Draft Genome Sequence of *Escherichia coli* UMB9246, Isolated from the Bladder of a Woman with Recurrent Urinary Tract Infection

Mikalina Belmonte,[a] Taylor Miller-Ensminger,[b] Adelina Voukadinova,[b] Alan J. Wolfe,[c] Catherine Putonti[a,b,c,d]

[a]Department of Biology, Loyola University Chicago, Chicago, Illinois, USA
[b]Bioinformatics Program, Loyola University Chicago, Chicago, Illinois, USA
[c]Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois, USA
[d]Department of Computer Science, Loyola University Chicago, Chicago, Illinois, USA

ABSTRACT *Escherichia coli* is a Gram-negative, motile, rod-shaped bacterium that causes the majority of uncomplicated urinary tract infections (UTIs). Here, we report the draft genome of *E. coli* strain UMB9246, an isolate from a woman with recurrent UTI.

Pathogenic strains of *Escherichia coli* can cause intestinal, urinary tract, and internal infections, and *E. coli* isolates from multiple body sites have been sequenced (1). Specifically, uropathogenic *E. coli* (UPEC) strains are responsible for the majority of infrequent urinary tract infections (UTIs) in humans (2, 3). Recurrence of UTIs is frequent (4). Recurrent UTIs (rUTIs) are considerably more complicated and nuanced than infrequent UTIs. Here, we isolated *E. coli* strain UMB9246 from a catheterized urine sample obtained from a woman with an rUTI.

E. coli strain UMB9246 was collected as part of a prior institutional review board (IRB)-approved study (University of California, San Diego, IRB no. 170077AW) using the expanded quantitative urine culture (EQUC) protocol (5). Briefly, 100 μl of urine was spread onto a 5% sheep blood agar plate and incubated overnight at 35°C with 5% CO_2 for 24 h. Each morphologically distinct colony was then further purified through successive plating. The genus and species for this isolate were determined via matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry as previously described (5). The isolate was then stored at −80°C until sequencing. From this freezer stock, the sample was streaked onto a Columbia nalidixic acid (CNA) agar plate and incubated at 35°C with 5% CO_2 for 24 h. A single colony was selected and incubated in LB broth at 37°C with shaking for 24 h. DNA was extracted using the DNeasy blood and tissue kit following the manufacturer's protocol for Gram-positive bacteria with the following alterations. We used 230 μl of lysis buffer (180 μl of 20 mM Tris-Cl, 2 mM sodium EDTA, and 1.2% Triton X-100 and 50 μl of lysozyme) in step 2, and in step 5 of the protocol, we incubated the sample at 56°C for 10 min. DNA was quantified using the Qubit fluorometer and sent to the Microbial Genome Sequencing Center (MiGS) at the University of Pittsburgh for sequencing. The DNA was enzymatically fragmented using an Illumina tagmentation enzyme, and indices were attached using PCR. The DNA was sequenced using an Illumina NextSeq 550 flow cell, producing 1,408,737 pairs of 150-bp reads. The raw reads were trimmed using Sickle v1.33 (https://github.com/najoshi/sickle) and assembled using SPAdes v3.13.0 with the "only-assembler" option for k values of 55, 77, 99, and 127 (6). The genome coverage is 71× and was calculated using BBMap v38.47 (https://sourceforge.net/projects/bbmap/). The genome assembly was first annotated using PATRIC v3.6.3 (7). The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (8) was used to annotate the publicly

available genome. The genome assembly was also examined using PHASTER (9) and CRISPRCasFinder (10) to identify prophage sequences and CRISPR spacer arrays, respectively. Default parameters were used for each software tool, unless previously stated.

The *E. coli* UMB9246 draft genome is 5,057,732 bp long in 90 contigs with an N_{50} score of 159,145 bp and a GC content of 50.8%. The PGAP annotation identified 4,691 protein-coding genes, 5 rRNA operons, and 77 tRNAs. PHASTER predicted 1 questionable, 2 incomplete, and 1 intact phage for this strain. CRISPRCasFinder found 2 CRISPR arrays. PATRIC identified in the genome assembly both virulence factors and antibiotic resistance genes, including those for tetracycline resistance. Further analysis of this strain and other isolates from individuals suffering from uncomplicated UTIs and rUTIs will give us insight into the mechanisms for recurrence.

Data availability. This whole-genome shotgun project has been deposited in GenBank under the accession no. JAAUVS000000000. The version described in this paper is the first version, JAAUVS010000000. The raw sequencing reads have been deposited in SRA under the accession no. SRR11441020.

REFERENCES

1. Moriel DG, Tan L, Goh KGK, Phan M-D, Ipe DS, Lo AW, Peters KM, Ulett GC, Beatson SA, Schembri MA. 2016. A novel protective vaccine antigen from the core Escherichia coli genome. mSphere 1:e00326-16. https://doi.org/10.1128/mSphere.00326-16.
2. Johnson JR. 1991. Virulence factors in Escherichia coli urinary tract infection. Clin Microbiol Rev 4:80–128. https://doi.org/10.1128/cmr.4.1.80.
3. Terlizzi ME, Gribaudo G, Maffei ME. 2017. Uropathogenic Escherichia coli (UPEC) infections: virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial strategies. Front Microbiol 8:1566. https://doi.org/10.3389/fmicb.2017.01566.
4. Foxman B, Gillespie B, Koopman J, Zhang L, Palin K, Tallman P, Marsh JV, Spear S, Sobel JD, Marty MJ, Marrs CF. 2000. Risk factors for second urinary tract infection among college women. Am J Epidemiol 151:1194–1205. https://doi.org/10.1093/oxfordjournals.aje.a010170.
5. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. J Clin Microbiol 52:871–876. https://doi.org/10.1128/JCM.02876-13.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.
7. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535–D542. https://doi.org/10.1093/nar/gkw1017.
8. 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.
9. 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.
10. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. https://doi.org/10.1093/nar/gky425.