Draft Genome Sequence of *Escherichia coli* ARS-CC7049, a Sequence Type 38 Strain Isolated from Dry Cow Feces on a Commercial Dairy Operation

Seon Woo Kim,[a] Jo Ann S. Van Kessel,[a] (ID) Bradd J. Haley[a]

[a]Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland, USA

ABSTRACT *Escherichia coli* is a diverse species of commensal and pathogenic strains, of which some can cause extraintestinal infections, such as sequence type 38 (ST38) strains. Here, we report the genome sequence of *E. coli* ARS-CC7049, an ST38 strain that was isolated from a composite fecal sample on a dairy farm.

Annually there are approximately 8 to 10 million diagnosed urinary tract infections (UTI) in the United States. Most of these infections are caused by a relatively small number of extraintestinal pathogenic *Escherichia coli* (ExPEC) sequence types (STs), with ST38 being one of the leading disease-causing ExPEC STs globally (1). In a larger study conducted between 2013 and 2015, *E. coli* was isolated from composite fecal samples collected from preweaned and postweaned calves and dry and lactating dairy cows on 80 commercial dairy operations in Pennsylvania (2). Here, we present the genome sequence of an *E. coli* ST38 isolate collected from dry cow feces in 2015.

Six fecal samples were collected from the floor of a dry cow pen on a commercial dairy farm in Pennsylvania and combined (approximately 120 g of combined feces). The sample was transported on ice to the laboratory, and approximately 5 g was homogenized in 45 mL of buffered peptone water which was subsequently streaked onto CHROMagar EC agar (Paris, France). Colonies were confirmed as *E. coli* after incubation for 24 h at 37°C on Simmons citrate agar (BD Diagnostics), MacConkey agar (Remel, Lenexa, KS), MacConkey sorbitol agar, and Luria-Bertani agar. DNA was extracted from a purified colony grown overnight in Luria-Bertani broth using a QIAcube with a QIAmp DNA mini kit (Qiagen, Hilden, Germany). Paired-end libraries (2×150 bp) were constructed using the Nextera XT library prep kit (Illumina, La Jolla, CA), which were then sequenced using a high-output v 1.0 flow cell on a NextSeq 500 platform (Illumina). Sequence data were demultiplexed using the BCL2FastQ program, and PhiX reads were removed using DeconSeq (PhiX NCBI accession NC_001422) (3). Reads were further trimmed using Trimmomatic v 0.35 (LEADING: 20 TRAILING:20 SLIDINGWINDOW4:20 MINLEN:36) (4) and assessed using FastQC v 0.11.6. The genome sequence was assembled *de novo* using SPAdes v 3.8.0 with the –careful option (5). Genome quality was assessed using QUAST (6).

The *E. coli* ARS-CC7049 genome was assembled into 135 contigs with a genome size of 4,877,373 bp and a GC content of 50.7%. The total number of reads was 1,705,268, the coverage was $45\times$, and the N_{50} value was 141,319 bp. Curated sequencing reads and the assembled genome were analyzed using the Center for Genomic Epidemiology Web server with parameters set to default (http://www.genomicepidemiology.org/services/). Using the MLST 2.0 program, the isolate was identified as ST38 of the ST38 Cplx within phylotype D (7). Using ResFinder 4.1, antimicrobial resistance genes or resistance-conferring single nucleotide polymorphisms (SNPs) were not identified among the reads or the assembled genome (8). No sequences with plasmid replicons were identified among the reads or the assembled genome using PlasmidFinder (9). In the assembled genome, *kps* genes of the K1 capsule

and the *chu* heme uptake genes (*chuA*, *eilA*, *gad*, *iss*, *kpsE*, *ompT*, and *terC*) were identified, indicating the presence of some virulence factors known to play a role in extraintestinal pathogenic *E. coli* infections (10).

Data availability. This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession number JAIQGV000000000 and BioProject number PRJNA761377. The version described in this paper is the first version, JAIQGV000000000.1. The raw sequence data are available in the Sequence Read Archive under the accession number SRR15793117.

REFERENCES

1. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. 2019. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. Clin Microbiol Rev 32:e00135-18. https://doi.org/10.1128/CMR.00135-18.

2. Cao H, Pradhan AK, Karns JS, Hovingh E, Wolfgang DR, Vinyard BT, Kim SW, Salaheen S, Haley BJ, Van Kessel JS. 2019. Age-associated distribution of antimicrobial-resistant *Salmonella enterica* and *Escherichia coli* isolated from dairy herds in Pennsylvania, 2013–2015. Foodborne Pathog Dis 16:60–67. https://doi.org/10.1089/fpd.2018.2519.

3. Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288. https://doi.org/10.1371/journal.pone.0017288.

4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.

6. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

7. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.

8. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500. https://doi.org/10.1093/jac/dkaa345.

9. Carattoli A, Hasman H. 2020. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol 2075:285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.

10. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, Choroszy-Krol I. 2019. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. Gut Pathog 11:10. https://doi.org/10.1186/s13099-019-0290-0.