Genome Sequence and Annotation of a Human Infection Isolate of *Escherichia coli* O26:H11 Involved in a Raw Milk Cheese Outbreak

Wessam Galia,[a,c,d] Patricia Mariani-Kurkdjian,[g] Estelle Loukiadis,[a,b] Stéphanie Blanquet-Diot,[d] Françoise Leriche,[c] Hubert Brugère,[e,f] Ayaka Shima,[e,f] Eric Oswald,[e,f] Benoit Cournoyer,[a] Delphine Thevenot-Sergentet[a,b]

Université de Lyon, Research Group on Bacterial Opportunistic Pathogens and Environment, UMR 5557 Ecologie Microbienne, CNRS, VetAgro Sup, and Université Lyon 1, Lyon, France[a]; Université de Lyon, VetAgro Sup, Campus Vétérinaire de Lyon, Laboratoire d'Étude des Microorganismes Alimentaires Pathogenes, French National Reference Laboratory for *Escherichia coli* including Shiga Toxin-Producing *E. coli*, Marcy l'Etoile, France[b]; Université de Clermont-Ferrand, VetAgro Sup, Campus Agronomique de Lempdes, Unité CALYTISS, Lempdes, France[c]; Clermont Université, Université d'Auvergne, Centre de Recherche en Nutrition Humaine Auvergne, EA 4678 CIDAM, BP 10448, Clermont-Ferrand, France[d]; INRA, USC1360, Toulouse, France[e]; INSERM UMR1043/CNRS UMR5282/Université Toulouse III, Centre de Physiopathologie de Toulouse Purpan (CPTP), Toulouse, France[f]; Laboratory associated with the National Reference Centre for *Escherichia coli* and *Shigella*, Robert Debré Hospital, Paris, France[g]

The consumption of raw milk cheese can expose populations to Shiga toxin-producing *Escherichia coli* (STEC). We report here the genome sequence of an *E. coli* O26:H11 strain isolated from humans during the first raw milk cheese outbreak described in France (2005).

Address correspondence to Wessam Galia, wessam.galia@vetagro-sup.fr.

Enterohemorrhagic *Escherichia coli* can cause foodborne diseases ranging from simple diarrhea to hemorrhagic colitis and life-threatening complications, such as hemolytic-uremic syndrome. *E. coli* O157:H7 is implicated in the majority of outbreaks and hemolytic-uremic syndrome (HUS) cases [1]. However, other serotypes, such as *E. coli* O26:H11, have also been implicated in outbreaks [2, 3]. Yet, little is known about the genomic diversity that exists among *E. coli* O26:H11 pathogen populations or how various genotypes of this pathogen relate to the development and severity of the clinical manifestations in infected patients. The genome sequence data bring more clarity to both the evolutionary processes and virulence factors of this important pathogen.

Genomic DNA was extracted from an overnight culture using the DNeasy blood and tissue kit (Qiagen). A sequencing library was prepared with Nextera XT version 3 chemistry (Illumina). Following fragmentation, end reparation, and sample tagging, the Illumina MiSeq platform sequencer produced 300-bp paired-end reads that were obtained from 700-bp inserts, yielding an average coverage of 1,000×. The program FastQC version 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was applied to check the quality of the generated 16,638,445 reads. This data set was employed for the *de novo* sequence assembly by various assemblers, including SPAdes version 3.0.0 [4], Minia version 1.6088 [5, 6], Velvet version 1.2.10 [7], and MaZuRCA version 2.2.1 [8]. The sequence contigs obtained from the assemblers were subsequently input into the Contig Integrator for Sequence Assembly (CISA) to generate an integrated set of contigs [9]. Compared with the assemblies generated by each assembler, a hybrid set of assemblies of superior contiguity and accuracy was obtained by CISA. Next, 102 hybrid contigs were given, resulting in a total genome size of 5,686,107 bp and an N_{50} of 107,803 bp. The resulting contigs were ordered by alignment to the reference genome of *E. coli* O26:H11 strain 11368 [10] using Mauve Contig Mover version 2.3.1 [11, 12]. Functional annotation was carried out using the tools of the MicroScope platform [13], and the annotated genome was implemented in the public EscherichiaScope database (see https://www.genoscope.cns.fr/agc/microscope/home/index.php). The assembly contains 5,781 genomic objects, among which are 5,611 coding sequences, 10 rRNA genes, 92 tRNA genes, and 68 noncoding RNAs. The genome G+C contents were relatively close to those of the reference genomes (50.6%).

A detailed report of the phylogenetic analysis of this draft sequence will be included in a future publication.

Nucleotide sequence accession numbers. The draft genome sequence of the *E. coli* O26:H11 strain 21765 in this study has been deposited as a whole-genome shotgun project at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view) under the contig accession numbers CDLB01000001 to CDLB01000102.

ACKNOWLEDGMENTS

REFERENCES

1. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11**:142–201.
2. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis **43**:1587–1595. http://dx.doi.org/10.1086/509573.
3. **Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL.** 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. Emerg Infect Dis **11**:603–609. http://dx.doi.org/10.3201/eid1104.040739.

4. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

5. **Chikhi R, Rizk G.** 2013. Space-efficient and exact de Bruijn graph representation based on a Bloom filter. Algorithm Mol Biol **8:**22. http://dx.doi.org/10.1186/1748-7188-8-22.

6. **Chikhi R, Rizk G.** 2012. Space-efficient and exact de Bruijn graph representation based on a Bloom filter, p 236–248. *In* Raphael B, Tang J (ed), Algorithms in bioinformatics. Springer Verlag, Heidelberg, Germany.

7. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

8. **Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA.** 2013. The MaSuRCA genome assembler. Bioinformatics **29:**2669–2677. http://dx.doi.org/10.1093/bioinformatics/btt476.

9. **Lin S-H, Liao Y-C.** 2013. CISA: Contig Integrator for Sequence Assembly of bacterial genomes. PLoS One **8:**e60843. http://dx.doi.org/10.1371/journal.pone.0060843.

10. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 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. http://dx.doi.org/10.1073/pnas.0903585106.

11. **Darling ACE, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394–1403. http://dx.doi.org/10.1101/gr.2289704.

12. **Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT.** 2009. Reordering contigs of draft genomes using the mauve aligner. Bioinformatics **25:**2071–2073. http://dx.doi.org/10.1093/bioinformatics/btp356.

13. **Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D, Rouy Z, Salvignol G, Scarpelli C, Thil Smith AA, Weiman M, Médigue C.** 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res **41:**D636–D647. http://dx.doi.org/10.1093/nar/gks1194.