**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 N50 read length of 10,315 bp. The complete genome sequence was deposited in GenBank (accession no. MCAI00000000). The assembly contained 5,341,334 bp in 10 scaffolds. The average coverage was 100×, with a minimum of 20× and a maximum of 40×.

filtered reads were assembled de novo using the Hierarchical Genome Assembly Process (version 3) yielding 4 contigs. Using the genomes of *Escherichia 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 [10c 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 confering antimicrobial resistance were identified in the 20R2R chromosome or plasmid using ResFinder-3.2 [12]. Shiga toxin genes *stxA* and *stxB* were identified on the 20R2R chromosome with VirulenceFinder 2.0 (13) but *stxA* and *stxB* 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 with 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**


