# Complete Genome Sequence of *Escherichia coli* NCM3722

**Steven D. Brown, Suckjoon Jun**

Department of Physics and Section of Molecular Biology, University of California–San Diego, La Jolla, California, USA

*Escherichia coli* NCM3722 is a prototrophic K-12 strain with robust physiologic phenotypes. We report the complete 4,678,045-bp chromosome and 67,545-bp F-like plasmid of this unique model organism.

Address correspondence to Steven D. Brown, sdbrown@gmail.com, or Suckjoon Jun, suckjoon.jun@gmail.com.

*E*scherichia coli is a landmark pathogen, and the K-12 strains are ubiquitous model organisms. *E. coli* NCM3722 is a prototrophic K-12 strain first described by Sydney Kustu and colleagues for studies of microbial physiology (1, 2). This strain exhibits unique phenotypes, including galactose metabolism and response to spectinomycin, that are more robust than those of the popular MG1655/W3110 lineage. To enable development of this strain as a model organism, we have sequenced the NCM3722 genome as well as a derivative, SJ358, which has been cured of the F plasmid.

*E. coli* NCM3722 was obtained from the Coli Genetic Stock Center (CGSC) (strain 12355). Cells were prepared for sequencing by streaking on nonselective media and culturing a single colony in LB supplemented with 0.2% glucose and 1× M salts (3). Genomic DNA was prepared by collection of cells by centrifugation from an exponential-phase culture and extraction with a Life Technologies PureLink genomic DNA minikit. A library for sequencing by Pacific Biosciences SMRT technology was prepared and sequenced with P6-C4 chemistry by the UCSD Institute for Genomic Medicine. Illumina 250-cycle paired-end libraries were similarly prepared from single-colony cultures collected in late exponential phase. The resulting single-molecule real-time (SMRT) and paired-end reads were jointly assembled with SPAdes v. 3.1.0 (4), which produced a contig of 4,678,045 bp corresponding to the chromosome and a contig of 67,545 bp corresponding to an F-like plasmid. Samplings of individual single-nucleotide variants (SNVs) were verified by PCR and Sanger sequencing. The F-like plasmid was modified by λRed recombineering to sequentially remove the *ccdB* gene and the predicted *rep-srnB′* genes. The resulting strain was serially subcultured in LB plus 10% SDS, and samples were periodically streaked on LB for single clones (5). Genomic DNA from one clone, SJ358, was prepared for short-read sequencing identically to NCM3722 and aligned to the NCM3722 contigs using breseq 0.26.0 (6) and Bowtie2 2.2.5. SJ358 is available from Addgene (http://www.addgene.org/67755/).

The chromosomal and F sequences corroborate SNVs and large differences revealed by prior 454-based sequencing, including phage lambda lysogeny and absence the Qin prophage and a CP4-6 prophage fragment. The *ilvG* and *rph-1* mutations of MG1655 are absent, and the *rpoS* gene is the 33Am variant from EMG2. We found a variety of new SNVs within the rRNA genes *rrsH*, *rrlG*, *rrlE*, *rrlB*, *rrlA*, and *rrlC*. A deletion that includes the *rfb* operon has orphaned the *rfbA* gene, resulting in an accumulation of SNVs. Deletion of the *rfb* operon has been associated with increased resistance to antibiotics, which may contribute to the phenotype of the strain (7).

**Nucleotide sequence accession numbers.** The complete genome sequence of the *E. coli* NCM3722 chromosome was deposited in GenBank under the accession number CP011495, and the F-like plasmid was deposited under the accession number CP011496.

## ACKNOWLEDGMENTS

## REFERENCES

1. **Soupene E, van Heeswijk WC, Plumbridge J, Stewart V, Bertenthal D, Lee H, Prasad G, Paliy O, Charernnoppakul P, Kustu S.** 2003. Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. J Bacteriol **185:**5611–5626. http://dx.doi.org/10.1128/JB.185.18.5611-5626.2003.

2. **Lyons E, Freeling M, Kustu S, Inwood W.** 2011. Using genomic sequencing for classical genetics in *E. coli* K-12. PLoS One **6:**e16717. http://dx.doi.org/10.1371/journal.pone.0016717.

3. **Studier FW.** 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif **41:**207–234. http://dx.doi.org/10.1016/j.pep.2005.01.016.

4. **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. http://dx.doi.org/10.1089/cmb.2012.0021.

5. **Salisbury V, Hedges RW, Datta N.** 1972. Two modes of "curing" transmissible bacterial plasmids. J Gen Microbiol **70:**443–452. http://dx.doi.org/10.1099/00221287-70-3-443.

6. **Deatherage DE, Barrick JE.** 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol **1151:**165–188. http://dx.doi.org/10.1007/978-1-4939-0554-6_12.

7. **Liu A, Tran L, Becket E, Lee K, Chinn L, Park E, Tran K, Miller JH.** 2010. Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. Antimicrob Agents Chemother **54:**1393–1403. http://dx.doi.org/10.1128/AAC.00906-09.