Draft Genome Sequences of Enterohemorrhagic *Escherichia coli* Encoding Extended-Spectrum Beta-Lactamases

Charlotte Valat,*a Robert J. Goldstone, b Édouard Hirchaud, c Marisa Haenni, a David G. E. Smith, d Jean-Yves Madec a

Anses, Laboratoire de Lyon, Unité Antibiorésistance et Virulence Bactériennes, Lyon, France; c School of Life Sciences, Heriot-Watt University, Edinburgh, United Kingdom; d Anses, Laboratoire de Ploufragan-Plouzané, Unité Génétique Virale et Biosécurité, Plouzané, France

* Present address: Charlotte Valat, Independent Researcher, Lyon, France.

Extended-spectrum beta-lactamases (ESBLs) have rarely been observed among Shiga toxigenic *Escherichia coli* (STEC), and, to our best knowledge, only three ESBL-positive isolates of the enterohemorrhagic *E. coli* (EHEC) subpathotype have been reported. Here, we present the first draft genome sequences of two ESBL-positive EHEC isolates belonging to serotypes O111:H8 and O151:H16.

Shiga toxigenic *Escherichia coli* (STEC), including the subpathotype of enterohemorrhagic *E. coli* (EHEC) carrying *stx* and *eae* genes, is frequently associated with human infection. Ruminants are the major source of STEC, and they are also a reservoir of extended-spectrum beta-lactamase (ESBL)-producing *E. coli*, mostly of the CTX-M type (1). ESBLs have been associated with STEC in rare cases, either in humans (O104:H4/CTX-M-15 [2], O26/TEM-52 [3], O64/CTX-M-3 [4], O157:H7/CTX-M-1 [5], and O26:H11/CTX-M-18 [6]), in chickens (O157/CTX-M-2 [7]), or in cattle (O111:H8/CTX-M-15 [8] and O145:NM/CTX-M-1 [9]); however, to the best of our knowledge, only 3 ESBL-producing *E. coli* strains belonging to the subpathotype of EHEC (STx and Eae) have been identified so far (6, 8, 9). The aim of this study was to present the first draft genome sequences of two ESBL-producing EHEC, including those belonging to the serotype O111:H8 (sequence type 16 [ST16]) first characterized using a DNA array and PCR in a previous study (8), and a new isolate of serotype O151:H16 (ST21) producing CTX-M-1 and carrying the *stxlα* and *eaeB1* genes. These two ESBL-producing EHEC isolates were isolated from diseased cattle in France through the nationwide surveillance network of antimicrobial resistance in animal pathogens (Resapath [http://www.resapath.anses.fr]).

Genomic DNA was extracted from overnight culture using the MasterPure DNA purification kit (Epicentre). Sequencing was performed using an Illumina MiSeq sequencer at the Wolfson Wohl Cancer Research Centre, United Kingdom. A multiplex sequencing approach was used, involving 12 separately tagged libraries sequenced simultaneously in two lanes. The standard Illumina indexing protocol involved the fragmentation of 2 μg of genomic DNA by acoustic shearing to enrich for 200-bp fragments. A-tailing, adapter ligation, and an overlap extension PCR using the Illumina 3-primer set were performed to introduce specific tag sequences between the sequencing and flow cell binding sites of the Illumina adapter. DNA cleanup was carried out after each step to remove DNA sequences <150 bp using AMPure paramagnetic beads (Beckman Coulter, Inc., USA), and a quantitative PCR (qPCR) was used for final DNA quantification. The raw reads were trimmed by the removal of ambiguous nucleotides from the read ends and trimmed for reads on which quality scores were <0.001. Reads <15 nucleotides were also removed. For de novo assembly using CLC Genomics Workbench (version 6.5.2), scaffolding was performed, and paired distances were automatically detected. The minimum contig length was set to 200 bp. The de novo assembly produced 822 contigs for *E. coli* O111:H8 (isolate #22207) and 818 contigs for *E. coli* O151:H16 (isolate #22593). The median coverages of the assemblies were 110× for #22207 and 106× for #22593, with an N50 of 97 kb and 85 kb, and a genome size of 5.607 and 5.868 Mb, respectively. The contigs were annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) at http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. In the genomes of #22207 and #22593, 5,567 and 5,837 coding sequences were identified, respectively. Future comparative analysis of whole-genome sequencing data from these isolates will provide insights into the attributes of emerging ESBL-producing EHEC.

**Nucleotide sequence accession numbers.** The whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession numbers LMBJ00000000 and LMBK00000000. The versions described in this paper are versions LMBJ01000000 and LMBK01000000, respectively.

**ACKNOWLEDGMENTS**

This study was supported by Anses, Laboratoire de Lyon, Unité Antibiorésistance et Virulence Bactériennes. This work was partially funded by BBSRC and Defra (project references BB/1017283/1 and OD1718). The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.
FUNDING INFORMATION
European Commission (EC) provided funding to David G. E. Smith under grant number BB/1017283/1.

REFERENCES