RESEARCH ARTICLE

*J Anim Sci Technol* 2021;63(2):461-464
https://doi.org/10.5187/jast.2021.e38

JAST
*Journal of Animal Science and Technology*
pISSN 2672-0191 eISSN 2055-0391

# Complete genome sequence of *Escherichia coli* K_EC180, a bacterium producing shiga-like toxin isolated from swine feces

Hyeri Kim[1#], Jae Hyoung Cho[1#], Jin Ho Cho[2#], Minho Song[3#], Hakdong Shin[4], Sheena Kim[1], Eun Sol Kim[1], Hyeun Bum Kim[1*] and Ju-Hoon Lee[5*]

[1]*Department of Animal Resources Science, Dankook University, Cheonan 31116, Korea*
[2]*Division of Food and Animal Science, Chungbuk National University, Cheongju 28644, Korea*
[3]*Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Korea*
[4]*Department of Food Science and Biotechnology, College of Life Science, Sejong University, Seoul 05006, Korea*
[5]*Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul 08826, Korea*

Received: Oct 22, 2020
Revised: Jan 20, 2021
Accepted: Jan 21, 2021

#These authors contributed equally to this work.

***Corresponding author**
Hyeun Bum Kim
Department of Animal Resources Science, Dankook University, Cheonan 31116, Korea.
Tel: +82-41-550-3653
E-mail: hbkim@dankook.ac.kr

Ju-Hoon Lee
Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul 08826, Korea.
Tel: +82-2-880-4854
E-mail: juhlee@snu.ac.kr



## Abstract

*Escherichia coli* normally colonizes the lower intestine of animals and humans, but some serotypes are foodborne pathogens. The *Escherichia coli* K_EC180 was isolated from swine feces that were collected from a weaner pig. In this genome announcement, *E. coli* K_EC180 was sequenced using PacBio RS II and Illumina NextSeq 500 platforms. The complete chromosome of *E. coli* K_EC180 is composed of one circular chromosome (5,017,281 bp) with 50.4% of guanine + cytosine (G + C) content, 4,935 of coding sequence (CDS), 88 of tRNA, and 22 of rRNA genes. The complete genome of *E. coli* K_EC180 contains the toxin genes such as shiga-like toxins (stxA and stxB).

**Keywords:** *Escherichia coli* K_EC180, Swine feces, Whole genome sequencing, Shiga-like toxin

## INTRODUCTION

*Escherichia coli* is a facultative anaerobic bacterium which is commonly spread on biosphere. *E. coli* normally colonizes the lower intestine of animals and humans (1). However, Some of the serotypes such as Enterohemorrhagic *E. coli* (EHEC), Enterotoxigenic *E. coli* (ETEC), Enteropathogenic *E. coli* (EPEC) and Shiga toxin-producing *E. coli* (STEC) can cause foodborne illnesses in people.

*E. coli* K_EC180 was isolated from swine feces that were collected from a livestock farm in Haenam-gun, Jeollanam-do, Korea. *E. coli* K_EC180 was streaked to Luria-Bertani (LB) agar and incubated at 37℃ for 24 h. The suspected colony in LB agar was inoculated into LB broth and incubated at 37℃ for 24 h. To analyze the complete genome, the *E. coli* K_EC180 genome was sequenced by PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA) at Insilicogen (Yongin, Korea) and Illumina NextSeq 500 (Illumina, San Diego, CA, USA) platform at LabGenomics (Seongnam, Korea). The genomic DNA of *E. coli* K_EC180 for PacBio and Illumina sequencing was extracted using the MagAttract

**ORCID**
Hyeri Kim
https://orcid.org/0000-0002-6560-2390
Jae Hyoung Cho
https://orcid.org/0000-0002-1128-3451
Jin Ho Cho
https://orcid.org/0000-0001-7151-0778
Minho Song
https://orcid.org/0000-0002-4515-5212
Hakdong Shin
https://orcid.org/0000-0001-7615-9809
Sheena Kim
https://orcid.org/0000-0002-5410-1347
Eun Sol Kim
https://orcid.org/0000-0001-8801-421X
Hyeun Bum Kim
https://orcid.org/0000-0003-1366-6090
Ju-Hoon Lee
https://orcid.org/0000-0003-0405-7621

**Competing interests**
No potential conflict of interest relevant to this article was reported.

**Funding sources**
The present study was supported by the research fund (19162MFDS037) from the Ministry of Food and Drug Safety, Republic of Korea, and by the University Innovation Support Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Dankook University 2019).

**Acknowledgements**
We thank Mo Re Kim (Brandeis University, MA, USA) for the English grammar corrections.

**Availability of data and material**
Upon reasonable request, the datasets of this study can be available from the corresponding author.

**Authors' contributions**
Conceptualization: Cho Jin Ho, Song MH, Kim HB, Lee JH.
Data curation: Kim H, Shin H, Kim S, Kim ES.
Formal analysis: Kim H, Shin H, Kim S, Kim ES.
Methodology: Kim H, Cho Jae Hyoung, Song MH.
Software: Kim H, Cho Jae Hyoung, Song MH.
Validation: Kim H, Shin H, Kim S, Kim ES.
Investigation: Kim H, Cho Jin Ho, Song MH, Kim HB, Lee JH.
Writing - original draft: Kim H, Cho Jae Hyoung, Song MH, Kim HB, Lee JH.
Writing - review & editing: Kim H, Cho Jin Ho, Song MH, Kim HB, Lee JH.

**Ethics approval and consent to participate**
This article does not require IRB/IACUC approval because there are no human and animal participants.

HMW DNA Kit (QIAGEN), and NucleoSpin® Microbial DNA kit (TAKARA) according to the manufacturer's instructions. Library preparation was conducted using SMRTbell™ Template Prep Kit 1.0 for Pacbio (Pacific Biosciences) and TruSeq DNA Sample Preparation Kit for Illumina (Illumina) according to the manufacturer's instructions. PacBio sequencing yielded 1,131,537,370 base pairs and 145,423 long reads after filtering, and 9,199,306 paired-end reads with 1,389,095,206 bp were obtained with Illumina sequencing. *De novo* assembly was conducted using the hierarchical genome assembly process (HGAP v2.3.0) workflow (Chin et al., 2013) and polished using Quiver. Subsequently, Illumina NextSeq reads were aligned to the PacBio RSII assembly using Burrows-Wheeler Aligner (BWA)-MEM v0.7.17-r1188, and the errors were corrected by using Pilon version 1.23 (2, 3). The quality of genome assembly and the validaty of the final genome were assessed using Quality Assessment Tool for Genome Assemblies (QUAST) v5.0.2 and Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (4, 5). Open reading

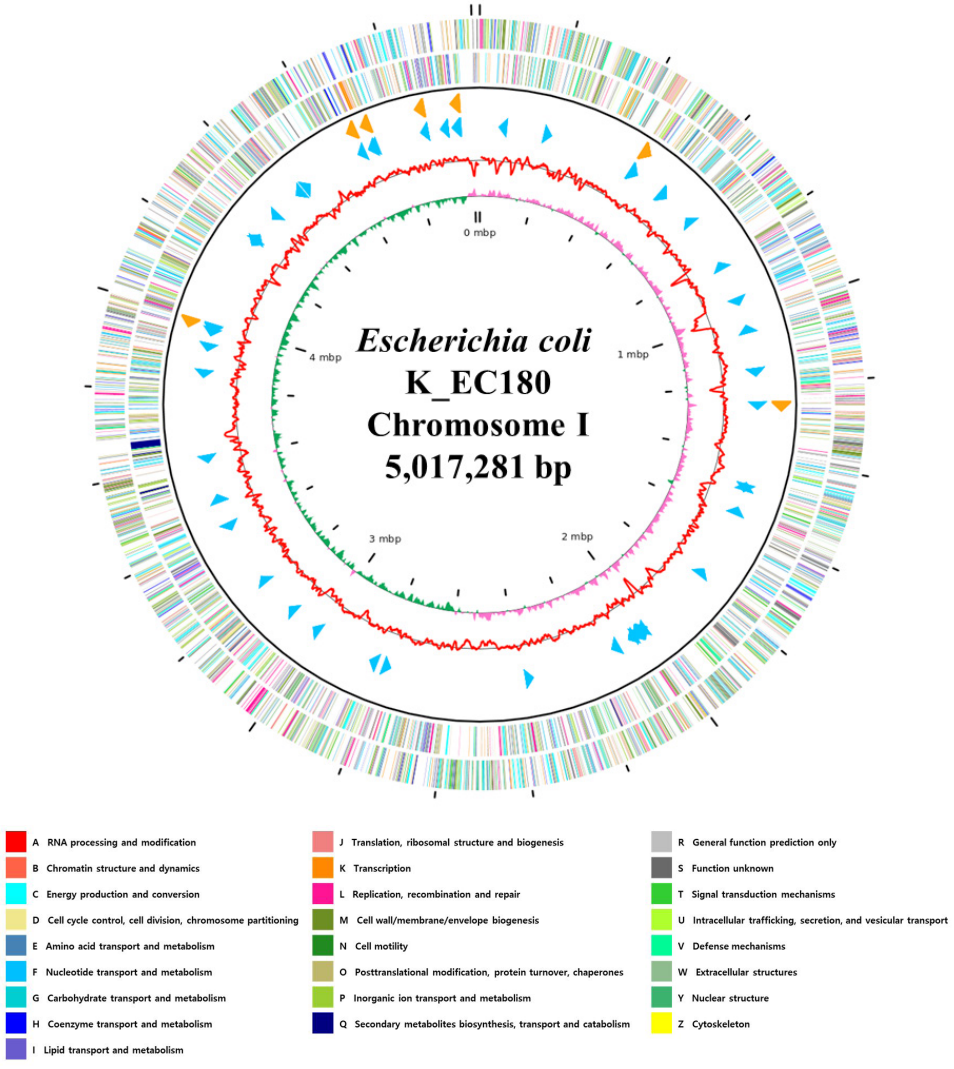


**Fig. 1. Genome map of *Escherichia coli* K_EC180.** The outer circle denotes the locations of all annotated ORFs, and the inner circle with the red denotes GC content. Pink, and green peaks denote GC skew. The orange arrows denote rRNAs, and the sky blue arrows denote the tRNA operons. All annotated ORFs are colored differently based on the COG assignments. ORFs, open reading frames; G, guanine; C, cytosine; COG, clusters of orthologous groups.


https://doi.org/10.5187/jast.2021.e38

**Table 1.** Genome features of *Escherichia coli* K_EC180

| Property | Term |
|---|---|
| Libraries used | PacBio SMRTbell™ library<br>TruSeq DNA Sample Preparation Kit |
| Sequencing platforms | PacBio RS II sequencer<br>Illumina NextSeq 500 |
| Assemblers | PacBio SMRT analysis v2.3.0 HGAP.3 |
| Annotation method | PROKKA v1.14.5 and RAST v2.0 |
| Average genome coverage | 100× |
| Chromosome length (bp) | 5,017,557 bp |
| No. of contigs | 1 |
| Guanine + cytosine (G + C) content (%) | 50.4 |
| Protein–coding genes (CDSs) | 4,935 |
| rRNA genes | 22 |
| tRNA genes | 88 |
| Plasmids | 0 |
| Genbank Accession No. | CP062203 |

frames (ORFs) and RNA genes of *E. coli* K_EC180 were predicted and functionally annotated through rapid prokaryotic genome annotation (PROKKA) v1.14.5 (6) and Rapid Annotation using Subsystem Technology (RAST) v2.0 (7). The functional categorization and classification of all predicted ORFs were conducted using the RAST server-based SEED viewer and Clusters of Orthologous Groups (COG) – based EggNOG. The putative virulence factors and Antimicrobial resistance were described using BLAST according to the Virulence Factor Database (VFDB) (8). The whole genome of *E. coli* K_EC180 is composed of one circular chromosome (5,017,281 bp) with 50.4% of G+C content, 4,935 of coding sequence (CDS), 88 of tRNA, and 22 of rRNA genes.

The complete genome of *E. coli* K_EC180 contains the toxin genes encoding shiga-like toxin (stx2e subunit A and stx2e subunit B), which may cause diseases in humans by damaging small blood vessels in places such as the digestive tract, kidneys and central nervous system (9, 10). *E. coli* K-EC180 also possessed *essC, escV, escR, escS, escV,* and *escJ* genes which involved in a type III secretion system. In addition, there were *fim* (A to H) genes encoding Type I fimbriae. We summarized the general properties of the *E. coli* K_EC180's complete genome in the Fig. 1 and Table 1.

## DATA AVAILABILITY

The complete genome sequences of *E. coli K*_EC180 were deposited in GeneBank under the accession numbers CP062203. The BioSample accession number is SAMN16277032, and BioProject accession number is PRJNA666028.

## REFERENCES

1. Guevarra RB, Lee JH, Lee SH, Seok MJ, Kim DW, Kang BN, et al. Piglet gut microbial shifts early in life: causes and effects. J Anim Sci Biotechnol. 2019;10:1. https://doi.org/10.1186/s40104-018-0308-3
2. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. https://arxiv.org/abs/1303.3997

3. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE. 2014;9:e112963. https://doi.org/10.1371/journal.pone.0112963
4. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210-2. https://doi.org/10.1093/bioinformatics/btv351
5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072-5. https://doi.org/10.1093/bioinformatics/btt086
6. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-9. https://doi.org/10.1093/bioinformatics/btu153
7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75. https://doi.org/10.1186/1471-2164-9-75
8. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucl Acids Res. 2019;47:D687-92.
9. Tesh VL, O'Brien AD. The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins. Mol Microbiol. 1991;5:1817-22. https://doi.org/10.1111/j.1365-2958.1991.tb00805.x
10. O'Brien AD, Holmes RK. Shiga and Shiga-like toxins. Microbiol Rev. 1987;51:206-20. https://doi.org/10.1128/MR.51.2.206-220.1987


https://doi.org/10.5187/jast.2021.e38