

# Complete Genome Assembly of *Escherichia coli* ATCC 25922, a Serotype O6 Reference Strain

T. D. Minogue,<sup>a</sup> H. A. Daligault,<sup>b</sup> K. W. Davenport,<sup>b</sup> K. A. Bishop-Lilly,<sup>c,d</sup> S. M. Broomall,<sup>e</sup> D. C. Bruce,<sup>a</sup> P. S. Chain,<sup>b</sup> O. Chertkov,<sup>b</sup> S. R. Coyne,<sup>a</sup> T. Freitas,<sup>a\*</sup> K. G. Frey,<sup>c,d</sup> H. S. Gibbons,<sup>e</sup> J. Jaissle,<sup>a</sup> C. L. Redden,<sup>c,d</sup> C. N. Rosenzweig,<sup>e</sup> Y. Xu,<sup>b</sup>  S. L. Johnson<sup>b</sup>

Diagnostic Systems Division (DSD), United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, USA<sup>a</sup>; Los Alamos National Laboratory, Los Alamos, New Mexico, USA<sup>b</sup>; Naval Medical Research Center (NMRC)-Frederick, Fort Detrick, Maryland, USA<sup>c</sup>; Henry M. Jackson Foundation, Bethesda, Maryland, USA<sup>d</sup>; United States Army Edgewood Chemical Biological Center (ECBC), Aberdeen Proving Ground, Maryland, USA<sup>e</sup>

\* Present address: T. Freitas, Med Fusion, Lewisville, Texas, USA.

**We present the complete genome assembly of *Escherichia coli* ATCC 25922 as submitted to NCBI under accession no. CP009072. This strain was originally isolated from a clinical sample in Seattle, Washington (1946), and is often used in quality control testing. The assembled genome is 5.20 Mb (50.4% G+C content) and includes two plasmids.**

Received 25 August 2014 Accepted 26 August 2014 Published 25 September 2014

**Citation** Minogue TD, Daligault HA, Davenport KW, Bishop-Lilly KA, Broomall SM, Bruce DC, Chain PS, Chertkov O, Coyne SR, Freitas T, Frey KG, Gibbons HS, Jaissle J, Redden CL, Rosenzweig CN, Xu Y, Johnson SL. 2014. Complete genome assembly of *Escherichia coli* ATCC 25922, a serotype O6 reference strain. *Genome Announc.* 2(5):e00969-14. doi:10.1128/genomeA.00969-14.

**Copyright** © 2014 Minogue et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to S. L. Johnson, shannonj@lanl.gov.

*Escherichia coli*, a seemingly ubiquitous Gram-negative bacterium, is best known for its ability to cause food-borne outbreaks. The strain ATCC 25922 is a commonly used quality control strain, particularly in antibody sensitivity assays and was originally isolated from a human clinical sample collected in Seattle and WA (1946). It is of serotype O6 and biotype 1. A prior assembly of the genome is publicly available, however, it is in 116 contigs (1).

High-quality genomic DNA was extracted from a purified isolate using a QIAGEN Genome Tip-500 at USAMRIID-Diagnostic Systems Divisions (DSD). Specifically, a 100-mL bacterial culture was grown to stationary phase and nucleic acid extracted as per manufacturer's recommendations. Sequence data generated includes both Illumina (standard unpaired 100-bp library at 300-fold genome coverage) and Roche 454 (7,847- ± 1,962-bp insert, 32-fold genome coverage) technologies (2, 3). Data from the two libraries were assembled together in Newbler (Roche) and the consensus sequences computationally shredded into 2-kbp overlapping fake reads (shreds). The raw reads were also assembled in Velvet and those consensus sequences computationally shredded into 1.5-kbp overlapping shreds (4). Draft data from all platforms were then assembled together with Allpaths and the consensus sequences were computationally shredded into 10-kbp overlapping shreds (5). We then integrated the Newbler consensus shreds, Velvet consensus shreds, Allpaths consensus shreds, and a subset of the long-insert read pairs using parallel Phrap (High Performance Software, LLC). Possible misassemblies were corrected and some gap closure accomplished with manual editing in Consed (6–8).

Automatic annotation for the *E. coli* ATCC 25922 genome utilized an Ergatis based workflow with minor manual curation. The complete annotated genome assembly is available in NCBI (accession no. CP009072) and raw data can be provided upon request.

This finished assembly includes one chromosome (5,130,767-bp) and two plasmids (48,488 and 24,185-bp, respectively). Preliminary review of the 5.20-Mbp (50.4% G+C content) genome finds 4,840 coding sequences, 21 rRNAs, and 85 tRNAs.

**Nucleotide sequence accession number.** The annotated genome assembly of *Escherichia coli* ATCC 25922 is available in GenBank under accession no. [CP009072](https://ncbi.nlm.nih.gov/assembly/CP009072).

## ACKNOWLEDGMENTS

Funding for this effort was provided by the Defense Threat Reduction Agency's Joint Science and Technology Office (DTRA J9-CB/JSTO).

This manuscript is approved by LANL for unlimited release (LA-UR-14-25958).

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

## REFERENCES

- Hein-Kristensen L, Franzyk H, Holch A, Gram L. 2013. Adaptive evolution of *Escherichia coli* to an  $\alpha$ -peptide/ $\beta$ -peptoid peptidomimetic induces stable resistance. *PLoS One* 8:e73620. <http://dx.doi.org/10.1371/journal.pone.0073620>.
- Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* 5:433–438. <http://dx.doi.org/10.1517/14622416.5.4.433>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <http://dx.doi.org/10.1038/nature03959>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read

- assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
5. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. *Genome Res.* 18:810–820. <http://dx.doi.org/10.1101/gr.7337908>.
  6. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 8:175–185. <http://dx.doi.org/10.1101/gr.8.3.175>.
  7. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* 8:186–194.
  8. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8:195–202. <http://dx.doi.org/10.1101/gr.8.3.195>.