AMERICAN SOCIETY FOR MICROBIOLOGY
Microbiology Resource Announcements
GENOME SEQUENCES

Escherichia coli Serotype O157:H7 PA20R2R Complete Genome Sequence

Gaylen A. Uhlich,[a] George C. Paoli,[a] Siddhartha Kanrar[a]

[a]Molecular Characterization of Foodborne Pathogens Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, Wyndmoor, Pennsylvania, USA

ABSTRACT *Escherichia coli* serotype O157:H7 strain 20R2R is a derivative of clinical isolate PA20. Prophage excision from the coding region of a PA20 transcription factor restored RpoS-dependent biofilm formation in 20R2R, providing a model for O157:H7 stress adaptation when transitioning between clinical and environmental settings. We report here the complete 20R2R genome sequence.

The survival of Shiga toxin-producing *Escherichia coli* moving from animals to the environment is enhanced by stress resistance genes. *E. coli* stress responses in stationary growth phase depend on the RpoS sigma factor and protect against conditions such as pH and temperature extremes, oxidative stress, osmotic stress, and nutrient deprivation (1). RpoS also controls the production of biofilms by driving the expression of *csgD*, a transcriptional regulator essential for producing biofilm components, curli fimbriae, and cellulose (2). The MlrA transcription factor is also required for maximal RpoS-dependent expression of *csgD* (3).

In contrast to the environment, curli production in the human host reduces, rather than enhances, O157:H7 survival by triggering immune clearance and inflammation (4). Clinically adapted O157:H7 represses curli using two major genome modifications that reduce *csgD* expression: (i) function-reducing *rpoS* mutations and (ii) prophage insertions in *mlrA* (5). Unlike *rpoS* mutations, *csgD* repression by interrupting *mlrA* reduces biofilm formation but spares other RpoS-dependent stress responses. Although clinical strains can carry both modifications, strains with only prophage insertions are clearly better adapted for transitioning back to the environment (5). We previously reported a strong biofilm-producing strain, 20R2R, isolated during agar passage of strain PA20, a clinical isolate with wild-type *rpoS* and a prophage in *mlrA* (6, 7). Partial 20R2R characterization revealed excision of the prophage from the parental *mlrA* making strain 20R2R a valuable strain for studying genetic responses of O157:H7 transitioning between the host and environment (8). To that end, we derived the complete 20R2R genome sequence using long-read Pacific Biosciences technology.

A single colony of *E. coli* 20R2R was grown overnight (~16 h) in Luria-Bertani medium at 37°C with shaking (180 rpm). The cells from 5 ml of the overnight culture were harvested by centrifugation, and genomic DNA was prepared using a Qiagen Genomic-tip 100/G kit. The genomic sequence was determined by the University of Delaware Sequencing and Genotyping Center (Newark, DE) using single-molecule real-time (SMRT) sequencing technology. The gDNA was sheared using g-TUBE (Covaris, Inc., Woburn, MA) and fractionated using the BluePippin size selection system (Sage, Beverly, MA) targeting 20-kb fragments. Template DNA was prepared using the SMRTbell template preparation kit (Pacific Biosciences, Menlo Park, CA), and sequencing was conducted using the PacBio RS II SMRT DNA sequencing system (Pacific Biosciences). All tools were run with default parameters unless otherwise specified. Sequencing using three SMRT cells yielded 450,876 raw reads. After quality control and raw read filtering, the sequencing runs yielded 203,605 reads with a mean length of 7,440 bp and an N_{50} read length of 10,315 bp. The

Citation Uhlich GA, Paoli GC, Kanrar S. 2020. *Escherichia coli* serotype O157:H7 PA20R2R complete genome sequence. Microbiol Resour Announc 9:e01143-20. https://doi.org/10.1128/MRA.01143-20.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Gaylen A. Uhlich, gaylen.uhlich@usda.gov.

Received 6 October 2020
Accepted 23 November 2020
Published 10 December 2020

filtered reads were assembled *de novo* using the Hierarchical Genome Assembly Process (version 3) yielding 4 contigs. Using the genomes of *E. coli* O157:H7 strains Sakai (NC_002695.2) and PA20 (CP017669.1) as a reference, we reordered the contigs into a single chromosome (5,488,557 bp, 50.5% GC content) and plasmid (pO157R2R, 104,844 bp, 47.1% GC content) without gaps using mauve contig mover (Mauve software, version 20150226 build [10]c 2003–2015) (9)] with a 256× genome coverage.

Genome sequences were annotated with RAST 2.0 (Rapid Annotations using Subsystems Technology) using default parameters (10). The software identified 5,619 chromosomal coding DNA sequences (CDSs) and 155 plasmid CDSs. The tRNAscan-SE 2.0 software predicted 112 tRNA genes in the 20R2R chromosome (11). No mutations or genes conferring antimicrobial resistance were identified in the 20R2R chromosome or plasmid using ResFinder-3.2 (12). Shiga toxin genes stx_{2a} and stx_{2b} were identified on the 20R2R chromosome with VirulenceFinder 2.0 (13) but stx_{1a} and stx_{1b} genes, which are located on the prophage inserted in *mlrA* of parent strain PA20, were not found (13). Virulence factors *ehxA*, *espP*, *etpD*, *katP*, and *tox* were identified on the pO157R2R plasmid using VirulenceFinder 2.0 (13). There were 20 chromosomal regions encoding prophage DNA, but no prophage, was identified in plasmid sequences using PHAST software (14). Strain 20R2R *mlrA* encoded no prophage, and the wild-type coding sequence was restored. Moreover, the 993-bp *rpoS* sequence was identical to the PA20 and Sakai reference strains, making 20R2R ideal for investigating stress-related genetic changes associated with the transition of O157:H7 between the host and environment.

Data availability. The genome sequence and annotation data for *Escherichia coli* O157:H7 strain 20R2R were deposited in DDBJ/GenBank under BioProject PRJNA294158, BioSample SAMN16202109, and the accession numbers CP062160 (chromosome) and CP062161 (plasmid).

ACKNOWLEDGMENT

This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture (USDA).

REFERENCES

1. Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annu Rev Microbiol 65:189–213. https://doi.org/10.1146/annurev-micro-090110-102946.
2. Olsén A, Arnqvist AA, Hammar M, Sukupolvi S, Normark S. 1993. The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csgA*, the subunit of fibronectin-binding curli in *Escherichia coli*. Mol Microbiol 7:523–536. https://doi.org/10.1111/j.1365-2958.1993.tb01143.x.
3. Brown PK, Dozois CM, Nickerson CA, Zuppardo A, Terlonge J, Curtiss R. III. 2001. MlrA, a novel regulator of curli (AgF) and extracellular matrix synthesis by *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. Mol Microbiol 41:349–363. https://doi.org/10.1046/j.1365-2958.2001.02529.x.
4. Richter AM, Povolotsky TL, Wieler LH, Hengge R. 2014. Cyclic-di-GMP signaling and biofilm-related properties of the Shiga toxin-producing 2011 German outbreak *Escherichia coli* O104:H4. EMBO Mol Med 6:1622–1637. https://doi.org/10.15252/emmm.201404309.
5. Uhlich GA, Chen CY, Cottrell BJ, Hofmann CS, Dudley EG, Strobaugh TP, Jr, Nguyen LH. 2013. Phage insertion in *mlrA* and variations in *rpoS* limit curli expression and biofilm formation in *Escherichia coli* serotype O157:H7. Microbiology (Reading) 159:1586–1596. https://doi.org/10.1099/mic.0.066118-0.
6. Chen CY, Nguyen LH, Cottrell BJ, Irwin PL, Uhlich GA. 2016. Multiple mechanisms responsible for strong Congo red-binding variants of *Escherichia coli* O157:H7 strains. FEMS Pathog Dis 74:ftv123. https://doi.org/10.1093/femspd/ftv123.
7. Uhlich GA, Paoli GC, Zhang X, Dudley EG, Figler HM, Cottrell BJ, Andreozzi E. 2017. Whole-genome sequence of *Escherichia coli* serotype O157:H7 strain PA20. Genome Announc 5:e01460-16. https://doi.org/10.1128/genomeA.01460-16.
8. Uhlich GA, Chen CY, Cottrell BJ, Hofmann CS, Yan X, Nguyen LH. 2016. Stx_1 prophage excision in *Escherichia coli* strain PA20 confers strong curli and biofilm formation by restoring native *mlrA*. FEMS Microbiol Lett 363: fnw123. https://doi.org/10.1093/femsle/fnw123.
9. 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. https://doi.org/10.1093/bioinformatics/btp356.
10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
11. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
12. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
13. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 52:1501–1510. https://doi.org/10.1128/JCM.03617-13.
14. Zhou Y, Liang Y, Lynch K, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.