Complete Closed Genome Sequence of the Extremely Heat-Resistant Strain *Escherichia coli* AW1.7

Manita Guragain,ᵃ James L. Bono,ᵃ Joseph M. Bosilevacᵃ

ᵃU.S. Department of Agriculture, Agricultural Research Service, Roman L. Hruska U.S. Meat Animal Research Center, Clay Center, Nebraska, USA

ABSTRACT *Escherichia coli* isolate AW1.7 is an extremely heat-resistant bacterium and has been widely used as a reference strain in extreme heat resistance studies for almost a decade. Here, we report its complete closed genome sequence.

AW1.7 is an extremely heat-resistant *Escherichia coli* strain recovered from an unpasteurized dressed beef carcass at a Canadian beef slaughter plant in 2001, isolated on SD-39 agar (42°C, 18 h) (1), passaged four times, and archived at −80°C. Here, we report the sequence of its complete closed genome.

An overnight culture of AW1.7 in Luria-Bertani (LB) broth was diluted 1:100 in LB broth and then shaken (150 rpm) at 37°C for 3 h. Genomic DNA (gDNA) was extracted using a Qiagen Genomic-tip 100/G kit (Valencia, CA) and sheared to an average size of 20 kb using a Covaris g-TUBE (Woburn, MA). A sequencing library was prepared using the SMRTbell template prep kit v1.0 (Pacific Biosciences, Menlo Park, CA) and size selected for 20-kb fragments using the BluePippin high pass Plus cassette (Sage Science, Inc., Beverly, MA). PacBio sequencing using the Sequel system (Menlo Park, CA) with v3.0 chemistry produced 21,618 subreads (331,276,790 total subread bases) with an average subread length of 15,220 bp. For the Illumina library, gDNA was sheared to an average size of 350 bp using a Covaris microtube; the library was prepared using the TruSeq DNA PCR-free library preparation (LP) kit (Illumina, San Diego, CA). An average genome coverage of 56× was obtained, with 847,900 paired-end reads (16,994,571 total bases) generated on a MiSeq device using a v2 kit (300 cycles) (Illumina). Quality trimming and adapter clipping of the raw reads was conducted using Trimmomatic v0.39 with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 MINLEN:36 (2). Default settings were used for all software unless otherwise noted. After assembly with the Microbial Assembly application (PacBio SMRT analysis pipeline v8.0), there were 3 contigs with an average coverage of 62×. One contig was circularized using the Microbial Assembly application and was determined to be the chromosome, while the remaining two were determined to be plasmids. The assembled genome sequence was analyzed using Geneious Prime v2020.1.2 (Biomatters Ltd., New Zealand). Further mapping of the circularized chromosome with the error-corrected PacBio and Illumina reads using the Geneious mapper confirmed the assembler's results. Iterative mapping using the error-corrected PacBio and Illumina reads produced overlapping sequence ends on the two contigs as visualized using LastZ v1.02.00 (3, 4). This allowed the overlapping sequences to be trimmed and the contigs concatenated to produce closed circular plasmids. Position 1 of the chromosome was set to the origin of replication using Ori-Finder v2.0 (5). Using the Geneious mapper, the error-corrected PacBio and Illumina reads produced an overall even coverage of the mapped reads with no single-nucleotide polymorphism (SNP) differences using Find variation/SNPs. Four additional plasmids were identified using the unused reads with the Geneious assembler. The replicon regions were identified by *in silico* analysis of the plasmid sequences using PlasmidFinder v2.1 (6). Finally, the closed chromosome and plasmids were polished using the Illumina reads with Pilon v1.23 (7).

TABLE 1 Characteristics of contigs from *E. coli* AW1.7 genome assembly

Contig name	Sequence length (bp)	% GC	GenBank accession no.[a]
Chromosome	4,947,577	50.7	CP072539
pAW1.7_1	103,129	50.1	CP072540
pAW1.7_2	66,909	54.2	CP072541
pAW1.7_3	4,842	51.5	CP072542
pAW1.7_4	1,934	51.3	CP072543
pAW1.7_5	3,211	49.1	CP072544
pAW1.7_6	2,454	48.8	CP072545

[a] Genome sequence annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.1 (8, 9).

The polished genome sequence was mapped against the shotgun sequences from the source lab to confirm the identity of the isolate. The characteristics of the chromosome and plasmids can be found in Table 1.

Data availability. The GenBank accession numbers for the PGAP-annotated genome sequence are provided in Table 1. The BioProject accession number for the raw sequence reads is PRJNA663878.

REFERENCES

1. Aslam M, Greer GG, Nattress FM, Gill CO, McMullen LM. 2004. Genotypic analysis of *Escherichia coli* recovered from product and equipment at a beef-packing plant. J Appl Microbiol 97:78–86. https://doi.org/10.1111/j.1365-2672.2004.02277.x.
2. 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.
3. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2004. Correction for Schwartz et al., "Human-mouse alignments with BLASTZ." Genome Res 14:786.
4. Harris RS. 2007. Improved pairwise alignment of genomic DNA. PhD thesis. Pennsylvania State University, State College, PA.
5. Luo H, Zhang C-T, Gao F. 2014. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front Microbiol 5:482. https://doi.org/10.3389/fmicb.2014.00482.
6. 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.
7. 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. https://doi.org/10.1371/journal.pone.0112963.
8. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.
9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.