



Species-Wide Collection of *Escherichia coli* Isolates for Examination of Genomic Diversity

Jayanthi Gangiredla, Mark K. Mammel, Tammy J. Barnaba, Carmen Tartera, Solomon T. Gebru, Isha R. Patel, Susan R. Leonard, Michael L. Kotewicz, Keith A. Lampel, Christopher A. Elkins, David W. Lacher

Division of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland, USA

ABSTRACT Pathogenic and nonpathogenic *Escherichia coli* strains present a vast genomic diversity. We report the genome sequences of 2,244 *E. coli* isolates from multiple animal and environmental sources. Their phylogenetic relationships and potential risk to human health were examined.

Most *Escherichia coli* strains are harmless commensals that are found as part of the gut microbiota. Some strains, however, have the ability to cause disease and are considered pathogenic. Phylogenetic analyses have shown that *E. coli* strains can be divided into several phylogroups (1, 2), with pathogenic and nonpathogenic strains randomly distributed among them. To facilitate the study of the genomic diversity of this species, we sequenced a collection of 2,244 *E. coli* isolates from multiple mammalian (including human), avian, and environmental sources. This diverse collection contains nonpathogenic strains as well as several different pathogenic types (i.e., pathotypes), including attaching and effacing (AEEC), enteroaggregative (EAEC), enteroinvasive (EIEC), enterotoxigenic (ETEC), and Shiga toxin-producing (STEC) *E. coli* (3). Pathotypes are often defined by differing sets of virulence-associated genes. Many of these genes are carried on mobile genetic elements that can be transferred among strains, resulting in new combinations and several hybrid pathotypes such as STEC/EAEC (4), STEC/ETEC (5–7), and AEEC/ETEC (8). In this report, we characterized 2,244 *E. coli* isolates based on phylogenetic relationships and their potential risk to human health. The information reported here will help to better understand the evolution of these emergent foodborne pathogens and improve the accuracy of trace-back investigations during outbreaks caused by them.

Pure cultures for each strain were grown aerobically overnight in Luria-Bertani broth at 37°C. Total genomic DNA was extracted from 1 ml of overnight culture using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). DNA extractions were performed with the Qiagen QIAcube instrument using the manufacturer's Gram-negative bacteria protocol. Sequencing libraries were prepared with 1 ng of DNA using the Nextera XT DNA sample prep kit (Illumina, San Diego, CA, USA) and sequenced on either the Illumina MiSeq or NextSeq platform. The resulting paired-end reads were quality controlled using FastQC ($Q > 30$) and *de novo* assembled using SPAdes 3.8.2 (9) or CLC Genomics Workbench 8.2.1 (CLC bio, Aarhus, Denmark).

Depth of coverage for the draft genomes ranged from 20× to 200× with the genome sizes ranging from 4,412,939 to 5,984,698 bp. The number of contigs ranged from 39 to 1,110, while the N_{50} values ranged from 14,741 to 699,676 bp. Each of the established *E. coli* phylogroups is represented in the sequenced strain collection as follows: A, 23%; B1, 47%; B2, 13%; D, 6%; E, 9%; and F, 2%. The strains were also screened for the presence of known or putative virulence factors, such as *aggR*, *eae*,

Received 24 October 2017 Accepted 27 October 2017 Published 14 December 2017

Citation Gangiredla J, Mammel MK, Barnaba TJ, Tartera C, Gebru ST, Patel IR, Leonard SR, Kotewicz ML, Lampel KA, Elkins CA, Lacher DW. 2017. Species-wide collection of *Escherichia coli* isolates for examination of genomic diversity. *Genome Announc* 5:e01321-17. <https://doi.org/10.1128/genomeA.01321-17>.

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 David W. Lacher, david.lacher@fda.hhs.gov.

ipaH, *LT*, *ST*, *stx1*, and *stx2*. Out of the 2,244 isolates, 394 can be classified as AEEC, 23 as EAEC, 9 as EIEC, 134 as ETEC, and 402 as STEC. Several strains were found to possess factors associated with hybrid pathotypes: STEC/ETEC ($n = 22$), AEEC/ETEC ($n = 2$), and STEC/EAEC ($n = 1$).

Accession number(s). The draft genome assemblies were deposited in DDBJ/ENA/GenBank through FDA's GenomeTrakr pipeline under BioProject [PRJNA230969](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA230969) with accession numbers [NJIZ000000000](https://www.ncbi.nlm.nih.gov/nuclink/NJIZ000000000) to [NJNL000000000](https://www.ncbi.nlm.nih.gov/nuclink/NJNL000000000), [NJRR000000000](https://www.ncbi.nlm.nih.gov/nuclink/NJRR000000000) to [NKAI000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKAI000000000), [NKDC000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKDC000000000) to [NKEV000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKEV000000000), [NKLT000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKLT000000000) to [NKPS000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKPS000000000), [NKUK000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKUK000000000) to [NKVR000000000](https://www.ncbi.nlm.nih.gov/nuclink/NKVR000000000), [NLFN000000000](https://www.ncbi.nlm.nih.gov/nuclink/NLFN000000000) to [NMOH000000000](https://www.ncbi.nlm.nih.gov/nuclink/NMOH000000000), [NNSX000000000](https://www.ncbi.nlm.nih.gov/nuclink/NNSX000000000) to [NOIE000000000](https://www.ncbi.nlm.nih.gov/nuclink/NOIE000000000), [NOMB000000000](https://www.ncbi.nlm.nih.gov/nuclink/NOMB000000000) to [NOUU000000000](https://www.ncbi.nlm.nih.gov/nuclink/NOUU000000000), [NOWO000000000](https://www.ncbi.nlm.nih.gov/nuclink/NOWO000000000) to [NOWP000000000](https://www.ncbi.nlm.nih.gov/nuclink/NOWP000000000), [NTND000000000](https://www.ncbi.nlm.nih.gov/nuclink/NTND000000000) to [NTPX000000000](https://www.ncbi.nlm.nih.gov/nuclink/NTPX000000000), [NVPS000000000](https://www.ncbi.nlm.nih.gov/nuclink/NVPS000000000), [NWNA000000000](https://www.ncbi.nlm.nih.gov/nuclink/NWNA000000000) to [NWQF000000000](https://www.ncbi.nlm.nih.gov/nuclink/NWQF000000000), and [NXMG000000000](https://www.ncbi.nlm.nih.gov/nuclink/NXMG000000000) to [NXNF000000000](https://www.ncbi.nlm.nih.gov/nuclink/NXNF000000000). The versions described in this announcement are the first versions.

ACKNOWLEDGMENTS

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration (FDA), or the U.S. Government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the FDA.

Part of this study was supported by the ORISE Fellowship Program.

REFERENCES

- Herzer PJ, Inouye S, Inouye M, Whittam TS. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* 172:6175–6181. <https://doi.org/10.1128/jb.172.11.6175-6181.1990>.
- Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5:58–65. <https://doi.org/10.1111/1758-2229.12019>.
- Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140. <https://doi.org/10.1038/nrmicro818>.
- Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK. 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 365:709–717. <https://doi.org/10.1056/NEJMoa1106920>.
- Monday SR, Keys C, Hanson P, Shen Y, Whittam TS, Feng P. 2006. Produce isolates of the *Escherichia coli* Ont:H52 serotype that carry both Shiga toxin 1 and stable toxin genes. *Appl Environ Microbiol* 72:3062–3065. <https://doi.org/10.1128/AEM.72.4.3062-3065.2006>.
- Nyholm O, Halkilahti J, Wiklund G, Okeke U, Paulin L, Auvinen P, Haukka K, Siitonen A. 2015. Comparative genomics and characterization of hybrid Shigatoxigenic and enterotoxigenic *Escherichia coli* (STEC/ETEC) strains. *PLoS One* 10:e0135936. <https://doi.org/10.1371/journal.pone.0135936>.
- Leonard SR, Mammel MK, Rasko DA, Lacher DW. 2016. Hybrid Shiga toxin-producing and enterotoxigenic *Escherichia coli* sp. cryptic lineage 1 strain 7v harbors a hybrid plasmid. *Appl Environ Microbiol* 82:4309–4319.
- Dutta S, Pazhani GP, Nataro JP, Ramamurthy T. 2015. Heterogenic virulence in a diarrheagenic *Escherichia coli*: evidence for an EPEC expressing heat-labile toxin of ETEC. *Int J Med Microbiol* 305:47–54. <https://doi.org/10.1016/j.ijmm.2014.10.006>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.