AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Whole-Genome Sequence of *Escherichia coli* Serotype O157:H7 Strain ATCC 43888

Gaylen A. Uhlich,[a] George C. Paoli,[a] Xinmin Zhang,[b] Elisa Andreozzi[a]

[a]Molecular Characterization of Foodborne Pathogens Research Unit, USDA, Agriculture Research Service, Eastern Regional Research Center, Wyndmoor, Pennsylvania, USA

[b]BioInfoRx, Inc., Madison, Wisconsin, USA

ABSTRACT *Escherichia coli* serotype O157:H7 strain ATCC 43888 is a Shiga toxin-deficient human fecal isolate. Due to its reduced toxicity and its availability from a curated culture collection, the strain has been used extensively in applied research studies. Here, we report the Illumina-corrected PacBio whole-genome sequence of *E. coli* O157:H7 strain ATCC 43888.

Hemorrhagic diarrhea caused by Shiga toxin (Stx)-producing *Escherichia coli* (STEC) serotype O157:H7 was first reported in the early 1980s (1, 2), and it remains the leading cause of STEC-related illnesses, deaths, and hospitalizations (3). Strains of serotype O157:H7 are genetically diverse due to variations in horizontally transferred regions (HTRs) that encode virulence factors, secretion systems, and host effector proteins (4, 5) and contribute to niche adaptation in the environment, in the host, and in foods (6, 7). The locus of enterocyte effacement (LEE) and *stx* genes, both encoded in HTRs, are considered definitive virulence factors. For safety reasons, studies of STEC O157:H7 strains often employ those deficient in different virulence factors, especially Stx. One such strain is B6914-MS1, a Stx-deficient human fecal isolate (8) that is available from the American Type Culture Collection (ATCC) as biosafety level 1 strain ATCC 43888. Due to genetic variability, the genome sequence from a single O157:H7 strain would not be considered representative for the serotype. Therefore, we determined the draft genome sequence of strain ATCC 43888 for use in comparative studies with sequenced genomes of O157:H7 strains whose fitness properties differ from those of ATCC 43888.

Strain ATCC 43888 was purchased from the ATCC and grown overnight in LB at 37°C. Large DNA fragments were extracted from a frozen cell pellet (−80°C) collected from 5 ml of the overnight culture using the Genomic-tip 100/G kit (Qiagen, Valencia, CA) and used for both PacBio and Illumina sequencing. PacBio sequencing and *de novo* genome assembly were done by the University of Delaware Sequencing Center (Dover, DE) as previously described (9). Illumina sequencing, sequence quality scoring, and sequence trimming were done by ProteinCT Biotechnologies, LLC (Madison, WI), as previously described (10). Default parameters were used for all software tools unless otherwise noted (9, 10). PacBio and Illumina sequencing yielded 111,360 reads with an average length of 8.45 kb (171-fold coverage) and 3.5 million 2 × 250-bp paired-end reads (327-fold coverage), respectively. The PacBio sequence assembly was corrected by mapping the Illumina sequence reads against the PacBio assembled contigs using BWA-MEM version 0.7.12-r1044 (https://github.com/lh3/bwa) to identify high-quality variants that were used to update the consensus sequences.

The closed ATCC 43888 genome consists of the following 3 circular contigs: a 5,375,715 bp chromosome with a G+C content of 49.9% and plasmids of 32,726 bp (51.2% G+C content) and 95,081 bp (48.8% G+C content) with sequence similarity to

Citation Uhlich GA, Paoli GC, Zhang X, Andreozzi E. 2019. Whole-genome sequence of *Escherichia coli* serotype O157:H7 strain ATCC 43888. Microbiol Resour Announc 8:e00906-19. https://doi.org/10.1128/MRA.00906-19.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to George C. Paoli, george.paoli@usda.gov.

Received 14 August 2019
Accepted 30 September 2019
Published 17 October 2019

E. coli O157:H7 plasmids p35K (GenBank accession number CP015845) and pO157, respectively, as determined by BLAST analysis using GenBank. Based on the high-quality PacBio reads, the average sequencing coverage for the 3 contigs was 90×. The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline version 4.8 (11). BLAST searches of individual genes or loci revealed that the genes for Stx1 and Stx2 are not present, and no prophage was present in the stx_1 insertion site within *mlrA*, nor was any prophage present in the stx_2 insertion sites within the *wrbA*, *sbcB*, *yecE*, or *Z2577* genes. An LEE pathogenicity island with 99.99% identity (100% coverage with no gaps) to the reference outbreak strain Sakai (GenBank accession number NC_002695) (4) was present. The PHAge Search Tool Enhanced Release (PHASTER) prophage locator and annotation tool (12, 13) identified 9 intact, 4 incomplete, and 3 questionable prophage regions, representing approximately 11% of the genome.

Data availability. The whole-genome sequence was deposited at DDBJ/ENA/GenBank (accession numbers CP041623, CP041624, and CP041625). The versions described in this paper are the first versions. The Illumina and PacBio reads can be found in the SRA under accession numbers SRR9694420 and SRR9694229, respectively.

ACKNOWLEDGMENTS

This research was supported in part by an appointment of Elisa Andreozzi to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA), managed by ORAU under contract number DE-SC0014664.

All opinions expressed in this paper are those of the authors and do not necessarily reflect the policies and views of USDA, ARS, DOE, or ORAU/ORISE. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

REFERENCES

1. Karmali MA, Steele BT, Petric M, Lim C. 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet 321:619–620. https://doi.org/10.1016/S0140-6736(83)91795-6.
2. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML. 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med 308:681–685. https://doi.org/10.1056/NEJM198303243081203.
3. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis 17:7–15. https://doi.org/10.3201/eid1701.09-1101p1.
4. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 8:11–22. https://doi.org/10.1093/dnares/8.1.11.
5. Perna NT, Plunkett G, 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533. https://doi.org/10.1038/35054089.
6. Wick LM, Qi W, Lacher DW, Whittam TS. 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. J Bacteriol 187:1783–1791. https://doi.org/10.1128/JB.187.5.1783-1791.2005.
7. Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, Fivian A, Younis R, Matthews S, Marches O, Frankel G, Hayashi T, Pallen MJ. 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc Natl Acad Sci U S A 103:14941–14946. https://doi.org/10.1073/pnas.0604891103.
8. Kim HH, Samadpour M, Grimm L, Clausen CR, Besser TE, Baylor M, Kobayashi JM, Neill MA, Schoenknecht FD, Tarr PI. 1994. Characteristics of antibiotic-resistant *Escherichia coli* O157:H7 in Washington State, 1984–1991. J Infect Dis 170:1606–1609. https://doi.org/10.1093/infdis/170.6.1606.
9. Uhlich GA, Reichenberger ER, Cottrell BJ, Fratamico P, Andreozzi E. 2017. Whole-genome sequence of *Escherichia coli* serotype O157:H7 strain B6914-ARS. Genome Announc 5:e01191-17. https://doi.org/10.1128/genomeA.01191-17.
10. Uhlich GA, Paoli GC, Chen C-Y, Cottrell BJ, Zhang X, Yan X. 2016. Whole-genome sequence of *Escherichia coli* serotype O157:H7 strain EDL932 (ATCC 43894). Genome Announc 4:e00647-16. https://doi.org/10.1128/genomeA.00647-16.
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. 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.
13. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.