Complete Genome Sequences of Seven Uropathogenic *Escherichia coli* Strains Isolated from Postmenopausal Women with Recurrent Urinary Tract Infection

Belle M. Sharon, Amber Nguyen, Amanda P. Arute, Neha V. Hulyalkar, Vivian H. Nguyen, Philippe E. Zimmern, Nicole J. De Nisco

**ABSTRACT** Uropathogenic *Escherichia coli* (UPEC) is the most common cause of urinary tract infection (UTI). This disease disproportionately affects women and frequently develops into recurrent UTI (rUTI) in postmenopausal women. Here, we report the complete genome sequences of seven UPEC isolates obtained from the urine of postmenopausal women with rUTI.

Urinary tract infection (UTI) is a serious public health concern (1). The most common etiologic agent of UTI is uropathogenic *Escherichia coli* (UPEC) (1, 2). The lifetime incidence rate of UTI in women is >60% (3). Approximately 50% of UTIs in postmenopausal women develop into recurrent UTI (rUTI), defined as ≥2 UTIs within 6 months or ≥3 UTIs within 12 months (4–6).

UPEC strains from postmenopausal women with rUTI are understudied. The availability of complete genomes enables analysis of the mobile genetic elements and resistance genes that complicate the management of rUTI and may inform new treatment strategies. We report the complete genomes of seven UPEC strains (Table 1) isolated from the urine of consenting postmenopausal women meeting criteria for uncomplicated rUTI as part of an institutional review board (IRB)-approved study (STU 032016-006, MR 17-120) (7).

Clean-catch midstream urine was collected from seven subjects and plated onto CHROMagar Orientation (BD) and blood agar media. After overnight incubation at 37°C, well-isolated single colonies were picked for species identification by Sanger sequencing of the 16S rRNA gene and a nonredundant/nucleotide (nr/nt) database query using MegaBLAST (BLAST v2.10.0) (7, 8). Identified UPEC isolates were grown in brain heart infusion broth overnight at 37°C, and genomic DNA was extracted using a DNeasy blood and tissue kit (Qiagen).

The genomic DNA, assessed by the 260/280-nm absorbance ratio and agarose gel electrophoresis, was sequenced using Illumina and Oxford Nanopore (ONT) technologies. Illumina library preparation and sequencing were performed using the Nextera DNA Flex library prep kit and the NextSeq 500 system (300 cycles) to generate 2 × 150-bp paired-end reads. ONT libraries were prepared using the ligation sequencing kit (SQK-LSK109) and barcode expansion kit 1-12 (EXP-NBD104); then, they were sequenced on the MinION instrument using R9 FLO-MIN106 flow cells. Live fast base calling, demultiplexing, and barcode trimming were performed using ONT MinKNOW software.

Illumina reads were quality assessed and trimmed using CLC Genomics Workbench v12.0.3 trimming reads with a Phred score below 20. Reads of <15 bp were discarded. ONT reads were assessed for quality using NanoStats v1.2.0 (9) and trimmed using...
| Strain | BioSample accession no. | SRA accession no. | Total no. of reads | N₀ (bp) | Read depth (×) | MLST a | GenBank accession no. | Type of contig (circular) | Total length (bp) | GC content (%) | No. of CDSs c | Plasmid replicon(s) |
|--------|------------------------|------------------|------------------|--------|---------------|--------|-----------------------|------------------------|------------------|---------------|---------------|-------------------|
| EcPF5  | SAMN15075992           | SRX8452297 (O)   | 270,764          | 8,643  | 273           | 73     | CP054236              | Chromosome            | 5,147,412        | 50.4          | 4,747         | NA b              |
|        | SRX8452051 (I)         |                  | 2,717,378        | 47     |               |        |                       |                        |                 |               |               |                   |
|        | SAMN15075993           | SRX8452298 (O)   | 261,756          | 3,872  | 122           | 2279   | CP054232              | Chromosome            | 4,996,527        | 50.7          | 4,585         | NA b              |
|        | SRX8452052 (I)         |                  | 2,614,770        | 74     |               |        |                       |                        |                 |               |               |                   |
| EcPF7  | SAMN15075994           | SRX8452299 (O)   | 526,610          | 4,967  | 338           | 73     | CP054230              | Chromosome            | 5,129,852        | 50.5          | 4,725         | NA b              |
|        | SRX8452053 (I)         |                  | 2,914,068        | 83     |               |        |                       |                        |                 |               |               |                   |
| EcPF14 | SAMN15075995           | SRX8452300 (O)   | 122,027          | 17,034 | 250           | 394    | CP054227              | Chromosome            | 4,796,742        | 50.5          | 4,342         | NA b              |
|        | SRX8452054 (I)         |                  | 2,373,392        | 70     |               |        |                       |                        |                 |               |               |                   |
| EcPF16 | SAMN15075996           | SRX8452301 (O)   | 390,622          | 4,192  | 197           | 216    | CP054224              | Chromosome            | 4,721,932        | 51.0          | 4,430         | NA b              |
|        | SRX8452055 (I)         |                  | 8,965,752        | 231    |               |        |                       |                        |                 |               |               |                   |
| EcPF18 | SAMN15075997           | SRX8452302 (O)   | 196,886          | 7,168  | 161           | 131    | CP054219              | Chromosome            | 5,010,549        | 50.7          | 4,679         | NA b              |
|        | SRX8452056 (I)         |                  | 12,568,424       | 310    |               |        |                       |                        |                 |               |               |                   |
| EcPF40 | SAMN15075998           | SRX8452303 (O)   | 426,633          | 5,656  | 315           | 1193   | CP054214              | Chromosome            | 5,025,664        | 50.6          | 4,704         | NA b              |
|        | SRX8452057 (I)         |                  | 2,296,172        | 64     |               |        |                       |                        |                 |               |               |                   |

a The technology used is shown in parentheses. O, ONT; I, Illumina.
b MLST, multilocus sequence type.
c CDSs, coding sequences.
d NA, not applicable.
if found. The hybrid assembly quality was assessed using QUAST v5.0.2 (14). Genome completeness was evaluated using Bandage v0.8.1 (15) and BUSCO v1 (16) with bacteria ortholog set on the gVolante server v1.2.1 (17). Coverage of core genes was 100% for all genomes. Annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline v4.11 with default parameters (18, 19). The GC content and number of coding sequences were determined with Geneious Prime v2020.0.5. Multilocus sequence typing was performed using MLST v2.0 (http://www.genomicepidemiology.org/) with Escherichia coli configuration 1 (20). Plasmid replicons were identified with PlasmidFinder v2.1 (21, 22) with 90% identity and 60% minimum coverage cutoffs. Nearly all strains contained incompatibility plasmids of 88 kb to 131 kb and Col plasmids of 1.5 kb to 6.4 kb. (Table 1).

Hybrid assembly enabled the complete resolution of chromosome and plasmid sequences. These data provide insight into the plasmids harbored by diverse UTI UPEC strains isolated from postmenopausal women.

Data availability. The complete sequences were deposited in GenBank under BioProject accession number PRJNA636382. The BioSample and SRA accession numbers for each isolate can be found in Table 1.

ACKNOWLEDGMENTS

We thank the Welch Foundation, award number AT-2030-20200401 to N.J.D.N., and by the Felecia and John Cain Chair in Women’s Health, held by P.E.Z.

REFERENCES

1. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13:269–284. https://doi.org/10.1038/nrmicro3452

2. Foxman B. 2014. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect Dis Clin North Am 28:1–13. https://doi.org/10.1016/j.idc.2013.09.003

3. Klein RD, Hultgren SJ. 2020. Urinary tract infections: microbial pathogenesis, host-pathogen interactions and new treatment strategies. Nat Rev Microbiol 18:211–226. https://doi.org/10.1038/s41579-020-0324-0

4. Dason S, Dason JT, Kapoor A. 2011. Guidelines for the diagnosis and management of recurrent urinary tract infection in women. Can Urol Assoc J 5:316–322. https://doi.org/10.5489/cuaq.687

5. Foxman B. 1990. Recurring urinary tract infection syndromes. Am J Public Health 80:331–333. https://doi.org/10.2105/ajph.80.3.331

6. Glover M, Moreira CG, Sperandio V, Zimmermann P. 2014. Recurrent urinary tract infections in healthy and nonpregnant women. Urol Sci 25:1–8. https://doi.org/10.1111/j.1345-598X.2013.11.007

7. De Nisco NJ, Neugent M, Mull J, Chen L, Kuprasertkul A, de Souza Santos M, Palmer KL, Zimmermann P, Orth K. 2019. Direct detection of tissue-resident bacteria and chronic inflammation in the bladder wall of postmenopausal women with recurrent urinary tract infection. J Mol Biol 431:4368–4379. https://doi.org/10.1016/j.jmb.2019.04.008

8. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214. https://doi.org/10.1089/1066527005818478

9. De Coster W, D’Hert S, Schultz DT, Cruts M, Van Breechoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149

10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pyrbyskiel AD, Syrzhkina AV, Syrotik 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

11. Vaser R, Sovic I, Nagarajan N, Sikic 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

12. Walker BJ, Abeel T, Shea T, Priest M, Abouelhail S, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pillow: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e9112963. https://doi.org/10.1371/journal.pone.0112963

13. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Uncycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595

14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1073. https://doi.org/10.1093/bioinformatics/btt086

15. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383

16. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

17. Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics 33:3635–3637. https://doi.org/10.1093/bioinformatics/btx445

18. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O’Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanick N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068
19. 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.

20. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.

21. 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.

22. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, 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.