Korean Journal of Microbiology (2024) Vol. 60, No. 1, pp. 50-53
DOI https://doi.org/10.7845/kjm.2024.3103

eISSN 2383-9902
pISSN 0440-2413

Complete genome sequence of enteropathogenic *Escherichia coli* MFDS1019811 isolated from donut

Jina Ha, Doo Won Seo, Woojung Lee, Eun Sook An, Seung Hwan Kim, and Soon Han Kim*

Food Microbiology Division, Ministry of Food and Drug Safety, Osong, Cheongju 28159, Republic of Korea

식품에서 분리된 장병원성대장균(enteropathogenic *Escherichia coli*) MFDS1019811의 유전체 서열분석

하진아 · 서두원 · 이우정 · 안은숙 · 김승환 · 김순한*

식품의약품안전처 식품의약품안전평가원 미생물과

(Received November 28, 2023; Revised March 15, 2024; Accepted March 25, 2024)

Escherichia coli, a foodborne pathogen that causes food poisoning, is implicated in intestinal infections. This study presents the complete genome sequence of enteropathogenic *E. coli* strain MFDS1019811, isolated from donut sample in 2021, which triggered a foodborne disease outbreak in Osan, South Korea. The complete genome sequence of *E. coli* strain MFDS 1019811 comprised 5,077,766 bp length chromosomal DNA and a 50,011 bp plasmid, with GC contents of 50.67% and 50.02%, respectively. Gene prediction revealed 92 tRNAs, 22 rRNAs and 5,048, and 90 CDSs on the genome. Several virulence genes were identified including bundle-forming pili (*bfp*) and intimin (*eae*).

Keywords: *Escherichia coli*, complete genome, donut

Enteropathogenic *Escherichia coli* (EPEC) are a major bacterial cause of diarrhea worldwide. EPEC infection causes acute watery diarrhea accompanied by dehydration, vomiting, and fever (Kaper *et al*., 2004). The primary mechanism of EPEC infection is attachment to epithelial cells *via* bundle-forming pili (*bfp*) and tight attachment with the help of translocated intimin receptor (*tir*) and intimin (*eae*), resulting in accumulation of actin and formation of pedestal structures (Ledwaba *et al*., 2020). This study presents the complete genome sequence of an EPEC strain collected from Osan, South Korea, in 2021.

EPEC MFDS1019811 was identified and obtained during food poisoning investigation conducted by the Ministry of Food and Drug Safety (MFDS, South Korea) following a reported outbreak at a high school in Osan in 2021. The strain was isolated from donut and incubated on tryptic soy agar medium at 37°C for 24 h. Genomic DNA was isolated using the Bioneer Genomic DNA Extraction Kit. Using a NanoDrop 2000 UV-visible spectrophotometer (Thermo Fisher Scientific) and a Qubit 3.0 Fluorometer (Invitrogen), the extracted DNA was quantified. Complete genome sequencing was performed using two sequencing facilities: Illumina MiSeq and Oxford Nanopore MinION. The Nextera DNA Flex Library Prep Kit (Illumina) was used for library preparation on Miseq. The size of the prepared library was confirmed using a Bioanalyzer 2100 (Agilent Technologies). The paired-end library sequencing was performed using the MiSeq system with MiSeq Reagent Kit v3 (600-cycles). Individual sequence reads were analyzed using FastQC-v.0.11.8 post-sequencing. For Nanopore sequencing, a library was constructed using a Native Barcoding Kit24 V14 (Oxford Nanopore Technologies) and basecalling performed on guppy_barcoder v6.0.1. Illumina and nanopore sequencing

***For correspondence.** E-mail: lambndog@korea.kr;
Tel.: +82-43-719-4301; Fax: +82-43-719-4300

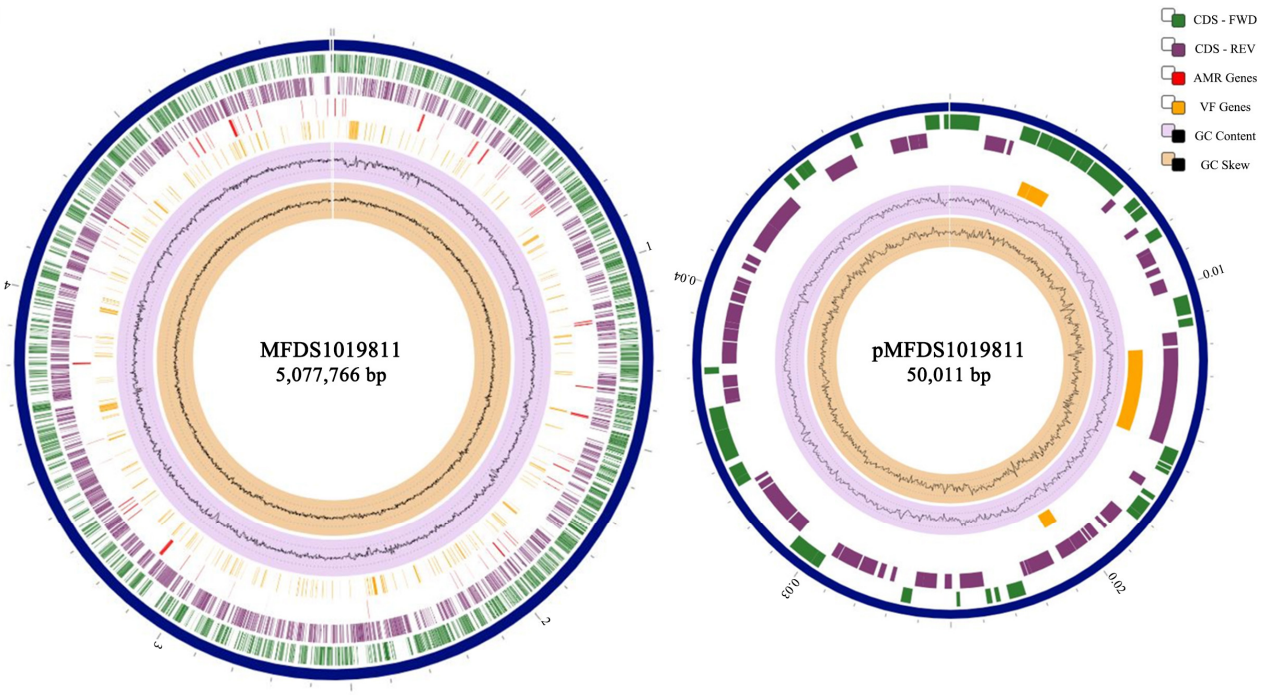

Fig. 1. Genome map of enteropathogenic *Escherichia coli* MFDS1019811 chromosomal DNA and plasmid. The graphical circular maps are shown of the alignment of genes and other genomic data. The tracks on the viewer are displayed as concentric rings, from outermost to innermost: CDS, AMR genes, VF genes, GC content, GC skew. Genes with specialized functions are labeled with different colors. Virulence-related genes in yellow, and antibiotic resistance in red.

Table 1. Genome features of EPEC MFDS1019811

Genomic features	Chromosome	Plasmid
Contig	2	1
Genome size (bp)	5,077,766	50,011
GC content (%)	50.67	50.02
Total number of CDSs	5,048	90
tRNA genes	92	0
rRNA genes	22	0

data were processed and hybrid-assembled using Unicycler v0.4.9. The assembled genome was then annotated using the RASTtk, through the BV-BRC web service (v3.32.13a). The Virulence Factor Database (VFDB) (Liu *et al*., 2022) and PATRIC_VF (Snyder *et al*., 2007) were used to predict the virulence-associated genes. MFDS1019811 genome comprises of 5,077,766 bp (50.67% GC contents) chromosome and 50,011 bp (50.02% GC contents) plasmid. 5,048 and 90 genes were identified, respectively (Table 1 and Fig. 1).

Formation of attaching and effacing (A/E) lesion in gut epithelium necessitates the presence of enterocyte effacement (LEE) in the bacterial genome. This locus encodes intimin (*eae*), a type III secretion system (T3SS) and six effectors, including the indispensable translocated intimin receptor (Tir) (Kaper *et al*., 2004). Additionally, typical EPEC possess *E. coli* adherence factor (EAF) associated the *bfp* gene, whereas atypical EPEC lack this factor (Schmidt, 2010). MFDS1019811 harbored *eae*, *tir* and *bfpA*. These results suggested that this strain is typical EPEC. Moreover, MFDS1019811 contains several virulence-associated genes related to biofilm formation (*agaBC*, *csgB*, *gatBCZ*, *kbaY*), capsule biosynthesis (*etk*, *etp*, *gfcABCDE*), invasion (*aslA*, *dam*, *ibeB*, *map*, *ompA*, *relA*, *yijP*), and typical III secretion system (*cesD*, *escCDFJNRSTUV*, *espB*, *sepDLQ*) (Table 2). This information will prove valuable for investigating food-borne pathogens and offers a genetic foundation for a more detailed analysis of virulence factors.

Nucleotide sequence accession number(s)

Accession numbers for nucleotide sequence. The complete genome sequence of *Escherichia coli* MFDS1019811 has been deposited at the NCBI GenBank database under accession numbers JAWWMY000000000, and the strain has been

Table 2. Virulence associated genes of EPEC MFDS1019811

Classification	Gene(s)	Predicted function	References
Biofilm formation	*agaBC*	PTS system, galactosamine-specific II component	Berlyn *et al.* (1998)
	csgB	Minor curlin subunit CsgB, nucleation component of curlin monomers	Berlyn *et al.* (1998)
	gatBCZ	PTS system, galactitol-specific II component (EC 2.7.1.200)	Berlyn *et al.* (1998)
	kbaY	D-tagatose-1,6-bisphosphate aldolase subunit KbaY (EC 4.1.2.40)	Berlyn *et al.* (1998)
Capsule biosynthesis	*etk*	Tyrosine-protein kinase (EC 2.7.10.2)	Ilan *et al.* (1999)
	etp	Low molecular weight protein-tyrosine-phosphatase (EC 3.1.3.48) => Etp	Vincent *et al.* (2000)
	gfcABCDE	Threonine-rich inner membrane protein Gfc	Larson *et al.* (2021)
Invasion	*aslA*	Arylsulfatase (EC 3.1.6.1)	Berlyn *et al.* (1998)
	dam	Methyl-directed repair DNA adenine methylase (EC 2.1.1.72)	Berlyn *et al.* (1998)
	ibeB	Copper/silver efflux RND transporter, outer membrane protein CusC	Berlyn *et al.* (1998)
	map	Putative chaperone (IpgB2)	Berlyn *et al.* (1998)
	ompA	Outer membrane protein A precursor	Berlyn *et al.* (1998)
	relA	Inactive (p)ppGpp 3'-pyrophosphohydrolase domain / GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase I	Berlyn *et al.* (1998)
	tir	Hypothetical protein	Berlyn *et al.* (1998)
	yijP	Phosphoethanolamine transferase EptC [*E. coli*], specific for LPS heptose I residue	Berlyn *et al.* (1998)
Type III secretion system (TTSS)	*cesD*	Type III secretion chaperone protein for YopD (SycD)	Parkbin *et al.* (2021)
	escC	Type III secretion outermembrane pore forming protein (YscC, MxiD, HrcC, InvG)	Parkbin *et al.* (2021)
	escD	Type III secretion protein SsaD	Parkbin *et al.* (2021)
	escF	Type III secretion protein SsaG	Parkbin *et al.* (2021)
	escJ	Type III secretion bridge between inner and outermembrane lipoprotein (YscJ, HrcJ, EscJ, PscJ)	Parkbin *et al.* (2021)
	escN	Type III secretion cytoplasmic ATP synthase (EC 3.6.3.14, YscN, SpaL, MxiB, HrcN, EscN)	Parkbin *et al.* (2021)
	escR	Type III secretion inner membrane protein (YscR, SpaR, HrcR, EscR, homologous to flagellar export components)	Parkbin *et al.* (2021)
	escS	Type III secretion inner membrane protein (YscS, homologous to flagellar export components)	Parkbin *et al.* (2021)
	escT	Type III secretion inner membrane protein (YscT, HrcT, SpaR, EscT, EpaR1, homologous to flagellar export components)	Parkbin *et al.* (2021)
	escU	Type III secretion inner membrane protein (YscU, SpaS, EscU, HrcU, SsaU, homologous to flagellar export components)	Parkbin *et al.* (2021)
	escV	Type III secretion inner membrane channel protein (LcrD, HrcV, EscV, SsaV)	Parkbin *et al.* (2021)
	espB	Secretion system effector SseD	Parkbin *et al.* (2021)
	sepD	Type III secretion system protein SepD	Parkbin *et al.* (2021)
	sepL	Type III secretion cytoplasmic protein (YscL)	Parkbin *et al.* (2021)

deposited in the Korean Culture Collection for foodborne Pathogens under strain number MFDS1019811.

적 요

대장균은 장관감염증과 같은 식중독을 유발하는 식품매개 병원균 중 하나이다. 본 연구에서는 2021년 오산의 한 고등학교 급식에서 일어난 식중독 사고의 원인 식품으로 추정되는 도넛으로부터 분리된 *Escherichia coli* (MFDS1019811)의 유전체 분석을 수행하였다. *Escherichia coli* MFDS1019811은 5,077,766 bp 길이의 chromosomal DNA와 50,011 bp 길이의 plasmid로 구성되었으며, 각각의 G + C contents는 50.67%와 50.02%로 확인되었다. 유전자 예측 결과, 92개 tRNA, 22개 rRNA, 그리고 chromosome과 plasmid 각각 5,048개와 90개의 단백질 코딩유전자가 동정되었다. 또한, bundle-forming pili (*bfp*)와 intimin (*eae*)를 포함한 여러 병원성 유전자들이 확인되었다.

Acknowledgments

This research was financially supported by the Ministry of Food and Drug Safety, Republic of Korea (23194MFDS016).

Conflict of Interest

The authors have no conflict of interest to report.

References

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

Ilan O, Bloch Y, Frankel G, UIlrich 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.

Kaper JB, Nataro JP, and Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, **2**, 123–140.

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.

Ledwaba SE, Costa DVS, Bolick DT, Giallourou N, Medeiros PHQS, Swann JR, Traore AN, Potgieter N, Nataro JP, and Guerrant RL. 2020. Enteropathogenic *Escherichia coli* infection induces diarrhea, intestinal damage, metabolic alterations, and increased intestinal permeability in a murine model. *Front. Cell. Infect. Microbiol.* **10**, 595266.

Liu B, Zheng D, Zhou S, Chen L, and Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917.

Pakbin B, Brück WM, and Rossen JWA. 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**, D401–D406.

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.