Complete Genome Sequence of an Enterohemorrhagic Escherichia coli O111:H8 Strain Recovered from a Large Outbreak in Japan Associated with Consumption of Raw Beef

Tsuyoshi Sekizuka,a Kenichi Lee,b Keiko Kimata,c Junko Isobe,c Makoto Kuroda,a Sunao Iyoda,b Makoto Ohnishi,b Tetsutaro Sata,c Masanori Watahiki

aPathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan
bDepartment of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan
cDepartment of Bacteriology, Toyama Institute of Health, Toyama, Japan

ABSTRACT We present the complete genome sequence of an enterohemorrhagic Escherichia coli O111:H8 strain. This strain was isolated from a hemolytic-uremic syndrome patient and was responsible for a large outbreak associated with the consumption of raw beef in 2011.

Enterohemorrhagic Escherichia coli (EHEC) is a serious etiological agent of outbreaks of foodborne illness worldwide. More than 3,000 cases of EHEC infection are reported annually in Japan (1). In 2011, a large outbreak due to the consumption of raw beef was reported. A total of 181 patients, including 21 patients with acute encephalopathy (AE) and five deaths among 34 patients with hemolytic-uremic syndrome (HUS), were recorded (2, 3). EHEC O111:H8 and O157:H7 were isolated from the patients; however, the Shiga toxin 2a gene (stx2)-positive EHEC O111:H8 strain was the primary cause of the severe infection, according to serological testing of the patients (4). To reveal the characteristics of the strain, we determined its complete genome sequence, as the complete genome of the stx2-positive EHEC O111 strain has not been reported.

EHEC O111:H8 110512 was isolated as one of a few colonies grown on deoxycholate-hydrogen sulfide-lactose agar (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) from the fecal sample of a 6-year-old boy with HUS and AE. The strain was stored in 0.5% NaCl casein peptone medium (Kohjin Bio Co., Ltd., Saitama, Japan) until use. For DNA extraction, a single colony was inoculated into buffered peptone water (Nissui). The overnight culture was subjected to genomic DNA extraction by a DNaseasy blood and tissue kit (Qiagen, Venlo, Netherlands). The genomic DNA libraries were prepared using a Nextera DNA sample prep kit (Epicentre Biotechnologies, Madison, WI, USA), followed by paired-end sequencing (2 × 150-mer) with a MiSeq instrument (Illumina, San Diego, CA). Additionally, the strain was sequenced on a PacBio RS II instrument (Pacific Biosciences, Menlo Park, CA). The genomic DNA was fragmented prior to PacBio sequencing using the Megaruptor (Diagenode, Denville, NJ). PacBio sequencing was performed by using the SMRTbell template prep kit 1.0 and polymerase binding kit P6 after size selection using BluePippin (Sage Science, Beverly, MA) with a cutoff value of 20 kb. In bioinformatic analysis, default parameters were used for all software. The complete genome of the strain was obtained by using both PacBio and Illumina reads. PacBio reads were assembled using the Hierarchical Genome Assembly Process version 3 in SMRT Analysis software (5). The minimum seed length for the assembly was 6,000. Short plasmids (±10 kb) were assembled using A5 MiSeq software version 20140604 with Illumina short reads (6). The tentative circular contigs were subjected to error correction in three iterations using Pilon version 1.18 with default parameters and
Illumina short reads (7). Annotation was performed in Prokka version 1.11 (8), InterPro version 49.0 (9), and NCBI BLASTP/BLASTX.

In Illumina and PacBio sequencing, the read numbers were 5,894,924 and 195,083 (mean read length, 7,798 bp), respectively. The total read lengths for Illumina and PacBio sequencing were 732,012,018 bp and 1,521,220,270 bp, respectively; the mean coverage was 132-fold and 273-fold, respectively. The complete genome sequence consists of a chromosome and seven plasmids as shown in Table 1. The strain carried a novel Stx2 prophage but did not carry a Stx1 prophage. These data would be helpful for the characterization of stx₂-positive EHEC O111 and improve the resolution of the surveillance using whole-genome sequencing (10, 11).

**Data availability.** The complete genome sequence of EHEC O111:H8 strain 110512 was submitted to DDBJ/ENA/GenBank under the accession numbers AP019761 through AP019768. The primary data were deposited in the NCBI primary data archive, SRA, under the reference number DRP005302.

**ACKNOWLEDGMENTS**

This work was partially supported by a special research grant (H23-TOKUBETU-SHITEI-004) and a Health Labour Sciences research grant (H24-SHINKO-IPPAN-012) from the Ministry of Health, Labour, and Welfare (Japan) and by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development (AMED) (grant number JP18fk0108065).

**REFERENCES**


<table>
<thead>
<tr>
<th>Statistic</th>
<th>Data form</th>
<th>pO111-110512_1</th>
<th>pO111-110512_2</th>
<th>pO111-110512_3</th>
<th>pO111-110512_4</th>
<th>pO111-110512_5</th>
<th>pO111-110512_6</th>
<th>pO111-110512_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (kb)</td>
<td></td>
<td>5,302,257</td>
<td>118,827</td>
<td>78,470</td>
<td>46,350</td>
<td>6,795</td>
<td>6,674</td>
<td>5,423</td>
</tr>
<tr>
<td>No. of CDSs(^a)</td>
<td></td>
<td>5,119</td>
<td></td>
<td>81</td>
<td>57</td>
<td>6</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>No. of rRNAs</td>
<td></td>
<td>22</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No. of rRNAs</td>
<td></td>
<td>107</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No. of prophages</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC content (%)</td>
<td></td>
<td>50.6</td>
<td>53.3</td>
<td>50.0</td>
<td>40.7</td>
<td>50.8</td>
<td>50.3</td>
<td>47.2</td>
</tr>
<tr>
<td>Chromosomal sequence type or plasmid Inc type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimicrobial resistance gene(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession no.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP019761</td>
<td>AP019762</td>
<td>AP019763</td>
<td>AP019764</td>
<td>AP019765</td>
<td>AP019766</td>
<td>AP019767</td>
</tr>
</tbody>
</table>

\(^a\)ND, not detected.
\(^b\)CDS, coding DNA sequences.
