyl acetate and imaging by transmission electron microscopy (TEM) at the Texas A&M University Microscopy and Imaging Center (Fig. 1). The completed 159,384-bp myophage Marzo genome has 24-fold sequencing coverage and a G+C content of 54%. PhageTerm was unable to predict phage termini from the raw sequencing reads, but due to its similarities in terms of morphology and genome size to the canonical phage T4, Marzo likely uses headful

Editor Simon Roux, DOE Joint Genome Institute

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

The authors declare no conflict of interest.

Received 17 December 2021
Accepted 10 February 2022
Published 28 February 2022

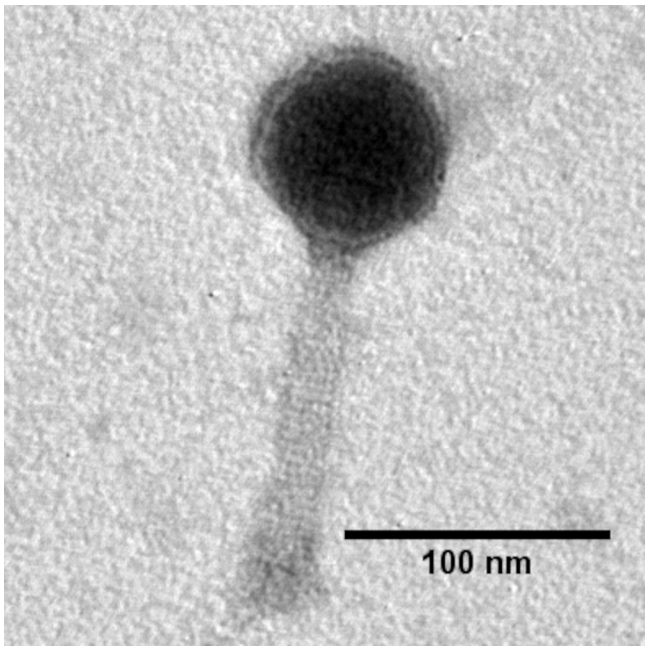

FIG 1 TEM of phage Marzo. Phage particles were diluted with TEM buffer (20 mM NaCl, 10 mM Tris-HCl [pH 7.5], 2 mM $MgSO_4$) and captured on freshly glow-discharged, Formvar carbon-coated grids. The grids were stained with 2% (wt/vol) uranyl acetate and observed on a JEOL 1200 EX TEM at 100-kV accelerating voltage at the Microscopy and Imaging Center at Texas A&M University.

packaging. Twenty-three tRNA genes, 1 tmRNA gene, and 268 protein-coding genes were found, with a coding density of 93%. The tRNA genes were found in two clusters, one with 3 tRNAs and the tmRNA and the other with 20 tRNAs. The Marzo tmRNA was highly similar to the SsrA *Betaproteobacter*-class tmRNA of *S. maltophilia*, as determined by sequence analysis at RNAcentral (21). Comparative genomics revealed that Marzo has ≥92% nucleotide identity to other *S. maltophilia* myophages, namely, IME-SM1 (GenBank accession number KR560069), YB07 (GenBank accession number NC_048755), and Mendera (GenBank accession number NC_048804). Some structural genes could be identified, encoding tail completion scaffold, portal, major capsid, baseplate wedge, and tail tube proteins. In addition, although no holin or endolysin genes could be identified, two spanin gene pairs were identified, one of the overlapping class and the other of the embedded class.

Data availability. The Marzo genome was deposited in GenBank with accession number MZ326868. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR14095257, and SAMN18509700, respectively.

ACKNOWLEDGMENTS

Funding was provided by the National Science Foundation (awards EF-0949351 and DBI-1565146), the CPT (an initial university multidisciplinary research initiative supported by Texas A&M University and Texas AgriLife), and the Department of Biochemistry and Biophysics (https://cpt.tamu.edu).

We are grateful for the advice and support of the CPT members, especially Carlos Gonzalez, who provided the *Stenotrophomonas* strain.

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

REFERENCES

1. Brooke JS. 2012. *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. Clin Microbiol Rev 25:2–41. https://doi.org/10.1128/CMR.00019-11.
2. Adams MH. 1959. Bacteriophages. Interscience Publishers, New York, NY.
3. 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.
4. 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.
5. 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.
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. UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
16. 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.
17. 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.
18. 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.
19. 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.
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. RNAcentral Consortium. 2019. RNAcentral: a hub of information for non-coding RNA sequences. Nucleic Acids Res 47:D221–D229. https://doi.org/10.1093/nar/gky1034. (Erratum, 47:D1250–D1251).