Investigating the Whole-Genome Sequence of a New Locus of Enterocyte Effacement-Positive Shiga Toxin-Producing *Escherichia coli* O157:H7 Strain Isolated from River Water

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

[a]College of Food Science & 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 Diverse Shiga toxin-producing *Escherichia coli* (STEC) strains have been isolated from several environmental samples. Rivers are associated with the distribution of STEC pathogens in the environment. Thus, we report the complete genome sequence of a locus of enterocyte effacement (LEE)-positive STEC O157:H7 strain isolated from the Mississippi River.

Shiga toxin-producing *Escherichia coli* (STEC) is one of the major bacterial pathogens associated with numerous foodborne outbreaks around the world (1, 2). *E. coli* O157:H7 has been the most persistent serotype in STEC-associated outbreaks causing severe human illnesses, such as hemolytic uremic syndrome (HUS), and high mortality among immunocompromised patients (3). Previous studies showed that diverse STEC strains have been isolated from several environmental samples, such as animal feces, leafy greens, and soil (4; https://www.cdc.gov/nationalsurveillance/ecoli-surveillance.html). Moreover, rivers were considered a distributing source of STEC due to its association with the aforementioned environmental factors (5–8). Thus, whole-genome sequencing was conducted on a STEC O157:H7 strain isolated from river water to unveil the pathogenicity of the strain.

Escherichia coli O157:H7 strain RM19259 was previously isolated from a sample collected from the Mississippi River watershed in 2016 using Moore swabs. For strain isolation, the environmental sample was enriched with tryptic soy broth (TSB) and subsequently incubated at 25°C for 2 hours, followed by 42°C for 8 hours prior to the further isolation process as previously described (9). Genomic DNA was extracted from the cultures grown to mid-exponential phase in 10 ml tryptic soy broth (TSB; Difco, Becton, Dickinson, Sparks, MD) using a Quick-DNA Miniprep Plus kit (Zymo Research, Irvine, CA) according to the manufacturer's instructions. The DNA library was constructed using an Express template prep kit 2.0 (Pacific Biosciences) and subsequently sequenced using PacBio Sequel II with V1 reagents (Pacific Biosciences), resulting in 398,925 single-end reads. *De novo* assembly was performed using Flye 2.4.1 with the default parameters, resulting in 3 contigs with an N_{50} value of 5,511,015 bp (10). Two contigs were identified as circular contigs represented by a complete chromosome and a plasmid using BUSCO 3 (11) and Blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides). The genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The predictions of serotype, prophages, and virulence genes were performed using SerotypeFinder 2.0 (12), PHASTER (13), and VirulenceFinder 2.0 (14), respectively. The genomic island was analyzed with Island-Viewer 4 using the IslandPath-DIMOB method (15). Default parameters were used for all software unless otherwise specified.

The strain contains a 5,511,015-bp chromosome and a 98,304-bp plasmid. The

chromosome has an average GC content of 50.5% and carries a total of 5,440 predicted protein-coding sequences (CDSs), 22 rRNAs, and 106 tRNAs. The strain is an *E. coli* O157:H7 strain that harbors two sets of *stx* genes, including one stx_{2a} located on an 83,266-bp prophage (chromosome position, base pairs 260410 to 343675) and one stx_{2c} located on a 35,769-bp prophage (chromosome position, base pairs 5403730 to 5439498). Additionally, a 41,255-bp pathogenicity island—locus of enterocyte effacement (LEE)—was identified between base pairs 1729493 and 1770747 of the chromosome. This strain also carried enterohemorrhagic *E. coli* (EHEC)-associated non-LEE-encoded type III translocated virulence factors (*nleA*, *nleB*, and *nleC*). In addition, the strain contains a virulence gene, EAST-1 heat-stable toxin (*astA*), commonly related to many *E. coli* pathovars, such as enteroaggregative *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), and EHEC (16). The plasmid pRM19259 is a typical F-like plasmid pO157, which contains several virulence-related genes, such as *ehxA* (enterohemolysin), *espP* (serine protease), *etpD* (type II secretion protein), *katP* (catalase-peroxidase), and *toxB* (toxin B) (17). The findings of this study provide valuable insights into the epidemiological surveillance of *E. coli* O157:H7 infections.

Data availability. The sequence described in this study is available under BioProject accession number PRJNA573729. The GenBank accession numbers of the RM19259 chromosome and plasmid pRM19259 are CP046527 and CP046526, respectively. The raw reads of the strain are available under Sequence Read Archive (SRA) accession number SRR10598574.

REFERENCES

1. White A, Cronquist A, Bedrick EJ, Scallan E. 2016. Food source prediction of Shiga toxin-producing *Escherichia coli* outbreaks using demographic and outbreak characteristics, United States, 1998–2014. Foodborne Pathog Dis 13:527–534. https://doi.org/10.1089/fpd.2016.2140.
2. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Muller L, King LA, Rosner B, Buchholz U, Stark K, Krause G, HUS Investigation Team. 2011. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. N Engl J Med 365:1771–1780. https://doi.org/10.1056/NEJMoa1106483.
3. Jenkins C, Dallman TJ, Launders N, Willis C, Byrne L, Jorgensen F, Eppinger M, Adak GK, Aird H, Elviss N, Grant KA, Morgan D, McLauchlin J. 2015. Public health investigation of two outbreaks of Shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. Appl Environ Microbiol 81:3946–3952. https://doi.org/10.1128/AEM.04188-14.
4. CDC. 2018. Multistate outbreak of Shiga toxin-producing *Escherichia coli* O157:H7 infections linked to leafy greens (final update). https://www.cdc.gov/ecoli/2017/o157h7-12-17/index.html. Accessed 9 September 2019.
5. Gyles CL. 2007. Shiga toxin-producing *Escherichia coli*: an overview. J Anim Sci 85:E45–E62. https://doi.org/10.2527/jas.2006-508.
6. Adams NL, Byrne L, Smith GA, Elson R, Harris JP, Salmon R, Smith R, O'Brien SJ, Adak GK, Jenkins C. 2016. Shiga toxin-producing *Escherichia coli* O157, England and Wales, 1983–2012. Emerg Infect Dis 22:590–597. https://doi.org/10.3201/eid2204.151485.
7. Gelting RJ, Baloch MA, Zarate-Bermudez M, Hajmeer MN, Yee JC, Brown T, Yee BJ. 2015. A systems analysis of irrigation water quality in an environmental assessment of an *E. coli* O157:H7 outbreak in the United States linked to iceberg lettuce. Agric Water Manag 150:111–118. https://doi.org/10.1016/j.agwat.2014.12.002.
8. Olaimat AN, Holley RA. 2012. Factors influencing the microbial safety of fresh produce: a review. Food Microbiol 32:1–19. https://doi.org/10.1016/j.fm.2012.04.016.
9. Cooley MB, Jay-Russell M, Atwill ER, Carychao D, Nguyen K, Quiñones B, Patel R, Walker S, Swimley M, Pierre-Jerome E, Gordus AG, Mandrell RE. 2013. Development of a robust method for isolation of Shiga toxin-positive *Escherichia coli* (STEC) from fecal, plant, soil and water samples from a leafy greens production region in California. PLoS One 8:e65716. https://doi.org/10.1371/journal.pone.0065716.
10. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
11. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. https://doi.org/10.1093/molbev/msx319.
12. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426. https://doi.org/10.1128/JCM.00008-15.
13. Arndt D, Marcu A, Liang Y, Wishart DS. 2019. PHAST, PHASTER and PHASTEST: tools for finding prophage in bacterial genomes. Brief Bioinform 20:1560–1567. https://doi.org/10.1093/bib/bbx121.
14. 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.
15. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman F. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. https://doi.org/10.1093/nar/gkx343.
16. Coura FM, Diniz S de A, Silva MX, de Oliveira CHS, Mussi JMS, de Oliveira CSF, Lage AP, Heinemann MB. 2019. Virulence factors and phylotyping of *Escherichia coli* isolated from non-diarrheic and diarrheic water buffalo calves. Cienc Rural 49:e20180998. https://doi.org/10.1590/0103-8478cr20180998.
17. Lim JY, Yoon JW, Hovde CJ. 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. J Microbiol Biotechnol 20:5–14. https://doi.org/10.4014/jmb.0908.08007.