



High-Quality Whole-Genome Sequences for 21 Enterotoxigenic *Escherichia coli* Strains Generated with PacBio Sequencing

Peyton Smith,^{a,b} Rebecca L. Lindsey,^a Lori A. Rowe,^a Dhvani Batra,^a Devon Stripling,^a Lisle Garcia-Toledo,^{a,b} Daniel Drapeau,^{a,b} Kristen Knipe,^a Nancy Strockbine^a

^aCenters for Disease Control and Prevention, Atlanta, Georgia, USA

^bOak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

ABSTRACT Enterotoxigenic *Escherichia coli* (ETEC) is an important diarrheagenic pathogen. We report here the high-quality whole-genome sequences of 21 ETEC strains isolated from patients in the United States, international diarrheal surveillance studies, and cruise ship outbreaks.

Enterotoxigenic *Escherichia coli* (ETEC) infections are the leading cause of travelers' diarrhea and the most common cause of diarrhea among children in developing countries (1). ETEC causes approximately 210 million infections and 380,000 deaths per year (2). ETEC infections in humans are characterized by the production of either or both heat-stable (ST) and heat-labile (LT) enterotoxins (3). Here, we report the availability of high-quality genome sequences for 21 ETEC strains generated by PacBio sequencing. Each of the 21 genomes contained one chromosomal sequence, and 17 of these were determined to be circular with overlapping ends that were trimmed from one end.

ETEC genomic DNA was extracted using Archive Pure according to the manufacturer's protocol (5 Prime, Gaithersburg, MD, USA). The genomic DNA was sheared to 20 kb using needle shearing. These libraries were further size selected using BluePippin (Sage Scientific, Beverly, MA, USA). The sheared DNA was used to generate large SMRTbell libraries using the standard library protocols of the Pacific Biosciences DNA template preparation kit (Pacific Biosciences, Menlo Park, CA, USA). All strains were sequenced using one single-molecule real-time (SMRT) cell, except strain 9276-90, which used two. The finished libraries were bound to proprietary P6 v2 polymerase and sequenced on a PacBio RSII system using C4 chemistry for 360-min movies. Sequence reads were filtered and assembled *de novo* utilizing the PacBio Hierarchical Genome Assembly Process version 3 (4).

Table 1 lists the accession numbers, detected serotype (genotypic), assembly metrics, and reference (if available) for each ETEC whole-genome sequence. A single chromosomal sequence was obtained for all genomes with an average coverage of 102.6× (range, 42.3× to 188.6×). The average G+C content for each chromosomal sequence was 50.68%, ranging from 50.37% to 50.82%. The chromosomal sequences are circular with overlapping ends for all but four genomes. The single chromosomal sequences for isolates 90-9276, 90-9269, F5176C6, and 00-3279 could not be circularized because of unresolved or collapsed repeats. Each ETEC genome contained between one and five plasmids. Table 1 notes whether overlap was found to form circular chromosomal or plasmid contigs.

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions.

Received 15 November 2017 **Accepted** 20 November 2017 **Published** 11 January 2018

Citation Smith P, Lindsey RL, Rowe LA, Batra D, Stripling D, Garcia-Toledo L, Drapeau D, Knipe K, Strockbine N. 2018. High-quality whole-genome sequences for 21 enterotoxigenic *Escherichia coli* strains generated with PacBio sequencing. *Genome Announc* 6:e01311-17. <https://doi.org/10.1128/genomeA.01311-17>.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Rebecca L. Lindsey, rebecca.lindsey@cdc.hhs.gov.

TABLE 1 Accession numbers and assembly metrics of 21 ETEC whole-genome sequences

<i>E. coli</i> strain no. (reference)	Serotype	Chromosomal GenBank accession no.	Average coverage (×)	Chromosome size (bp)	Associated plasmid size (bp) (GenBank accession no.)
2014EL-1345-2	O169:H41	CP024223	92.3	4,943,397 ^a	145,086 ^a (CP024227) 85,864 (CP024226) 27,947 (CP024225) 20,005 (CP024224)
ATCC 43886 or E2539C1 (5, 6)	O25:H16	CP024256	148.1	4,914,654 ^a	95,515 ^a (CP024255) 107,732 ^a (CP024254)
D181 (7)	O182:H21	CP024252	101.6	4,891,230 ^a	46,427 (CP024253) 99,099 ^a (CP024250) 167,230 ^a (CP024249) 35,752 (CP024251)
90-9281 (8)	O128:H27	CP024243	117.5	4,978,613 ^a	152,012 ^a (CP024244)
90-9276 (8)	O114:H49	CP024299	79.8	5,004,571	167,764 (CP024297) 110,577 (CP024298)
90-9280 (8)	O114:H49	CP024240	188.6	4,966,338 ^a	74,644 ^a (CP024242) 104,674 (CP024241)
90-9269 (8)	OUND:H4	CP024661	150.6	4,759,941	160,351 ^a (CP024662) 12,610 (CP024663) 52,270 ^a (CP024664) 96,720 (CP024665) 20,753 (CP024666)
90-9272 (8)	O15:H11	CP024239	48.3	4,906,680 ^a	274,465 (CP024238)
ATCC 43896 (or TX1) (9, 10)	O78:H12	CP024278	107.7	5,088,038 ^a	84,894 ^a (CP024281) 52,655 (CP024280) 28,860 (CP024279)
M9682-C1 (11)	O6:H16	CP024275	156.6	4,778,550 ^a	100,184 ^a (CP024277) 38,177 (CP024276)
B4103-1 (12)	O27:H7	CP024245	83.1	4,708,118 ^a	138,289 ^a (CP024248) 68,864 (CP024247) 36,948 (CP024246)
F6326-C1	O169:H41	CP024263	78.1	4,934,701 ^a	150,389 ^a (CP024265) 72,060 (CP024264)
F5176C6 (13)	O167:H5	CP024667	86.4	5,069,317	167,071 ^a (CP024668) 110,612 ^a (CP024669) 46,919 ^a (CP024670) 31,494 (CP024671)
2014EL-1343-2	O25:H16 (genotypic), O25:NM (phenotypic)	CP024228	91.4	4,848,034 ^a	82,510 ^a (CP024231) 73,915 (CP024230) 38,467 (CP024229)
F5656C1 (14)	O6:H16	CP024260	63.8	4,733,683 ^a	119,846 ^a (CP024262) 45,056 ^a (CP024261)
2014EL-1346-6	O6:H16	CP024232	98.8	4,872,840 ^a	152,713 ^a (CP024237) 226,119 ^a (CP024236) 62,188 ^a (CP024235) 40,223 (CP024234) 30,162 (CP024233)
F9792 (14, 15)	O169:H41	CP024273	92.2	4,875,605 ^a	145,089 ^a (CP024274)
F5505-C1	O25:H16	CP024257	94.3	4,886,938 ^a	94,817 ^a (CP024259) 96,607 ^a (CP024258)
F6699	O6:H16	CP024266	42.3	4,881,899 ^a	95,350 (CP024268) 34,042 (CP024267)
F8111-1SC3 (14)	O169:H41	CP024269	78.6	4,905,023 ^a	147,766 ^a (CP024272) 103,618 (CP024271) 35,769 (CP024270)
00-3279	O78:H12	CP024293	155.0	5,116,480	101,907 ^a (CP024294) 103,995 ^a (CP024295) 97,297 ^a (CP024296)

^aA sequence that is circular with overlapping ends.

ACKNOWLEDGMENTS

This work was funded by federal appropriations to the Centers for Disease Control and Prevention, through the Advanced Molecular Detection Initiative line item.

The findings and conclusions of this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. Use

of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services.

REFERENCES

1. Isidean SD, Riddle MS, Savarino SJ, Porter CK. 2011. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine* 29:6167–6178. <https://doi.org/10.1016/j.vaccine.2011.06.084>.
2. Wennerås C, Erling V. 2004. Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world. *J Health Popul Nutr* 22:370–382.
3. Sack RB. 1975. Human diarrheal disease caused by enterotoxigenic *Escherichia coli*. *Annu Rev Microbiol* 29:333–353. <https://doi.org/10.1146/annurev.mi.29.100175.002001>.
4. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
5. Wachsmuth K, Wells J, Shipley P, Ryder R. 1979. Heat-labile enterotoxin production in isolates from a shipboard outbreak of human diarrheal illness. *Infect Immun* 24:793–797.
6. Centers for Disease Control and Prevention (CDC). 1976. Diarrheal illness on a cruise ship caused by enterotoxigenic *Escherichia coli*. *MMWR Morbid Mortal Wkly Rep* 25:229–230.
7. Youssef M, Shurman A, Bougnoux M, Rawashdeh M, Bretagne S, Strockbine N. 2000. Bacterial, viral and parasitic enteric pathogens associated with acute diarrhea in hospitalized children from northern Jordan. *FEMS Immunol Med Microbiol* 28:257–263. <https://doi.org/10.1111/j.1574-695X.2000.tb01485.x>.
8. Strockbine NA, Faruque SM, Kay BA, Haider K, Alam K, Alam AN, Tzipori S, Wachsmuth IK. 1992. DNA probe analysis of diarrhoeagenic *Escherichia coli*: detection of EAF-positive isolates of traditional enteropathogenic *E. coli* serotypes among Bangladeshi paediatric diarrhoea patients. *Mol Cell Probes* 6:93–99. [https://doi.org/10.1016/0890-8508\(92\)90052-Y](https://doi.org/10.1016/0890-8508(92)90052-Y).
9. Ryder RW, Wachsmuth IK, Buxton AE, Evans DG, DuPont HL, Mason E, Barrett FF. 1976. Infantile diarrhea produced by heat-stable enterotoxigenic *Escherichia coli*. *N Engl J Med* 295:849–853. <https://doi.org/10.1056/NEJM197610142951601>.
10. Wachsmuth IK, Falkow S, Ryder RW. 1976. Plasmid-mediated properties of a heat-stable enterotoxin-producing *Escherichia coli* associated with infantile diarrhea. *Infect Immun* 14:403–407.
11. Rosenberg ML, Koplan JP, Wachsmuth IK, Wells JG, Gangarosa EJ, Guerrant RL, Sack DA. 1977. Epidemic diarrhea at Crater Lake from enterotoxigenic *Escherichia coli*: a large waterborne outbreak. *Ann Intern Med* 86:714–718. <https://doi.org/10.7326/0003-4819-86-6-714>.
12. Dalton CB, Mintz ED, Wells JG, Bopp CA, Tauxe RV. 1999. Outbreaks of enterotoxigenic *Escherichia coli* infection in American adults: a clinical and epidemiologic profile. *Epidemiol Infect* 123:9–16. <https://doi.org/10.1017/S0950268899002526>.
13. Daniels NA, Neimann J, Karpati A, Parashar UD, Greene KD, Wells JG, Srivastava A, Tauxe RV, Mintz ED, Quick R. 2000. Traveler's diarrhea at sea: three outbreaks of waterborne enterotoxigenic *Escherichia coli* on cruise ships. *J Infect Dis* 181:1491–1495. <https://doi.org/10.1086/315397>.
14. Beatty ME, Bopp CA, Wells JG, Greene KD, Puhr ND, Mintz ED. 2004. Enterotoxin-producing *Escherichia coli* O169:H41, United States. *Emerg Infect Dis* 10:518–521. <https://doi.org/10.3201/eid1003.030268>.
15. Devasia RA, Jones TF, Ward J, Stafford L, Hardin H, Bopp C, Beatty M, Mintz E, Schaffner W. 2006. Endemically acquired foodborne outbreak of enterotoxin-producing *Escherichia coli* serotype O169:H41. *Am J Med* 119:168.e7–168.e10. <https://doi.org/10.1016/j.amjmed.2005.07.063>.