

# Finished Genome Sequence of the Laboratory Strain *Escherichia coli* K-12 RV308 (ATCC 31608)

Peter M. Krempf<sup>a,c</sup> Juergen Mairhofer<sup>b,c</sup> Gerald Striedner<sup>b,c</sup> Gerhard G. Thallinger<sup>a,c</sup>

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

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

***Escherichia coli* strain K-12 substrain RV308 is an engineered descendant of the K-12 wild-type strain. Like its ancestor, it is an important organism in biotechnological research and is heavily used for the expression of single-chain variable fragments. Here, we report the complete genome sequence of *E. coli* K-12 RV308 (ATCC 31608).**

Received 23 August 2014 Accepted 10 October 2014 Published 20 November 2014

**Citation** Krempf PM, Mairhofer J, Striedner G, Thallinger GG. 2014. Finished genome sequence of the laboratory strain *Escherichia coli* K-12 RV308 (ATCC 31608). *Genome Announc.* 2(6):e00971-14. doi:10.1128/genomeA.00971-14.

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

Address correspondence to Gerhard G. Thallinger, [Gerhard.Thallinger@tugraz.at](mailto:Gerhard.Thallinger@tugraz.at).

The prototrophic *Escherichia coli* K-12 RV308 strain was derived from the RV strain by P1 transduction in the Ptashne lab (1). The original *E. coli* K-12 RV strain was engineered by P. Roy Vagelos in 1962 during his stay at the Institut Pasteur, Paris, France, by mating the *E. coli* K-12 3000X74V2 Hfr strain with the *E. coli* K-12 C600 strain (M. Malamy, personal communication). According to Maurer et al. (1), *E. coli* K-12 RV308 carries the following modifications: *su*<sup>-</sup>,  $\Delta$ *lacX74*, *gal IS II::OP308*, and *strA*. However, the actual genome sequence remains to be elucidated.

The genome of *E. coli* strain K-12 substrain RV308 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, 779,412 reads, for a total of 123.1 Mb; 26-fold coverage; ZMF, Medical University of Graz, Austria) generated with Newbler 2.6 consisted of 126 contigs, 71 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 4,583,642 bp. To improve the quality of the sequence by eliminating the 454 sequencing errors in homopolymer stretches, the genome was subsequently sequenced using the Illumina paired-end method (Illumina HiSeq 2000; 16,875,234 reads, for a total of 1.69 Gb; 368-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 RV308 was compared with that of reference *E. coli* K-12 strains, especially MG1655, using Mauve (2), showing that RV308 does not contain genomic regions without a homolog in MG1655. Functional annotation was performed by matching the genes of MG1655 based on sequence similarity, using EcoCyc (3) as a primary data source. Additionally, the origin of replication was predicted with OriginX (4). Clustered regularly interspaced short palindromic repeats (CRISPRs) were predicted with CRT (5).

The final sequence includes 4,585,620 bases, with a G+C content of 50.79%. The genome consists of 4,463 putative genes, 4,245 of which are protein coding. There are seven instances of the ribosomal 5S-23S-16S cluster, an additional 5S rRNA genes, 86 predicted tRNAs, one transfer-messenger RNA (tmRNA), and 91 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* MG1655 (6). In relation to MG1655 (GenBank accession no. U00096.3), the RV308 genome shows 8 deletions (65 to 36,218 bp) and 6 insertions (181 to 1336 bp). Specifically, the *e14* and CPZ-55 prophages are missing, and the *yah*, *prp*, *cod*, *cyn*, *lac*, and *mhp* operons and one *IS1* element are deleted. The insertions comprise 5 additional IS elements (3 *IS1*, 1 *IS2*, and 1 *IS3*).

**Nucleotide sequence accession number.** The genome sequence for *E. coli* strain K-12 substrain RV308 has been deposited at the European Nucleotide Archive under the accession no. [LM995446](http://www.ebi.ac.uk/ena/submit/submitter/LM995446).

## 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

- Maurer R, Meyer B, Ptashne M. 1980. Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor. *J. Mol. Biol.* 139:147–161. [http://dx.doi.org/10.1016/0022-2836\(80\)90302-2](http://dx.doi.org/10.1016/0022-2836(80)90302-2).
- 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>.
- 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.

- ogy. *Nucleic Acids Res.* 39:D583–D590. <http://dx.doi.org/10.1093/nar/gkq1143>.
4. 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>.
  5. 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>.
  6. 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>.