

Genome sequences of 15 *Escherichia coli* O77g strains related to an emerging STEC lineage

Mai-Lan Tran,^{1,2} Fabien Vorimore,² Patrick Fach,^{1,2} Sabine Delannoy^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 4.

ABSTRACT We report the genome sequences of 15 *Escherichia coli* O77g strains belonging to the O17, O44, O73, O77, and O106 antigen cluster. Strains of *E. coli* O77g have been associated with hemolytic uremic syndrome (HUS) cases in France in early 2025.

KEYWORDS Shiga toxins, *Escherichia coli*

Shiga toxin-producing *Escherichia coli* (STEC) are major foodborne pathogens responsible for sporadic and epidemic cases of hemolytic uremic syndrome (HUS) worldwide (1). In January 2025, HUS cases were reported in France and linked to unpasteurized cheese contaminated with an emerging STEC O77g:H18 belonging to sequence type 69 (ST69) (*stx2d*-positive, *eae*-negative) (2). To provide comparative genomics and molecular epidemiology of this lineage, 15 *E. coli* O77g isolates from our lab collection were sequenced and assembled. These strains belong to serogroups O17, O44, O73, O77, and O106, which share the same O-antigen gene cluster (3). *E. coli* strains were originally isolated on trypticase soy broth with yeast extract (TSYE) plates overnight at 37°C. Prior to DNA extraction, a colony was grown overnight in brain heart infusion (BHI) broth at 37°C with agitation. Genomic DNA was extracted from 1 mL of culture using the DNeasy Blood and Tissue Kit (Qiagen, France) according to the manufacturer's instructions. All 15 isolates were sequenced on R10.4.1 flow cells using a MinION Mk1D device (Oxford Nanopore Technologies, Oxford, UK). Libraries were prepared with 200 ng extracted DNA using the Rapid Barcoding Kit SQK-RBK114.24 (Oxford Nanopore Technologies) following the manufacturer's protocol. DNA was not sheared nor size selected prior to library preparation. Sequencing runs were performed for 72 h using MinKNOW v25.09.16, and raw data were live basecalled using the Super-accurate model v5.2.0-400bps. The option "trim barcodes," "barcode both ends," and "Min Q score 10" was enabled on the MinKNOW (4). Raw reads from the pass folder were trimmed using chopper v0.10.1 (5) with "-l 1000" parameter to filter reads below 1,000 bp length. The min quality parameter was set to 20 and 25 for the *E. coli* O17:H18 reference strain K12a and the LP-26-E1-261 strain, respectively. Assemblies of the strains were generated using autocycler v0.5.0 (6) with the custom script provided by the author (https://github.com/rrwick/Autocycler/blob/main/pipelines/Automated_Autocycler_Bash_script_by_Ryan_Wick/autocycler_full.sh).

Assemblies were corrected using Medaka v2.0.1 (<https://github.com/nanopore-tech/medaka>) using the "-bacteria" option (model r1041_e82_400bps_bacterial_methylation). Statistics on raw reads were obtained using NanoPlot v1.45.0 (4). A total of 37,760 to 447,721 reads were generated with N_{50} values ranging from 6,125 to 14,398 bp (mean, 12,130.40 ± 2,189.38 bp). The number of contigs per assembly ranged from one to nine with total assembly lengths between 4,879,178 and 5,580,222 bp. The mean GC content was 50.58% (±0.12). Assembly N_{50} values ranged from 4,879,178 to 5,244,335 bp. All genomes submitted were circular, chromosome, and plasmids, if any, but they were

Editor David Rasko, University of Maryland Baltimore, Baltimore, Maryland, USA

Address correspondence to Mai-Lan Tran, mai-lan.tran@anses.fr.

The authors declare no conflict of interest.

See the related article at <https://doi.org/10.1128/aem.02449-21>.

Received 24 February 2026

Accepted 30 March 2026

Published 27 April 2026

Copyright © 2026 Tran et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Assembly metrics and accession numbers for the 15 *Escherichia coli* O77g isolates sequenced.

Strain	Source niche	Source type	Source details	Collection year	Continent	Country	ST ^a	ST complex ^a	Serotype ^b	ee ^c	stx ^d	Phylogroup ^e	raw reads	length (N ₅₀)	MinION read No. of contigs	Genome size (bp)	Assembly N ₅₀ (bp)	GC (%)	Genome coverage	NCBI accession no. of assembly	NCBI accession no. of reads
LP-26-E1-261	Environment	Water	River	2018	Europe	France	69	ST69 Cplk	O77g:H18	– ^g	–	D	161,560	6,125	9	5,580,222	5,039,650	50.76	130	JBVDMR000000000	SRR37276457
ECA64	Livestock	Bovine	Dairy product	1995	Europe	France	663	–	O77g:H18	–	stx2d	D	264,842	9,098	3	5,233,331	5,059,963	50.71	296	JBVDMQ000000000	SRR37276456
CB209	Human	Human	Homo sapiens; stool	1987	Europe	Germany	449 ^f	–	O77g:H18	–	–	D	116,374	12,099	3	5,222,438	5,089,256	50.53	170	JBVDMF000000000	SRR37276450
CB6837	Human	Human	Homo sapiens	1997	Europe	Germany	663	–	O77g:H18	–	stx1a, stx2a	D	235,219	11,945	3	5,314,391	5,135,902	50.7	329	JBVDMO000000000	SRR37276449
CB6886	Human	Human	Homo sapiens	1997	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	113,334	12,507	5	5,250,473	5,040,535	50.76	160	JBVDMN000000000	SRR37276448
CB7286	Human	Human	Homo sapiens	1997	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	142,169	11,726	5	5,484,097	5,244,335	50.67	174	JBVDMM000000000	SRR37276447
CB7453	Human	Human	Homo sapiens	1998	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	104,279	14,351	4	5,131,675	4,923,747	50.53	176	JBVDMJ000000000	SRR37276446
CB8671	Human	Human	Homo sapiens	2001	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	145,483	11,161	2	5,046,815	4,921,061	50.62	195	JBVDMK000000000	SRR37276445
DG341-1	Poultry	Poultry	Chicken	1990	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	139,310	13,069	4	5,410,318	5,067,659	50.45	207	JBVDMJ000000000	SRR37276444
DG342-2	Poultry	Poultry	Chicken	1990	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	37,760	13,518	4	5,410,318	5,067,659	50.45	56	JBVDMJ000000000	SRR37276443
DG343-1	Poultry	Poultry	Chicken	1990	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	126,418	14,176	4	5,410,318	5,067,659	50.45	198	JBVDMH000000000	SRR37276455
DG351-2	Poultry	Poultry	Chicken	1990	Europe	Germany	69	ST69 Cplk	O77g:H18	–	–	D	160,270	11,356	4	5,410,318	5,067,659	50.45	204	JBVDMG000000000	SRR37276454
KK05-19	Human	Human	Homo sapiens; stool	1986	Europe	Germany	130	ST31 Cplk	O77g:H18	–	–	D	104,452	13,741	1	4,879,178	4,879,178	50.61	176	JBVDMF000000000	SRR37276453
KK19-02	Human	Human	Homo sapiens; stool	1986	Europe	Germany	130 ^f	–	O77g:H4	–	–	G	144,217	12,686	2	5,324,074	5,137,557	50.57	209	JBVDMO000000000	SRR37276452

(Continued on next page)

TABLE 1 Assembly metrics and accession numbers for the 15 *Escherichia coli* O77g isolates sequenced. (Continued)

Strain	Source	niche	Source type	Source details	Collection year	Continent	Country	ST ^a	ST complex ^a	Serotype ^b	eae ^c	stx ^d	Phylogroup ^e	Number of raw reads	MinION read length (N ₅₀) (bp)	No. of contigs	Genome size (bp)	Assembly size (N ₅₀) (bp)	GC (%)	Genome coverage	NCBI accession no. of assembly	NCBI accession no. of reads
O17_K12a	Human	Human	Homo sapiens; stool		1940	Europe	Denmark	925	–	O17:H18	–	–	D	447,721	14,398	3	5,272,206	5,093,858	50.47	751	JBYDMD000000000	SRR37276451.00

^aDetermined using MLST v2.0 (with the *E. coli* #1 MLST scheme and database version 2023-06-19) on the CGE website (<https://www.genomepidemiology.org/>).

^bDetermined using Abricate (<https://github.com/tseemann/abricate>) with the EcoH database (8).

^cDetermined using Abricate (<https://github.com/tseemann/abricate>) with VirulenceFinder.

^dDetermined using Abricate and a custom database (https://bitbucket.org/genomicepidemiology/virulencefinder_db/src/master/stx.fasta).

^eDetermined using EzClermont v0.7.0 (<https://github.com/nickp60/EzClermont>).

^fPresence of a novel allele. Nearest ST is displayed.

^g–, not available.

not rotated to a specific base. The number of contigs was defined using the output of autocycler “consensus_assembly.yaml” that counts the number of contigs and indicates the topology. All contigs were complete and circular, so the number of contigs in Table 1 indicates one chromosome and the rest are plasmids, if any. Detailed strain information, including serotype, presence of *stx* and *eae* genes, genome coverage, and accession numbers is reported in Table 1. All genomes were annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) v6.10 (7) at https://www.ncbi.nlm.nih.gov/refseq/annotation_prok/.

ACKNOWLEDGMENTS

We are grateful to Sandrine Baron and Laëtitia le Devendec (Anses, Ploufragan-Plouzané-Niort Laboratory, Mycoplasma-Bacteriology and Antimicrobial Resistance Unit) for providing strain LP-26-E1-261. The strain was isolated in the framework of the Resist3A project funded by the European Maritime and Fisheries Fund.

AUTHOR AFFILIATIONS

¹Pathogenic *E. coli* Unit (COLiPATH), Laboratory for Food Safety, Anses, Maisons-Alfort, France

²IdentityPath Genomics Platform (IDPA), Laboratory for Food Safety, Anses, Maisons-Alfort, France

AUTHOR ORCIDS

Mai-Lan Tran  <http://orcid.org/0000-0003-3161-8113>

Fabien Vorimore  <https://orcid.org/0000-0001-7678-5551>

Patrick Fach  <https://orcid.org/0000-0002-9396-6327>

Sabine Delannoy  <https://orcid.org/0000-0002-9208-5659>

AUTHOR CONTRIBUTIONS

Mai-Lan Tran, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Fabien Vorimore, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Patrick Fach, Conceptualization, Resources, Writing – review and editing | Sabine Delannoy, Conceptualization, Formal analysis, Investigation, Resources, Writing – review and editing

DATA AVAILABILITY

The genome assemblies and corresponding raw reads have been deposited in the NCBI database under BioProject accession number [PRJNA1424694](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1424694).

ETHICS APPROVAL

This study was conducted in accordance with the Declaration of Helsinki (2013).

REFERENCES

1. Karmali MA. 2004. Infection by shiga toxin-producing *Escherichia coli*: an overview. *Mol Biotechnol* 26:117–122. <https://doi.org/10.1385/MB:26:2:117>
2. de Larminat J, La K, Bidet P, Birgy A, Liguori S, Philipaux P, Courroux C, Crombé F, Blanco J, Coppo P, Dossier A, Armand-Lefevre L, Weill FX, Nodari CS, Jones G, Cointe A, Bonacorsi S. 2026. Hemolytic uremic syndrome outbreak in adults and shiga toxin-producing *Escherichia coli* negative for locus of enterocyte effacement, France, 2025. *Emerg Infect Dis* 32:592–602. <https://doi.org/10.3201/eid3204.251417>
3. Iguchi A, Iyoda S, Kikuchi T, Ogura Y, Katsura K, Ohnishi M, Hayashi T, Thomson NR. 2015. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 22:101–107. <https://doi.org/10.1093/dnares/dsu043>
4. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>

5. De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39:btad311. <https://doi.org/10.1093/bioinformatics/btad311>
6. Wick RR, Howden BP, Stinear TP. 2025. Autocycler: long-read consensus assembly for bacterial genomes. *Bioinformatics* 41:btaf474. <https://doi.org/10.1093/bioinformatics/btaf474>
7. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
8. Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-Browne RM, Holt KE. 2016. *In silico* serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb Genom* 2:e000064. <https://doi.org/10.1099/mgen.0.000064>