Shiga toxin-producing *Escherichia coli* causes bloody diarrhea and hemolytic-uremic syndrome and serious outbreaks worldwide. Here, we report the draft genome sequence of *E. coli* NCCP15657 isolated from a patient. The genome has virulence genes, many in the locus of enterocyte effacement (LEE) island, encoding a metalloprotease, the Shiga toxin, and constituents of type III secretion.

**Shiga toxin-producing *Escherichia coli* (STEC) strains are the major cause of bacterial foodborne outbreaks worldwide** (9). They cause a disease involving diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome (12). Most of these symptoms are caused by serotypes O157, O26, O103, and O111 (7), and their genomes have been sequenced and analyzed (9). Recently, there was a serious outbreak caused by serotype O104:H4 that occurred in Germany (8). In Korea, STEC infections have been frequently reported since 2002 (2, 4, 5). However, there are no genome sequences available for STEC strains isolated from Korean patients. NCCP15657 is an STEC originally isolated from a fecal sample of a female patient suffering from bloody diarrhea and abdominal pain in 2005. A verotoxin from the strain was detected by PCR and reverse passive latex agglutination methods. The genome sequence of NCCP15657 was determined by a massive parallel sequencing technology. 2.1 Gb and 2.8 Gb of paired-end sequences (∼420/558-fold genome coverage) produced from 600-bp and 3-kb genomic libraries with in-house Illumina/Solexa Genome Analyzer IIX were assembled with CLC Genomics Workbench (CLC bio, Inc.). A total of 161 contigs were ordered, and gaps were closed to 113 contigs by SSPACE (3) and IMAGE (11). Also, high-depth contigs that contain multicyclop rRNA operons and Rh elements were resolved into several contigs by mapping to the *E. coli* IA11 genome. Total contigs from *de novo* assembly and reference mapping were merged with Pherd/Phrap/Consed. A total of 5,195 coding sequences were predicted by Glimmer and GeneMarkS and annotated using the information from MicroScope, RAST (1), Pfam, and TIGRfam. The average nucleotide identity (ANI) value was calculated with JSpecies (10).

The final assembly consists of 93 contigs of 5,015,691 bp (50.6% of G+C content). A total of 4,303 protein-coding sequences were assigned functions. The average nucleotide identity values between *E. coli* NCCP15657 and the B1 phylogroup ranged from 98.74 to 99.42%. The genome is most similar to that of *E. coli* IA11. Phylogenetic analysis of the sequences encoding O-antigen polymerase/flipase and flagellin indicate that NCCP15657 belongs to the O123:H19 serotype.

Unlike *E. coli* IA11, *E. coli* NCCP15657 has many virulence-related factors and genes encoding the Shiga toxin. The genome contains virulent genes encoding a StcE metalloprotease, which is positioned next to the locus of a type II secretion system, and a locus of enterocyte effacement (LEE) that includes intimin, Tir, a type III secretion system, and regulator and effector proteins. Stx genes are homologs of type 1 toxins, and the intimin gene is an ε type. These results correspond to known features of the O123 serotype. Interestingly, LEE is flanked by the selC tRNA locus. This structure is similar to that of O157 unlike other non-O157 strains (9). This is the first genome sequence of serotype O123, and its availability will provide a better-defined genetic background for understanding the pathogenic mechanism to compare *E. coli* strains and benefit disease control efforts.

**Nucleotide sequence accession number.** The draft genome sequence was deposited in GenBank under accession number AJLU00000000. The sequence is also available from the Genome Encyclopedia of Microbes (GEM) (http://www.gem.re.kr) (6).

**ACKNOWLEDGMENTS**

This project was conducted in part as the Microbial Genome Sequencing and Analysis course at the University of Science and Technology in the fall of 2011. We are grateful to Seong Keun Kim and students in the class for critical comments on the manuscript and members of GEM and the KRIBB sequencing team for technical help.

This work was supported by the National Research Foundation of Korea (2011-0017670 to J.F.K.) of the Ministry of Education, Science and Technology, the KRIBB Research Initiative Program, and the National Institute of Health (NIH 4800-4845-300 to S.H.C.), Republic of Korea.

**REFERENCES**