

Finished Genome Sequence of *Escherichia coli* K-12 Strain HMS174 (ATCC 47011)

Juergen Mairhofer,^{b,c} Peter M. Kreml,^{a,c} Gerhard G. Thallinger,^{a,c} Gerald Striedner^{b,c}

Bioinformatics Group, Institute for Knowledge Discovery, Graz University of Technology, Graz, Austria^a; Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria^b; Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria^c

J.M. and P.M.K. contributed equally.

***Escherichia coli* strain K-12 substrain HMS174 is an engineered descendant of the *E. coli* K-12 wild-type strain. Like its ancestor, it is an important organism in biotechnological research and is used in fermentation processes for heterologous protein production. Here, we report the complete genome sequence of *E. coli* HMS174 (ATCC 47011).**

Received 22 August 2014 Accepted 10 October 2014 Published 20 November 2014

Citation Mairhofer J, Kreml PM, Thallinger GG, Striedner G. 2014. Finished genome sequence of *Escherichia coli* K-12 strain HMS174 (ATCC 47011). *Genome Announc.* 2(6): e00975-14. doi:10.1128/genomeA.00975-14.

Copyright © 2014 Mairhofer 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 Gerald Striedner, gerald.striedner@boku.ac.at.

The prototrophic *Escherichia coli* HMS174 was derived by mating a *thy-1* version of *E. coli* strain K-12 substrain W3110 with *Escherichia coli* strain K-12 substrain KL16-99 Hfr by Campbell et al. (1, 2), and it carries the following modifications: F^- , λ^- , *recA1*, *rpoB331*, and *hsdR19*. HMS174 was originally used for the refinement of the structural and genetic map of the T7 phage (1) in the Studier lab at Brookhaven and later on became a popular strain for the expression of recombinant proteins after lysogenization with the DE3 prophage. *E. coli* HMS174 provides a *recA1* mutation in a K-12 background. Due to this mutation, certain target genes whose products may cause the loss of the DE3 prophage become stabilized. However, the actual genome sequence remains to be elucidated.

The genome of *E. coli* strain K-12 substrain HMS174 was sequenced using a combination of next-generation sequencing methods. A first draft assembly based on the sequences of an 8-kbp paired-end library (Roche 454 GS FLX Titanium; 718,718 reads, for a total of 108.8 Mb; 23-fold coverage; ZMF, Medical University of Graz, Austria) generated with Newbler 2.6 consisted of 165 contigs, 81 of which were joined into a single circular scaffold. The gaps resulting from repetitive sequences were resolved by *in silico* gap filling. The remaining gaps were closed by PCR, followed by Sanger sequencing, yielding a draft genome of 4,583,580 bp. To improve the quality of the sequence by eliminating the 454 sequencing errors in the homopolymer stretches, the genome was subsequently sequenced using the Illumina paired-end method (Illumina HiSeq 2000; 16,836,098 reads, for a total of 1.68 Gb; 367-fold coverage; Ambry Genetics, Aliso Viejo, CA). The Illumina reads were aligned to the draft genome with CLC Genomics Workbench 4.7.1 (CLC bio, Aarhus, Denmark). Conflicts in the consensus sequence were resolved by voting.

The genomic structure of HMS174 was compared with that of the reference *E. coli* K-12 strains MG1655 and W3110 using Mauve (3), showing that HMS174 does not contain genomic regions without homologs in one of those strains. Functional annotation was performed by matching the genes of these reference

strains based on sequence similarity, using EcoCyc (4) as the primary data source. Additionally, the origin of replication was predicted with OriginX (5). Clustered regularly interspaced short palindromic repeats (CRISPRs) were predicted with CRT (6).

The final genome includes 4,584,860 bases, with a G+C content of 50.82%. The annotation totals 4,437 putative genes, 4,241 of which are protein coding. There are seven instances of the ribosomal 5S-23S-16S cluster, an additional 5S rRNA gene, 85 tRNAs, one transfer-messenger RNA (tmRNA), and 88 RNA genes of regulatory or miscellaneous function. Two CRISPRs were detected with 12 and 6 spacers, respectively.

The closest sequenced K-12 strain is *E. coli* W3110 (7). In relation to W3110, the HMS174 genome shows 13 deletions (1,198 to 20,110 bp) and 7 insertions (543 to 6,790 bp). Specifically, 9 IS elements and the *ymf*, *abg*, *yda*, *ynf*, and *rbs* operons are deleted. In addition, HMS174 carries the CPZ-55 prophage, 5 additional IS elements (1 IS3, 1 IS4, 2 IS5, and 1 IS150), and a lyase containing HEAT-repeat.

Nucleotide sequence accession number. The genome sequence for *E. coli* strain K-12 substrain HMS174 has been deposited at the European Nucleotide Archive under the accession no. [LM993812](https://www.ebi.ac.uk/ena/submit/).

ACKNOWLEDGMENTS

This work was supported in part by the Austrian Ministry of Science and Research GEN-AU project BIN (FFG grant 820962) and the Austrian BMWFJ, BMVIT, SFG, Standortagentur Tirol, and ZIT through the Austrian FFG-COMET-Funding Program (FFG grant 824186).

REFERENCES

1. Campbell JL, Tamanoi F, Richardson CC, Studier FW. 1979. Cloning of the T7 genome in *Escherichia coli*: use of recombination between cloned sequences and bacteriophage T7 to identify genes involved in recombination and a clone containing the origin of T7 DNA replication. *Cold Spring Harb. Symp. Quant. Biol.* 43:441–448.
2. Campbell JL, Richardson CC, Studier FW. 1978. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of

- T7 DNA. Proc. Natl. Acad. Sci. U. S. A. 75:2276–2280. <http://dx.doi.org/10.1073/pnas.75.5.2276>.
3. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403. <http://dx.doi.org/10.1101/gr.2289704>.
 4. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 39:D583–D590. <http://dx.doi.org/10.1093/nar/gkq1143>.
 5. Worning P, Jensen LJ, Hallin PF, Staerfeldt HH, Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. Environ. Microbiol. 8:353–361. <http://dx.doi.org/10.1111/j.1462-2920.2005.00917.x>.
 6. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics 8:209. <http://dx.doi.org/10.1186/1471-2105-8-209>.
 7. Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, Horiuchi T. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol. Syst. Biol. 2:2006.0007. <http://dx.doi.org/10.1038/msb4100049>.