Genome Sequence of a T4-like Phage, *Escherichia* Phage vB_EcoM-Sa45lw, Infecting Shiga Toxin-Producing *Escherichia coli* Strains

Yen-Te Liao,[a] Yujie Zhang,[a,b] Alexandra Salvador,[a] Vivian C. H. Wu[a]

aProduce Safety and Microbiology Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, California, USA

bCollege of Food Science and Technology, Shanghai Ocean University, Shanghai, China

ABSTRACT *Escherichia* phage vB_EcoM-Sa45lw, a new member of the T4-like phages, was isolated from surface water in a produce-growing area. The phage, containing double-stranded DNA with a genome size of 167,353 bp and 282 predicted open reading frames (ORFs), is able to infect generic *Escherichia coli* and Shiga toxin-producing *E. coli* O45 and O157 strains.

Shiga toxin-producing *Escherichia coli* (STEC) is notorious for causing hemolytic uremic syndrome through consumption of a variety of contaminated food items (1, 2). The use of lytic bacteriophages for controlling bacterial pathogens has been increasingly considered, as it is less likely to lead to the development of antibiotic-resistant bacteria (3). However, the number of genomically characterized coliphages is relatively scarce compared to their bacterial hosts. In this study, the complete genome sequence of a T4-like coliphage is described.

Escherichia phage vB_EcoM-Sa45lw, with morphology belonging to the *Myoviridae* family, was previously isolated from surface water using generic *E. coli* (ATCC 13706) as the primary host strain (4). The host range was determined against O157 and the top 6 non-O157 STEC strains using a spot test. A phage DNA extraction kit (Norgen Biotek, Ontario, Canada) was used to extract phage DNA, followed by DNA library construction using a TruSeq Nano DNA library prep kit (Illumina, San Diego, CA) prior to sequencing on an Illumina MiSeq sequencer. A total of 5,069,865 2 $\times$ 250-bp paired-end reads were generated. The quality control of the raw sequence reads was performed using FastQC (Galaxy v0.72), and poor reads were trimmed using Trimmomatic Galaxy v0.36.6 (5) with the average quality set at Q30. The good-quality reads were assembled *de novo* using Unicycler Galaxy v0.4.6.0 (SPAdes v2.5.1) with the default settings (6). The resulting contig ($n = 1$) was annotated using the Prokka pipeline Galaxy v1.13 (7), and the sequences of predicted open reading frames (ORFs) were confirmed with PHASTER (8) and Artemis v1.0 (9) and were manually curated with UniProt (10) and GB2sequin (11) using Geneious v11.0.4. The packaging mechanisms and the termini of the phage genome were identified using PhageTerm Galaxy v1.0.12 (12). tRNA was predicted using the tRNAscan-SE (v2.0) server (13).

Escherichia phage vB_EcoM-Sa45lw, or Sa45lw, has double-stranded DNA with a genome size of 167,353 bp (5,466$\times$ coverage), an average G+C content of 35.4%, and 11 tRNAs. These genomic characteristics are closely similar to T4 or T4-like phages (14, 15). Phage Sa45lw shares 97.17% average nucleotide identity based on BLAST (ANIb) with phage wV7 (96.26% query coverage; GenBank accession number HM997020), 96.07% ANIb with phage AR1 (93.49% query coverage; GenBank accession number NC_027983), and 96.10% ANIb with phage PE37 (92.54% query coverage; GenBank accession number KU925172), calculated using the JSpeciesWS Web server (16). Addi-

tionally, the packaging mechanism of phage Sa45lw was predicted to be headful packaging, similar to that of phage T4 (12, 17). These findings show that phage Sa45lw belongs to the *Tequatrovirus* genus of the subfamily *Tevenvirinae* within the family *Myoviridae* (18).

A total of 282 ORFs were predicted in the Sa45lw genome, including 137 categorized into the following functional groups: virion structure, host lysis, DNA replication and transcription regulation, which includes anti-sigma factors commonly present in T-even phages, and host cell regulation and metabolism (15). No ORFs contain *stx* or antibiotic resistance genes. Furthermore, ORF258 of Sa45lw encodes a large distal long-tail-fiber subunit, which is responsible for recognizing the receptor on the bacterial hosts, and shares 92.12%, 86.11%, and 83.41% amino acid identity with the counterparts in coliphages wV7, AR1, and PE37, respectively. This finding likely accounts for the different host ranges between phage Sa45lw (ATCC 13706, STECO157, and O45) and the three T4-like phages (STEC O157).

Data availability. The complete genome sequence of *Escherichia* phage vB_EcoM-Sa45lw has been deposited in GenBank under the accession number MK977694. The sequencing reads have been deposited under the accession number PRJNA544313. The version of the phage genome described in this paper is the first version.

REFERENCES

1. Herman KM, Hall AJ, Gould LH. 2015. Outbreaks attributed to fresh leafy vegetables, United States, 1973–2012. Epidemiol Infect 143:3011–3021. https://doi.org/10.1017/S0950268815000047.
2. Liao YT, Brooks JC, Martin JN, Echeverry A, Loneragan GH, Brashears MM. 2015. Antimicrobial interventions for O157:H7 and non-O157 Shiga toxin-producing *Escherichia coli* on beef subprimal and mechanically tenderized steaks. J Food Prot 78:511–517. https://doi.org/10.4315/0362-028X.JFP-14-178.
3. Mahony J, McAuliffe O, Ross RP, van Sinderen D. 2011. Bacteriophages as biocontrol agents of food pathogens. Curr Opin Biotechnol 22:157–163. https://doi.org/10.1016/j.copbio.2010.10.008.
4. Liao Y-T, Quintela IA, Nguyen K, Salvador A, Cooley MB, Wu VCH. 2018. Investigation of prevalence of free Shiga toxin-producing *Escherichia coli* (STEC)-specific bacteriophages and its correlation with STEC bacterial hosts in a produce-growing area in Salinas, California. PLoS One 13:e0190534. https://doi.org/10.1371/journal.pone.0190534.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
7. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
8. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
9. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for analysis and visualization of high-throughput sequence-based experimental data. Bioinformatics 28:464–469. https://doi.org/10.1093/bioinformatics/btr703.
10. The UniProt Consortium. 2007. The Universal Protein Resource (UniProt). Nucleic Acids Res 35:D193–D197. https://doi.org/10.1093/nar/gkl929.
11. Lehwark P, Greiner S. 2018. GB2sequin: a file converter preparing custom GenBank files for database submission. Genomics. https://doi.org/10.1016/j.ygeno.2018.05.003.
12. 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.
13. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
14. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. 2003. Bacteriophage T4 genome. Microbiol Mol Biol Rev 67:86–156. https://doi.org/10.1128/MMBR.67.1.86-156.2003.
15. Liao WC, Ng WV, Lin IH, Syu W Jr, Liu T-T, Chang CH. 2011. T4-like genome organization of the *Escherichia coli* O157:H7 lytic phage AR1. J Virol 85:6567–6578. https://doi.org/10.1128/JVI.02378-10.
16. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a Web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931. https://doi.org/10.1093/bioinformatics/btv681.
17. Casjens SR, Gilcrease EB. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Mol Biol 502:91–111. https://doi.org/10.1007/978-1-60327-565-1_7.
18. Adriaenssens EM, Brister JR. 2017. How to name and classify your phage: an informal guide. Viruses 9:70. https://doi.org/10.3390/v9040070.