Genome Sequences of 228 Shiga Toxin-Producing Escherichia coli Isolates and 12 Isolates Representing Other Diarrheagenic E. coli Pathotypes

Eija Trees,a Nancy Stockbime,a Shankar Changayil,b Satishkumar Ranganathan,b Kun Zhao,b Ryan Weil,b Duncan MacCannell,b Ashley Sabol,a Amber Schmidtke,a Haley Martin,a Devon Stripling,a Efrain M. Ribot,a Peter Gerner-Smidtab

Division of Foodborne, Waterborne, and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; Office of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Shiga toxin-producing Escherichia coli (STEC) are a common cause for food-borne diarrheal illness outbreaks and sporadic cases. Here, we report the availability of the draft genome sequences of 228 STEC strains representing 32 serotypes with known pulsed-field gel electrophoresis (PFGE) types and epidemiological relationships, as well as 12 strains representing other diarrheagenic E. coli pathotypes.

Received 24 June 2014 Accepted 16 July 2014 Published 7 August 2014

Citation Trees E, Strockbine N, Changayil S, Ranganathan S, Zhao K, Weil R, MacCannell D, Sabol A, Schmidtke A, Martin H, Stripling D, Ribot EM, Gerner-Smidt P. 2014. Genome sequences of 228 Shiga toxin-producing Escherichia coli isolates and 12 isolates representing other diarrheagenic E. coli pathotypes. Genome Announc. 2(4):e00718-14. doi: 10.1128/genomeA.00718-14.

Copyright © 2014 Trees et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.
Address correspondence to Eija Trees, eih9@cdc.gov.

The rapidly decreasing cost of next-generation sequencing (NGS) will facilitate its application for real-time surveillance in the near future. PulseNet, the molecular subtyping network for food-borne disease surveillance, currently relies on pulsed-field gel electrophoresis (PFGE) to define clusters of illness (1). In order to use NGS as a primary method for cluster detection, a thorough understanding of the genetic diversity in the target population is needed. Shiga toxin-producing Escherichia coli (STEC) are among the pathogens tracked by PulseNet. In this report, we announce the availability of the draft sequences of a carefully selected set of STEC strains that should enable us to gain insights into the sequence diversity within an outbreak or a carrier state and among epidemiologically unrelated isolates within a serotype and between serotypes.

We sequenced 228 STEC strains representing 32 serotypes with known PFGE types and epidemiological relationships. The strain set included a total of 50 isolates from five outbreaks, 11 isolates from a long-term carrier, and epidemiologically unrelated strains. Twelve strains of other diarrheagenic E. coli pathotypes were included as outliers. Genomic DNA from each strain was isolated using the ArchivePure DNA cell/tissue kit (5Prime, Hamburg, Germany). All 240 strains were sequenced to a minimum depth of 100× with the HiSeq 2000 or GAIIx (Illumina, San Diego, CA, USA) using the TrueSeq DNA LT sample prep kit (Illumina) for DNA library preparation and 100-bp paired-end read chemistry. Additionally, 82 strains were sequenced with the PacBio RS system (Pacific Biosciences, Menlo Park, CA) using C2 chemistry and four single-molecule real-time (SMRT) cells per genome.

Raw read quality checks were performed on the 240 samples using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) and in-house Perl scripts/Java programs. Primary analysis for the Illumina data was performed using CLC Genomics Workbench 5.5.1 (Aarhus, Denmark). The raw read files for each sample were trimmed with length (minimum, 50 bp) and quality score (0.02) filters. The trimmed reads were assembled into contigs with specific parameter settings (length fraction, 0.8; similarity fraction, 0.8; minimum contig length, 450 bp), and assembly statistics were parsed out in a table format using in-house scripts. The PacBio data analysis was performed using the whole-genome sequencing (WGS) assembler toolkit (2). Error correction of the filtered subreads was performed with the paired-end Illumina data (~60× data was used) using the WGS toolkit PacBioToCA script, followed by de novo assembly using the runCA script. The best assembly for each of these 82 samples was chosen based on the number of contigs, N50 value, and genome length.

The average genome size for the sequenced strains was 5,282,291 bp (range, 4,527,885 to 5,712,627). For the 240 Illumina assemblies, the average number of contigs was 211 (range, 31,094 to 1,414,730). The average genome size for the 82 PacBio hybrid assemblies was 128,850 (range, 26,435 to 230,877). For the 82 PacBio hybrid assemblies, the average number of contigs was 207 (range, 31 to 207), and the average N50 was 172,854 (range, 31,094 to 1,414,730).

Nucleotide sequence accession numbers. The draft genome sequences for these 240 diarrheagenic E. coli strains have been deposited in DDBJ/ENA/GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENT
No external funding was received for this project.

REFERENCES
1. Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyytia-Trees E, Ribot EM, Swaminathan B. 2006. PulseNet USA: a five-year update. Foodborne Pathog. Dis. 3:9–19. http://dx.doi.org/10.1089/fpd.2006.3.9.
2. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. 30:693–700. http://dx.doi.org/10.1038/nbt.2280.
| Strain ID | Serotype | NCBI accession no. |
|----------|----------|-------------------|
| 00-3279 | O78:H12 | JFBE00000000000 |
| 01-3076 | O111:NM | JFPJ00000000000 |
| 02-3147 | O45:H2 | JHOC00000000000 |
| 02-3012 | O81:NM | JHNZ00000000000 |
| 02-3140 | O28ac:NM | JHYN00000000000 |
| 03-3227 | O121:H19 | JHXN00000000000 |
| 03-3269 | O174:H21 | JHWO00000000000 |
| 03-3458 | O119:H4 | JHNV00000000000 |
| 03-3484 | O111:NM | JHUN00000000000 |
| 03-3500 | O26:H11 | JHTN00000000000 |
| 04-3023 | O103:H11 | JHOD00000000000 |
| 04-3038 | O174:H8 | JHOC00000000000 |
| 04-3211 | O111:NM | JHN50000000000 |
| 05-3646 | O26:H11 | JHOC00000000000 |
| 06-3003 | O121:H19 | JHN90000000000 |
| 06-3256 | O118:H16 | JHQC00000000000 |
| 06-3325 | O69:H11 | JHNP00000000000 |
| 06-3464 | O26:H11 | JHNC00000000000 |
| 06-3484 | O145:NM | JHNN00000000000 |
| 06-3501 | O79:H7 | JHN00000000000 |
| 06-3555 | O55:H7 | JHN00000000000 |
| 06-3612 | O118:H16 | JHNX00000000000 |
| 06-3691 | O91:H14 | JHN00000000000 |
| 06-3745 | O157:H7 | JHN10000000000 |
| 06-3822 | O121:H19 | JHN10000000000 |
| 06-4039 | O157:H7 | JHN00000000000 |
| 07-3091 | O157:NM | JHN00000000000 |
| 07-3391 | O157:H7 | JHN00000000000 |
| 07-4224 | O113:H21 | JHOB00000000000 |
| 07-4281 | O69:H11 | JHLA00000000000 |
| 08-3037 | O157:H7 | JHKO00000000000 |
| 08-3527 | O157:H7 | JHKO00000000000 |
| 08-3651 | O118:H16 | JHKO00000000000 |
| 08-4169 | O157:H7 | JHKO00000000000 |
| 08-4270 | O145:NM | JHKO00000000000 |
| 08-4487 | O111:NM | JHKO00000000000 |
| 08-4529 | O157:H7 | JHH00000000000 |
| 08-4540 | O157:NM | JHHP00000000000 |
| 08-4661 | O69:H11 | JHHG00000000000 |
| 2009C-3227 | O91:H14 | JHHP00000000000 |
| 2009C-3279 | O103:H2 | JHHP00000000000 |
| 2009C-3292 | O145:H28 | JHHP00000000000 |
| 2009C-3299 | O121:H7 | JHHP00000000000 |
| 2009C-3307 | O123:H11 | JHHP00000000000 |
| 2009C-3601 | O69:H11 | JHHA00000000000 |
| 2009C-3612 | O26:H11 | JHGC00000000000 |
| 2009C-3686 | O45:H2 | JHGYO00000000000 |
| 2009C-3689 | O26:H11 | JHGX00000000000 |
| 2009C-3745 | O91:NM | JHGW00000000000 |
| 2009C-3996 | O26:H11 | JHGV00000000000 |
| 2009C-4006 | O111:NM | JHGU00000000000 |
| 2009C-4052 | O111:NM | JHGS00000000000 |
| 2009C-4126 | O111:H8 | JHGR00000000000 |
| 2009C-4258 | O157:H7 | JHGO00000000000 |
| 2009C-4446 | O118:H16 | JHGP00000000000 |
| 2009C-4646 | O91:H21 | JHGO00000000000 |
| 2009C-4659 | O121:H19 | JHGN00000000000 |
| 2009C-4747 | O26:H11 | JHGM00000000000 |
| 2009C-4750 | O121:H19 | JHGL00000000000 |
| 2009C-4760 | O26:H11 | JHGL00000000000 |
| 2009C-4780 | O45:H2 | JHGO00000000000 |
| 2009C-4826 | O26:H11 | JHGO00000000000 |
| 2009EL1302 | O121:H19 | JHGO00000000000 |
| 2009EL1412 | O121:H19 | JHGO00000000000 |

(Continued on following page)
| Strain ID    | Serotype   | NCBI accession no. |
|-------------|------------|--------------------|
| 2011C-3387  | O26:H11    | JHLV000000000      |
| 2011C-3453  | O111:H8    | JHLU000000000      |
| 2011C-3500  | O121:H19   | JHLT000000000      |
| 2011C-3506  | O26:H11    | JHSL000000000      |
| 2011C-3537  | O111:NM    | JHLQ000000000      |
| 2011C-3573  | O121:H19   | JHLR000000000      |
| 2011C-3602  | O26:H11    | JHLS000000000      |
| 2011C-3632  | O111:NM    | JHLO000000000      |
| 2011C-3655  | O111:NM    | JHLM000000000      |
| 2011C-3679  | O103:H2    | JHLN000000000      |
| 2011C-3750  | O157:H7    | JHLK000000000      |
| 2011EL-1107 | O157:H7    | JHLJ000000000      |
| 2011EL-1675A| O104:H4    | JHLI000000000      |
| 2011EL-2090 | O157:H7    | JHLG000000000      |
| 2011EL-2091 | O157:H7    | JHLH000000000      |
| 2011EL-2092 | O157:H7    | JHLG000000000      |
| 2011EL-2093 | O157:H7    | JHLF000000000      |
| 2011EL-2094 | O157:H7    | JHLE000000000      |
| 2011EL-2096 | O157:H7    | JHLD000000000      |
| 2011EL-2097 | O157:H7    | JHLC000000000      |
| 2011EL-2098 | O157:H7    | JHLB000000000      |
| 2011EL-2099 | O157:H7    | JHLA000000000      |
| 2011EL-2101 | O157:H7    | JHJZ000000000      |
| 2011EL-2102 | O157:H7    | JHJY000000000      |
| 2011EL-2103 | O157:H7    | JHJX000000000      |
| 2011EL-2104 | O157:H7    | JHJW000000000      |
| 2011EL-2105 | O157:H7    | JHJW000000000      |
| 2011EL-2106 | O157:H7    | JHJW000000000      |
| 2011EL-2107 | O157:H7    | JHJW000000000      |
| 2011EL-2108 | O157:H7    | JHJW000000000      |
| 2011EL-2109 | O157:H7    | JHJW000000000      |
| 2011EL-2111 | O157:H7    | JHJW000000000      |
| 2011EL-2112 | O157:H7    | JHJW000000000      |
| 2011EL-2113 | O157:H7    | JHJW000000000      |
| 2011EL-2114 | O157:H7    | JHJW000000000      |
| 2011EL-2128 | O157:H7    | JHJW000000000      |
| 2011EL-2286 | O157:H7    | JHJW000000000      |
| 2011EL-2287 | O157:H7    | JHJW000000000      |
| 2011EL-2288 | O157:H7    | JHJW000000000      |
| 2011EL-2289 | O157:H7    | JHJW000000000      |
| 2011EL-2290 | O157:H7    | JHJW000000000      |
| 2011EL-2312 | O157:H7    | JHJW000000000      |
| 2011EL-2313 | O157:H7    | JHJW000000000      |
| 94-3025     | O103:H2    | JHJW000000000      |
| 94-1133     | O104:H4    | JHJW000000000      |
| 96-0024     | O105:H7    | JHJW000000000      |
| 96-1133     | O105:H7    | JHJW000000000      |
| 96-1134     | O105:H7    | JHJW000000000      |
| 96-1135     | O105:H7    | JHJW000000000      |
| 99-3165     | O6:H16     | JHJW000000000      |
| E2539C1     | O25:NM     | JHJW000000000      |
| F5656C1     | O6:H16     | JHJW000000000      |
| F6142       | O157:H7    | JHJW000000000      |
| F6627       | O103:H2    | JHJW000000000      |
| F6714       | O105:H7    | JHJW000000000      |
| F6749       | O157:H7    | JHJW000000000      |
| F6750       | O157:H7    | JHJW000000000      |
| F6751       | O157:H7    | JHJW000000000      |
| F7384       | O157:H7    | JHJW000000000      |
| F7410       | O157:H7    | JHJW000000000      |
| F9792       | O169:H41   | JHJW000000000      |
| G3503       | O157:H7    | JHJW000000000      |
| H2495       | O157:H7    | JHJW000000000      |
| H2498       | O157:H7    | JHJW000000000      |
| K1420       | O157:H7    | JHJW000000000      |
| K156        | O157:H18   | JHJW000000000      |
| K1792       | O157:H7    | JHJW000000000      |
| K1793       | O157:H7    | JHJW000000000      |