Complete Genome Sequence of a Shiga Toxin-Converting Bacteriophage, *Escherichia* Phage Lys12581Vzw, Induced from an Outbreak Shiga Toxin-Producing *Escherichia coli* Strain

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

[a]College of Food Science and Technology, Shanghai Ocean University, Shanghai, China
[b]Produce Safety and Microbiology Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, California, USA

ABSTRACT Although numerous Shiga toxin (Stx)-producing *Escherichia coli* (STEC) strains have been sequenced, genomic information on Stx-converting phages, highly related to the primary virulence factors of STEC, is scarce. Here, we report the complete genome sequence of a Stx-converting phage induced from an outbreak STEC O145 strain.

Shiga toxin (Stx)-producing *Escherichia coli* (STEC), as one of the major foodborne pathogens, has been widely associated with multiple foodborne outbreaks in the United States (1, 2). Stx-converting phages, carrying the major virulence genes of STEC, are usually induced from STEC strains and have been related to the emergence of new STEC strains through the transfer of *stx* genes to other bacteria (3, 4). However, there is a limited number of complete whole-genome sequences of Stx-converting phages. In this study, the complete genome sequence of a Stx2-converting phage isolated from an outbreak STEC O145 strain is described.

The Stx-converting phage Lys12581Vzw was induced from *E. coli* O145:H28 (RM12581), associated with the U.S. romaine lettuce outbreak in 2010 (5), by using mitomycin C (0.5 μg/ml) in LB broth at 37°C overnight. Phage purification was performed using a double-layer plaque assay against *E. coli* strain WG5 (ATCC 700078) from a single plaque picking, followed by CsCl gradient concentration to get rid of bacterial DNA and debris. The phage DNA was extracted using a phage DNA extraction kit (Norgen Biotek, Ontario, Canada). The DNA library (2 × 250 bp) was constructed using a TruSeq Nano DNA library prep kit (Illumina, San Diego, CA) and was subsequently sequenced on an Illumina MiSeq sequencer, resulting in 6,779,298 paired-end reads. The raw sequence reads were quality filtered and trimmed using FastQC and Trimmomatic (6), with the average quality set at Q30. *De novo* assembly of the remaining quality reads was performed using SPAdes version 3.13.0 on the KBase server (7, 8) with the minimum contig length set to >10,000 bp. The resulting contig was annotated using both Prokka (9) and RAST server (10) pipelines with default settings. The annotations were subsequently compared and curated with UniProt (11) and BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), with the identity and query coverage greater than 95%, using Geneious (version 11.0.4). The packaging mechanisms and genome termini were determined using PhageTerm (12). tRNAs were predicted using the tRNAscan-SE search server (13). ResFinder (version 3.0) and VirulenceFinder were used to screen for antibiotic resistance genes and virulence genes (14, 15).

Phage Lys12581Vzw, belonging to the family *Podoviridae*, has double-stranded DNA with a genome size of 62,668 bp (3,257× coverage) and an average G+C content of 50.1%. The phage has a Mu-like packaging mechanism; the phage genome lacks

obvious termini and carries pieces of the host sequences. There are 81 open reading frames (ORFs) predicted, of which 45 are annotated with functional proteins, such as DNA replication and modification (restriction endonuclease, DNA primase, and methylase), membrane-related proteins (lysis proteins, membrane proteins, and cell division inhibitors), antirepressors, virion structures, integrases, Shiga toxin, and tRNAs. Most functional ORFs contribute to the virulence and propagation of Stx-converting phages. None of the ORFs contain antibiotic resistance genes. The whole-genome sequence of Lys12581Vzw shares 95% average nucleotide identity with two published Stx2-converting phages—*Enterobacteria* phage VT2-Sakai (GenBank accession number NC_000902) and bacteriophage 933W (GenBank accession number AF125520)—induced from different *E. coli* O157:H7 strains. This study provides valuable insights into the diversity of Stx-converting phages which are associated with the evolution of STEC strains.

Data availability. The complete genome sequence of *Escherichia* phage Lys12581Vzw has been deposited in GenBank under the accession number MN067333. The sequencing reads have been deposited under the BioSample accession number SAMN12086808. The version of the phage genome described in this paper is the first version.

ACKNOWLEDGMENT

REFERENCES

1. Bennett SD, Sodha SV, Ayers TL, Lynch MF, Gould LH, Tauxe RV, Infection. 2018. Produce-associated foodborne disease outbreaks, USA, 1998–2013. Epidemiol Infect 146:1397–1406. https://doi.org/10.1017/S0950268818001620.
2. CDC. 2019. Food safety alert: outbreak of E. coli infections linked to ground beef. CDC, Atlanta, GA. https://www.cdc.gov/ecoli/2019/o103-04-19/index.html. Accessed 19 June 2019.
3. Silva CJ, Brandon DL, Skinner CB, He X. 2017. Shiga toxins: a review of structure, mechanism, and detection. Springer, Dordrecht, The Netherlands.
4. Grau-Leal F, Quirós P, Martínez-Castillo A, Muniesa M. 2015. Free Shiga toxin 1-encoding bacteriophages are less prevalent than Shiga toxin 2 phages in extraintestinal environments. Environ Microbiol 17:4790–4801. https://doi.org/10.1111/1462-2920.13053.
5. Cooper KK, Mandrell RE, Louie JW, Korlach J, Clark TA, Parker CT, Huynh S, Chain PS, Ahmed S, Carter MQ. 2014. Complete genome sequences of two Escherichia coli O145:H28 outbreak strains of food origin. Genome Announc 2:e00482-14. https://doi.org/10.1128/genomeA.00482-14.
6. 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.
7. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714–737. https://doi.org/10.1089/cmb.2013.0084.
8. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM,

Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566. https://doi.org/10.1038/nbt.4163.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
11. The UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. https://doi.org/10.1093/nar/gku989.
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. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
15. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510. https://doi.org/10.1128/JCM.03617-13.