

Genome Sequence of the Shiga Toxin-Producing *Escherichia coli* Strain NCCP15657

Byung Kwon Kim,^a Geun Cheol Song,^{a,b} Gun Hyong Hong,^{a,b} Won-Keun Seong,^c Seon-Young Kim,^{d,e} Haeyoung Jeong,^{a,b} Sung Gyun Kang,^{f,g} Soon-Kyeong Kwon,^{a,b} Choong Hoon Lee,^a Ju Yeon Song,^a Dong Su Yu,^a Mi-Sun Park,^c Seung-Hak Cho,^c and Jihyun F. Kim^{a,h}

Systems and Synthetic Biology Research Center,^a and BioMedical Genomics Research Center,^d Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea; Biosystems and Bioengineering Program,^b Department of Functional Genomics,^e and Department of Marine Biotechnology,^f University of Science and Technology, Daejeon, Republic of Korea; Division of Enteric Bacterial Infections, Center for Infectious Diseases, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Republic of Korea^g; Korea Ocean Research and Development Institute, Ansan, Republic of Korea^g; and Department of Systems Biology, Yonsei University, Seoul, Republic of Korea^h

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 food-borne 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 multicopy rRNA operons and Rhs element repeats were resolved into several contigs by mapping to the *E. coli* IAI1 genome. Total contigs from *de novo* assembly and reference mapping were merged with Phred/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 4,033 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 commensal *E. coli* IAI1. Phylogenetic analysis of the sequences encoding O-antigen polymerase/flippase and flagellin indicate that NCCP15657 belongs to the O123:H19 serotype.

Unlike *E. coli* IAI1, *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 AJLU000000000. 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

1. Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi:10.1186/1471-2164-9-75.

Received 18 April 2012 Accepted 30 April 2012

Address correspondence to Jihyun F. Kim, jfk1@yonsei.ac.kr, or Seung-Hak Cho, skcho38@korea.kr.

B.K.K., G.C.S., and G.H.H. contributed equally to this work.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00654-12

2. Bae WK, et al. 2006. A case of hemolytic uremic syndrome caused by *Escherichia coli* O104:H4. *Yonsei Med. J.* 47:437–439.
3. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579.
4. Cho SH, et al. 2006. Surveillance of bacterial pathogens associated with acute diarrheal disease in the Republic of Korea during one year, 2003. *J. Microbiol.* 44:327–335.
5. Cho YH, et al. 2002. A case of hemolytic uremic syndrome caused by *Escherichia coli* O8: case report. *Korean J. Gastrointest. Endosc.* 25:213–216.
6. Jeong H, Yoon SH, Yu DS, Oh TK, Kim JF. 2008. Recent progress of microbial genome projects in Korea. *Biotechnol. J.* 3:601–611.
7. Madic J, et al. 2011. Detection of Shiga toxin-producing *Escherichia coli* serotypes O26:H11, O103:H2, O111:H8, O145:H28, and O157:H7 in raw-milk cheeses by using multiplex real-time PCR. *Appl. Environ. Microbiol.* 77:2035–2041.
8. Mellmann A, et al. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 6:e22751. doi:10.1371/journal.pone.0022751.
9. Ogura Y, et al. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 106:17939–17944.
10. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106: 19126–19131.
11. Tsai IJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 11:R41. doi:10.1186/gb-2010-11-4-r41.
12. Whittam TS, et al. 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* 61:1619–1629.