Draft Genome Sequence of Enterohemorrhagic *Escherichia coli* O157 NCCP15739, Isolated in the Republic of Korea

Taesoo Kwon,[a] Seung-Hak Cho[b]

Division of Biosafety Evaluation and Control, Korea National Institute of Health, Osong, Republic of Korea[a]; Division of Enteric Bacterial Infections, Center for Infectious Diseases, Korea National Institute of Health, Osong, Republic of Korea[b]

Enterohemorrhagic *Escherichia coli* (EHEC) is the main cause of the recent outbreaks of diarrhea, hemolytic-uremic syndrome (HUS), and hemorrhagic colitis worldwide. Herein, we present the draft genome sequence of the NCCP15739 isolate from a patient in the Republic of Korea.

Address correspondence to Seung-Hak Cho, skcho38@korea.kr.

Although most *Escherichia coli* strains in the environment or in the intestinal tract of mammals are nonpathogenic, the following serotypes can cause severe foodborne outbreaks: enterohemorrhagic *Escherichia coli* (EHEC) O26, O103, O104, O111, O118, O121, O145, and O157 (1). The Shiga toxin-producing *E. coli* strain O157:H7 is one of the strains responsible for food poisoning. Previous studies have shown the genome information of *E. coli* O157:H7 EDL933 (2), Sakai (3), TW14349 (4), and EC4115 (5).

Genomic DNA of the NCCP15739 strain isolated from human feces was purified using the whole-genome shotgun protocol for whole-genome shotgun sequencing. The paired-end sequencing library had 500 bp, and the DNA sequencing platform used was Illumina-HiSeq 2000. From whole-genome shotgun sequencing, we generated 555,640,000 reads with 74-fold coverage. The reads were assembled into 479 contigs in 162 scaffolds ($N_{50} = 146,745$) with the SOAPdenovo program (version 1.05) (6), and we determined that the NCCP15739 strain has an approximately 5.3-Mb genome with a 50% G+C content from the assembly. Using Rapid Annotation using Subsystem Technology (RAST) (7) analysis, we identified 5,190 putative open reading frames (ORFs) and 36 RNA genes and can functionally annotate 4,181 (80.6%) ORFs. Functional comparison of the genome sequences available on the RAST server revealed that the *E. coli* O157:H7 EDL933 strain was similar to NCCP15739 (score 520). Comparison of NCCP15739 and EDL933 based on sequence similarity demonstrated that most of the functional genes of NCCP15739 were conserved in *E. coli* O157:H7 EDL933, but 723 genes were unique. The unique genes in NCCP15739, such as the protein secretion system, transcriptional regulator, resistance to antibiotics and toxic compounds, flagella motility, and prophage cluster, explain that phenotypic difference originated due to environmental adaptation. The NCCP15739 strain is related to the enterohemorrhagic human pathogen and produces the Shiga and Shiga toxin-like proteins. It also possesses characteristics such as intimin, invasion, and adhesion and can produce hemolysins in the genome. The genome includes plasmid-borne sequences similar to the plasmid sequence, pSD107, of *Salmonella* enteric substrains. This study will provide information that will enhance the understanding of the O157:H7 type of EHEC, which in turn will aid in the study of the evolution of pathogenic mechanisms through a broader comparative genomic approach of EHEC genomes.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no. ASHA00000000.

ACKNOWLEDGMENT

REFERENCES

1. **Tzschoppe M, Martin A, Beutin L**. 2012. A rapid procedure for the detection and isolation of enterohaemorrhagic *Escherichia coli* (EHEC) serogroup O26, O103, O111, O118, O121, O145 and O157 strains and the aggregative EHEC O104:H4 strain from ready-to-eat vegetables. Int J Food Microbiol **152**:19–30. http://dx.doi.org/10.1016/j.ijfoodmicro.2011.10.009.

2. **Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR**. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**:529–533. http://dx.doi.org/10.1038/35054089.

3. **Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H**. 2001. Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res **8**:11–22. http://dx.doi.org/10.1093/dnares/8.1.11.

4. **Kulasekara BR, Jacobs M, Zhou Y, Wu Z, Sims E, Saenphimmachak C, Rohmer L, Ritchie JM, Radey M, McKevitt M, Freeman TL, Hayden H, Haugen E, Gillett W, Fong C, Chang J, Beskhlebnaya V, Waldor MK, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI**. 2009. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. Infect Immun **77**:3713–3721. http://dx.doi.org/10.1128/IAI.00198-09.

5. **Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA**. 2011.

Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. Proc Natl Acad Sci U S A **108**:20142–20147. http://dx.doi.org/10.1073/pnas.1107176108.

6. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J.** 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res **20**:265–272. http://dx.doi.org/10.1101/gr.097261.109.

7. **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. http://dx.doi.org/10.1186/1471-2164-9-75.