Genome Sequence of *Escherichia coli* Isolated from an Adult in Kibera, an Urban Informal Settlement in Nairobi, Kenya

Gilbert K. Kikwai,[a] Bonventure Juma,[d] Fredrick Nindo,[b,c] Caroline Ochieng,[a] Newton Wamola,[a] Kevin Mbogo,[f] Douglas R. Call,[e] Elizabeth Hunsperger[d]

[a]Diagnostic and Laboratory System Programme, Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya
[b]Bioinformatics and Informatics Platform, Institute of Research on Cancer and Aging, Medical School, University of Cote d'Azur, Nice, France
[c]Division of Medical Microbiology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[d]Division of Global Health Protection, Centers for Disease Control and Prevention-Kenya, Nairobi, Kenya
[e]Paul G. Allen School for Global Health, Washington State University, Pullman, Washington, USA
[f]Biochemistry Department, College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

ABSTRACT An *Escherichia coli* strain (sequence type 636) was isolated from an adult residing in an urban informal settlement in Nairobi, Kenya, and was sequenced using the Illumina MiSeq platform. The draft genome was 5,075,726 bp, with a Col(BS512) plasmid plus *aph(6)-Id*, *bla*$_{TEM-1B}$, and *dfrA7* genes, which encode kanamycin, ampicillin, and trimethoprim resistance proteins, respectively.

Escherichia coli is a facultative anaerobic bacterium that is able to exist within a mammalian tract as a harmless commensal or as a pathogen. The bacterial genome can acquire or lose genetic information, which can offer competitive advantages for individual strains (1, 2).

We report the genome sequence of an *Escherichia coli* sequence type 636 (ST636) strain that was isolated from an adult residing in Kibera, an urban informal settlement in Nairobi, Kenya. This community is characterized by poor sanitation, unregulated antibiotic use, outdoor food vending, and poor water supplies (3, 4). Sample collection was performed using protocol SSC 2998 of the Scientific and Ethical Review Unit of the Kenya Medical Research Institute and protocol 14413 of the Washington State University institutional review board. The strain was isolated by spread-plating a stool sample on a MacConkey plate. After 12 to 18 h of incubation at 37°C, a single pink/rose colony was transferred to Luria-Bertani broth and stored at −20°C.

The selected *E. coli* isolate was revived on Trypticase soy agar for 24 h at 37°C, and genomic DNA was then extracted by using a Qiagen DNeasy blood and tissue kit. The concentration was determined by using a Qubit double-stranded DNA (dsDNA) assay kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Genomic DNA libraries were prepared with a Nextera XT library preparation kit (Illumina, USA), and paired-end sequencing was completed by using a MiSeq sequencer (Illumina, Inc., San Diego, CA, USA) with 2 × 250 cycles.

Raw reads were quality filtered and subsequently trimmed with Trimmomatic v0.38 (5) using the settings sliding window: 4:15, leading: 3, and minlen: 50. *De novo* assembly was performed with SPAdes v3.13.0 using default assembly parameters (6). The contigs obtained were filtered using in-house scripts (https://doi.org/10.6084/m9.figshare.17126777) to retain those that met minimum length and coverage criteria of 500 bp and 1.5× coverage, respectively, and the quality of the assembly was assessed using QUAST v5.0.2 (7). Multilocus sequence typing (MLST), plasmid, and resistome profiles were generated using staramr v0.7.1 (8–10), and a phage profile was generated using the PHASTER 2016 update (11). The assembled genomes were annotated using NCBI PGAP v5.3 (12). Default parameters were used for all software unless otherwise specified.

A total of 1,038,300 raw paired-end reads were obtained from sequencing, with 1,021,121 reads being retained after quality trimming, and the GC content (50.52%) was as expected. *De novo* assembly yielded 480 contigs, with a combined length of 5,140,290 bases and an N_{50} value of 552,852 bp. After removal of contigs that were <500 bp and lacked 1.5× coverage, the remaining data set was composed of 325 contigs, with an N_{50} value of 552,851 bp and 100× genome coverage. The total draft genome length was 5,075,726 bp.

The identified ST was ST636, and a Col(BS512) plasmid was present. Phage analysis identified three prophage regions, of which two regions were intact; these included PHAGE_Salmon_Fels_2_NC_010463 (35.9 kbp), PHAGE_Yersin_L_413C_NC_004745 (32.3 kbp), and PHAGE_Entero_lambda_NC_001416 (31.7 kbp). Annotation revealed that the ST636 strain contains 4,647 coding sequences, 4,739 putative genes, 9 rRNAs, 81 tRNAs, and 1 transfer-messenger RNA. Antimicrobial resistance genes, including *aph(6)-Id*, *bla*$_{TEM-1B}$, and *dfrA7*, which encode proteins for resistance to kanamycin, ampicillin, and trimethoprim, respectively, were found on the chromosome.

Data availability. The whole-genome sequence of the *Escherichia coli* ST636 strain has been deposited in GenBank under accession number GCA_020005225.1, and raw reads can be found under SRA accession number SRX13373163.

REFERENCES

1. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev 15:589–594. https://doi.org/10.1016/j.gde.2005.09.006.
2. Tettelin H, Riley D, Cattuto C, Medini D. 2008. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472–477. https://doi.org/10.1016/j.mib.2008.09.006.
3. Njuguna HN, Cosmas L, Williamson J, Nyachieo D, Olack B, Ochieng JB, Wamola N, Oundo JO, Feikin DR, Mintz ED, Breiman RF. 2013. Use of population-based surveillance to define the high incidence of shigellosis in an urban slum in Nairobi, Kenya. PLoS One 8:e58437. https://doi.org/10.1371/journal.pone.0058437.
4. Omulo S, Lofgren ET, Lockwood S, Thumbi SM, Bigogo G, Ouma A, Verani JR, Juma B, Njenga MK, Kariuki S, McElwain TF, Palmer GH, Call DR. 2021. Carriage of antimicrobial-resistant bacteria in a high-density informal settlement in Kenya is associated with environmental risk-factors. Antimicrob Resist Infect Control 10:18. https://doi.org/10.1186/s13756-021-00886-y.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. 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.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
8. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903. https://doi.org/10.1128/AAC.02412-14.
9. Jolley KA, Maiden MCJ. 2014. Using MLST to study bacterial variation: prospects in the genomic era. Future Microbiol 9:623–630. https://doi.org/10.2217/fmb.14.24.
10. Lokody I. 2014. RiboSNitches reveal heredity in RNA secondary structure. Nat Rev Genet 15:219. https://doi.org/10.1038/nrg3700.
11. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
12. 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.