AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Whole-Genome Analysis of a Shiga Toxin-Producing *Escherichia coli* O103:H2 Strain Isolated from Cattle Feces

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

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

ABSTRACT Shiga toxin-producing *Escherichia coli* (STEC) serotype O103 is one of the primary pathogenic contaminants of beef products, contributing to several foodborne outbreaks in recent years. Here, we report the whole-genome sequence of a STEC O103:H2 strain isolated from cattle feces that contains a locus of enterocyte effacement (LEE) pathogenicity island.

Shiga toxin-producing *Escherichia coli* (STEC) O103 strains have been frequently associated with foodborne outbreaks (1, 2). During 2019 to 2020, three STEC O103-associated outbreaks were reported and related to the contamination of beef products and clover sprouts, causing 272 illnesses in the United States (3). Additionally, previous studies showed that cattle are the primary reservoir of STEC strains, which could be further disseminated in beef production and processing environments through direct and indirect contamination with feces (4–6). Thus, the complete genome sequence and the virulence factors of a STEC O103:H2 strain isolated from cattle feces are reported here to unveil the potential pathogenicity of this pathogen.

Strain RM13322 was previously isolated from cattle feces collected in California and confirmed as *E. coli* serotype O103 via an *E. coli* isolation procedure, including the use of selective media, enzyme-linked immunosorbent assay (ELISA), and PCR, as previously described (7). The bacterial culture was grown to the mid-exponential phase in 10 ml tryptic soy broth (TSB; Difco, Sparks, MD) before being subjected to genomic DNA extraction 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 following the manufacturer's procedure without DNA shearing (Pacific Biosciences, Menlo Park, CA). Subsequently, the pooled library was subjected to size selection using BluePippin with a cutoff of 8 kb (Sage Science, Beverly, MA) and sequenced on a PacBio Sequel II instrument with v1 reagents, resulting in a total of 100,000 single-end reads with an N_{50} value of 14,710 bp. The raw reads were filtered and demultiplexed using PacBio SMRTLink v7.0 before being assembled into a circular 5,622,341-bp contig and a 76,269-bp linear contig utilizing Flye v2.4.1 (8). The two contigs were identified as a complete chromosome and a plasmid using BUSCO v3 (9), blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides), and PlasmidFinder 2.0 (10). The complete genome was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (11). The prediction of serotype, prophages, virulence genes, and antibiotic resistance genes was performed via SerotypeFinder 2.0 (12), PHASTER (13), VirulenceFinder 2.0 (14), and ResFinder 3.2 (15), respectively. The genomic island was analyzed via IslandViewer4 using IslandPath-DIMOB methods (16). Default parameters were used for all software unless otherwise specified.

The result indicated that this *E. coli* O103:H2 strain contains a 5,622,341-bp circular chromosome (201× coverage) and a 76,269-bp linear plasmid (63× coverage). The

Citation Zhang Y, Liao Y-T, Wu VCH. 2020. Whole-genome analysis of a Shiga toxin-producing *Escherichia coli* O103:H2 strain isolated from cattle feces. Microbiol Resour Announc 9:e00896-20. https://doi.org/10.1128/MRA.00896-20.

Editor Vincent Bruno, University of Maryland School of Medicine

Address correspondence to Vivian C. H. Wu, vivian.wu@usda.gov.

Received 30 July 2020
Accepted 17 October 2020
Published 5 November 2020

chromosome has an average GC content of 50.8% and encodes a total of 5,534 coding DNA sequences (CDSs), 22 rRNAs, and 98 tRNAs. The genomic characterization showed that this strain harbors a 117,904-bp Stx1a prophage and a 71,689-bp locus of enterocyte effacement (LEE) pathogenicity island. Additionally, several non-LEE-encoded type III translocated virulence factors (*nleA*, *nleB*, *espJ*, and *cif*) and adherence-associated genes (*iha* and *efa1*) were detected in the bacterial chromosome. Plasmid pRM13322 has a GC content of 45.9% and contains several phage-associated essential CDSs encoding functional proteins, such as structurally related proteins and lysis-related proteins. The findings of this study provide valuable insights into the potential virulence factors of STEC O103:H2 of bovine origin.

Data availability. The sequence described in this study is available under BioProject accession number PRJNA573729. The GenBank accession numbers of the *E. coli* O103:H2 strain RM13322 chromosome and plasmid pRM13322 are CP050498 and CP050499, respectively. The raw read of the strain is available under Sequence Read Archive (SRA) accession number SRR12005555.

ACKNOWLEDGMENT

This research was funded by the U.S. Department of Agriculture, Agricultural Research Service, CRIS project 2030-42000-050-00D.

REFERENCES

1. CDC. 2020. National Shiga toxin-producing *Escherichia coli* (STEC) surveillance. https://www.cdc.gov/ecoli/surveillance.html. Accessed 12 June 2020.
2. CDC. 2020. National Outbreak Reporting System (NORS). https://wwwn.cdc.gov/norsdashboard/. Accessed 12 June 2020.
3. CDC. 2020. Reports of selected *E. coli* outbreak investigations. https://www.cdc.gov/ecoli/outbreaks.html. Accessed 12 June 2020.
4. Arthur TM, Brichta-Harhay DM, Bosilevac JM, Kalchayanand N, Shackelford SD, Wheeler TL, Koohmaraie M. 2010. Super shedding of *Escherichia coli* O157:H7 by cattle and the impact on beef carcass contamination. Meat Sci 86:32–37. https://doi.org/10.1016/j.meatsci.2010.04.019.
5. Caprioli A, Morabito S, Brugère H, Oswald E. 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. Vet Res 36:289–311. https://doi.org/10.1051/vetres:2005002.
6. Bach SJ, McAllister TA, Veira DM, Gannon VPJ, Holley RA. 2002. Transmission and control of *Escherichia coli* O157:H7: a review. Can J Anim Sci 82:475–490. https://doi.org/10.4141/A02-021.
7. 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.
8. 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.
9. 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.
10. Carattoli A, Zankari E, Garcìa-Fernandez A, Larsen M, Lund O, Villa L, Aarestrup F, Hasman H. 2014. PlasmidFinder and pMLST: *in silico* detection and typing of plasmids. Antimicrob Agents Chemother 58:3895–3903. https://doi.org/10.1128/AAC.02412-14.
11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
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. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder Web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4:e27943. https://doi.org/10.4161/bact.27943.
16. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FSL, Simon Fraser University Research Computing Group. 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.