AMERICAN SOCIETY FOR MICROBIOLOGY | **Microbiology** Resource Announcements

Draft Genome Sequences of *Escherichia coli* Strains FP2 and FP3, Isolated from the Canada Goose (*Branta canadensis*)

Allison L. Denny,[a] Susan E. Arruda[a]

[a]Department of Biology, Franklin Pierce University, Rindge, New Hampshire, USA

ABSTRACT Draft genomes of two strains of *Escherichia coli*, FP2 and FP3, isolated from the feces of the Canada goose (*Branta canadensis*), were sequenced. Genome sizes were 5.26 Mb with a predicted G+C content of 50.54% (FP2) and 5.07 Mb with a predicted G+C content of 50.41% (FP3).

Escherichia coli bacteria are rod-shaped, Gram-negative, facultative aerobes that are often encapsulated (1–3). The *E. coli* genome consists of a single circular chromosome that ranges from 4.5 to 5.9 million base pairs and comprises 4,000 to 5,500 genes (2, 4, 5). Like some other members of the *Enterobacteriaceae* family (e.g., *Serratia marcescens*), *E. coli* bacteria have pathogenic and nonpathogenic strains (6). Pathogenic strains can be categorized into different pathotypes based on which strategy is used to interact with the host cell and the degree of virulence (2, 7). Various strains of nonpathogenic *E. coli* are known to be the principal inhabitants of mammalian gut microbiomes and have also been found in the intestinal tract of birds, reptiles, and fish (2, 8).

E. coli strains were obtained as environmental isolates on the Franklin Pierce University campus in Rindge, New Hampshire. Samples of Canada goose feces were streaked onto lysogeny broth agar and incubated overnight at 37°C. Single colonies were used to inoculate lysogeny broth from which genomic DNA was isolated with the QIAamp DNA purification minikit (Qiagen, Bethesda, MD, USA). Fragmented genomic DNA was tagged with adapters using the KAPA HyperPlus kit (Wilmington, MA, USA) and then loaded on an Illumina HiSeq 2500 instrument by the Hubbard Center for Genome Studies (University of New Hampshire, Durham, NH, USA) for sequencing (FP2, 8,418,266 reads, 188× coverage; FP3, 8,418,266 reads, 225× coverage). The 250-bp paired-end reads were trimmed using Trimmomatic version 3.5 (default settings), and genome sequences underwent *de novo* assembly using SPAdes version 3.9.0 (using default settings) (9, 10). Contigs less than 500 bp were removed before the genome was submitted for annotation (2 July 2018) with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (11).

The FP2 genome is approximately 5,257,000 bp, distributed in 274 contigs (the largest is 330,378 bp), with an overall G+C content of 50.54% and an N_{50} of 148,984 bp. PGAP predicted 5,127 protein-coding genes, 21 rRNA genes, 82 tRNA genes, and 245 pseudogenes. The FP3 genome is approximately 5,068,010 bp, distributed in 297 contigs (the largest is 721,880 bp), with an overall G+C content of 50.41% and an N_{50} of 600,909 bp. PGAP identified 4,836 protein-coding genes, 13 rRNA genes, 84 tRNA genes, and 261 pseudogenes. The similarity between these genomes and other known *E. coli* genomes (e.g., GenBank accession numbers CP022393 and LT883142), along with the manner in which the samples were isolated, suggests that FP2 and FP3 are likely residents of the *Branta canadensis* gut microbiome.

Data availability. The FP2 and FP3 whole-genome shotgun sequences were deposited in DDBJ/ENA/GenBank under the accession numbers QNRA00000000 and

Address correspondence to Susan E. Arruda, arrudas@franklinpierce.edu.

QNRB00000000, respectively. The versions described in this paper are versions QNRA01000000 and QNRB01000000, respectively.

REFERENCES

1. Jang J, Hur H-G, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. 2017. Environmental *Escherichia coli*: ecology and public health implications—a review. J Appl Microbiol 123:570–581. https://doi.org/10.1111/jam.13468.
2. Blount ZD. 2015. The unexhausted potential of *E. coli*. Elife 4:e05826. https://doi.org/10.7554/eLife.05826.
3. Whitfield C, Roberts IS. 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. Mol Microbiol 31:1307–1319. https://doi.org/10.1046/j.1365-2958.1999.01276.x.
4. Lukjancenko O, Wassenaar TM, Ussery DW. 2010. Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol 60:708–720. https://doi.org/10.1007/s00248-010-9717-3.
5. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genome analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190:6881–6893. https://doi.org/10.1128/JB.00619-08.
6. Zhang Q, Melcher U, Zhou L, Najar FZ, Roe BA, Fletcher J. 2005. Genomic comparison of plant pathogenic and nonpathogenic *Serratia marcescens* strains by suppressive subtractive hybridization. Appl Environ Microbiol 71:7716–7723. https://doi.org/10.1128/AEM.71.12.7716-7723.2005.
7. Sousa CP. 2006. The versatile strategies of *Escherichia coli* pathotypes: a mini review. J Venom Anim Toxins Incl Trop Dis 12. https://doi.org/10.1590/S1678-91992006000300002.
8. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. Science 308:1635–1638. https://doi.org/10.1126/science.1110591.
9. 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.
10. 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.
11. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W. 2013. Prokaryotic Genome Annotation Pipeline. *In* The NCBI handbook, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.