

## Complete Genome Sequences of Extraintestinal Pathogenic *Escherichia coli* Clinical Isolates from Danish Ulcerative Colitis Patients

Meski, Simon; Struve, Carsten; Mirsepasi-Lauridsen, Hengameh C.; Petersen, Andreas M.; Jelsbak, Lotte; Skovgaard, Ole; Krogfelt, Karen A.

*Published in:*  
Microbiology Resource Announcements

*DOI:*  
[10.1128/mra.01186-22](https://doi.org/10.1128/mra.01186-22)

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Meski, S., Struve, C., Mirsepasi-Lauridsen, H. C., Petersen, A. M., Jelsbak, L., Skovgaard, O., & Krogfelt, K. A. (2023). Complete Genome Sequences of Extraintestinal Pathogenic *Escherichia coli* Clinical Isolates from Danish Ulcerative Colitis Patients. *Microbiology Resource Announcements*, 12(2).  
<https://doi.org/10.1128/mra.01186-22>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact [rucforsk@kb.dk](mailto:rucforsk@kb.dk) providing details, and we will remove access to the work immediately and investigate your claim.



# Complete Genome Sequences of Extraintestinal Pathogenic *Escherichia coli* Clinical Isolates from Danish Ulcerative Colitis Patients

 Simon Meski,<sup>a</sup>  Carsten Struve,<sup>b</sup>  Hengameh C. Mirsepasi-Lauridsen,<sup>a,b,c</sup>  Andreas M. Petersen,<sup>c,d,e</sup>  Lotte Jelsbak,<sup>a</sup>  
 Ole Skovgaard,<sup>a</sup>  Karen A. Krogfelt<sup>a,b,f</sup>

<sup>a</sup>Department of Science and Environment, Roskilde University, Roskilde, Denmark

<sup>b</sup>Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark

<sup>c</sup>Department of Clinical Microbiology, Copenhagen University Hospital—Amager and Hvidovre, Copenhagen, Denmark

<sup>d</sup>Department of Gastroenterology, Copenhagen University Hospital—Amager and Hvidovre, Copenhagen, Denmark

<sup>e</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

<sup>f</sup>Department of Virus and Microbiological Special Diagnostics, Statens Serum Institut, Copenhagen, Denmark

**ABSTRACT** Extraintestinal pathogenic *Escherichia coli* (ExPEC) is a potential factor in ulcerative colitis etiology. We report here the complete genome and plasmid sequences of three *Escherichia coli* isolates, C 237-04 (p7), C 236-04A (p10A), and C 691-04A (p19A), obtained from fecal samples from ulcerative colitis patients in Copenhagen, Denmark.

Ulcerative colitis (UC) is characterized by periods of colonic mucosal inflammation, including signs and symptoms such as diarrhea, rectal bleeding, and stomach pain, followed by periods of remission. The disease etiology remains unsolved, but several host and environmental factors have been implicated (1, 2), including an association with virulent and pathogenic *Escherichia coli* (3–6). Studying *E. coli* strains isolated from UC patients helps us understand their role in UC etiology. Three clinical *E. coli* strains were isolated from patients' fecal samples as part of a Danish clinical case-control study on inflammatory bowel disease (IBD), with informed written consent from participants and permission from the Regional Ethics Committee for Copenhagen County Hospitals (permission number KA03019).

The *E. coli* strains were isolated by suspending feces in phosphate-buffered saline, plating them on SSI selective enteric medium (number 724; SSI Diagnostica, Hillerød, Denmark) (7, 8), and incubating the plates at 37°C overnight. The colonies were assessed visually as *E. coli*, the species were confirmed using a Minibact E kit (9), and the cultures were stored in glycerol at –80°C. DNA was purified from fresh cultures, incubated at 37°C overnight in LB medium, using the phenol-chloroform method as previously described (10). Separate DNA batches were prepared for short- and long-read sequencing. The purity and concentrations were measured using a NanoDrop One spectrophotometer.

Sequencing and postprocessing were performed as described by Lallement et al. (11) using a combined short- and long-read sequencing approach. The read length of long-read sequencing was favored by omitting shearing and size selection of DNA before library preparation using an SQK-LSK109 ligation kit and sequencing on an Oxford Nanopore MinION instrument with a R9.4.1 flow cell. Base calling of the fast5 files to fastq files was performed using Guppy v5.0.7, with the dna\_r9.4.1\_450bps\_sup model. Reads with a length of more than 10,000 nucleotides (nt) were selected using Filtrlong v0.2.1 (<https://github.com/rwrick/Filtrlong>) and assembled into scaffolds using Flye v2.9 (12) with -g 5 -nano-hq arguments or with Minimap2 v2.24 Miniasm v0.3 (13) assembly and two iterations of Minimap2 Racon v1.5.0 polishing (MMR) (14). The Minimap2 settings used were -x ava-ont for assembly and -x asm5 for polishing.

**Editor** David Rasko, University of Maryland School of Medicine

**Copyright** © 2023 Meski 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 Karen A. Krogfelt, karenak@ruc.dk, or Ole Skovgaard, olesk@ruc.dk.

The authors declare no conflict of interest.

**Received** 9 November 2022

**Accepted** 8 January 2023

**Published** 30 January 2023

**TABLE 1** Sequencing and assembly statistics

Isolate	Replicon	No. of reads		Coverage (x)		ONT $N_{50}$ (bp)	Size (bp)	GC content (%)	No. of predicted CDS <sup>c</sup>	SRA <sup>d</sup> accession no.	
		BGI <sup>a</sup>	ONT <sup>b</sup>	BGI	ONT					BGI	ONT
C 237-04	p7 (Chr) <sup>e</sup>	11,632,828	41,675	349	66	25,238	4,918,902	50.57	4,630	CP109921	SRR21857883
	pP7_1	11,433,789		1,203	0		6,230	49.31	8	CP109922	SRR21857880
	pP7_2	49,975		4,136	0		1,459	50.58	1	CP109923	
C 236-04A	p10A (Chr)	11,648,518	8,522	347	13	29,467	4,987,582	50.85	4,734	CP109924	SRR21857884
	pP10A_1	11,521,884		1,010	0		4,109	45.44	5	CP109925	
	pP10A_2	27,678		1,097	0		3,086	46.01	4	CP109926	
	pP10A_3	22,567		978	0		2,455	48.80	4	CP109927	
	pP10A_4	16,005		1,464	0		1,240	46.13	1	CP109928	
C 691-04A	p19A (Chr)	11,649,154	174,807	328	45	4,904	5,176,167	50.46	4,826	CP109929	SRR21857885
	pP19A_1	11,309,943		582	123		77,976	52.26	100	CP109930	SRR21857882

<sup>a</sup> BGI, BGISEQ sequencing; no. of reads mapped to each replicon using Bowtie2 v2.4.5 (19).

<sup>b</sup> ONT, Oxford Nanopore Technologies.

<sup>c</sup> CDS, coding DNA sequences.

<sup>d</sup> SRA, Sequence Read Archive.

<sup>e</sup> Chr, chromosome.

Paired-end 150-bp short-read sequencing was performed by BGI Europe A/S on the BGISEQ-500 platform, and the reads were filtered and trimmed using fastp (15) with the quality filter settings -q 25 and -u 10. Hybrid assemblies from these scaffolds and short sequence reads were constructed using SPAdes v3.15, with the -isolate and -trusted-contigs settings (16), and Unicycler v0.5.0 (17). These hybrid assemblies further identified the small circular plasmids in C 237-04 and C 236-04A solely from short sequence reads; using Unicycler, all replicon sequences were rotated to begin with *dnaA* (for chromosomes) or *repA* (for plasmids). Sequence annotation was performed by NCBI using PGAP v6.3 (18). Default parameters were used for all software unless otherwise specified. Sequence and assembly statistics are presented in Table 1.

Serotyping and virulence characterization (7) were confirmed *in silico* using Web applications provided by the Center for Genomic Epidemiology, DTU (20), and EZClermont v0.6.3 (21). Isolates p7 and p19A were found to be genotypically more virulent than p10A, corresponding to the inflammatory stage of the patients from active and inactive colitis, respectively (7).

**Data availability.** This whole-genome sequencing project has been deposited at GenBank under accession number [PRJNA882345](https://www.ncbi.nlm.nih.gov/PRJNA882345). The version described in this paper is the first version. The accession numbers for the raw reads and assemblies are provided in Table 1.

## ACKNOWLEDGMENT

We thank Jerry Wells of Wageningen University and Research for providing a preliminary sequence for isolate p19A. LJ was supported by a grant from the Novo Nordisk Foundation, Grant number NNF19OC0058547.

## REFERENCES

- Ristow LC, Welch RA. 2016. Hemolysin of uropathogenic *Escherichia coli*: a cloak or a dagger? *Biochim Biophys Acta* 1858:538–545. <https://doi.org/10.1016/j.bbame.2015.08.015>.
- Lee M, Chang EB. 2021. Inflammatory bowel diseases (IBD) and the microbiome—searching the crime scene for clues. *Gastroenterology* 160:524–537. <https://doi.org/10.1053/j.gastro.2020.09.056>.
- Mirsepasi-Lauridsen HC, Halkjaer SI, Mortensen EM, Lydolph MC, Nordgaard-Lassen I, Krogfelt KA, Petersen AM. 2016. Extraintestinal pathogenic *Escherichia coli* are associated with intestinal inflammation in patients with ulcerative colitis. *Sci Rep* 6:31152. <https://doi.org/10.1038/srep31152>.
- Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA, Petersen AM. 2019. *Escherichia coli* pathobionts associated with inflammatory bowel disease. *Clin Microbiol Rev* 32:e00060-18. <https://doi.org/10.1128/CMR.00060-18>.
- Kamali DR, Feizi A, Halaji M, Fazeli H, Adibi P. 2021. The prevalence of adherent-invasive *Escherichia coli* and its association with inflammatory bowel diseases: a systematic review and meta-analysis. *Front Med (Lausanne)* 8:730243. <https://doi.org/10.3389/FMED.2021.730243>.
- Petersen AM, Halkjaer SI, Gluud LL. 2015. Intestinal colonization with phylogenetic group B2 *Escherichia coli* related to inflammatory bowel disease: a systematic review and meta-analysis. *Scand J Gastroenterol* 50:1199–1207. <https://doi.org/10.3109/00365521.2015.1028993>.
- Petersen AM, Nielsen EM, Litrup E, Brynskov J, Mirsepasi H, Krogfelt KA. 2009. A phylogenetic group of *Escherichia coli* associated with active left-sided inflammatory bowel disease. *BMC Microbiol* 9:171. <https://doi.org/10.1186/1471-2180-9-171>.
- Blom M, Meyer A, Gerner-Smith P, Gaarslev K, Espersen F. 1999. Evaluation of Statens Serum Institut enteric medium for detection of enteric pathogens. *J Clin Microbiol* 37:2312–2316. <https://doi.org/10.1128/JCM.37.7.2312-2316.1999>.
- Kjaeldgaard P, Nissen B, Lange N, Laursen H. 1986. Evaluation of Minibact, a new system for rapid identification of Enterobacteriaceae. Comparison of Minibact, Micro-ID and API 20E with a conventional method as reference. *Acta Pathol Microbiol Immunol Scand B* 94:57–61.
- Rasool FN, Saavedra MA, Pamba S, Perold V, Mmochi AJ, Maalim M, Simonsen L, Buur L, Pedersen RH, Syberg K, Jelsbak L. 2021. Isolation and characterization of human pathogenic multidrug resistant bacteria associated with plastic litter collected in Zanzibar. *J Hazard Mater* 405:124591. <https://doi.org/10.1016/j.jhazmat.2020.124591>.
- Lallement C, Krogfelt KA, Skovgaard O, Jelsbak L. 2022. Complete genome sequence of a multidrug-resistant *Klebsiella pneumoniae* environmental isolate from Zanzibar, Tanzania, harboring novel insertion elements and two *bla*<sub>CTX-M-15</sub> genes. *Microbiol Resour Announc* 11:e0026322. <https://doi.org/10.1128/MRA.00263-22>.
- 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>.
- Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
- 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>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
- 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 W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin 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>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Technical University of Denmark. 2011. Center for Genomic Epidemiology. <http://www.genomic epidemiology.org/services/>. Accessed 30 November 2022.
- Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5:58–65. <https://doi.org/10.1111/1758-2229.12019>.