



Complete Genome Sequence of a Canadian Strain of *Escherichia coli* with Multiple Metal and Antimicrobial Resistance Genes That Was Isolated from Municipal Biosolids

Mingsong Kang,^a Philippe Charron,^a Emily Hoover,^a Jiewen Guan,^a Isaac Firth,^a  Sohail Naushad,^a  Hongsheng Huang^a

^aCanadian Food Inspection Agency, Ottawa Laboratory-Fallowfield, Ottawa, Ontario, Canada

ABSTRACT This announcement reports the complete genome sequence of a non-Shiga toxin-producing *Escherichia coli* strain that was isolated from municipal biosolids collected from a Canadian wastewater treatment plant. This strain contains multiple metal, antimicrobial, and heat resistance genes, as determined by genome sequencing, and could be a useful bacterial model for future studies.

Escherichia coli is a Gram-negative bacterium, either harmless or pathogenic, that is commonly found in the intestines of humans and animals and in environmental habitats, including agricultural soils and biosolids (1–3). This article reports the genome sequence of an *E. coli* strain (HH107) that was isolated from a posttreated (biosolids) sample collected from a Canadian wastewater treatment plant in 2010. The strain was isolated by enrichment in lauryl sulfate tryptose broth at 35°C for 24 h, followed by *E. coli* broth at 45°C for 24 h, and isolation using Levine's eosin methylene blue agar at 35°C for 24 h (4). The strain was identified as *E. coli* using API 20 E (bioMérieux Canada Inc.) and confirmed by genome sequencing.

Genomic DNA of HH107 was extracted from an overnight culture that had been grown from a single colony in tryptic soy broth using the NanoBind CBB kit (Pacific Biosciences [PacBio], USA), with subsequent treatment using the Short Read Eliminator XS kit (PacBio). Illumina sequencing was conducted by library preparation using a DNA preparation kit (Illumina, USA) and sequencing on a MiSeq platform (Illumina) using the MiSeq reagent kit v3 to generate a total of 2,684,745 paired-end (300-bp) reads, which were filtered and trimmed using Fastp v0.23.2 (5). Nanopore sequencing was performed by MinION library preparation using a ligation sequencing-native barcoding kit (SQK-NBD112.24; Oxford Nanopore Technologies, UK) without shearing and sequencing using a FLO-MIN112 (R10.4.1) flow cell on a MinION Mk1B device. A total of 40,436 reads (N_{50} of 17,401 bp) were obtained, followed by base calling using Guppy v6.1.2, trimming using Porechop v0.2.4, and filtering using NanoFilt v2.8.0 (6). Illumina and MinION reads were hybrid assembled using Unicycler v0.5.0 (7). The circularity of the genome and genome rotation using *dnaA* as the starting point were determined by Unicycler v0.5.0. The coverage depth (262×) was assessed using SAMtools v1.13 (8). Gene predictions and annotations were performed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.4 (9). Metal resistance, heat resistance, and antimicrobial resistance (AMR) genes were identified using AMRFinderPlus v3.11.2 with database v2022-12-19.1 (10). The plasmids were identified by PlasmidFinder v2.0.1 using database v2023-01-12 (11), and prophage sequences were analyzed using the PHASTER web server (12). The serotype was identified using SeroTypeFinder v2.0.1 (13), and pathogenicity was predicted using PathogenFinder v1.1 (14) and VirulenceFinder v2.0.3 (15). Default parameters were used for all of the bioinformatic tools.

The isolate HH107 was predicted to be non-Shiga toxin-producing *E. coli* O18ab:H14. Its genome contains a single chromosome and three plasmids. Table 1 presents detailed information on total length, chromosome size, GC content, protein count, prophage

Editor David Rasko, University of Maryland School of Medicine

© Crown copyright 2023. 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/).

Address correspondence to Mingsong Kang, mingsong.kang@inspection.gc.ca, or Hongsheng Huang, hongsheng.huang@inspection.gc.ca.

The authors declare no conflict of interest.

Received 6 February 2023

Accepted 24 March 2023

Published 17 April 2023

TABLE 1 Genomic characteristics of the *E. coli* strain isolated from a biosolids sample

Strain	No. of contigs	No. of plasmids	Total length (Mb)	Chromosome size (bp)	GC content (%)	No. of proteins	Genes related to AMR ^a	Genes related to metal resistance ^a	Genes related to heat resistance ^a	Virulence genes ^b	No. of intact prophages
HH1107	4	3	5.05	4,795,218	50.9	4,632	<i>acrF</i> , <i>mdtM</i> , <i>blaEC</i> , <i>gfpT</i> (4 genes)	<i>silE</i> , <i>silS</i> , <i>silR</i> , <i>silC</i> , <i>silF</i> , <i>silB</i> , <i>silA</i> , <i>silP</i> , <i>pcoA</i> , <i>pcoB</i> , <i>pcoC</i> , <i>pcoD</i> , <i>pcoR</i> , <i>pcoS</i> , <i>pcoE</i> , <i>arsR</i> , <i>arsD</i> , <i>arsA</i> , <i>arsC</i> (19 genes)	<i>hsp20</i> , <i>clpK</i> , <i>shsP</i> , <i>yfdX1</i> , <i>yfdX2</i> , <i>hdeD-GI</i> , <i>trxLHR</i> , <i>kerB-GI</i> , <i>pst-GI</i> (9 genes)	<i>ami</i> , <i>clpKI</i> , <i>csqA</i> , <i>fdxG</i> , <i>fimH</i> , <i>gad</i> , <i>hha</i> , <i>hlyA</i> , <i>hlyE</i> , <i>nlpI</i> , <i>terC</i> , <i>traI</i> , <i>traT</i> , <i>yehA</i> , <i>yehB</i> , <i>yehC</i> , <i>yehD</i> (17 genes)	4
Median GenBank sequences ^c			5.10		50.6	4,724					

^a AMR, heavy metal resistance, and heat resistance genes were predicted by AMRFinderPlus.

^b Virulence genes were predicted by VirulenceFinder.

^c Data summarized on 27 February 2023 using 33,383 genome assemblies are available in a genome assembly and annotation report (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/167>).

count, AMR, heavy metal resistance, and heat resistance genes, and virulence genes. The median total length, number of coding sequences (CDSs), and GC content of *E. coli* genome assemblies in GenBank are similar to those of this *E. coli* strain (Table 1).

Data availability. The chromosome and plasmid sequences of strain HH107 have been deposited in GenBank under accession numbers [CP116405](#), [CP116406](#), [CP116407](#), and [CP116408](#). MinION and MiSeq raw data are available in the NCBI Sequence Read Archive (SRA) under SRA accession numbers [SRR23100672](#) and [SRR23100675](#), respectively.

ACKNOWLEDGMENT

This work was funded by the Canadian Food Inspection Agency.

REFERENCES

- Ingham SC, Losinski JA, Andrews MP, Breuer JE, Breuer JR, Wood TM, Wright TH. 2004. *Escherichia coli* contamination of vegetables grown in soils fertilized with noncomposted bovine manure: garden-scale studies. *Appl Environ Microbiol* 70:6420–6427. <https://doi.org/10.1128/AEM.70.11.6420-6427.2004>.
- Mays C, Garza GL, Waite-Cusic J, Radniecki TS, Navab-Daneshmand T. 2021. Impact of biosolids amendment and wastewater effluent irrigation on enteric antibiotic-resistant bacteria: a greenhouse study. *Water Res X* 13:100119. <https://doi.org/10.1016/j.wroa.2021.100119>.
- Luna-Guevara JJ, Arenas-Hernandez MMP, Martinez de la Pena C, Silva JL, Luna-Guevara ML. 2019. The role of pathogenic *E. coli* in fresh vegetables: behavior, contamination factors, and preventive measures. *Int J Microbiol* 2019:2894328. <https://doi.org/10.1155/2019/2894328>.
- Kang M, Chmara J, Naushad S, Huang H. 2021. Complete genome sequence of a Canadian strain of *Raoultella planticola* with metal and antimicrobial resistance genes. *Microbiol Resour Announc* 10:e00415-21. <https://doi.org/10.1128/MRA.00415-21>.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- 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>.
- Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 11:12728. <https://doi.org/10.1038/s41598-021-91456-0>.
- 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.
- 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>.
- Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 53:2410–2426. <https://doi.org/10.1128/JCM.00008-15>.
- Cosentino S, Voldby Larsen M, Moller Aarestrup F, Lund O. 2013. Pathogen-Finder: distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 8:e77302. <https://doi.org/10.1371/journal.pone.0077302>.
- Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* 4:e27943. <https://doi.org/10.4161/bact.27943>.