

Complete Genome Sequence of *Escherichia coli* Strain BL21

Haeyoung Jeong,^{a,b} Hyun Ju Kim,^{b,c} Sang Jun Lee^{b,c}

Super-Bacteria Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea^a; Biosystems and Bioengineering Program, University of Science and Technology (UST), Daejeon, Republic of Korea^b; Infection and Immunity Research Center, KRIBB, Daejeon, Republic of Korea^c

***Escherichia coli* strain BL21 is one of the widely used bacterial hosts for high-level recombinant protein production and for other applications. Here, we present the complete genome sequence of a commercial version of the *Escherichia coli* BL21 strain.**

Received 3 February 2015 Accepted 10 February 2015 Published 19 March 2015

Citation Jeong H, Kim HJ, Lee SJ. 2015. Complete genome sequence of *Escherichia coli* strain BL21. *Genome Announc* 3(2):e00134-15. doi:10.1128/genomeA.00134-15.

Copyright © 2015 Jeong 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 Haeyoung Jeong, hyjeong@kribb.re.kr, or Sang Jun Lee, leesj@kribb.re.kr.

Escherichia coli BL21 and BL21(DE3), created by F. Studier and Barbara A. Moffatt (1), are common laboratory strains for recombinant protein production. The lack of *lon* and *ompT* proteases, often regarded as common characteristics among B lineage, has driven the development of those strains for protein expression hosts. *E. coli* BL21(DE3), a derivative of BL21, is probably the most widely used in high-level expression of recombinant proteins, and it harbors a prophage DE3 derived from a bacteriophage λ , which carries the T7 RNA polymerase gene under the control of the *lacUV5* promoter. *E. coli* BL21 has been routinely used for non-T7 expression, and it was also recently modified to produce a plasmid DNA vaccine, due to its better performance in high-cell-density fed-batch cultures compared to K-12 strains (2).

The genome sequence of *E. coli* BL21(DE3) was previously determined by our institute (3). In principle, the genome sequence of BL21 would be identical with that of BL21(DE3), except for the DE3 prophage. However, stock-to-stock variations may occur that can have serious implications for experiments (4). With *E. coli* BL21(DE3) genome information as the reference, the complete genome sequence of BL21 was constructed using only next-generation sequencing, and base-by-base genomic differences were identified.

E. coli BL21 strain was purchased from TaKaRa Bio (code number 9126, lot number K142). Genome sequencing was carried out at the Human Derived Material Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), using the Illumina HiSeq 2000 system. The complete genome sequence of BL21 was produced both by elaborated reference mapping and *de novo* assembly of 101-nt Illumina reads (47,123,524 paired-end reads from an ~150-bp library) using the CLC Genomics Workbench version 6.5.1 (CLC bio).

We had expected that a modified BL21(DE3) genome sequence (CP001509.3)—from which the prophage region was deleted, leaving one copy of the lambda attachment site (*attB*)—would serve as the best reference sequence. However, coverage analysis from the first mapping showed that the *attB* of BL21 was not empty but occupied by a 12.1-kb-long λ^* B prophage as in *E. coli* REL606 (3), which was reported to be a common characteristic among the B lineage (5). A second trial of reference mapping, using a modified sequence having a λ^* B prophage sequence, revealed another low-coverage region between the *ybhC* gene and

the left boundary of the lambda attachment site, *attL*. By sequence comparison, the best matching sequence was identified from *de novo* assembled contigs and was 23 bp longer than the corresponding region from BL21(DE3). An updated reference sequence was then prepared by replacing the low coverage region and its neighbor (~200 bp), and mapping was carried out again. Uniform read coverage on the entire final reference sequence was confirmed by visual investigation, and no other difference was found by quality-based SNV/structural variation detection or by breakpoint analysis. Assembly using unmapped reads did not result in any contigs that can support the presence of BL21-specific regions. Genome annotation was carried out by the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) service.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [CP010816](https://www.ncbi.nlm.nih.gov/nuccore/CP010816).

ACKNOWLEDGMENTS

This work was supported by the KRIBB Research Initiative Program and the Basic Science Research Program through the National Research Foundation of Korea (NRF-2012R1A1A2042114), Ministry of Science, ICT and Future Planning, Republic of Korea.

REFERENCES

1. Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113–130. [http://dx.doi.org/10.1016/0022-2836\(86\)90385-2](http://dx.doi.org/10.1016/0022-2836(86)90385-2).
2. Phue JN, Lee SJ, Trinh L, Shiloach J. 2008. Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5 α). *Biotechnol Bioeng* 101:831–836. <http://dx.doi.org/10.1002/bit.21973>.
3. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol* 394:644–652. <http://dx.doi.org/10.1016/j.jmb.2009.09.052>.
4. Freddolino PL, Amini S, Tavazoie S. 2012. Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. *J Bacteriol* 194: 303–306. <http://dx.doi.org/10.1128/JB.06087-11>.
5. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. 2009. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B & K-12 genomes. *J Mol Biol* 394:653–680. <http://dx.doi.org/10.1016/j.jmb.2009.09.021>.