Korean Journal of Microbiology (2022) Vol. 58, No. 4, pp. 324-326
DOI https://doi.org/10.7845/kjm.2022.2070

pISSN 0440-2413
eISSN 2383-9902

Complete genome sequence of enteropathogenic *Escherichia coli* MFDS1001074 isolated from food

Jiyoung Park[1†], Heeyeon Ha[2†], Eunsu Ha[2], Woojung Lee[1], Eun Sook An[1], Seung Hwan Kim[1], Jinho Choi[2*], and Soon Han Kim[1*]

[1]*Food Microbiology Division, Ministry of Food and Drug Safety, Cheongju 28159, Republic of Korea*
[2]*Sanigen Co., Ltd, Anyang 14059, Republic of Korea*

식품에서 분리된 장병원성대장균(enteropathogenic *Escherichia coli*) MFDS1001074의 유전체 서열 분석

박지영[1†] · 하희연[2†] · 하은수[2] · 이우정[1] · 안은숙[1] · 김승환[1] · 최진호[2*] · 김순한[1*]

[1]식품의약품안전처 식품의약품안전평가원 미생물과, [2](주)세니젠 R&D센터 진단기술연구소

(Received October 14, 2022; Revised December 15, 2022; Accepted December 16, 2022)

Virulent strains of *Escherichia coli* can cause gastroenteritis, urinary tract infections, neonatal meningitis, hemorrhagic colitis, and Crohn's disease in humans. This study elucidates the complete genome sequence of *Escherichia coli* strain MFDS1001074 isolated in 2012 from the ham of a high school meals in Cheongju, South Korea. The complete genome sequence of *Escherichia coli* strain MFDS1001074 consisted of a 4,807,852-bp chromosome and a 59,536-bp plasmid, with 50.49% and 49.08% G + C content, respectively. Gene predictions revealed that this strain contained 4,836 coding sequences (CDSs). Especially biofilm formation (*csgB*), capsule biosynthesis (*etp*), endotoxin (*gtrAB*), and type III secretion system (*espADG*) genes were identified.

Keywords: *Escherichia coli*, complete genome, ham

Enteropathogenic *Escherichia coli* (EPEC) is a bacterial pathogen that induces epithelial cell actin rearrangement, resulting in pedestal formation beneath adherent bacteria (Ochoa and Contreras, 2011). Its main mechanism is to introduce attaching-and-effacing (A/E) lesions on intestinal cells, characterized by microvillus destruction, intimate adherence of strains to the intestinal epithelium, pedestal formation, and aggregation of polarized actin at sites of bacterial attachment (Frankel *et al*., 1998). In this study, we present the complete genome sequence of an EPEC strain collected from Cheongju, South Korea, in 2012.

EPEC MFDS1001074 was detected and isolated in a food poisoning investigation managed by the Ministry of Food and Drug (MFDS, South Korea) after it was reported at a high school in Cheongju in 2012. This strain was detected in ham. After isolation, MFDS1001074 was incubated anaerobically in tryptic soy agar medium at 37°C for 24 h. The total genomic DNA of MFDS1001074 was extracted using the Genomic DNA Prep Kit (Bioneer). Genomic DNA was qualitatively and quantitatively measured using a NanoDrop 2000 UV-visible spectrophotometer (Thermo Fisher Scientific) and a Qubit 3.0 Fluorometer (Invitrogen). Complete genome sequencing was performed using the Illumina MiSeq and Oxford Nanopore MinION platform. For MiSeq sequencing, library preparation was performed using the Nextera DNA Flex Library Prep kit (Illumina) according to the manufacturer's instructions. The

[†]These authors contributed equally to this work.
***For correspondence.** (J. Choi) E-mail: cjh@sanigen.kr;
Tel.: +82-10-3211-8406; Fax: +82-2-573-3134 /
(S.H. Kim) E-mail: lambndog@korea.kr;
Tel.: +82-43-719-4301; Fax: +82-43-719-4300

Table 1. Genome features of EPEC MFDS1001074

Genomic feature	Chromosome	Plasmid
Contig	1	1
Genome size (bp)	4,807,852	59,536
GC content (%)	50.49	49.08
Genes	4,686	100
Protein-coding genes (CDSs)	4,829	100
tRNA genes	95	0
rRNA genes	22	0

size of the libraries was confirmed using a Bioanalyzer 2100 (Agilent Technologies). Illumina sequencing was performed using MiSeq system on a paired-end library with MiSeq Reagent Kit v2 (300-cycles). After sequencing, the individual sequence reads were analyzed using FastQC-v.0.11.8. The Illumina sequencing yielded 8,705,702 reads, including 1,190 Mb, with 239.7× coverage. The sequencing quality values Q20 and Q30 were 93.14% and 88.6%, respectively. For nanopore sequencing, the Oxford Nanopore Rapid Barcoding Kit (Oxford Nanopore Technologies) was used to construct libraries. Nanopore sequencing data were basecalled on guppy_barcoder v6.0.1. The Nanopore sequencing yielded 24,306 reads, including 104 Mb, with 20.1× coverage (N50, 8,128 bp). The sequencing mean quality was 13.5. The complete genome sequence of MFDS1001074 was 4,867,388 bp with a GC content of 50.47% (Table 1).

Illumina and nanopore sequencing data were processed and *de novo* assembled with Unicycler v0.4.8, using the Patho-systems Resource Integration Center (PATRIC) v3.6.12 web server; the genome was then annotated using RASTtk (Brettin *et al*., 2015). Virulence associated genes were predicted using the Virulence Factor Database (VFDB) (Liu *et al*., 2019) and PATRIC_VF (Snyder *et al*., 2007). A total of 4,793 genes were predicted. EPEC were defined as containing intimin (*eae*) and lacking Shiga toxins. Additionally, typical EPEC possessed a virulence plasmid known as EPEC adherence factor (EAF) plasmid including *bfpA*, *perA*, *perB*, and *perC* (Schmidt, 2010). However, atypical EPEC did not contain EAF plasmid (Schmidt, 2010). MFDS1001074 was detected *eae* and not detected EAF plasmid, suggesting that the strain was atypical EPEC. MFDS 1001074 was found to contain virulence associated genes, involved in biofilm formation (*agaBC*, *csgB*, *gatBCZ*, *kbaY*), capsule biosynthesis (*etk*, *etp*, *gfcABDE*), endotoxins (*gtrAB*), and a type III secretion system (*cesD*, *cif*, *escCDFJNQRSTUV*, *espADG*, *map*) (Table 2). The complete genome information will be useful for investigating potential possibilities for understanding foodborne pathogens and providing a genetic basis for a more detailed analysis of virulence factors.

Table 2. VAGs of EPEC MFDS1001074

Classification	Gene(s)	Predicted function	References
Biofilm formation	*agaBC*	Dihydroxyacetone phosphate synthesis from galactitol and galactosamine	Domka *et al*. (2007)
	csgB	Curli fibres	Antão *et al*. (2009)
	gatBCZ	Dihydroxyacetone phosphate synthesis from galactitol and galactosamine	Domka *et al*. (2007)
	kbaY	Ketose−bis−phosphate aldolase, tagatose−bisphosphate aldolase; part of aga cluster for K1 transport	Berlyn *et al*. (1998)
Capsule biosynthesis	*etk*	An inner membrane protein that catalyses tyrosine autophosphorylation and phosphorylation of a synthetic co-polymer poly	Ilan *et al*. (1999)
	etp	Phosphotyrosine−protein phosphatase and involved in capsule formation	Vincent *et al*. (2000)
	gfcABDE	Unknown functions specific to group 4 capsule export	Larson *et al*. (2021)
Endotoxin	*gtrA*	Synthesis of undecaprenyl phosphate-D-glucose (UndP-D-Glc) from UDP-D-Glc and UndP	Zdorovenko *et al*. (2018)
	gtrB	Translocation of UndP-D-Glc through the cytoplasmic membrane	Zdorovenko *et al*. (2018)
Type III secretion system (TTSS)	*espAD*	Translocator protein, involved in intimate adherence and regulatory elements, secreted by TTSS of the locus of enterocyte effacement (LEE).	Pakbin *et al*. (2021)
	espG	Effector protein, involved in intimate adherence and regulatory elements, secreted by TTSS of the LEE	Pakbin *et al*. (2021)
	map	Effector protein, involved in intimate adherence and regulatory elements, secreted by TTSS of the LEE	Pakbin *et al*. (2021)

Nucleotide sequence accession numbers

Nucleotide sequence accession numbers. The complete genome sequence of *Escherichia coli* MFDS1001074 has been deposited at the NCBI GenBank database under the accession numbers CP106929 (chromosome, MFDS1001074), CP106930 (plasmid, pMFDS1001074). The strain has been deposited in the Korean Culture Collection for foodborne Pathogens under the strain number MFDS1001074.

적 요

대장균의 독성 균주는 인간에서 위장염, 요로 감염, 신생아 뇌수막염, 출혈성 대장염 및 크론병을 유발할 수 있다. 이 연구는 2012년 대한민국 청주의 한 고등학교 급식의 햄에서 분리된 *Escherichia coli* (MFDS1001074)의 유전체 분석을 수행했다. *Escherichia coli* MFDS1001074는 4,807,852 bp 길이의 chromosomal DNA와 59,536 bp 길이의 plasmid로 구성되었으며, G + C contents는 각각 50.49% 및 49.08%였다. 유전자 예측에 따르면 이 균주는 염색체에 4,836개의 CDS를 나타냈다. 특히, biofilm formation (*csgB*), capsule biosynthesis (*etp*), endotoxin (*gtrAB*), type III secretion system (*espADG*) 관련 유전자를 확인했다.

Acknowledgments

This research was financially supported by the Ministry of Food and Drug Safety, Republic of Korea (20161MFDS030, 21162MFDS027).

Conflict of Interest

There are no conflicts of interest to declare.

References

Antão EM, Wieler LH, and Ewers C. 2009. Adhesive threads of extraintestinal pathogenic *Escherichia coli*. *Gut Pathog.* **1**, 22.

Berlyn MK. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. *Microbiol. Mol. Biol. Rev.* **62**, 814–984.

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, *et al.* 2015. *RASTtk*: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365.

Domka J, Lee J, Bansal T, and Wood TK. 2007. Temporal gene-expression in *Escherichia coli* K-12 biofilms. *Environ. Microbiol.* **9**, 332–346.

Frankel G, Phillips AD, Rosenshine I, Dougan G, Kaper JB, and Knutton S. 1998. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol. Microbiol.* **30**, 911–921.

Ilan O, Bloch Y, Frankel G, Ullrich H, Geider K, and Rosenshine I. 1999. Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. *EMBO J.* **18**, 3241–3248.

Larson MR, Biddle K, Gorman A, Boutom S, Rosenshine I, and Saper MA. 2021. *Escherichia coli* O127 group 4 capsule proteins assemble at the outer membrane. *PLoS ONE* **16**, e0259900.

Liu B, Zheng D, Jin Q, Chen L, and Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **47**, 687–692.

Ochoa TJ and Contreras CA. 2011. Enteropathogenic *E. coli* (EPEC) infection in children. *Curr. Opin. Infect. Dis.* **24**, 478–483.

Pakbin B, Brück WM, and Rossen JW. 2021. Virulence factors of enteric pathogenic *Escherichia coli*: a review. *Int. J. Mol. Sci.* **22**, 9922.

Schmidt MA. 2010. LEEways: tales of EPEC, ATEC and EHEC. *Cell. Microbiol.* **12**, 1544–1552.

Snyder EE, Kampanya N, Lu J, Nordberg EK, Karur HR, Shukla M, Soneja J, Tian Y, Xue T, Yoo H, *et al.* 2007. PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res.* **35**, 401–406.

Vincent C, Duclos B, Grangeasse C, Vaganay E, Riberty M, Cozzone AJ, and Doublet P. 2000. Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in Gram-negative bacteria. *J. Mol. Biol.* **304**, 311–321.

Zdorovenko EL, Wang Y, Shashkov AS, Chen T, Ovchinnikova OG, Liu B, Golomidova AK, Babenki VV, Letarov AV, and Knirel YA. 2018. O-Antigens of *Escherichia coli* strains O81 and HS3-104 are structurally and genetically related, except O-antigen glucosylation in *E. coli* HS3-104. *Biochemistry* **83**, 534–541.