



# Hybrid Nanopore-Illumina Assemblies for Five Extraintestinal Pathogenic *Escherichia coli* Isolates

D. Mattrasingh,<sup>a\*</sup> A. Hinz,<sup>a</sup> L. Phillips,<sup>a</sup> A. C. Carroll,<sup>a</sup>  A. Wong<sup>a</sup>

<sup>a</sup>Department of Biology, Carleton University, Ottawa, Canada

**ABSTRACT** Extraintestinal pathogenic *Escherichia coli* (ExPEC) is an important source of multidrug-resistant infections, particularly in hospitals. We report hybrid Nanopore-Illumina assemblies for 5 ExPEC isolates with various drug resistance profiles.

Extraintestinal pathogenic *Escherichia coli* (ExPEC) causes serious illnesses, including blood and urinary tract infections (UTIs), and has a wide variety of antibiotic resistance profiles (1). Five ExPEC isolates with various degrees of documented antibiotic resistance were collected as part of the CANWARD survey of antibiotic-resistant pathogens in Canada (2, 3) (kindly shared by the Zhanel laboratory). Two isolates were derived from UTIs (PB3 and PB4), two from blood infections (PB29 and PB35), and one from a respiratory infection (PB33). Isolates were collected under University of Manitoba Research Ethics Board approval (H2009:O59).

Strains were cultured overnight at 37°C in lysogeny broth (1% tryptone, 0.5% yeast extract, and 1% NaCl), and DNA was extracted using the One-4-All genomic DNA mini-prep kit (BioBasic, Markham, Canada). Short reads were generated on the Illumina NextSeq platform using 150-bp paired-end reads with Nextera XT library preparation, generating a total of 162,925,724 clusters passing filter. Long reads were generated on the Nanopore MinION platform using the rapid barcoding kit and Guppy v3.1.5 base calling, generating a total of 697,371 reads and a mean read length of 1,176 bases. Quality scoring before and after trimming was performed using FastQC v0.11.7 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and Illumina reads were trimmed using Trimmomatic v0.38 (4) with the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Porechop v0.2.4 (<https://github.com/rwick/Porechop>) was used to remove adapter sequences from the Nanopore reads. Hybrid assemblies were generated using the SPAdes optimizer Unicycler v0.4.8 (5). QUAST v4.6 (6) was used to produce summary statistics for the assembly. Genome annotations were carried out with the NCBI Prokaryotic Genome Annotation Pipeline v4.12 (7), serotypes were predicted using SeroTypeFinder v2.0.1 (8), and antimicrobial resistance (AMR) profiles were predicted with ResFinder v3.1 (9). Default parameter settings were used for all software unless otherwise indicated.

The genomes varied between 4,880,873 bp (PB4) and 5,309,474 bp (PB29) in length and maintained a GC content from 50.6 to 50.85 (Table 1). The strains harbored similar numbers of coding DNA sequences (CDS), rRNAs, and tRNAs. PB4 and PB29 carried two predicted CRISPR sequences each, while the others carried none. The strains carried between 0 and 6 chromosomal resistance mutations and between 1 and 14 resistance genes (Table 1). PB33 and PB35 both belong to serotype O25:H4-ST131, a major epidemic clone of ExPEC (10, 11).

We have provided *de novo* hybrid genome assemblies of five extraintestinal pathogenic *E. coli* isolates. Three of the isolates exhibit a genomic signature of multidrug resistance characterized by 14 or more resistance mutations/genes (PB29, PB33, and PB35). The remaining two isolates have relatively fewer resistance determinants (PB3 and PB4). These

**Citation** Mattrasingh D, Hinz A, Phillips L, Carroll AC, Wong A. 2021. Hybrid Nanopore-Illumina assemblies for five extraintestinal pathogenic *Escherichia coli* isolates. *Microbiol Resour Announc* 10:e01027-20. <https://doi.org/10.1128/MRA.01027-20>.

**Editor** Christina A. Cuomo, Broad Institute

**Copyright** © 2021 Mattrasingh 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/).

Address correspondence to A. Wong, alex.wong@carleton.ca.

\* Present address: D. Mattrasingh, University of Guelph, Guelph, Canada.

**Received** 3 September 2020

**Accepted** 26 January 2021

**Published** 11 February 2021

**TABLE 1** Assembly and genome characteristics for 5 ExPEC isolates

Strain	Total length (bp)	No. of contigs	Largest contig (bp)	Contig $N_{50}$	Coverage <sup>a</sup>	No. of protein CDS <sup>b</sup>	No. of ARG <sup>c</sup>	No. of ARM <sup>d</sup>	ESBL <sup>e</sup>	Serotype
PB3	5,224,675	48	4,043,831	4,043,831	104; 23	4,793	2	0	None	O6:H1
PB4	4,880,873	23	1,168,174	688,147	70; 31	4,414	1	1	None	O—:H9
PB29	5,309,474	55	1,155,588	473,125	57; 15	4,909	13	0	CTX-M-15	O—:H9
PB33	5,168,956	27	1,099,130	710,413	25; 45	4,772	8	6	CTX-M-15	O25:H4
PB35	5,236,721	41	871,437	497,725	34; 28	4,792	14	5	CTX-M-15	O25:H4

<sup>a</sup> Mean fold coverage for Illumina reads; Nanopore reads, rounded to the nearest whole number.

<sup>b</sup> Predicted protein coding sequences, excluding pseudogenes.

<sup>c</sup> Number of predicted antibiotic resistance genes (ARG).

<sup>d</sup> Number of predicted antibiotic resistance mutations (ARM).

<sup>e</sup> Type (if any) of extended-spectrum beta-lactamase (ESBL) gene present in the genome.

genomes will support future investigations into *E. coli* pathogenesis and resistance evolution.

**Data availability.** The assemblies and raw sequence data have been deposited in GenBank under BioProject [PRJNA648312](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA648312), with accession numbers [JACFYA000000000](https://www.ncbi.nlm.nih.gov/assembly/GCA_014042275.1) (assembly number [GCA\\_014042275.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_014042275.1)), [JACFYB000000000](https://www.ncbi.nlm.nih.gov/assembly/GACFYB000000000) ([GCA\\_014042265.1](https://www.ncbi.nlm.nih.gov/assembly/GACFYB000000000)), [JACFYC000000000](https://www.ncbi.nlm.nih.gov/assembly/GACFYC000000000) ([GCA\\_014042305.1](https://www.ncbi.nlm.nih.gov/assembly/GACFYC000000000)), [JACFYD000000000](https://www.ncbi.nlm.nih.gov/assembly/GACFYD000000000) ([GCA\\_014042325.1](https://www.ncbi.nlm.nih.gov/assembly/GACFYD000000000)), and [JACFYE000000000](https://www.ncbi.nlm.nih.gov/assembly/GACFYE000000000) ([GCA\\_014265895.1](https://www.ncbi.nlm.nih.gov/assembly/GACFYE000000000)). The versions described in this paper are the first versions, [JACFYA010000000](https://www.ncbi.nlm.nih.gov/assembly/JACFYA010000000) to [JACFYE010000000](https://www.ncbi.nlm.nih.gov/assembly/JACFYE010000000).

## ACKNOWLEDGMENTS

We appreciate the comments of an anonymous reviewer in improving the manuscript. We thank Catherine Carrillo and Paul Manninger at the Canadian Food Inspection Agency for assistance with Nanopore sequencing.

This work was funded by the Natural Sciences and Engineering Research Council (NSERC Discovery Grant RGPIN-2018-05340).

## REFERENCES

- Dale AP, Woodford N. 2015. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): disease, carriage and clones. *J Infect* 71:615–626. <https://doi.org/10.1016/j.jinf.2015.09.009>.
- Zhanel GG, Adam HJ, Baxter MR, Fuller J, Nichol KA, Denisuik AJ, Lagace-Wiens PR, Walkty A, Karlowsky JA, Schweizer F, Hoban DJ, Canadian Antimicrobial Resistance Alliance. 2013. Antimicrobial susceptibility of 22746 pathogens from Canadian hospitals: results of the CANWARD 2007–11 study. *J Antimicrob Chemother* 68(Suppl 1):i7–i22. <https://doi.org/10.1093/jac/dkt022>.
- Basra P, Alsaadi A, Bernal-Astrain G, O'Sullivan ML, Hazlett B, Clarke LM, Schoenrock A, Pitre S, Wong A. 2018. Fitness tradeoffs of antibiotic resistance in extraintestinal pathogenic *Escherichia coli*. *Genome Biol Evol* 10:667–679. <https://doi.org/10.1093/gbe/evy030>.
- 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>.
- 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>.
- 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>.
- Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34:1037–1039. <https://doi.org/10.1093/bioinformatics/btx713>.
- Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheut 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>.
- 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>.
- Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, Baquero F, Canton R, Nordmann P. 2008. Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis* 14:195–200. <https://doi.org/10.3201/eid1402.070350>.
- Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V, Demarty R, Alonso MP, Canica MM, Park YJ, Lavigne JP, Pitout J, Johnson JR. 2008. Intercontinental emergence of *Escherichia coli* clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* 61:273–281. <https://doi.org/10.1093/jac/dkm464>.