

Whole-Genome Draft Sequences of 26 Enterohemorrhagic *Escherichia coli* O157:H7 Strains

Mark Eppinger,^a Sean Daugherty,^b Sonia Agrawal,^b Kevin Galens,^b Naomi Sengamalay,^b Lisa Sadzewicz,^b Luke Tallon,^b Thomas A. Cebula,^c Mark K. Mammel,^d Peter Feng,^d Robert Soderlund,^e Phillip I. Tarr,^f Chitrita DebRoy,^g Edward G. Dudley,^h Claire M. Fraser,^b Jacques Ravel^b

University of Texas at San Antonio, Department of Biology, and South Texas Center for Emerging Infectious Diseases, San Antonio, Texas, USA^a; Institute for Genome Sciences (IGS), University of Maryland, School of Medicine, Department of Microbiology and Immunology, Baltimore, Maryland, USA^b; Johns Hopkins University, Department of Biology, Baltimore, Maryland, USA^c; Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Laurel, Maryland, USA^d; National Veterinary Institute, Department of Bacteriology, SVA, Uppsala, Sweden^e; Washington University School of Medicine, St. Louis, Missouri, USA^f; Penn State University, Department of Veterinary and Biomedical Sciences, University Park, Pennsylvania, USA^g; Penn State University, Department of Food Science, University Park, Pennsylvania, USA^h

First identified in 1982, *Escherichia coli* O157:H7 is the dominant enterohemorrhagic serotype underlying food-borne human infections in North America. Here, we report the genomes of twenty-six strains derived from patients and the bovine reservoir. These resources enable detailed whole-genome comparisons and permit investigations of genotypic and phenotypic plasticity.

Received 14 December 2012 Accepted 23 January 2013 Published 28 February 2013

Citation Eppinger M, Daugherty S, Agrawal S, Galens K, Sengamalay N, Sadzewicz L, Tallon L, Cebula TA, Mammel MK, Feng P, Soderlund R, Tarr PI, DebRoy C, Dudley EG, Fraser CM, Ravel J. 2013. Whole-genome draft sequences of 26 enterohemorrhagic *Escherichia coli* O157:H7 strains. *Genome Announc.* 1(2):e00134-12. doi:10.1128/genomeA.00134-12.

Copyright © 2013 Eppinger et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Mark Eppinger, mark.eppinger@utsa.edu.

The Shiga-toxin-producing, non-sorbitol-fermenting, and β -glucuronidase-negative *Escherichia coli* (STEC) O157:H7 strain is thought to have evolved from an O55:H7-like progenitor (1, 2). The O157:H7 serotype is now the most common enterohemorrhagic *E. coli* (EHEC) found in North America. Although it causes disease in humans, *E. coli* O157:H7 asymptotically colonizes cattle, which is the major reservoir for this organism (3, 4). *E. coli* O157:H7 is distinguished from other serotypes by its genetically homogenous population structure and it exhibits a high degree of proteome conservation and a syntenic chromosomal backbone disrupted by interspersed phages (5, 6). The rapid emergence of *E. coli* O157:H7 from being an unknown strain in 1982 to being the dominant hemorrhagic serotype in the United States and the cause of widespread outbreaks of human food-borne illness highlights a need to critically evaluate the extent to which the genomic plasticity of this important enteric pathogen contributes to human disease severity and bovine niche adaptation. An estimated 15 to 20% of infected patients present with indications severe enough to require hospitalization. Symptoms may progress to hemolytic uremic syndrome, renal failure, hemorrhagic colitis, and central nervous system failure, with potentially lethal outcomes. Yet, little is known about the genomic diversity that exists among extant *E. coli* O157:H7 populations or how various genotypes of this pathogen relate to the development and severity of the clinical manifestations in infected patients.

Strains for sequencing were selected to represent phylogenetically diverse isolates within the O157:H7 lineage based on multiple typing assays (1, 5–7). Genomic DNA was subjected to next-generation Illumina HiSeq 2000 (300-bp insert size, 100-bp paired-end reads) or Illumina and 454 FLX/XLR (3-kb insert size) hybrid sequencing followed by assembly as described previously

TABLE 1 Genome features of the 26 EHEC strains

Strain	Genome size (Mbp)	Contigs	Accession no.
EC1734	5.42	54	AKMO00000000
EC1738	5.47	62	AKMN00000000
EC4013	5.35	463	AKMH00000000
EC4203	5.39	552	AKMB00000000
EC4402	5.44	785	AKMI00000000
EC4421	5.31	437	AKMF00000000
EC4422	5.31	393	AKMG00000000
EC4436	5.41	456	AKMK00000000
EC4439	5.4	534	AKMJ00000000
EC4448	5.49	566	AKMM00000000
FDA505	5.35	540	AKKW00000000
FDA517	5.52	533	AKKX00000000
FR1K1985	5.54	730	AKKZ00000000
FR1K1990	5.51	629	AKLA00000000
FR1K1996	5.43	580	AKKY00000000
PA3	5.36	508	AKLC00000000
PA5	5.35	554	AKLD00000000
PA9	5.42	427	AKLE00000000
TW07945	5.36	404	AKLU00000000
TW09098	5.48	486	AKLX00000000
TW09109	5.57	66	AKLY00000000
TW09195	5.46	1288	AKLZ00000000
TW10119	5.55	56	AKMA00000000
TW10246	5.45	47	AKLV00000000
TW11039	5.6	71	AKLW00000000
TW14301	5.29	485	AKME00000000

(8). Hybrid and Illumina assemblies were generated using the Celera and Velvet assemblers, respectively (9, 10), and all chromosomes and plasmids were manually annotated using the Manatee system (<http://manatee.sourceforge.net/>). Genome architectures and gene inventory were compared using Mauve and BLAST Score Ratio analysis (11, 12). Strains were further characterized

using phylogenomic assays of the genomic backbone and mobilome to investigate plasticity in architectures, prophage profiles, and single nucleotide polymorphisms (5, 6). Access to these high-quality genome sequences and their comparative analyses with relatives or other serotypes will facilitate additional comprehensive bioinformatics and phylogenetic analyses, thus expanding our understanding of the pathogenomic evolution of this major public health problem caused by these pathogens. These data should also prove useful for the development of a refined phylogenomic framework for forensic, diagnostic, and epidemiological studies in order to better prepare for future outbreaks and for better risk assessment in response to novel and emerging *E. coli* O157:H7 biotypes.

Nucleotide sequence accession numbers. The genome sequences are deposited in GenBank under the accession no. listed in Table 1. Cultures are available from the Biodefense and Emerging Infections Research Resources Repository (<http://www.beiresources.org/>).

ACKNOWLEDGMENTS

This project was funded with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract no. HHSN272200900007C.

REFERENCES

1. Leopold SR, Magrini V, Holt NJ, Shaikh N, Mardis ER, Cagno J, Ogura Y, Iguchi A, Hayashi T, Mellmann A, Karch H, Besser TE, Sawyer SA, Whittam TS, Tarr PI. 2009. A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 106:8713–8718.
2. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 106:17939–17944.
3. Lowe RM, Baines D, Selinger LB, Thomas JE, McAllister TA, Sharma R. 2009. *Escherichia coli* O157:H7 strain origin, lineage, and Shiga toxin 2 expression affect colonization of cattle. *Appl. Environ. Microbiol.* 75:5074–5081.
4. Beutin L. 2006. Emerging enterohaemorrhagic *Escherichia coli*, causes and effects of the rise of a human pathogen. *J. Vet. Med. B Infect. Dis. Vet. Public Health* 53:299–305.
5. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. 2011. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 108:20142–20147.
6. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. 2011. Genome signatures of *Escherichia coli* O157:H7 isolates from the bovine host reservoir. *Appl. Environ. Microbiol.* 77:2916–2925.
7. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS. 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 105:4868–4873.
8. Eppinger M, Worsham PL, Nikolich MP, Riley DR, Sebastian Y, Mou S, Achtman M, Lindler LE, Ravel J. 2010. Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* 192:1685–1699.
9. Huson DH, Reinert K, Kravitz SA, Remington KA, Delcher AL, Dew IM, Flanigan M, Halpern AL, Lai Z, Mobarrry CM, Sutton GG, Myers EW. 2001. Design of a compartmentalized shotgun assembler for the human genome. *Bioinformatics* 17(Suppl 1):S132–S139.
10. Zerbino DR. 2010. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Curr. Protoc. Bioinformatics* 11:11.5.
11. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* 25:2071–2073.
12. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.