JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of Enterohemorrhagic *Escherichia coli* NCCP15658

Ju Yeon Song,[a] Ran Hee Yoo,[b,c] Song Yee Jang,[a,c] Won-Keun Seong,[d] Seon-Young Kim,[e,f] Haeyoung Jeong,[a,c] Sung Gyun Kang,[g,h] Byung Kwon Kim,[a] Soon-Kyeong Kwon,[a,c] Choong Hoon Lee,[a] Dong Su Yu,[a] Mi-Sun Park,[d] Seung-Hak Cho,[d] and Jihyun F. Kim[a,i]

Systems and Synthetic Biology Research Center,[a] Green Bio Research Center,[b] and BioMedical Genomics Research Center,[e] Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea; Biosystems and Bioengineering Program,[c] Department of Functional Genomics,[f] and Department of Marine Biotechnology,[g] University of Science & 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[d]; Korea Ocean Research and Development Institute, Ansan Republic of Korea[h]; and Department of Systems Biology, Yonsei University, Seoul, Republic of Korea[i]

Enterohemorrhagic *Escherichia coli* causes severe food-borne disease in the guts of humans and animals. Here, we report the high-quality draft genome sequence of *E. coli* NCCP15658 isolated from a patient in the Republic of Korea. Its genome size was determined to be 5.46 Mb, and its genomic features, including genes encoding virulence factors, were analyzed.

Escherichia coli strains are mainly found in the environment or in the guts of endotherms. Most of the *E. coli* strains existing in human guts are harmless commensals. However, several serotypes can cause severe food-borne infectious diseases (15). *E. coli* O157:H7 is one of the representative pathogenic *E. coli* strains, which are classified as Shiga toxin-producing enterohemorrhagic *E. coli* (EHEC). So far, the genome sequences of *E. coli* O157:H7 EDL933 (11), Sakai (7), TW14359 (9), and EC4115 (5) have been reported. Strain NCCP15658 was isolated from a female patient who presented with diarrhea, vomit, and stomach cramps in 2004; the isolate was provided to and deposited at the National Culture Collection for Pathogens, National Institute of Health, Republic of Korea.

The genome sequence of NCCP15658 was determined by using the Illumina/Solexa platforms. A total of 4.6 Gb of reads were produced from a 600-bp paired-end library and a 3-kb mate-pair library that were produced, respectively, in-house and at NICEM in the Republic of Korea. CLC Genomics Workbench version 4.8 (CLC bio, Inc.) was used for sequence trimming, *de novo* assembly, and read mapping. The contigs from *de novo* assembly and the consensus sequences from mapping to the *E. coli* TW14359 genome were adopted for the final assembly. SSPACE version 1.2 (3) was used to construct the scaffolds, and the assembly was improved by gap filling with IMAGE version 2.4 (14). Glimmer 3.02 and GeneMarkS were used for prediction of protein-coding sequences. Functional assignment of the predicted genes, performed through homology-based searches against the MicroScope, GenBank, Pfam, and TIGRFam databases, was complemented with the annotation result from the RAST server (1).

The final draft genome sequence in 51 contigs has 5,462,312 bp with a GC content of 50.45%, and 5,547 protein-coding genes were annotated. The genome sequence showed the highest similarity to *E. coli* TW14359 and EC4115, with an ANIb value of 99.93% as estimated by Jspecies version 1.2.1 (12). The genome has plasmid-borne sequences similar to the sequence of pO157 that is commonly harbored in O157:H7 strains. Serotypes of *E. coli* have been classified by specific genes, such as *wzx*, *wzy*, and *fliC* (2). Phylogenetic analysis of these genes in NCCP15658 suggests its identity as O157:H7. NCCP15658 contains effectors and structural proteins associated with type III secretion, most of which are encoded at the locus of enterocyte effacement (LEE) (4, 10, 13). The chromosome and pO157 virulence plasmid contain genes for Shiga toxin and a Shiga toxin-like protein that are related to significant enteric human pathogens (6). NCCP15658 also possesses genes involved in cell attachment and virulence, including intimin/invasion/adhesion, hemolysins, and pertactin. Thus, the NCCP15658 genome presents the genetic features for various virulence factors. This work provides additional genomic information for the O157:H7 type of EHEC, which will be useful to study the genome evolution of pathogenicity through comparative analysis of EHEC genomes.

Nucleotide sequence accession number. The draft genome sequence was deposited in GenBank under the accession number AJMD00000000. The sequence and annotation are also available from the Genome Encyclopedia of Microbes (GEM; http://www.gem.re.kr) (8).

ACKNOWLEDGMENTS

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 Korea National Institute of Health (NIH 4800-4845-300 to S.-H.C.), and the KRIBB Research Initiative Program.

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 to members of GEM and the KRIBB sequencing team for technical help.

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.
2. **Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H.** 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. J. Clin. Microbiol. **45**:370–379.
3. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**:578–579.

Received 18 April 2012 Accepted 30 April 2012

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

J.Y.S., R.H.Y., and S.Y.J. contributed equally to this work.

doi:10.1128/JB.00653-12

4. **Deng W, et al.** 2004. Dissecting virulence: systematic and functional analyses of a pathogenicity island. Proc. Natl. Acad. Sci. U. S. A. **101**:3597–3602.
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.
6. **Franz E, Klerks MM, De Vos OJ, Termorshuizen AJ, van Bruggen AH.** 2007. Prevalence of Shiga toxin-producing *Escherichia coli* stx_1, stx_2, *eaeA*, and *rfbE* genes and survival of *E. coli* O157:H7 in manure from organic and low-input conventional dairy farms. Appl. Environ. Microbiol. **73**:2180–2190.
7. **Hayashi T, et al.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157: H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**:11.
8. **Jeong H, Yoon SH, Yu DS, Oh TK, Kim JF.** 2008. Recent progress of microbial genome projects in Korea. Biotechnol. J. **3**:601–611.
9. **Kulasekara BR, et al.** 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.
10. **Mundy R, Jenkins C, Yu J, Smith H, Frankel G.** 2004. Distribution of *espI* among clinical enterohaemorrhagic and enteropathogenic *Escherichia coli* isolates. J. Med. Microbiol. **53**:1145–1149.
11. **Perna NT, et al.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**:529–533.
12. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106**:19126.
13. **Sandner L, Eguiarte LE, Navarro A, Cravioto A, Souza V.** 2001. The elements of the locus of enterocyte effacement in human and wild mammal isolates of *Escherichia coli*: evolution by assemblage or disruption? Microbiology **147**:3149.
14. **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.
15. **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.