Complete Genome Sequences of Curli-Negative and Curli-Positive Isolates of Foodborne *Escherichia coli* O157:H7 Strain 86-24

Vijay K. Sharma,[a] Darrell O. Bayles,[b] David P. Alt,[b] Torey Looft[a]

Food Safety and Enteric Pathogens Research Unit, National Animal Disease Center, ARS-USDA, Ames, Iowa, USA[a]; Infectious Bacterial Diseases Research Unit, National Animal Disease Center, ARS-USDA, Ames, Iowa, USA[b]

Escherichia coli O157:H7 strain 86-24 does not produce curli fimbriae, but gives rise to curli-positive isolates at a variable frequency. Here, we report the complete genome sequences of curli-negative and curli-positive isolates of strain 86-24.

Address correspondence to Vijay K. Sharma, vijay.sharma@ars.usda.gov.

*E*scherichia coli O157:H7 (O157) is an organism that causes foodborne diarrheal illnesses of varied spectra. The O157 strain EDL933 (1) genome has been used as a reference for studying other O157 strains. The genomes of O157 isolates (1, 2) are larger than the *E. coli* K12 strain MG1655 genome, a fact attributed to lateral gene transfer. This genome expansion correlates with broader host range and more robust environmental survival of O157. Population heterogeneity in O157 confers genotypic variants with advantages in colonization and survival in diverse niches (3–5). Here, we describe the complete genome sequences of a curli-negative (NADC 6564) and a curli-positive (NADC 6565) isolate of strain 86-24 linked to the 1986 Walla Walla (WA) outbreak (6, 7), and identify genetic differences responsible for the curli-positive phenotype.

Genomic DNA, prepared from midlogarithmic-phase bacterial cultures (Qiagen Genomic-tip G100), was used for generating 20-kb large-insert libraries and sequenced using the PacBio RS II platform (Yale Center for Genomic Analysis, West Haven, CT). The PacBio reads were assembled into contigs (PacBio SMRT Analysis 2.3.0 and CANU v. 1.3) (8). For Illumina MiSeq sequencing, genomic DNA libraries were prepared using the TruSeq Nano DNA Library Preparation kit. The PacBio assemblies were circularized using the AMOS v. 3.1.0 Minimus2 (9) assembler and the genomes subjected to one round of error correction and polishing with the Quiver algorithm by remapping the PacBio reads to their respective circularized genomes as the references. These assemblies were further error corrected and polished to their final version by mapping the Illumina reads to their respective genomes and performing consensus recalling and corrections using the Broad Institute's Pilon v 1.18 (10). Both genomes were broken to begin at the *dnaA* gene. The fully closed genomes and associated plasmids were annotated by the NCBI Prokaryotic Genome Annotation Pipeline.

The NADC 6564 genome consisted of a chromosome of 5,466,770 bp and a plasmid of 92,691 bp, contained a total of 5,676 predicted genes (5,578 for the chromosome and 98 for the plasmid), 5,542 coding sequences (CDS) including pseudogenes and 5,399 genes. The NADC 6565 genome comprised a chromosome of 5,467,107 bp and a plasmid of 92,690 bp, contained a total of 5,678 predicted genes (5,580 for the chromosome and 98 for the plasmid), 5,544 CDS including pseudogenes and 5,404 genes. Both genomes contained 103 tRNA, eight 5S RNA, seven each of 16S and 23S RNA, and nine noncoding RNA (ncRNA) genes and had G+C content of 50% and 48% for the chromosomes and plasmids, respectively. There were 143 predicted pseudogenes in NADC 6564 and 140 in NADC 6565. The NADC 6564 and NADC 6565 genomes aligned (Artemis Comparison Tool) (11) with the EDL933 genome except for some yet-uncharacterized inversions and indels. A tandem duplication of a 5-bp sequence was identified in *rcsB* of the *rcsBD* signal transduction system (12) and was determined to confer the curli-positive phenotype.

Accession number(s). The genome sequences have been deposited at GenBank under the GenBank accession numbers CP017251 and CP017252 (NADC 6564) and CP017249 and CP017250 (NADC 6565).

REFERENCES

1. **Perna NT, Plunkett G III, 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. http://dx.doi.org/10.1038/35054089.
2. **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. http://dx.doi.org/10.1093/dnares/8.1.11.

3. **Carter MQ, Louie JW, Huynh S, Parker CT.** 2014. Natural *rpoS* mutations contribute to population heterogeneity in *Escherichia coli* O157:H7 strains linked to the 2006 U.S. spinach-associated outbreak. Food Microbiol **44:**108–118. http://dx.doi.org/10.1016/j.fm.2014.05.021.

4. **Carter MQ, Parker CT, Louie JW, Huynh S, Fagerquist CK, Mandrell RE.** 2012. RcsB contributes to the distinct stress fitness among *Escherichia coli* O157:H7 curli variants of the 1993 hamburger-associated outbreak strains. Appl Environ Microbiol **78:**7706–7719. http://dx.doi.org/10.1128/AEM.02157-12.

5. **van Hoek AH, Aarts HJ, Bouw E, van Overbeek WM, Franz E.** 2013. The role of *rpoS* in *Escherichia coli* O157 manure-amended soil survival and distribution of allelic variations among bovine, food and clinical isolates. FEMS Microbiol Lett **338:**18–23. http://dx.doi.org/10.1111/1574-6968.12024.

6. **Griffin PM, Ostroff SM, Tauxe RV, Greene KD, Wells JG, Lewis JH, Blake PA.** 1988. Illnesses associated with *Escherichia coli* O157:H7 infections. A broad clinical spectrum. Ann Intern Med **109:**705–712. http://dx.doi.org/10.7326/0003-4819-109-9-705.

7. **Melton-Celsa AR, Rogers JE, Schmitt CK, Darnell SC, O'Brien AD.** 1998. Virulence of Shiga toxin-producing *Escherichia coli* (STEC) in orally infected mice correlates with the type of toxin produced by the infecting strain. Jpn J Med Sci Biol **51**(Suppl):S108–S114. http://dx.doi.org/10.7883/yoken1952.51.Supplement1_S108.

8. **Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM.** 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol **33:**623–630. http://dx.doi.org/10.1038/nbt.3238.

9. **Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M.** 2011. Next generation sequence assembly with AMOS. Curr Protoc Bioinformatics Chapter 11:Unit 11.18.

10. **Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM.** 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9:**e112963. http://dx.doi.org/10.1371/journal.pone.0112963.

11. **Carver TJ, RK, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: theol artemis comparison tool. Bioinformatics (Oxford, England) **21:**3422–3423.

12. **Vianney A, Jubelin G, Renault S, Dorel C, Lejeune P, Lazzaroni JC.** 2005. *Escherichia coli tol* and *rcs* genes participate in the complex network affecting curli synthesis. Microbiology **151:**2487–2497. http://dx.doi.org/10.1099/mic.0.27913-0.