Complete Genome Sequence of *Stenotrophomonas maltophilia* Siphophage Sonora

Michael Teve,ᵃ James Clark,ᵃ,ᵇ Tram Le,ᵃ,ᵇ Ben Burrowes,ᶜ Mei Liuᵃ,ᵇ

ᵃDepartment of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA
ᵇCenter for Phage Technology, Texas A&M University, College Station, Texas, USA
ᶜBB Phage Consultancy, LLC, Georgetown, Texas, USA

ABSTRACT Phage Sonora is a siphophage that was isolated against the opportunistic human pathogen *Stenotrophomonas maltophilia*. The genome of phage Sonora is 63,825 bp long and is not related to that of any phage at the nucleotide level. Sonora shares 46 of 97 total proteins with the *Bordetella* phages CN2, MW2, and FP1.

*S*tenotrophomonas maltophilia is an opportunistic human pathogen that is naturally found in water and soil environments (1). Some infections caused by *S. maltophilia* are difficult to treat due to the multidrug resistance of the causative strains. Because antibiotic resistance remains a significant public health issue, alternative treatments, such as phage therapy, to combat these infections are of interest. Bacteriophage Sonora's isolation and genome characteristics are described in this report.

Sonora was isolated in 2019 from a topsoil sample (collected ~3 in. from the surface) collected in Caldwell, Texas (coordinates: 30.612660, −96.551836). The phage was isolated and purified following the soft agar overlay method (2) using *S. maltophilia* (ATCC 51331) as the host, cultured aerobically at 30°C in tryptone nutrient (0.5% tryptone, 0.25% yeast extract, 0.1% glucose, 0.85% NaCl [wt/vol]) broth or agar. Sonora's genomic DNA was extracted using a modified Wizard DNA cleanup kit protocol as described (3). DNA libraries were prepared as 300-bp inserts using a Swift 2S Turbo kit and sequenced on an Illumina MiSeq sequencer with paired-end 150-bp reads using 300-cycle v2 chemistry to produce 242,366 raw reads. These reads were quality controlled using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc), trimmed with FASTX-Toolkit v0.11.6 (http://hannonlab.cshl.edu/fastx_toolkit/download .html), and assembled with SPAdes v3.5.0 (4) into a raw contig with 140-fold coverage. Since the raw contig assembled by SPAdes is usually opened at a random spot in the middle of the genome and the contig ends often have redundant or missing bases, PCR amplification of the genomic DNA using primers designed off the contig ends, followed by Sanger sequencing of the PCR product, allows verification of sequences in that region. Sonora's genome was completed by PCR and Sanger sequencing using the primers 5′-ACTACCACGGTCACGCATAC-3′ and 5′-CAGATCATCGAACATGCCGC-3′. Genome termini were predicted by PhageTerm (5). The genome was annotated with the Center for Phage Technology (CPT) Galaxy-Apollo platform (https://cpt.tamu.edu/ galaxy-pub) (6–9); structural annotation was performed using GLIMMER v3 (10) and MetaGeneAnnotator v1.0 (11) to predict protein-coding genes, with manual adjustments as needed. Rho-independent termination sites were identified using TransTerm (http:// transterm.cbcb.umd.edu). The functional annotation of Sonora utilized InterProScan v5.48 (12), BLAST v2.9.0 (13), TMHMM v2.0 (14), HHpred (15), LipoP v1.0 (16), SignalP v5.0 (17), and SwissProt (18) databases. Sonora's genome DNA sequence similarity to other phages was calculated by progressiveMauve v2.4 (19). All software was used with default settings.

Phage Sonora was determined to be a siphophage via negative staining with 2%
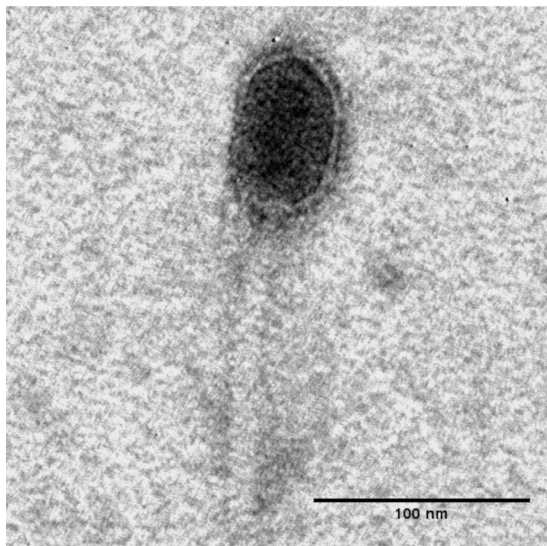
FIG 1 TEM of phage Sonora. Phage particles were diluted with TEM buffer (20 mM NaCl, 10 mM Tris-HCl [pH 7.5], 2 mM MgSO₄) and captured on a freshly glow-discharged, Formvar carbon-coated grid. The grids were stained with 2% (wt/vol) uranyl acetate and observed at 100-kV accelerating voltage on a JEOL 1200 EX system at the Microscopy and Imaging Center at Texas A&M University.

(wt/vol) uranyl acetate and imaging by transmission electron microscopy (TEM) at the Texas A&M Microscopy and Imaging Center (Fig. 1). Phage Sonora has a genome of 63,825 bp, with a coding density of 93.8% and a G+C content of 63.0%. PhageTerm predicted a *cos* sequence at position 46430, which is within a gene encoding a predicted membrane protein. However, this location is separated by ~18 kb from the gene encoding the terminase small subunit, suggesting that the packaging signal is significantly separated from the gene encoding the protein that recognizes it for packaging. A lysis cassette containing a transglycosylase endolysin with a signal arrest release (SAR) domain, a class I holin, and an inner and outer membrane spanin complex, with the o-spanin gene completely embedded within the i-spanin gene, was identified. Sonora is not closely related to any phages in the NCBI database, sharing at most only 33 to 34% genome-wide nucleotide sequence identity with several phages, such as *Pseudomonas* phage YuA (GenBank accession number AM749441) and *Bordetella* phage CN1 (GenBank accession number NC_047876), as determined by progressiveMauve (19). At the protein level (BLASTp, with E values of <0.001), Sonora shares 46 of 97 total proteins with each of the three *Bordetella* phages, namely, CN2 (GenBank accession number NC_047877.1), MW2 (GenBank accession number NC_047879.1), and FP1 (GenBank accession number NC_047878.1).

Data availability. Sonora's genome was deposited in GenBank with accession number MZ326860. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR14095252, and SAMN18509468, respectively.

ACKNOWLEDGMENTS

REFERENCES

1. Brooke JS, Di Bonaventura G, Berg G, Martinez JL. 2017. Editorial: a multidisciplinary look at *Stenotrophomonas maltophilia*: an emerging multidrug-resistant global opportunistic pathogen. Front Microbiol 8:1511. https://doi.org/10.3389/fmicb.2017.01511.
2. Adams M. 1956. 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. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.

19. 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.