



<b>For the corresponding author (responsible for correspondence, proofreading, and reprints)</b>	<b>Fill in information in each box below</b>
First name, middle initial, last name	Hyeun Bum Kim <sup>1</sup> , Ju-Hoon Lee <sup>2</sup>
Email address – this is where your proofs will be sent	1 hbkim@dankook.ac.kr 2 juhlee@snu.ac.kr
Secondary Email address	
Address	1 Department of Animal Resources Science, Dankook University, Cheonan, South Korea 2 Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul, South Korea
Cell phone number	
Office phone number	
Fax number	

1 Complete chromosome sequence of *Escherichia coli* K\_EC180, a bacterium producing shiga-  
2 like toxin isolated from swine feces

3

4 Hyeri Kim<sup>1#</sup>, Jae Hyoung Cho<sup>1#</sup>, Jin Ho Cho<sup>2#</sup>, Minho Song<sup>3#</sup>, Hakdong Shin<sup>4</sup>, Sheena Kim<sup>1</sup>,  
5 Eun Sol Kim<sup>1</sup>, Hyeun Bum Kim<sup>1\*</sup>, Ju-Hoon Lee<sup>5\*</sup>

6

7 <sup>1</sup> Department of Animal Resources Science, Dankook University, Cheonan, South Korea

8 <sup>2</sup> Division of Food and Animal Science, Chungbuk National University, Cheongju, South Korea

9 <sup>3</sup> Division of Animal and Dairy Science, Chungnam National University, Daejeon, South Korea

10 <sup>4</sup> Department of Food Science and Biotechnology, College of Life Science, Sejong University,  
11 Seoul, South Korea

12 <sup>5</sup> Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center  
13 for Food and Bioconvergence, Seoul National University, Seoul, South Korea

14

15 # Equal contributors

16

17 \* Corresponding authors

18 Hyeun Bum Kim

19 Department of Animal Resources Science, Dankook University, Cheonan, South Korea

20 Tel: +82-41-550-3653

21 Email: [hbkim@dankook.ac.kr](mailto:hbkim@dankook.ac.kr)

22

23 Ju-Hoon Lee

24 Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center  
25 for Food and Bioconvergence, Seoul National University, Seoul, South Korea

26 Tel: +82-2-880-4854

27 Email: [juhlee@snu.ac.kr](mailto:juhlee@snu.ac.kr)

28

29 **Abstract (up to 350 words)**

30 *Escherichia coli* normally colonizes the lower intestine of animals and humans, but some serotypes are  
31 foodborne pathogens. The *Escherichia coli* K\_EC180 was isolated from swine feces that were collected  
32 from a weaner pig. In this genome announcement, *E.coli* K\_EC180 was sequenced using PacBio RS II and  
33 Illumina NextSeq 500 platforms. The complete chromosome of *E. coli* K\_EC180 is composed of one  
34 circular chromosome (5,017,281 bp) with 50.4 % of G+C content, 4,935 of coding sequence (CDS), 88 of  
35 tRNA, and 22 of rRNA genes. The complete genome of *E. coli* K\_EC180 contains the toxin genes such as  
36 shiga-like toxins (stxA and stxB).

37 **Keywords (3 to 6):** *Escherichia coli* K\_EC180, Swine feces, Whole genome sequencing, shiga-like toxin

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

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

57 *Escherichia coli* K\_EC180 was isolated from swine feces that were collected from a livestock farm  
58 in Haenam-gun, Jeollanam-do, Republic of Korea. *E. coli* K\_EC180 was streaked to Luria-Bertani(LB)  
59 agar and incubated at 37°C for 24 h. The suspected colony in LB agar was inoculated into LB broth and  
60 incubated at 37°C for 24 h. To analyze the complete genome, the *Escherichia coli* K\_EC180 genome was  
61 sequenced by PacBio RS II (Pacific Biosciences, USA) at Insilicogen (Yong-in-si, Gyeonggi-do, Republic  
62 of Korea) and Illumina NextSeq 500 (Illumina, Inc., USA) platform at LabGenomics (Seongnam-si,  
63 Gyeonggi-do, Republic of Korea). The genomic DNA of *Escherichia coli* K\_EC180 for PacBio and  
64 Illumina sequencing was extracted using the MagAttract HMW DNA Kit (QIAGEN), and NucleoSpin®  
65 Microbial DNA kit (TAKARA) according to the manufacturer's instructions. Library preparation was  
66 conducted using SMRTbell™ Template Prep Kit 1.0 for Pacbio (Pacific Biosciences, USA) and TruSeq  
67 DNA Sample Preparation Kit for Illumina (Illumina, Inc., USA) according to the manufacturer's  
68 instructions. PacBio sequencing yielded 1,131,537,370 base pairs and 145,423 long reads after filtering,  
69 and 9,199,306 paired-end reads with 1,389,095,206 bp were obtained with Illumina sequencing. *De novo*  
70 assembly was conducted using the hierarchical genome assembly process (HGAP v2.3.0) workflow (Chin  
71 et al., 2013) and polished using Quiver. Subsequently, Illumina NextSeq reads were aligned to the PacBio  
72 RSII assembly using Burrows-Wheeler Aligner (BWA)-MEM v0.7.17-r1188, and the errors were corrected  
73 by using Pilon version 1.23 (2, 3). The quality of genome assembly and the validity of the final genome  
74 were assessed using Quality Assessment Tool for Genome Assemblies (QUAST) v5.0.2 and Benchmarking  
75 Universal Single-Copy Orthologs (BUSCO) v3.0.2 (4, 5). Open reading frames (ORFs) and RNA genes of  
76 *E. coli* K\_EC180 were predicted and functionally annotated through rapid prokaryotic genome annotation  
77 (PROKKA) v1.14.5 (6) and Rapid Annotation using Subsystem Technology (RAST) v2.0 (7). The  
78 functional categorization and classification of all predicted ORFs were conducted using the RAST  
79 server-based SEED viewer and Clusters of Orthologous Groups(COG) – based EggNOG. The  
80 putative virulence factors and Antimicrobial resistance were described using BLAST according to the  
81 Virulence Factor Database (VFDB)(8). The whole genome of *E. coli* K\_EC180 is composed of one circular  
82 chromosome (5,017,281 bp) with 50.4 % of G+C content, 4,935 of coding sequence (CDS), 88 of tRNA,  
83 and 22 of rRNA genes.

84 The complete genome of *E. coli* K\_EC180 contains the toxin genes encoding shiga-like toxin (stx2e  
85 subunit A and stx2e subunit B), which may cause diseases in humans by damaging small blood vessels in  
86 places such as the digestive tract, kidneys and central nervous system(9, 10). *E. coli* K-EC180 also

87 possessed *essC*, *escV*, *escR*, *escS*, *escV*, and *escJ* genes which involved in a type III secretion system. In  
88 addition, there were *fim*(A to H) genes encoding Type I fimbriae. We summarized the general properties of  
89 the *E. coli* K\_EC180's complete genome in the figure 1 and table 1.

#### 90 **Data availability**

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

94

95

96

97

98

99

ACC

100 **References**

- 101 1. Guevarra RB, Lee JH, Lee SH, Seok MJ, Kim DW, Kang BN, et al. Piglet gut microbial shifts  
102 early in life: causes and effects. *Journal of animal science and biotechnology*. 2019;10:1.
- 103 2. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.  
104 arXiv:13033997v1 [q-bioGN. 2013.
- 105 3. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated  
106 tool for comprehensive microbial variant detection and genome assembly improvement. *PloS*  
107 *one*. 2014;9(11):e112963.
- 108 4. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing  
109 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*.  
110 2015;31(19):3210-2.
- 111 5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome  
112 assemblies. *Bioinformatics*. 2013;29(8):1072-5.
- 113 6. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*.  
114 2014;30(14):2068-9.
- 115 7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server:  
116 rapid annotations using subsystems technology. *BMC genomics*. 2008;9:75.
- 117 8. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform  
118 with an interactive web interface. *Nucleic acids research*. 2019;47(D1):D687-D92.
- 119 9. Tesh VL, O'Brien AD. The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins.  
120 *Molecular microbiology*. 1991;5(8):1817-22.
- 121 10. O'Brien AD, Holmes RK. Shiga and Shiga-like toxins. *Microbiological reviews*.  
122 1987;51(2):206-20.

123

124

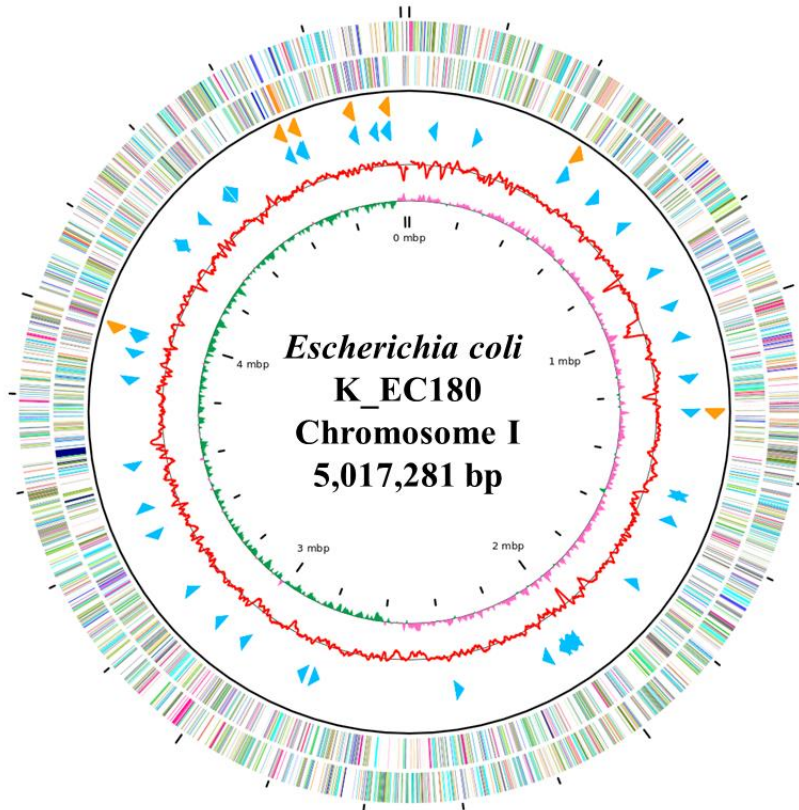
125 **Table 1. Genome features of *Escherichia coli* K\_EC180**

<b>Property</b>	<b>Term</b>
Libraries used	PacBio SMRTbell™ library TruSeq DNA Sample Preparation Kit
Sequencing platforms	PacBio RS II sequencer 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	100x
Chromosome length (bp)	5,017,557bp
No. of contigs	1
G+C content (%)	50.4
Protein-coding genes (CDSs)	4,935
rRNA genes	22
tRNA genes	88
Plasmids	0
Genbank Accession No.	CP062203

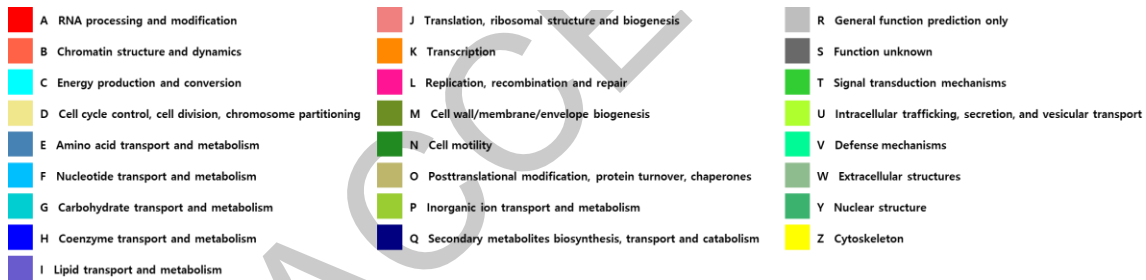
126

127





129



130

131 **Figure 1. Genome map of *Escherichia coli* K\_EC180**

132 The outer circle denotes the locations of all annotated ORFs, and the inner circle with the red  
 133 denotes GC content. Pink, and green peaks denote GC skew. The orange arrows denote rRNAs,  
 134 and the sky blue arrows denote the tRNA operons. All annotated ORFs are colored differently  
 135 based on the COG assignments.

136

137