

Complete genome sequences of two avian pathogenic *Escherichia coli* strains isolated from broilers exhibiting colibacillosis in Mississippi

Linan Jia,¹ Mark A. Arick II,² Chuan-Yu Hsu,² Daniel G. Peterson,² Jeffrey D. Evans,³ Kelsy Robinson,³ Pratima Adhikari,¹ Li Zhang¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the whole-genome sequences of *Escherichia coli* strains APEC-O2-MS1266 and APEC-O2-MS1657 isolated from the liver and heart of infected broilers in Mississippi State, US. The genomic information of these two causative strains may provide a valuable reference for comparative studies of avian pathogenic *E. coli*.

KEYWORDS avian pathogenic *Escherichia coli*, poultry farm, whole-genome sequencing

Avian pathogenic *Escherichia coli* is the causative agent of poultry colibacillosis, a disease with high mortality (1, 2). The genomic information of highly infectious strains can help veterinarians make efficient disease control decisions. Two *E. coli* strains were isolated previously from the hepatic and cardiac lesions in broilers that had experienced colibacillosis (3).

Lesion swabs were streaked on MacConkey agar in triplicate and incubated overnight at 37°C. Pink colonies were randomly selected and sub-cultured three times with the same conditions for pure isolation. Genomic DNA for both Nanopore and Illumina sequencing was extracted from cultures grown overnight in a 37°C LB broth using a GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, Waltham, MA, USA). The *ybbW* gene, a gene in *E. coli* core genome and have 100% inclusivity and 100% exclusivity for *E. coli*, was amplified using qPCR for *E. coli* confirmation (4). The quantity and quality of DNA were measured using Qubit fluorometer and electrophoresis with a 0.8% wt/vol agarose gel. For long-read sequencing, the genomic DNA was fragmented with g-Tube (Covaris, Woburn, MA, USA) following the manufacturer's procedure to generate the mean fragment size of 12–15 kb. The multiplexing library pool was prepared and barcoded using a Ligation Sequencing Kit (SQK-LSK109) and the Native Barcoding Kit (SQK-NBD104), and sequenced on an R9.4 MinION flow cell using the Nanopore GridION sequencer (Oxford Nanopore Technologies, Oxford, UK). Guppy v6.3.2 (Oxford Nanopore Technologies, Oxford, UK) was used for data base-calling. The same genomic DNA was used for short-read sequencing and a 350-bp short insert DNA-Seq library for each sample was prepared by using Illumina TruSeq DNA PCR-free Sample Prep Kit and sequenced with PE150 sequencing method on a HiSeq X-Ten sequencer (Illumina, San Diego, CA, USA). The short-reads were filtered using cutadapt (v4.3) (5) to remove any read with a TruSeq adapter sequence, and the long-reads were filtered using filtlong (v0.2.1) (6) to discard any read shorter than 1 kb and the worst 10% base on kmer overlap of the filtered short-reads. Flye (v2.9) (7) was used to create a preliminary assembly using the filtered long-reads. Unicycler (v0.5.0) (8) used the Flye assembly, filtered long-reads, and filtered short-reads to create a final, complete, circular assembly for both strains. Base coverage was calculated using SAMtools (v1.15.1) (9) with the bwa (v0.7.17) (10) aligned raw short-reads and minimap2 (v2.14) (11) aligned raw

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

Address correspondence to Li Zhang, lzhang@msstate.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 1 November 2023

Accepted 17 April 2024

Published 30 April 2024

Copyright © 2024 Jia 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 Summary of genome assembly and genotypic characteristics of two *E. coli* isolates

Label	Source	Serotype	ST ^a	Size (bp)	Avg. per-base coverage (x)		GC content (%)	Nanopore <i>N</i> ₅₀	CDSs (with protein)	GenBank accession no.	
					Nanopore	Illumina					
APEC-O2-MS1266 ^b	Liver	O2/O50:H5	355	4,735,863	300	304	51	8,936	4,571	CP135959	
pAPEC-O2-MS1266-1 ^c				134,102	124	139				49	CP135960
pAPEC-O2-MS1266-2 ^c				122,100	145	217				54	CP135961
pAPEC-O2-MS1266-3 ^c				1,552	6	8,066				52	CP135962
APEC-O2-MS1657 ^b	Heart	O2/O50:H1	429	5,059,841	144	319	51	8,866	4,883	CP135957	
pAPEC-O2-MS1657-1 ^c				191,500	176	389				50	CP135958

^aST: sequencing type.

^bChromosome.

^cPlasmid.

long-reads. Both the assemblies have BUSCO scores above 99.3% using BUSCO (v5.5.0) (12) and the enterobacterales_odb10 database. The sequences were annotated using the Prokaryotic Genome Annotation Pipeline (PGAP, v6.6) (13) at NCBI. Multilocus sequence typing (MLST) was identified using PubMLST (14), serotyping (O-antigen and flagellin genes) was performed using SerotypeFinder (v.2.0) (15). Default parameters were used for all software unless otherwise specified. Nanopore generated 194,007 reads with an average length of 7,731 bp for MS1266, and 103,532 reads with an average length of 7,626 bp for MS1657. Illumina generated 4,994,817 and 5,631,894 pairs for MS1266 and MS1657, respectively. Table 1 shows the genome assembly and genotypic characteristics of two *E. coli* isolates.

ACKNOWLEDGMENTS

The project was funded by USDA-ARS SCA no. 6064-13000-013-00D, USDA-ARS NACA 58-6066-0-064, USDA ARS CRIS project MIS-322430/NE-1942, and the Mississippi Agricultural and Forestry Experiment Station.

AUTHOR AFFILIATIONS

¹Poultry Science, Mississippi State University, Starkville, Mississippi, USA

²Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, Mississippi, USA

³Poultry Research Unit, USDA, Agriculture Research Service, Starkville, Mississippi, USA

AUTHOR ORCIDs

Li Zhang  <http://orcid.org/0000-0002-3933-5794>

FUNDING

Funder	Grant(s)	Author(s)
U.S. Department of Agriculture (USDA)	NACA 58-6064-2-014, CRIS MIS-322430/ NE-1942	Li Zhang
U.S. Department of Agriculture (USDA)	NACA 58-6066-0-064	Daniel G. Peterson
MSU Mississippi Agricultural and Forestry Experiment Station, Mississippi State University (MAFES, MSU)		Li Zhang

AUTHOR CONTRIBUTIONS

Linan Jia, Data curation, Formal analysis, Writing – original draft | Mark A. Arick II, Data curation, Methodology, Writing – review and editing | Chuan-Yu Hsu, Data curation, Methodology, Writing – review and editing | Daniel G. Peterson, Funding acquisition,

Project administration, Writing – review and editing | Jeffrey D. Evans, Funding acquisition, Project administration, Writing – review and editing | Kelsy Robinson, Funding acquisition, Project administration, Writing – review and editing | Pratima Adhikari, Funding acquisition, Writing – review and editing | Li Zhang, Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome sequences and raw data are available at NCBI under the BioProject [PRJNA839731](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA839731). The assembled genome sequences and annotations are available at GenBank under accessions [CP135957](https://www.ncbi.nlm.nih.gov/nuccore/CP135957)-[CP135962](https://www.ncbi.nlm.nih.gov/nuccore/CP135962) (Table 1). The raw data are available at the SRA (Sequence Read Archive, <https://www.ncbi.nlm.nih.gov/sra/>) under accessions [SRR26195173](https://www.ncbi.nlm.nih.gov/sra/SRR26195173), [SRR26195174](https://www.ncbi.nlm.nih.gov/sra/SRR26195174), [SRR26195175](https://www.ncbi.nlm.nih.gov/sra/SRR26195175), and [SRR26195176](https://www.ncbi.nlm.nih.gov/sra/SRR26195176).

REFERENCES

- Kim YB, Yoon MY, Ha JS, Seo KW, Noh EB, Son SH, Lee YJ. 2020. Molecular characterization of avian pathogenic *Escherichia coli* from broiler chickens with colibacillosis. *Poult Sci* 99:1088–1095. <https://doi.org/10.1016/j.psj.2019.10.047>
- Nolan LK, Barnes HJ, Vaillancourt J-P, Abdul-Aziz T, Logue CM. 2013. Colibacillosis, p 751–805. In *Diseases of poultry*
- JiaL, Devkota P, Arick IIMA, HsuC-Y, PetersonDG, EvansJD. 2022. Complete genome sequence of six multidrug resistance avian pathogenic *Escherichia coli* strains isolated from broilers exhibiting colibacillosis, Abstr 349P. In *Abstr Poult SCI Assoc Annu Mtg. Poultry Science Association, San Antonio, TX.*
- Walker DI, McQuillan J, Taiwo M, Parks R, Stenton CA, Morgan H, Mowlem MC, Lees DN. 2017. A highly specific *Escherichia coli* qPCR and its comparison with existing methods for environmental waters. *Water Res* 126:101–110. <https://doi.org/10.1016/j.watres.2017.08.032>
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
- Wick RR. 2017. *Filtlong*. Available from: <https://github.com/rrwick/Filtlong>
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Wick RR, Judd LM, Gorrie CL, Holt KE, Phillippy AM. 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>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>
- Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
- Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124. <https://doi.org/10.12688/wellcomeopenres.14826.1>
- Joensen KG, Tetzschner AMM, 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>