Loyola University Chicago

## Loyola eCommons

Biology: Faculty Publications and Other Works

Faculty Publications and Other Works by Department

# Draft Genome Sequence of Escherichia coli K-12 (ATCC 10798)

Catherine Putonti
*Loyola University Chicago*, cputonti@luc.edu

Daniela Dimitrova

Kathleen C. Engelbrecht

David W. Koenig

Alan J. Wolfe
*Loyola University Chicago*

## Recommended Citation

# Draft Genome Sequence of *Escherichia coli* K-12 (ATCC 10798)

Daniela Dimitrova,[a] Kathleen C. Engelbrecht,[b] Catherine Putonti,[c,d,e,f]
David W. Koenig,[b] Alan J. Wolfe[f]

Corporate Research and Engineering, Kimberly-Clark Corporation, Roswell, Georgia, USA[a]; Corporate Research and Engineering, Kimberly-Clark Corporation, Neenah, Wisconsin, USA[b]; Bioinformatics Program, Loyola University Chicago, Chicago, Illinois, USA[c]; Department of Biology, Loyola University Chicago, Chicago, Illinois, USA[d]; Department of Computer Science, Loyola University Chicago, Chicago, Illinois, USA[e]; Department of Microbiology and Immunology, Stritch School of Medicine, Health Sciences Division, Loyola University Chicago, Maywood, Illinois, USA[f]

**ABSTRACT** Here, we present the draft genome sequence of *Escherichia coli* ATCC 10798. *E. coli* ATCC 10798 is a K-12 strain, one of the most well-studied model microorganisms. The size of the genome was 4,685,496 bp, with a G+C content of 50.70%. This assembly consists of 62 contigs and the F plasmid.

$E$scherichia coli is commonly found in the human intestinal tract, and while most strains are harmless members of the microflora, some are the causative agents of a variety of illnesses. Here, we announce the genome sequence of *E. coli* strain ATCC 10798 ([Migula] Castellani and Chalmers) isolated from the stool sample of a diphtheria convalescent patient in 1922 (1, 2). Per ATCC documentation, the genotype is $\lambda^+$ $F^+$ (https://www.atcc.org/products/all/10798).

For genome extraction and sequencing, the purchased culture isolate was grown on 5% sheep blood agar (BD BBL prepared plated medium) under 5% $CO_2$ at 35°C for 48 h. To extract genomic DNA, cells were resuspended in 0.5 ml of DNA extraction buffer (20 mM Tris-HCl, 2 mM EDTA, 1.2% Triton X-100 [pH 8]), followed by the addition of 50 $\mu$l of lysozyme (20 mg/ml), 30 $\mu$l of mutanolysin, and 5 $\mu$l of RNase (10 mg/ml). After a 1-h incubation at 37°C, 80 $\mu$l of 10% SDS and 20 $\mu$l of proteinase K were added, followed by a 2-h incubation at 55°C. Two hundred ten microliters of 6 M NaCl and 700 $\mu$l of phenol-chloroform were added. After a 30-min incubation with rotation, the solutions were centrifuged at 13,500 rpm for 10 min, and the aqueous phase was extracted. An equivalent volume of isopropanol was added, and the solution was centrifuged at 13,500 rpm for 10 min after a 10-min incubation. The supernatant was decanted and the DNA pellet precipitated using 600 $\mu$l of 70% ethanol. Following ethanol evaporation, the DNA pellet was resuspended in Tris-EDTA and stored at −20°C.

Genomic DNA was diluted in water to a concentration of 0.2 ng/$\mu$l, as measured by a fluorometric-based method (Life Technologies, Inc.); 5 $\mu$l was used to obtain a total of 1 ng of input DNA. Library preparation of the isolated DNA was performed using the Nextera XT DNA library preparation kit. The library was sequenced on the MiSeq sequencer (Illumina) using the MiSeq reagent kit version 2 (500 cycles). The run produced 1,652,091 paired-end reads in total. BBDuk, part of the BBMap package (http://sourceforge.net/projects/bbmap/), was used to trim the reads. The trimmed reads were then assembled using SPAdes (version 3.5) (3), followed by scaffolding by SSPACE (4), which produced 63 contigs that varied in size from 680 bp to 507,483 bp ($N_{50}$, 149,239 bp), with an average coverage of 156.4×. One of these contigs is a complete plasmid, a copy of the F plasmid. Furthermore, the *E. coli* strain sequenced is a lysogen; the sequence of the bacteriophage $\lambda$ was identified within the genome

assembly through BLAST. Annotations were produced using the software tool Peasant (5). Eight rRNA genes, 80 tRNA genes, and 4,519 protein-coding sequences were detected. Two confirmed clustered regularly interspaced short palindromic repeat (CRISPR) arrays were found (6). The genome size was 4,685,496 bp, with an observed G+C content of 50.70%.

**Accession number(s).** The draft whole-genome project for *E. coli* ATCC 10798 has been deposited at DDBJ/EMBL/GenBank under accession number NARG00000000. Raw sequence reads are deposited at DDBJ/EMBL/GenBank under accession number SRR5364300.

## ACKNOWLEDGMENTS

## REFERENCES

1. Evans DJ, Evans DG. 1996. *Escherichia coli* in diarrheal disease. *In* Baron S (ed), Medical microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston, TX. https://www.ncbi.nlm.nih.gov/books/NBK7710/.
2. Bachmann BJ. 1996. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p 2460–2488. *In* Neidhardt FC (ed), Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed. ASM Press, Washington, DC.
3. 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.
4. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.
5. Brenner J, Kalesinskas L, Putonti C. 2017. Exploring the diversity of *Bacillus* whole genome sequencing projects using Peasant, the prokaryotic assembly and annotation tool. bioRxiv 132084. https://doi.org/10.1101/132084.
6. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.