AMERICAN SOCIETY FOR MICROBIOLOGY
Microbiology® Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of an Enterotoxigenic *Escherichia coli* Strain Carrying Genes for Colonization Surface Antigen 13 and a Heat-Labile Toxin

Samuel M. Njoroge,[a,b,f] Laure F. Madé,[c] Astrid von Mentzer,[d,j] Benard W. Kulohoma,[a] Timothy K. Kamanu,[e] Tom T. Ouko,[f] John Kiiru,[f,g] Melissa J. Ward,[h,i] Nicholas R. Thomson,[j] Eric M. Fèvre,[b,c] Mark Woolhouse,[i,k] Samuel Kariuki[l]

[a]University of Nairobi, Department of Biochemistry, Nairobi, Kenya
[b]International Livestock Research Institute, Nairobi, Kenya
[c]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom
[d]Chalmers University of Technology, Department of Mathematical Sciences, Gothenburg, Sweden
[e]University of Nairobi, School of Mathematics, College of Biological and Physical Sciences, Nairobi, Kenya
[f]Centre for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya
[g]Ministry of Public Health and Sanitation, National Public Health Laboratory Services, Nairobi, Kenya
[h]Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom
[i]Centre for Immunity, Infection, and Evolution, University of Edinburgh, Edinburgh, Scotland, United Kingdom
[j]Wellcome Sanger Institute, Hinxton, United Kingdom
[k]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom
[l]Kenya Medical Research Institute, Nairobi, Kenya

Samuel M. Njoroge, Laure F. Madé, Astrid von Mentzer, Benard W. Kulohoma, Timothy K. Kamanu, Tom T. Ouko, John Kiiru, Melissa J. Ward, Nicholas R. Thomson, Eric M. Fèvre, Mark Woolhouse, and Samuel Kariuki contributed equally to this work. Author order was determined on the basis of seniority.

ABSTRACT Here, we report the draft genome of ESEI_597, an enterotoxigenic *Escherichia coli* (ETEC) strain harboring genes encoding colonization surface antigen 13 (CS13) and a heat-labile toxin. The ESEI_597 strain was isolated from an 8-month-old child living in Korogocho, Kenya, in 2013.

Enterotoxigenic *Escherichia coli* (ETEC) strain ESEI_597 was isolated in 2013 from a stool sample from an 8-month-old child residing in Korogocho, an informal settlement northeast of Nairobi, Kenya. A pea-sized fecal sample was emulsified in 5 mL buffered peptone water (Oxoid, UK). Approximately 10 μL of the overnight broth inoculum was plated on MacConkey agar (Oxoid), and a biochemically identified colony was subcultured in Mueller-Hinton agar (Oxoid) before being stocked at −80°C in tryptone soy broth (Oxoid) with 15% glycerol. All incubations were performed at 37°C for 18 to 24 h. Labile toxin (LT) and colonization surface antigen 13 (CS13) conventional uniplex PCR screening was performed using previously described primers (1, 2). Genomic DNA extraction was performed using a TANBead system (Taiwan Advanced Nanotech, Inc., Taiwan). Library preparation was performed using the NEBNext Ultra DNA library preparation kit for Illumina (New England Biolabs) and standard Illumina multiplexing adapters, with minor modifications to the manufacturer's protocol (3). Paired-end reads (150 bp long) were generated using a HiSeq2500 platform (Illumina) at the Oxford Genomics Centre (https://www.well.ox.ac.uk/ogc) and preprocessed using an automated protocol developed by the Modernising Medical Microbiology (MMM) Oxford Group. Read trimming to remove remnant adaptor sequences was performed using BBDuk, part of the BBTools package (4) (parameters: minoverlap=12, k=19, mink=12, hdist=1, ktrim=r). Kraken v0.10.6-beta (5) was used for species identification analysis against an in-house database downloaded from the NCBI Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra), with an automated step for removal of contaminant reads. The remaining reads were mapped (using Stampy v1.0.23)

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Samuel M. Njoroge, samihnjoroge@gmail.com.

The authors declare no conflict of interest.

Received 29 April 2022
Accepted 22 August 2022
Published 12 September 2022

TABLE 1 *E. coli* ESEI_597 features

Characteristic	Finding
Genome size (Mbp)	4,937,264
No. of contigs	139
N_{50} (bp)	140,019
G+C content (%)	50.7
No. of coding sequences	4,795
No. of tRNAs	82
No. of complete rRNAs	1
AMR genes (% identity)	*aadA4* (100), $\mathit{bla}_{\text{TEM-1B}}$ (100), *dfrA21* (99.8), *sul2* (99.88), $\mathit{bla}_{\text{EC-18}}$ (99.21)
Replicons (% identity)	IncFII(pSE11) (99.62), IncQ1 (100)
MLST (Achtman scheme) ST	155
LT position[a]	4654598..4655766
CS13 operon position[a]	3066924..3074846
MLST (Pasteur scheme) ST	21
BioProject accession no.	PRJNA802315
BioSample accession no.	SAMN26012842
SRA accession no.	SRR18056643
GenBank accession no.	JALDSW000000000

[a] Genes are on the reverse strand.

to an *E. coli* reference genome (GenBank accession AE014075.1) (6). The quality of a total of 1,485,420 reads was assessed with FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), with all per-base-sequence quality flagged as passed.

The draft genome was assembled using SPAdes v3.12.0 with k-mers adjusted to 33, 55, and 91, and the coverage cutoff was set as auto (7). Contigs of $<$200 bp were removed. The assembled genome has 139 contigs, with an N_{50} value of 140,019 bp, a G+C content of 50.71%, and a total length of 4,937,264 bp. ABRicate v1.0.1 (https://github.com/tseemann/abricate) was used to detect acquired antimicrobial resistance (AMR) genes with ResFinder v2.1 (8). Multilocus sequence typing (MLST) was performed using MLST v2.19 (https://github.com/tseemann/mlst) and the PubMLST database (https://pubmlst.org) (9), and plasmid replicons were screened using PlasmidFinder (10). Automatic annotation was performed by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1, with manual curation of colonization factor genes. A total of 4,795 coding sequences were predicted. Default settings were used for all software unless otherwise specified.

ESEI_597 belongs to sequence type 155 (ST155) (Achtman scheme) (11); plasmid replicons IncQ1 and IncFII(pSE11) were identified, with 100% and 99.62% sequence identity, respectively. Based on the large-scale BLAST score ratio (LS-BSR), the *aat* operon, *etpBAC* operon, *eatA*, and *tleA*-like autotransporter are all absent. The *cexE* gene is present but truncated, with an LS-BSR value of 0.8893 (12). The CS13 operon (*cshABCDEFGH*) is 7,923 bp long, and LT in ESEI_597 has been assigned to allele 29.

The Kenya Medical Research Institute (KEMRI)/National Ethics Review Committee approved all procedures (protocol number 2507).

Data availability. The draft genome of *E. coli* ESEI_597 has been deposited in GenBank under the accession number JALDSW000000000. Other features of the sequence data are provided in Table 1.

ACKNOWLEDGMENTS

We thank Joyce Arua, RoseTabby Kiiru, Winnie Kitali, Christine Juma, Purity Karimi, and Hannah Njeri.

This work is published with permission from the Director of the KEMRI.

This work was supported by the Medical Research Council, the Natural Environment Research Council, the Economic and Social Research Council, and the Biotechnology and Biosciences Research Council (grant G1100783/1). M.J.W. was supported by a Sir Henry Wellcome Postdoctoral Fellowship from the Wellcome Trust (grant WT103953MA). S.M.N. was supported by the International Livestock Research Institute (ILRI) graduate fellowship program. The High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford, UK) (funded by Wellcome Trust grant 090532/Z/09/Z) generated the sequencing data.

REFERENCES

1. Fujioka M, Otomo Y, Ahsan CR. 2013. A novel single-step multiplex polymerase chain reaction assay for the detection of diarrheagenic *Escherichia coli*. J Microbiol Methods 92:289–292. https://doi.org/10.1016/j.mimet.2012.12.010.
2. Rodas C, Iniguez V, Qadri F, Wiklund G, Svennerholm A-M, Sjöling A. 2009. Development of multiplex PCR assays for detection of enterotoxigenic *Escherichia coli* colonization factors and toxins. J Clin Microbiol 47:1218–1220. https://doi.org/10.1128/JCM.00316-09.
3. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol 13:104. https://doi.org/10.1186/1472-6750-13-104.
4. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. https://www.osti.gov/biblio/1241166.
5. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46.
6. Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21:936–939. https://doi.org/10.1101/gr.111120.110.
7. 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. https://doi.org/10.1089/cmb.2012.0021.
8. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
9. Jolley KA, Maiden MCJ. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
10. Carattoli A, Hasman H. 2020. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol 2075:285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.
11. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60:1136–1151. https://doi.org/10.1111/j.1365-2958.2006.05172.x.
12. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale BLAST score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2:e332. https://doi.org/10.7717/peerj.332.