creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학석사학위논문

한국의 집단 식중독 사태에서 분리된 장독소대장균의 비교유전체 연구

Comparative genomic analysis of enterotoxigenic *Escherichia coli* strains isolated from diarrheal patients in Korea

2018년 8월

서울대학교 대학원

생명과학부

정 시 윤

ABSTRACT

Comparative genomic analysis of enterotoxigenic *Escherichia coli* strains isolated from diarrheal patients in Korea

Siyun Chung
School of biological sciences
The graduate school
Seoul National University

Enterotoxigenic *Escherichia coli* (ETEC) is a common cause of bacterial infection leading to diarrhea. ETEC is defined by the ability to produce enterotoxin, a heat-labile toxin (LT) and/or heat-stable toxin (ST). ETEC harbors one or more cell adhesion factors called colonization factors(CFs) to colonize on epithelial cells of intestinal surfaces of hosts. In a previous study, 258 ETEC isolates from patients with diarrhea in Korea were analyzed about CFs and multi-locus sequence typing (MLST). ST171 (24%) was identified as the most prevalent ETEC type in Korea, followed by ST955, ST964 and ST656. However, genomic and pathogenic characteristics of major MLST type of ETEC in Korea had not yet been investigated.

In this study, whole-genome sequencing was performed to reveal the genomic characteristics of five predominant MLST types of ETEC isolates in Korea during 2003-2011: *E. coli* O159 NCCP15731 and NCCP15733 (Chapter 1), *E. coli* O6 NCCP15732 (chapter 2), *E. coli* O169 NCCP15735 (chapter 3) and *E. coli* O25 NCCP15741 (chapter 4). Moreover, to identify their pathogenic characteristics, comparative genomic analysis was performed among same serotypes including reference ETEC strains reported over a similar time period.

In summary, new knowledges of ETEC strains in Korea were found. Domestic ETEC had unique genes and their combinations of colonization factors with enterotoxin type were consistent with that of globally distributes ETEC isolates. This thesis presents a basic study to understand the biological functions of ETEC pathogenicity and Korean specific pathogens.

Key words: pathogen, Whole-genome sequencing, comparative genomic analysis, *Escherichia coli*, ETEC, virulence factor, MLST

Student number: 2016-22634

CONTENTS

ABSTRACT · · · · · · · · · · i
CONTENTS · · · · · · · · · · iii
LIST OF FIGURES · · · · · · · · · · v
LIST OF TABLES · · · · · · · · · · vii
INTRODUCTION · · · · · · · · · · 1
MATERIALS AND METHODS · · · · · · · · · · 5
Bacteria and Strain isolation · · · · · · · · · · 5
Whole genome sequencing, assembly and annotation · · · · · · · · · · 5
Genomic analysis · · · · · · · · · · 6
Phylogenetic analysis · · · · · · · · · · 9
Virulence factor analysis · · · · · · · · · · 10
RESULTS · · · · · · · · · · 12
Chapter 1. Comparative genomic analysis of enterotoxigenic
***Escherichia coli* O159 strains** · · · · · · · · · · 12
Genomic analysis · · · · · · · · · · 12
Phylogenetic analysis · · · · · · · · · · 16
Virulence factor analysis · · · · · · · · · · 20
Comparison with other *E. coli* strains in phylo-group A · · · · · · · · · · 23
Chapter 2. Comparative genomic analysis of enterotoxigenic
***Escherichia coli* O6 strains** · · · · · · · · · · 27
Genomic analysis · · · · · · · · · · 27
Phylogenetic analysis · · · · · · · · · · 30
Virulence factor analysis · · · · · · · · · · 34
Chapter 3. Comparative genomic analysis of enterotoxigenic
***Escherichia coli* O169 strains** · · · · · · · · · · 41
Genomic analysis · · · · · · · · · · 41
Phylogenetic analysis · · · · · · · · · · 44
Virulence factor analysis · · · · · · · · · · 50

Chapter 4. Comparative genomic analysis of enterotoxigenic *Escherichia coli* O25 strains · · · 58
Genomic analysis · · · 58
Phylogenetic analysis · · · 61
Virulence factor analysis · · · 65
DISCUSSION · · · 71
CONCLUSION · · · 75
REFERENCES · · · 76
APPENDIX · · · 86
Appendix 1. Accession Number of reference strains in this study · · · 86
Appendix 2. Virulence factors of ETEC O159 strains · · · 89
Appendix 3. Virulence factors of ETEC O6 strains · · · 103
ABSTRACT (In Korean) · · · 114

LIST OF FIGURES

Fig. 1. Subsystem category distribution of NCCP15731 · · · · · · · · · · · · · · · · · · 13

Fig. 2. Subsystem category distribution of NCCP15733 · · · · · · · · · · · · · · · · · · 14

Fig. 3. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O159 strains ·18

Fig. 4. Maximum likelihood (ML) tree of MLST sequences of ETEC O159 strains · 19

Fig. 5. Hierarchical clustering of 16 ETEC strains according to virulence factors 21

Fig. 6. Comparison of virulence factors of 6 ETEC strains in phylo-group A · · · 26

Fig. 7. Subsystem category distribution of NCCP15732 · · · · · · · · · · · · · · · · · · 28

Fig. 8. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O6 strains ·31

Fig. 9. Dendrogram and heatmap of 16 ETEC O6 strains according to virulence factors · 35

Fig. 10. Comparison of virulence factors of NCCP15732, NCCP15740 and non-ETEC strain ATCC25922 · 40

Fig. 11. Subsystem category distribution of NCCP15735 · · · · · · · · · · · · · · · · · 42

Fig. 12. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O169 strains ·48

Fig. 13. Maximum likelihood (ML) tree of MLST sequences of ETEC O169 stra ins · 49

Fig. 14. Dendrogram and heatmap of 16 ETEC O169 strains according to virulence factors · 51

Fig. 15. Comparison of virulence factors of five ETEC O169 strains · · · · · · · · ·57

Fig. 16. Subsystem category distribution of NCCP15741 · · · · · · · · · · · · · · · · 59
Fig. 17. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O25 strains · 63
Fig. 18. Maximum likelihood (ML) tree of MLST sequences of ETEC O25 strains · 64
Fig. 19. Hierarchical clustering of 11 ETEC strains according to virulence factors · 66

LIST OF TABLES

Table 1. Enterotoxigenic *Escherichia coli* isolates studied in this thesis · · · · · · · ·4

Table 2. Primers used in this thesis · 7

Table 3. Forward and reverse sequences of MLST-primers · · · · · · · · · · · · · · · · · ·11

Table 4. Genomic characteristics of NCCP15731 and NCCP15733 · · · · · · · · · · 15

Table 5. Genomic features of whole genome datasets of ETEC O159 strains · · · ·17

Table 6. Average nucleotide identity values based on USEARCH · · · · · · · · · · ·24

Table 7. Genomic characteristics of NCCP15732 · 29

Table 8. Genomic features of whole genome datasets of ETEC O6 strains · · · · ·32

Table 9. Adherence system encoded by ETEC O6 strains genome · · · · · · · · · · 36

Table 10. Major virulence factors of ETEC O6 strains · · · · · · · · · · · · · · · · · · ·39

Table 11. Genomic characteristics of NCCP15735 · 43

Table 12. Genomic features of whole genome datasets of ETEC O169 strains · · 46

Table 13. Adherence system encoded by ETEC O169 strains · · · · · · · · · · · · · 53

Table 14. Secretion system encoded by ETEC O169 strains · · · · · · · · · · · · · · ·55

Table 15. Genomic characteristics of NCCP15741 · 60

Table 16 Genomic features of whole genome datasets of ETEC O25 strains · · · ·62

Table 17. Adherence system encoded by ETEC O25 strains · · · · · · · · · · · · · · 68

Table 18. Anti-phagocytosis system encoded by ETEC O25 strains · · · · · · · · · 68

Table 19. Secretion system encoded by ETEC O25 strains · · · · · · · · · · · · · · · ·70

INTRODUCTION

Enterotoxigenic *E. coli* (ETEC) is a common cause of bacterial infection that leads to diarrhea in infants, young children and traveler in developing countries (Wennerås and Erling 2004, Organization 2006). Because ETEC outbreaks are gained by consuming contaminated food or water, food contamination is an important concern for public health (Gómez-Aldapa et al. 2013, Shin et al. 2016). In Korea, outbreak of foodborne or waterborne diseases by ETEC was the second most common of *E. coli* outbreaks in 2014 (Oh et al. 2015). For examples, a massive ETEC outbreak reported in Korea in 2012 was caused by ETEC O169 (Shin et al. 2016).

ETEC is defined by the ability to produce enterotoxin, a heat-labile toxin (LT) and/or heat-stable toxin (ST, including STh and STp), that can disturb the intestinal secretory state, thereby causing watery diarrhea of infected patients (Nataro and Kaper 1998, Qadri et al. 2005, Fleckenstein et al. 2010). While STh is produced by ETEC isolated from humans, STp originally found in pig ETEC is also associated with disease in humans (Bölin et al. 2006). ETEC harbors one or more cell adhesion factors called colonization factors (CFs) to colonize on epithelial cells of intestinal surfaces of hosts. Currently, more than 30 colonization factors (CFs) have been identified in human ETEC, and co-expressed with one, two or three CFs and/or toxic factors, such as CS1 + CS3 (±CS21) with LT + STh, CS2 + CS3 (±CS21) with LT + STh, CS5 + CS6 with LT + STh, CS6 with STp, CFA/I (±CS21)

with STh and CS7 with LT, globally (Gaastra and Svennerholm 1996, Svennerholm and Lundgren 2012, von Mentzer et al. 2014).

Phylogenetic analysis is an important for investigating the evolution and diversity of *E. coli* and evaluating bacterial pathogenicity (Clermont et al. 2013). Most commensal and pathogenic *E. coli* strains belonged to phylo-groups A and B1 (Escobar-Páramo et al. 2004, Badouei et al. 2015) and about 90% of foodborne *E. coli* isolates in Korea belonged to phylo-groups A and B1 (Koo et al. 2012). In recent studies, results of whole genome sequencing of a representative collection of ETEC isolated globally showed that they shared distinct colonization factors and enterotoxin profiles and were divided into phylo-groups (von Mentzer et al. 2014). This suggests that researching long-term ETEC clades can find a combination that optimizes fitness and transmissibility, and affect the tracking and prevention of ETEC disease.

Up to now, a lot of researches about new colonization factors and phylogenetic analysis among pathogens with the same serotype have been reported (Hazen et al. 2017, Kwon et al. 2017, Von Mentzer et al. 2017). However, despite many ETEC pathogen studies in each country, there have been limited such study conducted on the pathogenicity of ETEC outbreaks in Korea and comparative analysis with ETEC distributed globally. Therefore, it is required to study pathogen by revealing specific characters of ETEC in Korea because these results will provide important information for the development of an ETEC vaccine candidate and play a key role in increasing public health.

Recently, 258 ETEC isolates from patients with diarrhea in Korea were analyzed about CFs and multi-locus sequence typing (MLST) (Oh et al. 2014). 65 different MLST types were identified from domestic human ETEC isolates and ST171 (24%) was the most prevalent MLST type of domestic ETEC, followed by ST955, ST964 and ST656. NCCP15740 (Kwon et al. 2017), representing the major MLST type ST171 of ETEC in Korea, was investigated about genomic features, CF genes, and virulence factors. However, other MLST types have not been investigated. In this study, whole-genome sequencing was performed to reveal the genomic characteristics of five predominant MLST types of ETEC isolates in Korea (Table. 1).

Comparative genomic analysis was performed among same serotypes including reference ETEC strains reported over a similar time period: *E. coli* O159 NCCP157 31 and NCCP15733 (Chapter 1), *E. coli* O6 NCCP15732 (chapter 2), *E. coli* O169 NCCP15735 (chapter 3), and *E. coli* O25 NCCP15741 (chapter 4). Searching essential genes among the same serotypes will suggest effective method for the prevention and treatment of ETEC outbreaks. In addition, finding unique virulence genes in Korea will be important information for understanding the pathogenicity of domestic ETEC.

Table. 1. Enterotoxigenic *Escherichia coli* isolates studied in this thesis.

Strain	MLST type	Serotype	Genome Size (Mbp)	GenBank Accession
NCCP15731	ST964	O159:H34	4.66	QICG00000000
NCCP15732	ST955	O6:H16	4.8	-
NCCP15733	ST656	O159:H34	4.64	QICF00000000
NCCP15735	ST737	O169:H41	5	-
NCCP15741	ST949	O25:H16	4.71	-

* In silico serotyping of *E. coli* isolates was performed using SerotypeFinder (version 1.1) (Joensen, Tetzschner et al. 2015)

* O n.i.: not identified

MATERIALS AND METHODS

Bacteria and Strain isolation

All samples were isolated from diarrheal patients using MacConkey agar and Trypticase Soy Broth containing vancomycin. Candidate colonies of *E. coli* strains were identified based on phenotypes and biochemical properties using the API20E system (Biomerieux, Marcy l'Etoile, France). The isolated strains were deposited in the National Culture Collection for Pathogens (NCCP) under the registration numbers NCCP15731, NCCP15732, NCCP15733, NCCP15735 and NCCP15741. Five *E. coli* strains were identified as five major MLST types of ETEC in Korea (Oh et al. 2014).

Whole genome sequencing, assembly and annotation

A sequencing library was created using TruSeq sample preparation kit (Illumina, San Diego, CA, USA). Whole-genome sequencing of five ETEC strains was performed using the Illumina HiSeq 2000 platform (Theragen Etex Bio Institute, Suwon, Republic of Korea). Low-complexity reads, quality scores <Q20, as well as duplicate reads were deleted. SOAP de novo (version 1.05) (Li et al. 2010) was used to assemble high-quality reads. Assembled contigs of each strains wer e annotated using the Rapid Annotation using Subsystem Technology (RAST version 4.0) (Aziz et al. 2008) server pipeline.

Genomic analysis

In silico SerotypeFinder (version 1.1) (Joensen et al. 2015) was used to identify serotype of five strains. Colonization factors and multi-locus sequence typing (MLST) were identified using CF primers (Rodas et al. 2009, Liu et al. 2017) and *E. coli* MLST database (Wirth et al. 2006). The seven MLST genes were *adk, fumC, gyrB, icd, mdh, purA,* and *recA* from ETEC strains. The primer sequences using this thesis were listed in Table. 2. The same method as described by Clermont et al. (von Mentzer et al. 2014) was performed *in silico* to identify *E. coli* phylo-groups (A, B1, B2, D, E and F) using primersearch program from the European Molecular Biology Open Software Suite (EMBOSS) (Rice et al. 2000).

Table. 2. Primers used in this thesis.

Target	Gene	Sequence (F: Forward, R: reverse)	Amplicon length(bp)
Enterotoxin	*LT*	F: ACGGCGTTACTATCCTCTC	273
		R: TGGTCTCGGTCAGATATGTG	
	STh	F: TGGTCTCGGTCAGATATGTG	120
		R: TGGTCTCGGTCAGATATGTG	
	STp	F: TGAATCACTTGACTCTTCAAAA	136
		R: GGCAGGATTACAACAAAGTT	
CFA/I	*cfaB*	F: GCTTATTCTCCCGCATCAAA	170
		R: ACTTGTCCTCCCCATGACAC	
CS1	*csoA*	F: TCCGTTCGGCTAAGTCAGTT	243
		R: CCGCACATTTCCTGTGTTCT	
CS2	*cotA*	F: AGTGGTGGCAGCGAAACTAT	368
		R: TTCCTCTGTGGGTTCTCAGG	
CS3	*cstA*	F: CTAGCTTTGCCACCACCATT	100
		R: GGCAACTGACTCCCATTTGT	
CS4	*csaB*	F: ACCTGCGGCAAGTCGTTT	198
		R: TCTGCAGGTTCAAAAGTCACA	
CS5	*csfA*	F: TCCGCTCCCGTTACTCAG	226
		R: GAAAAGCGTTCACACTGTTTATATT	
CS6	*cssB*	F: CTGTGAATCCAGTTTCGGGT	152
		R: CAGGAACTTCCGGAGTGGTA	
CS7	*csvA*	F: CGCCGGTTACACGTAGTGAT	154
		R: CCATTTAAAGTGATTGCGACTT	
CS8	*cofA*	F: ATCCGGATTATCAAGCTCCA	166
		R: GAAGATGTTATTGCACCACCAA	
CS12	*cswA*	F: CCAGTCTATGCCAGGTTGCT	137

		R: TGTGGGGTCACAGTTTACCA	
CS13	*cshE*	F: GGGACTGCCACAATGAATTT	178
		R: CAGCACCACCTGCTGATTTA	
CS14	*csuA1*	F: TTTGCAACCGACATCTACCA	162
		R: CCGGATGTAGTTGCTCCAAT	
CS15	*nfaA*	F: CGAAATTGGACAAGCGATG	130
		R: TCCAGCAGGGATATTATTCG	
CS17/19	*csbA-csdA*	F: CGGTGCGTTTAACACAGCTA	195
		R: CTCAGGCGCAGTTCCTTGT	
CS18	*fotA*	F: ATCCGTCAGGTGTTTGTGGT	362
		R: CACCTGAATTCCTCGACAGG	
CS21	*lngA*	F: TCATGAGCCTGCTGGAAGTTATCA	617
		R: TCCGGCTACCTAAAGTAATTGAGT	

Phylogenetic analysis

Reference ETEC strains with the same O serotype as each of five domestic strains used this study were collected for phylogenetic analysis. These reference strains have been isolated globally over a similar time period. Assembly sequences of reference strain were obtained from NCBI and Illumina short reads of reference strain were obtained from NCBI Sequence Read Archive (SRA). The Accession numbers were subscripted in Appendix 1. *De novo* assemblies of their Illumina short reading frame were performed with Spades (version 3.5.0) (Bankevich et al. 2012).

To identify evolutionary relationship with each of strains and reference strains with same O serotype, phylogenetic analysis was performed. Multiple sequence alignments obtained from whole genome sequences of each ETEC strains were performed with Mugsy (version 1.2.3) (Angiuoli and Salzberg 2010). The generalized time reversible + CAT model (Stamatakis 2006) was used for maximum-likelihood phylogenetic tree construction using FastTree (version 2.1.7) (Price et al. 2009). The resulting trees were visualized with FigTree (ver 1.3.1; http://tree.bio.ed.ac.uk/software/ figtree/).

In order to exclude the effect of HGT in phylogenetic analysis, multi-locus sequence analysis method was used (Khan et al. 2008, Glaeser and Kämpfer 2015). Seven housekeeping genes (*adk, fumC, gyrB, icd, mdh, purA* and *recA)* from each ETEC isolates were obtained according to the protocol described in *E. coli* MLST

database (Wirth et al. 2006). Housekeeping genes were selected from the E. coli database at the MLST site (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi_html) (Table. 3). A phylogenetic tree based on MLST genes was created using the method employed for whole-genome phylogenetic analysis.

Virulence factor analysis

To identify the genes encoding virulence factors, the total CDSs of five domestic ETEC strains and reference strains were analyzed using blastp against the virulence factor genes of *E. coli* listed in VFDB (bacterial virulence factors database) with an e-value of 1e-5 (Chen et al. 2005). Genes with coverage at least 60% were selected.

Table. 3. Forward and reverse sequences of MLST-primers.

Gene (function)	Primer sequence (5'-3') (F: forward, R: reverse)	Amplicon size (bp)
adk (adenylate kinase)	ATTCTGCTTGGCGCTCCGGG (F)	583
	CCGTCAACTTTCGCGTATTT (R)	
fumC (fumarate hydratase)	TCACAGGTCGCCAGCGCTTC (F)	806
	GTACGCAGCGAAAAAGATTC (R)	
icd (isocitrate/isopropylmalate dehydrogenase)	ATGGAAAGTAAAGTAGTTGTTCCGGCACA (F)	878
	GGACGCAGCAGGATCTGTT (R)	
purA (adenylosuccinate dehydrogenase)	CGCGCTGATGAAAGAGATGA (F)	816
	CATACGGTAAGCCACGCAGA (R)	
gyrB (DNA gyrase)	TCGGCGACACGGATGACGGC (F)	911
	ATCAGGCCTTCACGCGCATC (R)	
recA (ATP/GTP binding motif)	CGCATTCGCTTTACCCTGACC (F)	780
	TCGTCGAAATCTACGGACCGGA (R)	
mdh (malate dehydrogenase)	ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG (F)	932
	TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT (R)	

RESULTS

Chapter 1. Comparative genomic analysis of enterotoxigenic *Escherichia coli* O159 strains

Genomic analysis

The genome size of NCCP15731 was 4,663,459 bp with G+C content of 50.7%. The genome size of NCCP15733 was 4,645,336 bp with G+C content of 50.6%. Based on RAST analysis, 4,435 coding sequences and 19 tRNA genes were detected in genome of NCCP15731, of which 3,855 (87%) were functional (Fig. 1). A total number of 4,369 coding sequences and 23 tRNA genes were detected in genome of NCCP15733, of which 3,852 (88%) were functional (Fig. 2). *In silico* analysis revealed that serotype of both NCCP15731 and NCCP15733 was O159:H34 belonging to phylo-group A, one of major *E. coli* phylo-groups (Badouei et al. 2015). Genomic and phylogenetic characteristics of NCCP15731 and NCCP15733 were summarized in Table. 4. Toxin type and colonization factors of NCCP15731 and NCCP15733 were investigated. Both NCCP15731 and NCCP 15733 possessed ST enterotoxin but not LT enterotoxin.

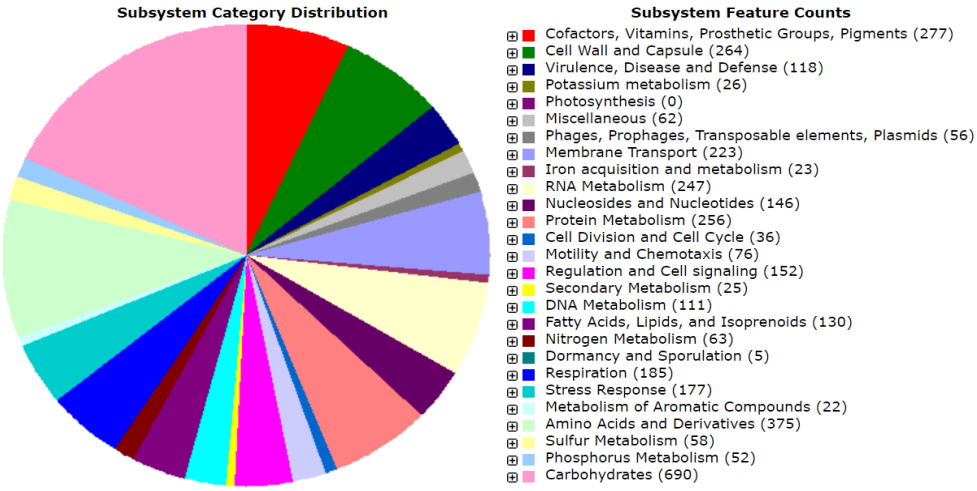

Fig. 1. Subsystem category distribution of NCCP15731. Based on RAST analysis, 4,435 coding sequences were detected in NCCP15731, of which 3,855 (87%) were functional.

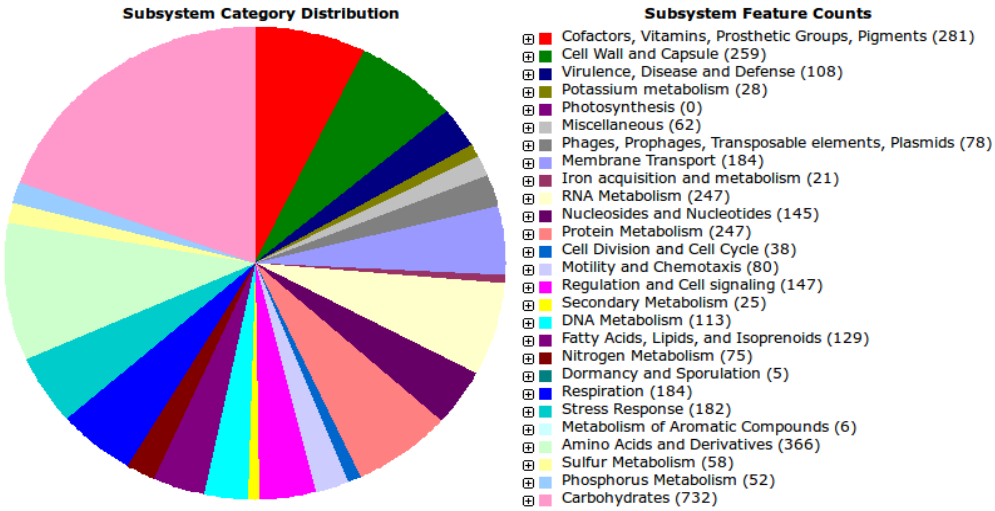

Fig. 2. Subsystem category distribution of NCCP15733. Based on RAST analysis, 4,369 coding sequences were detected in NCCP15733, of which 3,852 (88%) were functional.

Table. 4. Genomic characteristics of NCCP15731 and NCCP15733.

Strain	**NCCP15731**	**NCCP15733**
Genome size	4,663,459	4,645,336
% GC	50.7	50.6
N50	9,457	14,126
L50	147	105
Total contigs	986	694
Total subsystems	587	588
Total CDS	4,435	4,369
tRNAs	19	23
Serotype	O159:H34	O159:H34
Toxin type	STh	STh
Phylo-group	A	A

Phylogenetic analysis

Phylogenetic comparison of candidate genes, implemented in SEED (Overbeek et al. 2013), showed that both NCCP15731 and NCCP15733 were most close to *E. coli* O157:H7 str. 88.1467 (score: 531 and 530, respectively). 12 *E. coli* O159 strains were selected as reference strains with the same O serotype as NCCP15731 and NCCP15733. These 12 *E. coli* O159 strains have been isolated globally from 1980 to 2011 (von Mentzer et al. 2014) Their genomic features were shown in Table. 5.

Whole-genome phylogenetic analysis and MLST-based phylogenetic analysis were performed of sequence of NCCP15731, NCCP15733 and 12 *E. coli* reference strains. Whole-genome phylogenetic tree showed that NCCP15731 and NCCP15733 belonged to group A and both of two strains were closer to *E. coli* O159:H34 str. E159 with sequence type of ST218 (Fig. 3). Seven strains belonging to phylo-group B1 clustered into a single clade and, seven strains belonging to phylo-group A clustered into a single clade. However, in MLST phylogenetic analysis, phylo-group A clustered into a single clade whereas, phylo-group B1 did not cluster (Fig. 4). *E. coli* E159 strain was placed in a single clade with NCCP15731 and NCCP15733 in whole-genome and MLST phylogenetic tree.

Table. 5. Genomic features of whole genome datasets of ETEC O159 strains.

Strain	Genome (Mb)	Phylo-group	MLST type	Serotype	Toxin type	CF type [1]
E. coli NCCP15731	4.66	A	218	O159:H34	STh	CF NEG
E. coli NCCP15733	4.64	A	218	O159:H34	STh	CF NEG
E. coli E159	4.95	A	218	O159:H34	LT	CF NEG
E. coli E133	4.89	A	10	O159:H4	LT+STp	CS12
E. coli E1532	4.83	A	10	O159:H4	LT	CF NEG
E. coli E1573	4.95	A	10	O159:H4	LT	CF NEG
E. coli E1679sc	4.95	A	10	O159:H4	LT+STp	CS12
E. coli E391	4.91	B1	641	O159:H21	LT	CF NEG
E. coli E833	5.09	B1	641	O159:H34	LT	CS28a
E. coli E885	4.96	B1	641	O159:H21	LT	CF NEG
E. coli E1640	5.01	B1	641	O159:H34	LT+ST	CF NEG
E. coli E5085	5.06	B1	1490	O159:H34	STp	CS6
E. coli E5088	5.07	B1	1490	O159:H34	STp	CS6
E. coli E5051	5.25	B1	295	O159:H18	LT	CS6

[1] CF NEG : CF "negative"

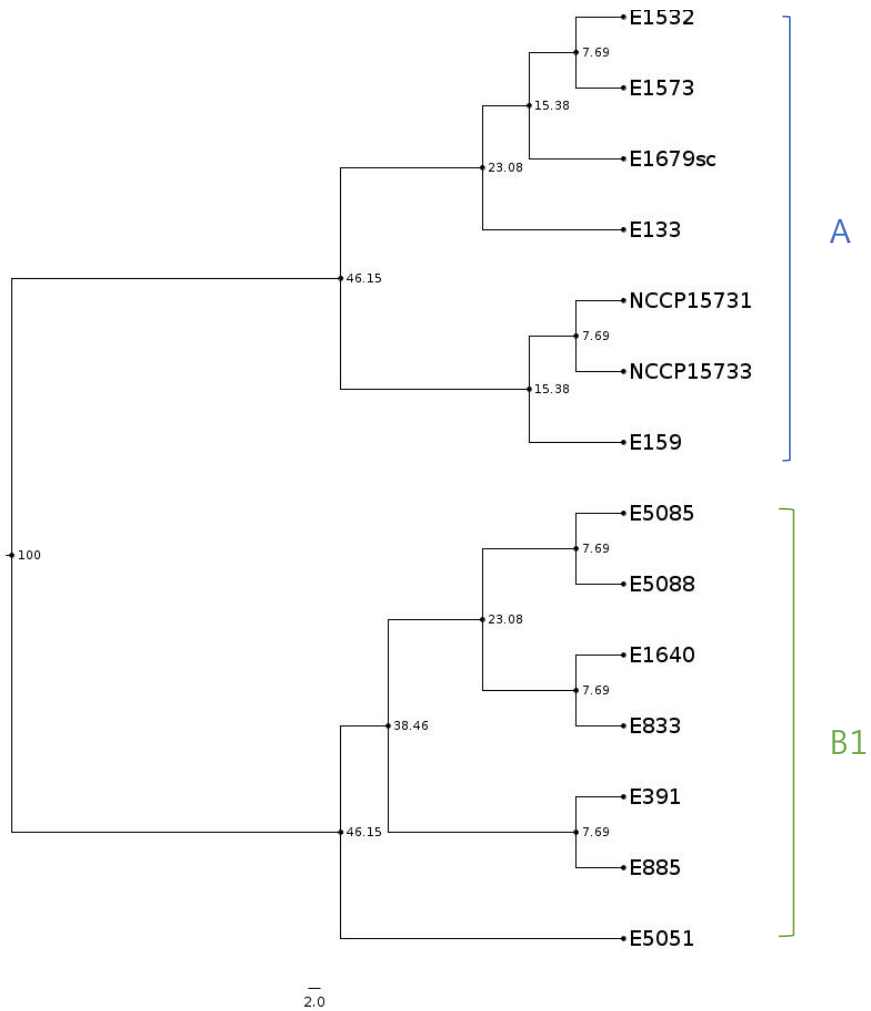

Fig. 3. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O159 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of whole-genomes, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site. The major *E. coli* phylo-groups are indicated.

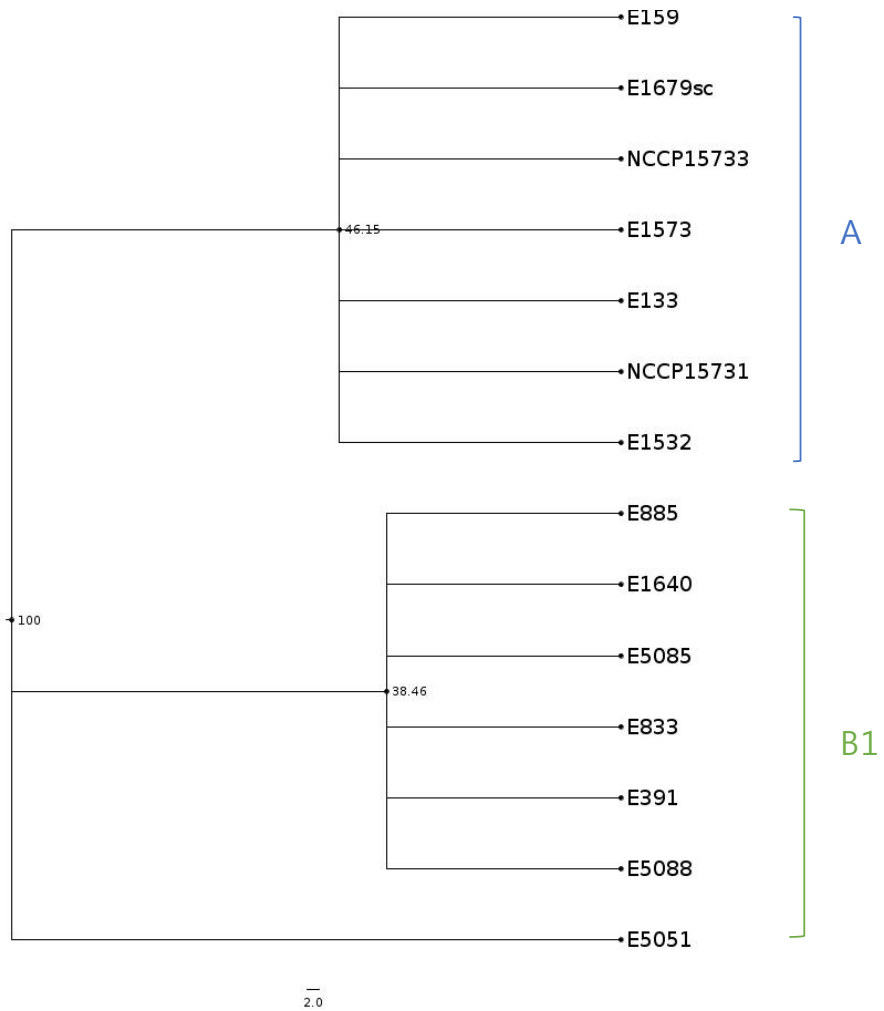

Fig. 4. Maximum likelihood (ML) tree of MLST sequences of ETEC O159 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of nucleotide sequences of seven MLST genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA)*, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site.

Virulence factor analysis

The acquisition of virulence factors has been suggested to be a major driving force for ETEC pathogenicity (Turner et al. 2006, Crossman et al. 2010). ETEC causes disease by colonizing the small bowel through attachment to the host epithelial lining by surface proteins called CFs and possibly other surface structures. Subsequently, adherent ETEC elaborates enterotoxins that cause typical clinical manifestations of ETEC-induced diarrhea. Thus, virulence factors can be used as important guide for understanding their pathogenicity.

Virulence factors of NCCP15731 and NCCP15733 were investigated and these factors were compared with those of 12 O159 strains. A total of 222 virulence factors grouped into 28 categories and 74 subcategories were identified (Appendix 2). Among these factors, 125 virulence factors were found in all of 14 ETEC strains. Each of NCCP15731 and NCCP15733 had 150 (67.6%) of these 222 virulence genes, respectively. They had the least number of virulence factors among 14 *E. coli* strains used in this study. *In silico* analysis revealed that major virulence factors belonged to the following six categories: adherence, auto transporter, iron uptake, non-LEE encoded TTSS effectors, toxin, and secretion system. Hierarchical clustering based on the presence or absence of 108 major virulence genes was constructed using R (version 3.4.3) (Fig. 5) (Team 2013). Phylo-group A and B1 strains completely clustered each other. These results suggest that virulence factors were related with phylo-group.

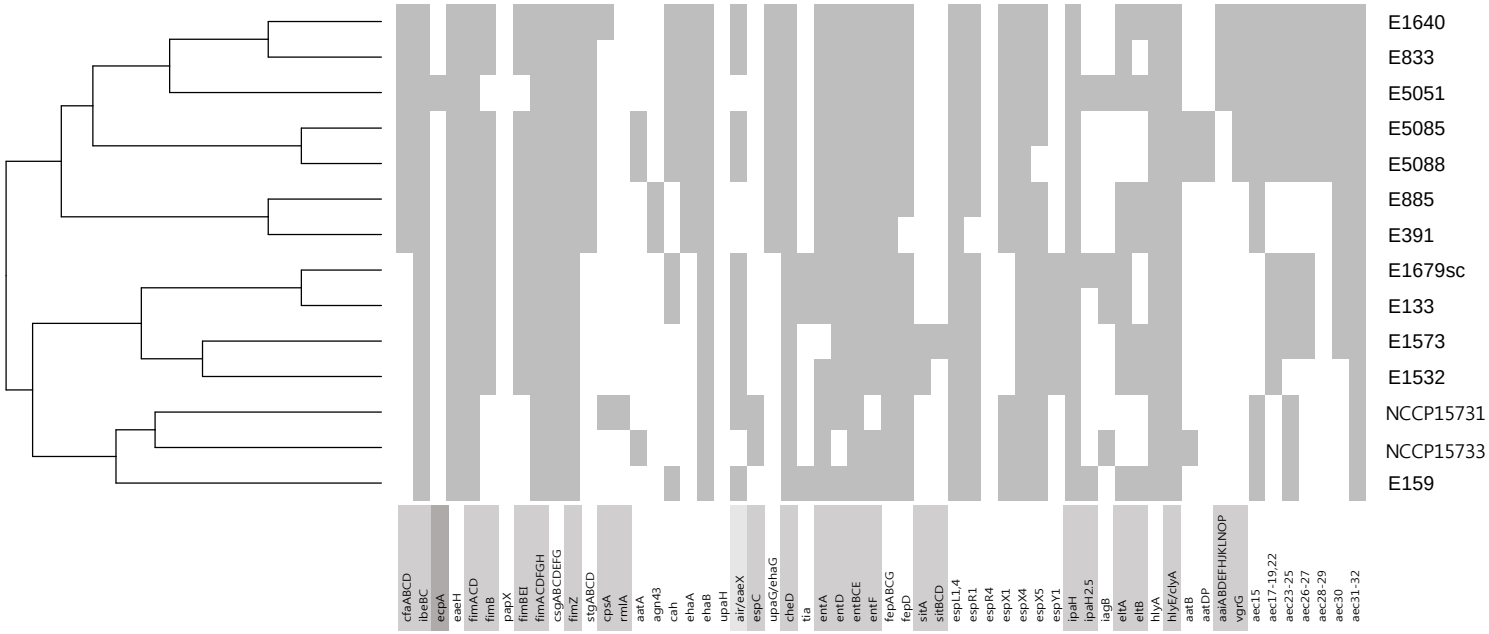

Fig. 5. Hierarchical clustering of 16 ETEC strains according to virulence factors. The dendrogram and associated heatmap were generated on the basis of gene presence/absence considering 108 genes involved in virulence factors. Grey/white color indicates gene presence / absence. Blue color indicates strains belonging to phylo-group A, and grey color indicates belonging to phylo-group B1.

In adherence and auto transporter categories, seven potentially functional systems : ECP, *EaeH*, Type I fimbriae, Curli fimbriae, Stg fimbriae, AIDA-I type and *UpaG* were found. These functional systems are known to be produced by pathogenic *E. coli* (McWilliams and Torres 2014) and suggest that ability of ETEC to attach to the host surface is the most important step in successful colonization (Antão, Wieler et al. 2009). Three virulence factors (*papX*, *upaH*, and *espC)* in these categories were only found in phylo-group A while *espC* was only identified in NCCP5731 and NCCP15733. Conversely, *ehaA* auto transporter gene and stg fimbriae systems were only found in phylo-group B1. It is known that stg fimbriae systems contribute to the attachment of human epithelial cells. They are associated with phylo-group B1 (Lymberopoulos et al. 2006).

Regarding toxins, enterobactin synthesis (*entABCE*) and ferric-enterobactin tansport (*fepABCG*) were present in all of the O159 strains. Alpha-hemolysin related gene (*hlyA)* plays a major pathogen role in ETEC and other pathogenic *E. coli* strains (Burgos and Beutin 2010). It was present in all O159 strains. Heat-labile enterotoxin (LT), i.e., *eltA* (10/14, 71.4%) and *eltB* (7/14, 50.0%) were found in O159 strains (Lasaro et al. 2008). However, none of these heat-labile genes were found in NCCP15731 or NCCP15733.

Comparison with other *E. coli* strains in phylo-group A

To calculate Average Nucleotide identity (ANI), five genomes of phylo-group A located in the same lineage as NCCP15731 and NCCP15733 were compared each other using ANIu calculator (Yoon et al. 2017). Results indicated that NCCP15731 was most similar to NCCP15733, with OrthoANIu value of 99.97% (Table. 6). Both NCCP15731 and NCCP15733 were most similar to *E. coli* strain E159 among reference strains in group A, with OrthoANIu values of 99.90% and 99.91%, respectively.

Table. 6. Average nucleotide identity values based on USEARCH. Numbers indicate OrthoANIu value (%) and Numbers in parentheses indicate the genome coverage (%)

	NCCP15731	NCCP15733	E159	E133	E1532
NCCP15731	/	99.97 (73.04)	99.90 (72.79)	99.15 (68.37)	99.23 (67.29)
NCCP15733	99.97 (71.21)	/	99.91 (73.34)	99.18 (70.64)	99.15 (68.57)
E159	99.90 (62.93)	99.91 (65.03)	/	98.99 (72.80)	99.00 (70.40)
E133	99.15 (59.53)	99.18 (63.08)	98.99 (73.32)	/	99.87 (73.95)
E1532	99.23 (59.38)	99.15 (62.06)	99.00 (71.87)	99.87 (74.95)	/

Venn diagram of the virulence factors of the five strains in phylo-group A was obtained using InteractiVenn (Heberle et al. 2015). They commonly had 137 genes encoding *E. coli* common pilus(ECP) proteins, flagella biosynthetic proteins, enterobactin transport proteins and proteins related Type II secretion system (Fig. 6). Comparison of NCCP15731, NCCP15733, and E159 (the most recent common ancestor of NCCP15731 and NCCP15733) showed that three strains commonly had 142 genes. Plus, both of NCCP15731 and NCCP15733 had two genes: *espC* gene related auto transporter and *orgA* gene encoding bsa T3SS secretion system. Compared to NCCP15733 and E159, NCCP15731 had three unique virulence genes: *aatA* encoding auto transporter protein, *aatB* encoding ABC transporter protein, and *iagB* encoding TTSS related secretion system. Compared to NCCP15731 and E159, NCCP15733 had two unique virulence genes: *cpsA* and *rmlA* encoding capsular polysaccharide. However, these two unique genes were not found in other *E. coli* O159 reference strains. In enterobactin synthesis of virulence systems, *entF* gene was found in NCCP15731, but not in NCCP15733 whereas *entD* gene was found in NCCP15733, but not in NCCP15731.

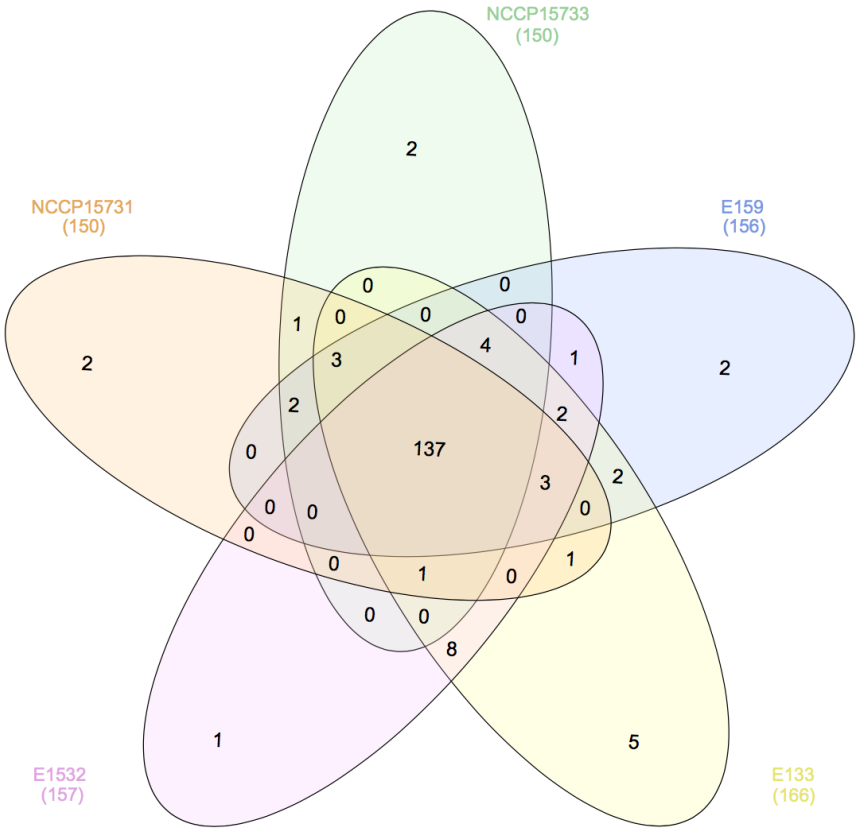

Fig. 6. Comparison of virulence factors of five ETEC strains in phylo-group A. NCCP15731 had three unique virulence genes, *aatA* encoding auto transporter protein, *aatB* encoding ABC transporter protein, and *iagB* encoding TTSS related secretion system. NCCP15733 had two unique virulence genes, *cpsA* and *rmlA* encoding Capsular polysaccharide.

Chapter 2. Comparative genomic analysis of enterotoxigenic *Escherichia coli* O6 strains

Genomic analysis

The NCCP15732 genome size was 4,801,812 bp with G+C content of 50.4%. Based on RAST analysis, 4,583 coding sequences and 18 tRNA genes were detected in NCCP15732, of which 3,907 (85%) were functional (Fig. 7). *In silico* analysis revealed that serotype of NCCP15732 was O6:H16 belonging to phylo-group A, which is one of major *E. coli* phylo-groups (Badouei et al. 2015). Genomic and phylogenetic characteristics of NCCP15732 were shown in Table. 7. Toxin types and colonization factors of NCCP15732 were investigated. NCCP15732 possessed both ST and LT enterotoxin.

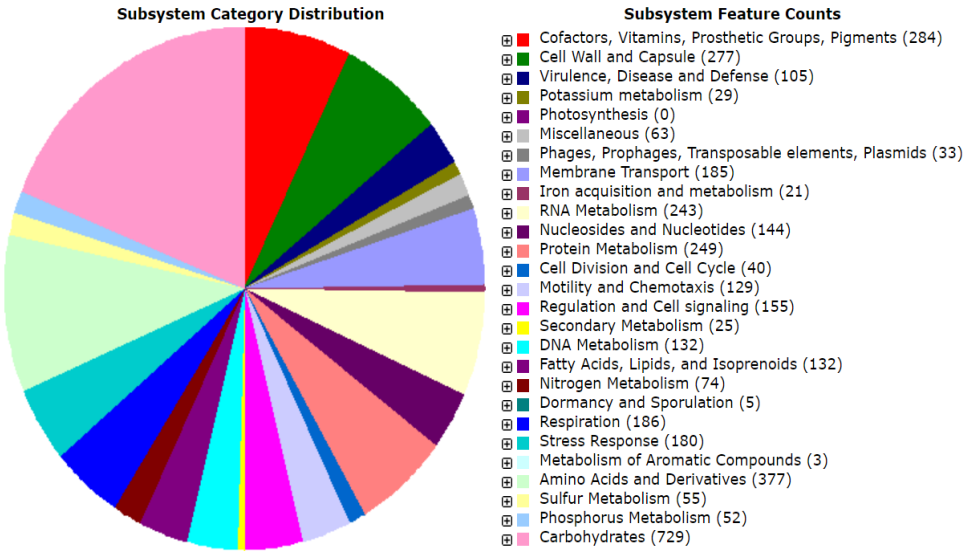

Fig. 7. Subsystem category distribution of NCCP15732. Based on RAST analysis, 4,873 coding sequences were detected in NCCP15732, of which 3,907 (85%) were functional.

Table. 7. Genomic characteristics of NCCP15732.

Strain	**NCCP15732**
Genome size	4,801,812
% GC	50.4
N50	910,848
L50	139
Total contigs	988
Total subsystems	605
Total CDS	4,873
tRNAs	18
Serotype	O6:H16
Toxin type	STh + LT
Phylo-group	A

Phylogenetic analysis

To compare the genome of NCCP15732, 16 whole-genome sequences of ETEC strains were selected as reference strains with same O6:H16 serotype as NCCP15732 over a similar time period. *E. coli* ATCC 25922 (Minogue et al. 2014) was used as serotype O6:H1 reference strain to investigate pathogenic characteristics of ETEC O6:H16 strains. *E. coli* NCCP15740 isolated in 2010 from diarrheal patients was selected as the same O6:H16 strain in Korea (Oh et al. 2014, Kwon et al. 2017). 12 *E. coli* O6 strains (O6:H16) (Pattabiraman and Bopp 2014, Pattabiraman and Bopp 2015) as reference strains had been isolated in the USA from 2011 to 2014. Two strains were used as reference strains: *E. coli* O6:H16: CFA/II str. B2C (traveler's diarrhea) (Madhavan et al. 2014) and *E. coli* O6:H16 str. 99-3165 (USA) (Trees et al. 2014). Genomic characteristics of 16 O6 strains were summarized in Table. 8.

Whole-genome phylogenetic analysis was constructed from multiple sequence alignments of the whole-genome sequence of 17 *E. coli* isolates (Fig. 8). Whole-genome phylogenetic tree showed that NCCP15732 belonged to a group of strains isolated in 2003-2004, and 2014 from USA. Another strain isolated from Korea, NCCP15740 belonged to a group of strains isolated in 2011 from USA. As results, two domestic strains clustered with USA strains whereas did not cluster each other into a single clade.

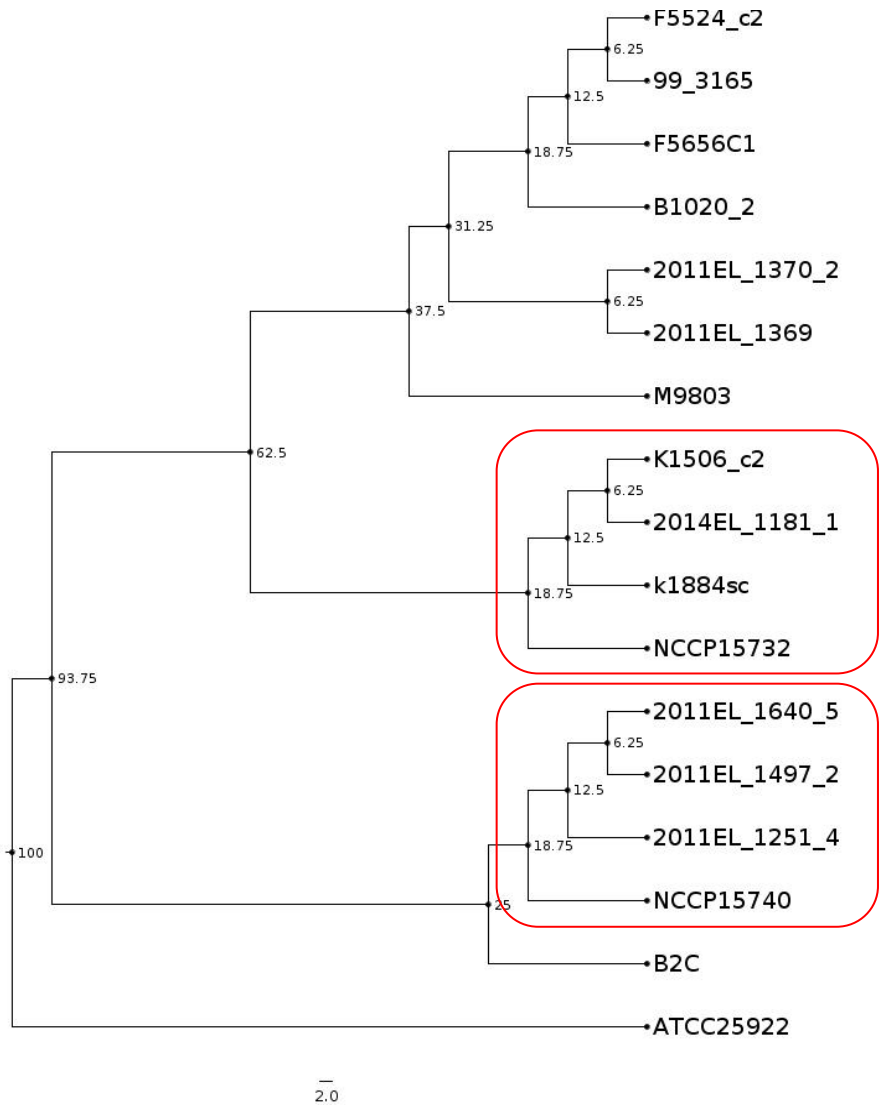

Fig. 8. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O6 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of whole-genomes, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site. ATCC25922 strain was out-group.

Table. 8. Genomic features of whole genome datasets of ETEC O6 strains.

Strain	Genome (Mb)	MLST type	Serotype	Toxin type	CF type
E. coli NCCP15732	4.80	ST4	O6:H16	STh+LT	CS1+CS3
E. coli NCCP15740	5.79	ST4	O6:H16	STh+LT	CS2+CS3
E. coli O6:H16: CFA/II str. B2C	5.02	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli O6:H16 str. 99-3165	4.75	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli O6:H16 str. F5656C1	4.81	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2011EL-1370-2	4.81	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2011EL-1369-1	4.80	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2011EL-1497-2	4.97	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain M9803	4.88	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2011EL-1640-5	4.87	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2011EL-1251-4	4.88	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain K1506-c2	4.94	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain 2014EL-1181-1	4.95	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain K1884 sc	4.99	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain B1020-2	4.77	ST4	O6:H16	STh+LT	CS2+CS3+CS21

E. coli strain F5524-c2	4.88	ST4	O6:H16	STh+LT	CS2+CS3+CS21
E. coli strain ATCC25922	5.20	ST73	O6:H1	None	None

Virulence factor analysis

To determine the casual mechanisms of pathogenicity, virulence factors in NCCP15732 and those of the reference strains were compared (Appendix 3). These strains except non-ETEC O6 strain ATCC25922 harbored total 163 virulence factors classified into 25 categories and 57 subcategories. NCCP15732 harbored 155 of these 163 virulence factors (95%). Most virulence factors were included in adherence, iron uptake, invasion and secretion system and these factors are the main contributors to ETEC diarrhea (Guerra et al. 2014) (Fig. 9).

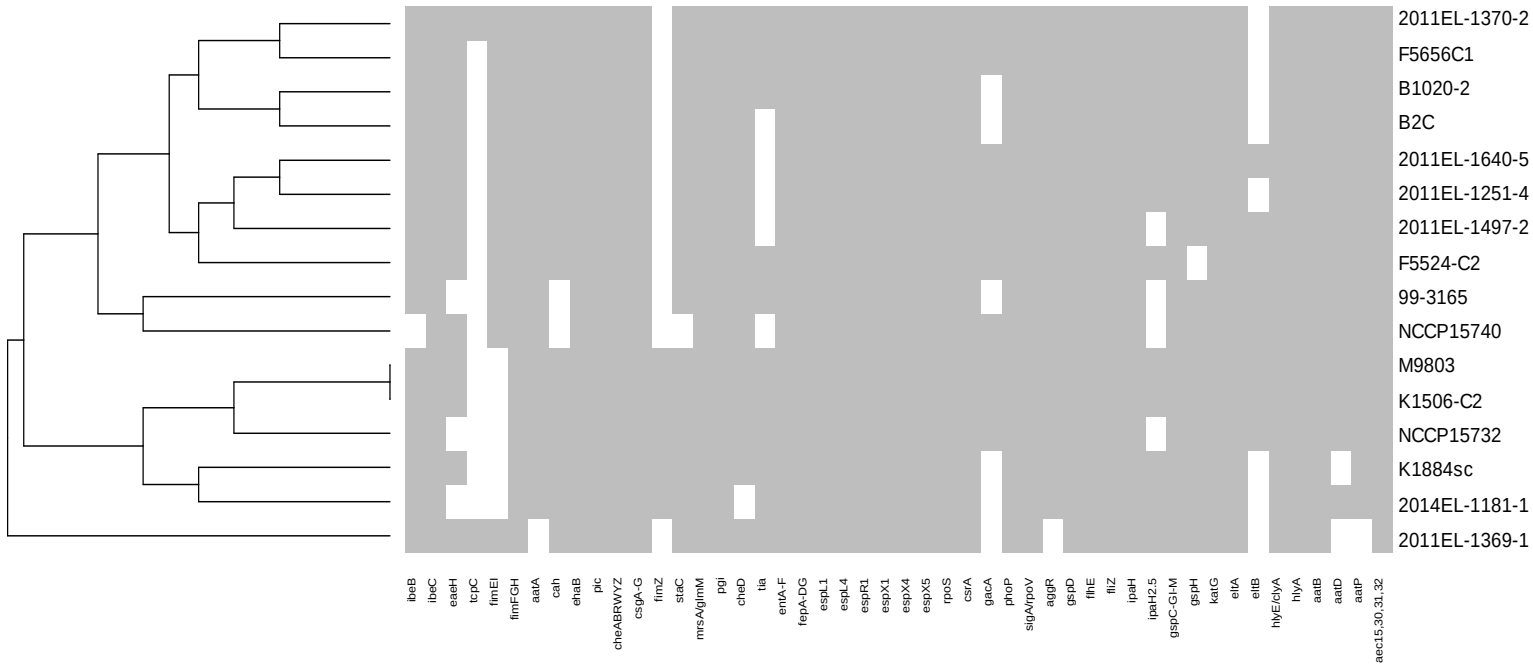

Fig. 9 Dendrogram and heatmap of 16 ETEC O6 strains according to virulence factors. The dendrogram and heatmap were generated on the basis of gene presence/absence considering major virulence factors, using R version 3.4.3. Grey color indicates gene presence, white gene absence. Pathogenicity of two domestic strains were different.

In adherence and auto transporter categories, eight potentially functional systems: CFA/I fimbriae, ECP (*E. coli* Common Pilus), *EaeH, etpA*, Type I fimbriae, Curli fimbriae, AIDA-I type and *fimZ* were found in virulence factors of ETEC O6 strains and these systems are known to be produced by pathogenic *E. coli* (Table. 9). Especially, *ibeB, etpA* and *fimZ*, are non-classical virulence factors known to contribute to diarrhea (Guerra et al. 2014), were found. NC CP15732 has *ibeB* and *fimZ* but not *etpA*.

Table. 9. Adherence system encoded by ETEC O6 strains genome.

Adherence system	Gene or Genes cluster
Curil fimbriae	*csgDEFG, csgBA*
ECP (*E. coli* Common Pilus)	*ecpRBCDE*
EtpA adhesin	*etpA*
EaeH adhesin	*eaeH*
CFA/I fimbriae	*ibeBC*
AIDA-I autotransporter	*aatA*, *cah* and *ehaB*
Fim	*fimZ*
Type I fimbriae	*fimEFGHI*

Among major virulence factors, two major enterotoxins, i.e., a heat-labile toxin (LT) and a heat-stable toxin (ST), induce watery diarrhea in ETEC. The LT toxin are encoded by the *eltA* and *eltB* gene. Another enterotoxin, ST toxins are classified into two types, STh encoded *estA* and STp encoded *st1* (Nataro and Kaper 1998). NCCP15732 carried both *eltA* and *eltB* enterotoxins. All of the ETEC O6 strains had *eltA* gene and STh toxin type but half of the ETEC O6 strains had *eltB* gene. Among all of 16 strains, CS2+CS3+CS21 was the most prevalent combination of colonization factors. CS21 was detected in 14 (88%) out of the 16 ETEC strains, but was not detected in two strains: both NCCP15732 and NCCP15740 isolated in Korea. In summary, 16 ETEC O6 strains carried both LT and ST enterotoxins.

In relation to toxins, enterobactin synthesis (*entABCE* and *entF*) and ferric-enterobactin tansport (*fepABCDG*) were present in all of the ETEC O6 strains. However, enterobactin is frequently produced by symbiotic *E. coli* isolated from healthy mammals (Searle et al. 2015). And these *ent* and *fep* genes were also found in virulence factors of non-ETEC ATCC25922 (data not shown). Thus, Pi et al. provide that intestinal bacteria play an important role in the colonization of healthy mouse gastrointestinal tract by pathogenic *E. coli* (Pi et al. 2012). Alpha-hemolysin related gene (*hlyA*) which plays a major pathogen role in pathogenic *E. coli* strains was present in all of 16 ETEC strains. And Invasion systems, *Tia* were found in NCCP15732 and presence of these gene was common in the ETEC O6 strains (11/16).

In comparison of the non-ETEC O6 strain ATCC 25922 and other strains, ATCC25922 strain had 202 virulence factors and 135 of these virulence factors of ATCC 25922 were found in all of ETEC O6 strains. Thus 28 virulence factors related to pathogenicity of ETEC O6 strains were identified (Table. 10).

Venn diagram of virulence factors of the two domestic strains, i.e., NCCP15732 and NCCP15740 in Korea and non-ETEC ATCC25922 were obtained using InteractiVenn (Heberle et al. 2015). A total of 125 genes encoding *E. coli* common pilus (ECP) proteins, flagella-biosynthetic proteins, enterobactin transport proteins and *ipaH* related TTSS effectors secretion system were found in three strains (Fig. 10). Compared with NCCP15732 and NCCP15740 isolated from diarrheal patients in Korea, *wza* related Capsular polysaccharide, *aatA* related auto transporter and most of genes related secretion systems were found in both strains. Comparing NCCP15740 and ATCC25922, NCCP15732 had two unique virulence genes: *fimZ* related adherence system and *tia* related invasion system. Interestingly, two genes are non-classical virulence factor known to contribute to diarrhea (Guerra et al. 2014).

Table. 10. Major virulence factors of ETEC O6 strains.

Category	Subcategory	Gene
Adherence	EtpA	*etpA*
Antiphagocytosis	Capsular polysaccharide	*wza*
Autotransporter	AatA, AIDA-I type	*aatA*
Fimbrial adherence determinants	Fim	*fimZ*
Invasion	Tia/Hek	*tia*
Non-LEE encoded TTSS effectors	EspL4	*espL4*
Non-LEE encoded TTSS effectors	EspR1	*espR1*
Non-LEE encoded TTSS effectors	EspX1	*espX1*
Non-LEE encoded TTSS effectors	EspX4	*espX4*
Non-LEE encoded TTSS effectors	EspX5	*espX5*
Others	VirK	*virK*
Regulator	AggR	*aggR*
Secretion system	EPS type II secretion system	*gspD*
Secretion system	T2SS (Type II secretion system)	*gspC*
Secretion system	T2SS (Type II secretion system)	*gspE*
Secretion system	T2SS (Type II secretion system)	*gspF*
Secretion system	T2SS (Type II secretion system)	*gspG*
Secretion system	T2SS (Type II secretion system)	*gspH*
Secretion system	T2SS (Type II secretion system)	*gspI*
Secretion system	T2SS (Type II secretion system)	*gspJ*
Toxins	Heat-labile enterotoxin	*eltA*
Toxins	Heat-labile enterotoxin	*eltB*
Type I secretion system	ABC transporter for dispersin	*aatB*
Type I secretion system	ABC transporter for dispersin	*aatD*
Type I secretion system	ABC transporter for dispersin	*aatP*
Type VI secretion system	ACE T6SS	*aec30*
Type VI secretion system	ACE T6SS	*aec31*
Type VI secretion system	ACE T6SS	*aec32*

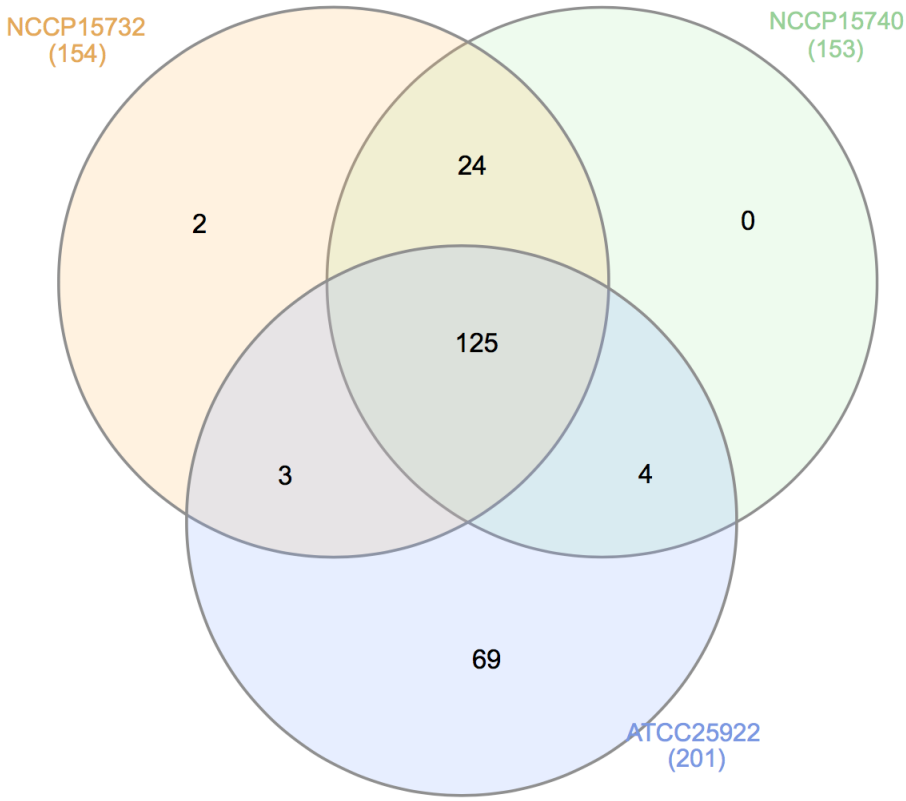

Fig. 10. Comparison of virulence factors of NCCP15732, NCCP15740 and non-ETEC strain ATCC25922. NCCP15732 had two unique non-classical virulence genes: *fimZ* related adherence system and *tia* related invasion system. Whereas NCCP15740 did not harbor unique gene.

Chapter 3. Comparative genomic analysis of enterotoxigenic *Escherichia coli* O169 strains

Genomic analysis

The NCCP15735 genome size was 5,001,083 bp with 50.3% G+C content. Based on RAST analysis, 4,761 coding sequences, 28 tRNA genes and 5 rRNA genes were detected in NCCP15735, of which 2,792 (59%) were functional (Fig. 11). *In silico* analysis revealed that the NCCP15735 serotype was O169:H41 and belonged to the phylo-group F (Clermont et al. 2013). Genomic and phylogenetic characteristics of NCCP15735 were shown in Table. 11.

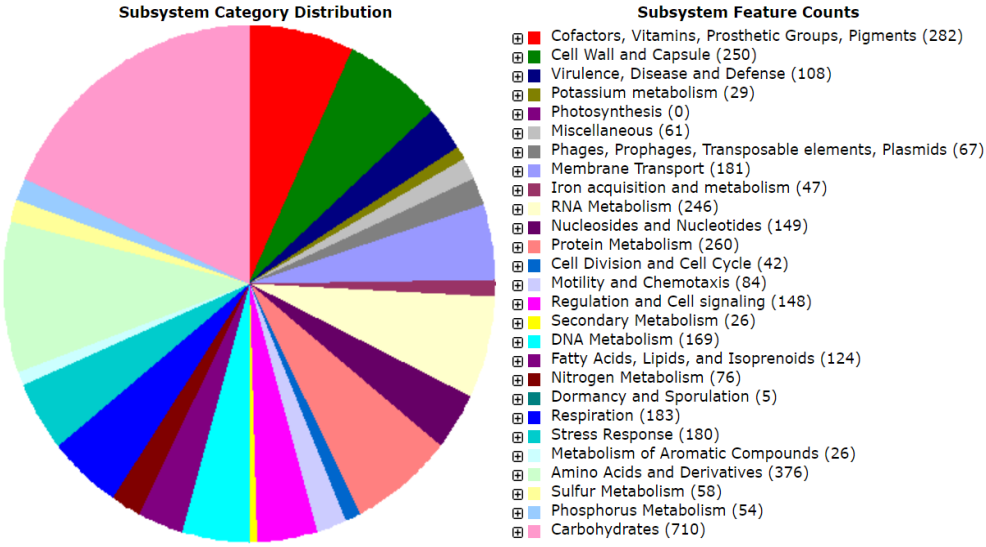

Fig. 11. Subsystem category distribution of NCCP15735. Based on RAST analysis, 4,761 coding sequences were detected in NCCP15735, of which 2,792 (59%) were functional.

Table. 11. Genomic characteristics of NCCP15735.

Strain	**NCCP15735**
Genome size	5,001,083
% GC	50.3
N50	60,074
Total contigs	319
Total subsystems	248
Total CDS	4,761
rRNAs	5
tRNAs	28
Serotype	O169:H41
Toxin type	STp
Phylo-group	F

Phylogenetic analysis

To infer evolutionary relationships among ETEC O169:H41 strains, 15 whole-genome sequences of ETEC strains reported as O169:H41 isolated over a similar time period were selected. The 11 *E. coli* O169 strains had been isolated globally from 1980 to 2011 (von Mentzer et al. 2014). 2014EL-1345-2 and F6326 were isolated from patients in the United States (Smith et al. 2018). Two additional strains, F9792 and F8111-1SC3 were used as reference isolated ETEC outbreaks in United states (Beatty et al. 2004, Devasia et al. 2006, Smith et al. 2018). Genomic characteristics of 16 O169 strains were summarized in Table. 12.

Whole-genome phylogenetic analysis and MLST-based phylogenetic analysis were performed of sequence of NCCP15735 and 15 E. coli reference strains (Fig. 11). E. coli O169 strains had simply phylogenetic relationships, according to both whole-genome and MLST data. The whole- genome phylogeny showed that strains isolated from same country did not cluster each other (Fig. 12). These results were inferred to be due to the similarity of 16 strains with same MLST type and serogroup. In the MLST-based phylogeny, all isolates except E520 strain with unknown CF types were clustered into a single clade (Fig. 13).

Concerning phylo-group, most pathogenic *E. coli* strains isolated from outbreaks belonged to pyhlogroup A and B1 (Escobar-Páramo et al. 2004, Badouei et al. 2015) and about 90% of ETEC outbreaks in Korea belonged to phylo-group A and B1(Koo et al. 2012). However, NCCP15735 isolated from ETEC outbreaks in Korea and four strains (2014EL-1345-2, F6326-C1, F8111-1SC3, and F9792) isolated from ETEC outbreaks in USA belonged to phylo-group F. These five strains belonging to phylo-group F did not cluster each other in whole-genome phylogeny whereas clustered into a single clade in MLST phylogeny.

Table. 12. Genomic features of whole genome datasets of ETEC O169 strains.

Strain	Phylo-group	MLST type	Serotype	Toxin type	CF type	Country
E. coli NCCP15735	F	ST182	O169:H41	STp	CS6	Korea
E. coli E344	D/E	ST182	O169:H41	STp	CS6	Mexico
E. coli E370	D/E	ST182	O169:H41	STp	CS6	Guatemala
E. coli E520	D/E	Unknown	O169:H41	STp	CS6	Mexico
E. coli E554	D/E	ST182	O169:H41	STp	CS6	Mexico
E. coli E1245	D/E	ST182	O169:H41	STp	CS6	Egypt
E. coli E1328	D	ST182	O169:H41	STp	CS6+CS21	Kenya
E. coli E1329	D	ST182	O169:H41	STp	CS6+CS21	Kenya
E. coli E1373	D/E	ST182	O169:H41	STp	CS6	Indonesia
E. coli E1484	D	ST182	O169:H41	STp	CS6	Guinea Bissau
E. coli E2256	D/E	ST182	O169:H41	STp	CS6	Thailand

E. coli E5084	D/E	ST182	O169:H41	STp	CS6	Bangladesh
E. coli 2014EL-1345-2	F	ST182	O169:H41	STp	CS6	United States
E. coli F6326-C1	F	ST182	O169:H41	STp	CS6	United States
E. coli F8111-1SC3	F	ST182	O169:H41	STp	CS6	United States
E. coli F9792	F	ST182	O169:H41	STp	CS6	United States

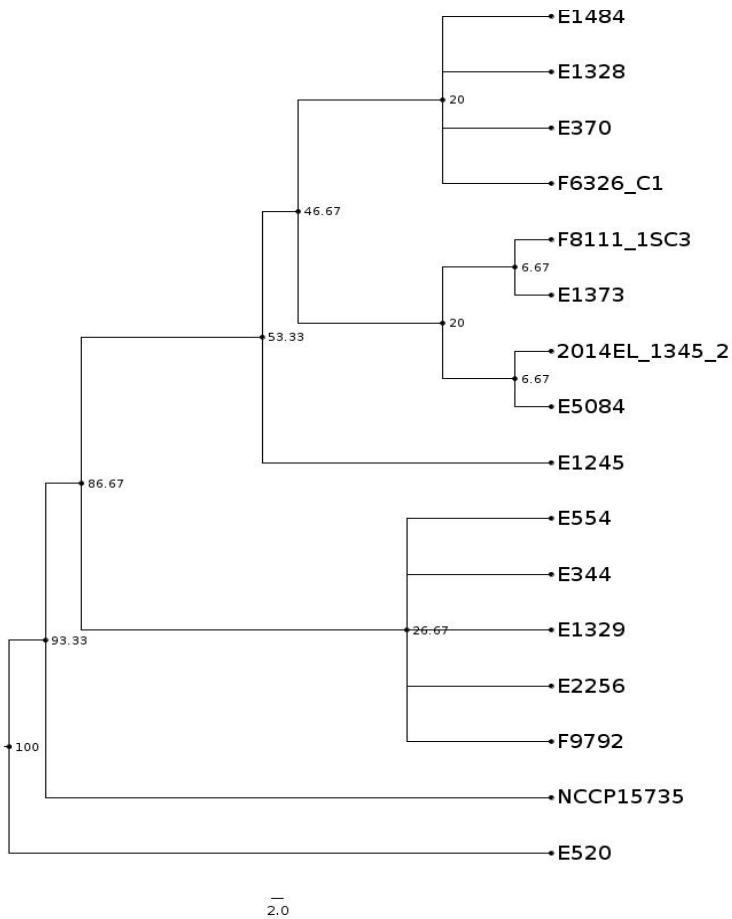

Fig. 12. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O169 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of whole-genomes, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site.

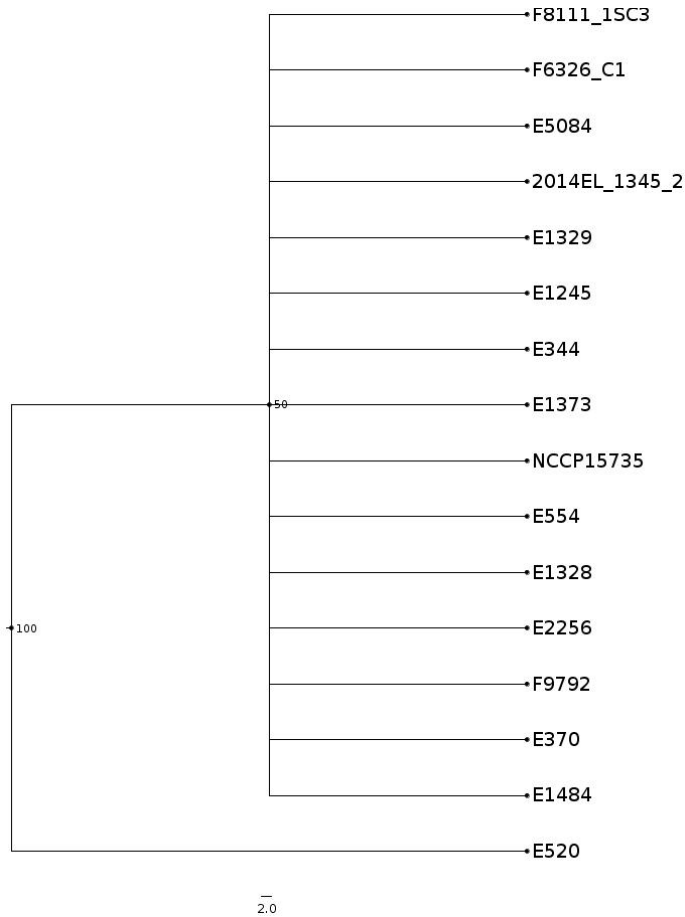

Fig. 13. Maximum likelihood (ML) tree of MLST sequences of ETEC O169 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of nucleotide sequences of seven MLST genes (*adk, fumC, gyrB, icd, mdh, purA,* and *recA),* using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site.

Virulence factor analysis

To determine causal mechanisms of pathogenicity, virulence factors of O169 strains were compared. 195 virulence factors grouped into 27 categories and 71 subcategories were detected. Among these virulence factors, 171 virulence genes were found in all of O169 strains. NCCP15735 harbored 177 (91%) of these 195 virulence factors. Most virulence factors were included in adherence, auto transporter, iron uptake and secretion system and these factors are the main contributors to ETEC diarrhea (Guerra et al. 2014) (Fig. 14). When clustering of the virulence factors, epidemiological features of domestic ETEC was close to the that of ETEC isolated in Mexico.

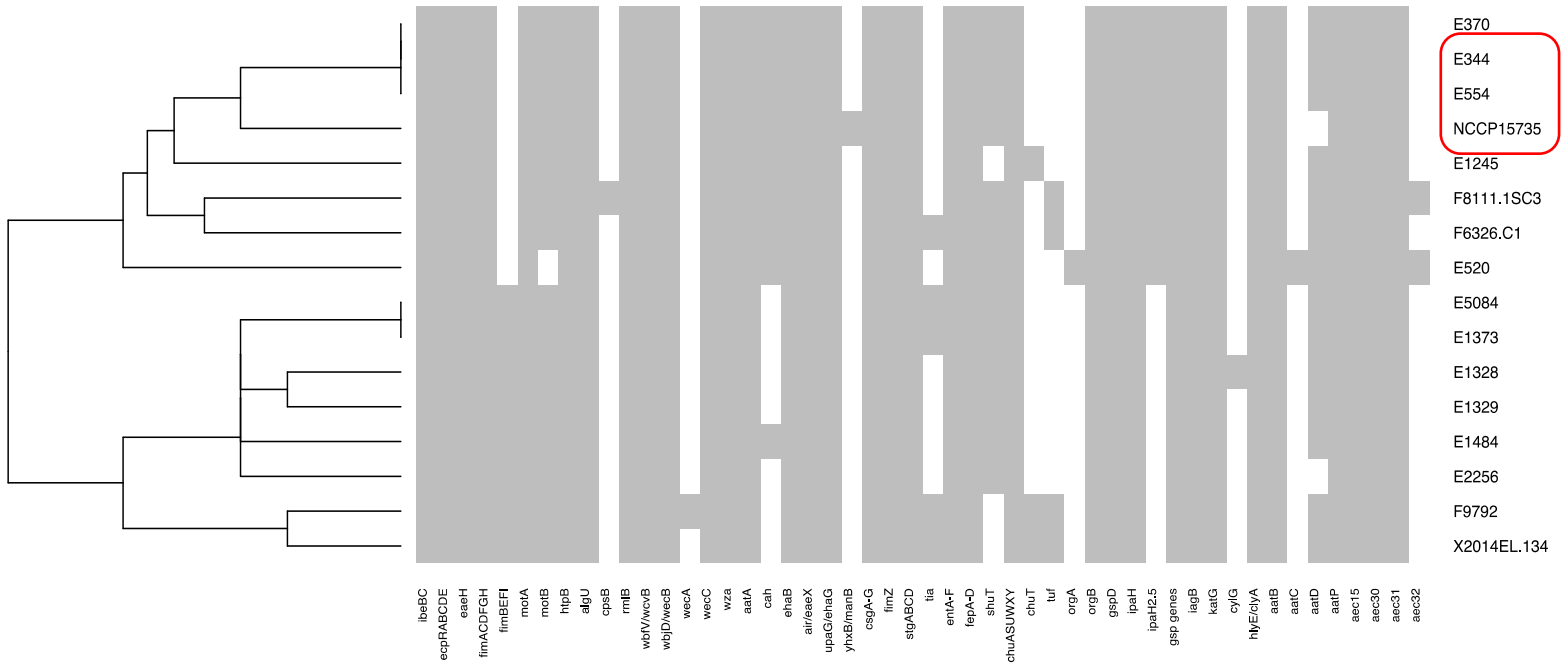

Fig. 14. Dendrogram and heatmap of 16 ETEC O169 strains according to virulence factors. Dendrogram and heatmap were generated on the basis of gene presence/absence considering major virulence factors, using R version 3.4.3. Grey color indicates gene presence, white gene absence. The Korea ETEC NCCP15735 was highlighted.

In adherence and auto transporter categories, some of 10 potentially functional systems: CFA/I fimbriae, ECP (*E. coli* Common Pilus), *EaeH*, *UpaG*, Flagella, Type I fimbriae, Curli fimbriae, Stg fimbriae, AIDA-I type and *fimZ* were found in virulence factors of ETEC O169 strains (Table. 13). These systems are known to be produced by pathogenic *E. coli* and their ability of ETEC to attach to the host surface is the most important step in successful colonization (Antão et al. 2009). *EhaB* and *UpaG* are known to bind to lamin / curil and involved in biofilm formation (Leatham-Jensen et al. 2012). Laminin is an extracellular matrix that is commonly found in mammalian intestines and acts as a linking molecule in connective tissue that promotes bacterial attachment and propagation into host tissues (Simon-Assmann et al. 2010). Moreover, Stg systems contribute to the attachment of human epithelial cells (Lymberopoulos et al. 2006).

Table. 13. Adherence system encoded by ETEC O169 strains.

Adherence system	Gene or Genes cluster
Curil fimbriae	*csgDEFG, csgCBA*
ECP (*E. coli* Common Pilus)	*ecpRABCDE*
Flagella adhesin	*motAB*
UpaG adhesin	*upaG/ehaG*
EaeH adhesin	*eaeH*
CFA/I fimbriae	*ibeBC*
AIDA-I autotransporter	*aatA, cah* and *ehaB*
Fim	*fimZ*
Stg fimbriae	*stgABCD*
Type I fimbriae	*fimABCDEFGHI*

In relation to toxins, enterobactin synthesis (*entABCDEF*) and ferric-enterobactin tansport (*fepABCDG*) were present in all of the O169 strains. Alpha-hemolysin related gene (*hlyE/clyA*) plays a major pathogen role in ETEC and these gene was present in all of the O169 strains (Burgos and Beutin 2010). Interestingly, *yhxB/manB* related LOS system of endotoxin was found only NCCP15735.

Heat-labile enterotoxin(LT), i.e., *eltA* and *eltB* gene were not found in all of O169 strains. Whereas one of genes related heat-stable enterotoxin(ST), STp was detected in all of O169 strains. Moreover, most of O169:H41 ETEC isolates co-expressed colonization factor called CS6 and STp toxin type (14 of 16 strains, 88%). These combination is known to be detected from Individual ETEC isolates, globally (Gaastra and Svennerholm 1996, Svennerholm and Lundgren 2012, von Mentzer et al. 2014). Only two strains (E1328 and E1329) co-expressed CS6+CS21 colonization factors and STp toxin type. In summary, 16 ETEC O169 strains carried STp enterotoxin and CS6 colonization factor.

In secretion categories, seven potentially functional systems were found in ETEC O169 strains (Table. 14). Most of gene were detected in all of ETEC O169 strains. But, *orgA* related Bsa T3SS system was only found in E554 strain. *IpaH2.5* gene related TTSS effector was found in half of all ETEC O169 strains, including NCCP15735. NCCP15735 harbored *aatB* and *aatP* gene related ABC transporter of Type I secretion system but not *aatC* and *aatD* gene.

Table. 14. Secretion system encoded by ETEC O169 strains genome.

Secretion system	Gene or Genes cluster
Bsa T3SS	*orgAB*
EPS type II secretion	*gspD*
Mxi-Spa TTSS effectors	*ipaH and ipaH2.5*
T2SS Type II secretion system	*gspCEFGHIJKLM*
TTSS (SPI-1 encode)	*iagB*
Type I secretion system	*aatBCDP*
Type VI secretion system	*aec15, 30, 31, 32*

Venn diagram of virulence factors of five strains was obtained using InteractiVenn (Fig. 15) (Heberle et al. 2015). NCCP15735 and E520, most close to NCCP15735 strain, were selected. Two strains (E1328 and E1329) were chosen because, unlike the other reference strains, they had two colonization factors, called CS6 and CS21. Also, F9792 isolated from ETEC outbreaks in USA was selected. All of five strains commonly had 171 genes encoding *E. coli* common pilus (ECP) proteins, flagella biosynthetic proteins, endotoxin proteins, enterobactin transport proteins, Non-LEE encoded TTSS effectors proteins and proteins related Type II / Type VI secretion system. Compared four reference strains, NCCP15735 had a unique gene: *yhxB/manB* related endotoxin. Moreover, *yhxB/manB* gene was not found in any *E. coli* O169 strains. Both E520 (most close to NCCP15735) and NCCP15735 had unique genes: *cah* related auto transporter and *ipaH2*.5 related secretion system. However, two strains isolated from patients of ETEC outbreaks in Korea and USA did not have specific virulence factors whereas, they had same colony factor and MLST sequence type.

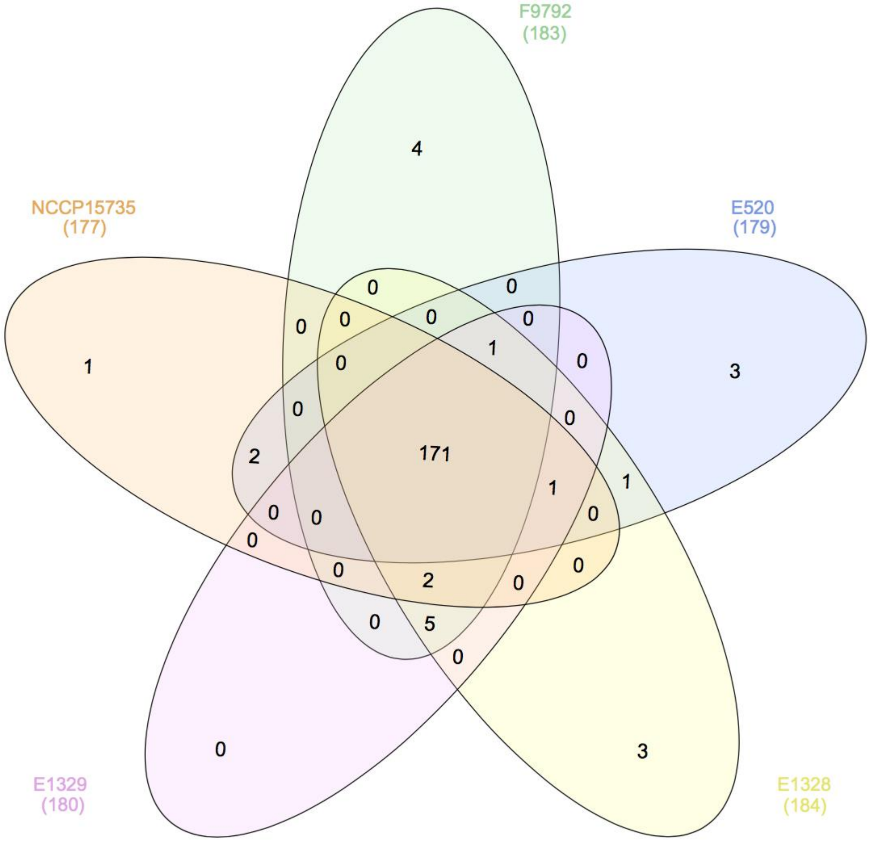

Fig. 15. Comparison of virulence factors of five ETEC O169 strains. NCCP15735 had a unique virulence gene, *yhxB/manB* related endotoxin. Both E520 (most close to NCCP 15735) and NCCP15735 had two genes: *cah* related auto transporter and *ipaH2*.5 related secretion system. Two strains isolated from patients of ETEC outbreaks in Korea and USA did not have specific virulence factors.

Chapter 4. Comparative genomic analysis of enterotoxigenic *Escherichia coli* O25 strains

Genomic analysis

Genome size of NCCP15741 was 4,718,388bp with 50.5% G+C content in the length. Based on RAST analysis, 4,512 coding sequences and 16 tRNA genes were detected in NCCP15741, of which 3,988 (88%) were functional (Fig. 16). *In silico* analysis revealed that the NCCP15741 serotype was O25:H16 and belonged to the phylo-group A, which is one of major *E. coli* phylo-groups (Badouei et al. 2015). Genomic and phylogenetic characteristics of NCCP15732 were shown in Table. 15. Toxin types and colonization factors of NCCP15741 were investigated. NCCP15741 possessed LT enterotoxin and CS6 colonization factor.

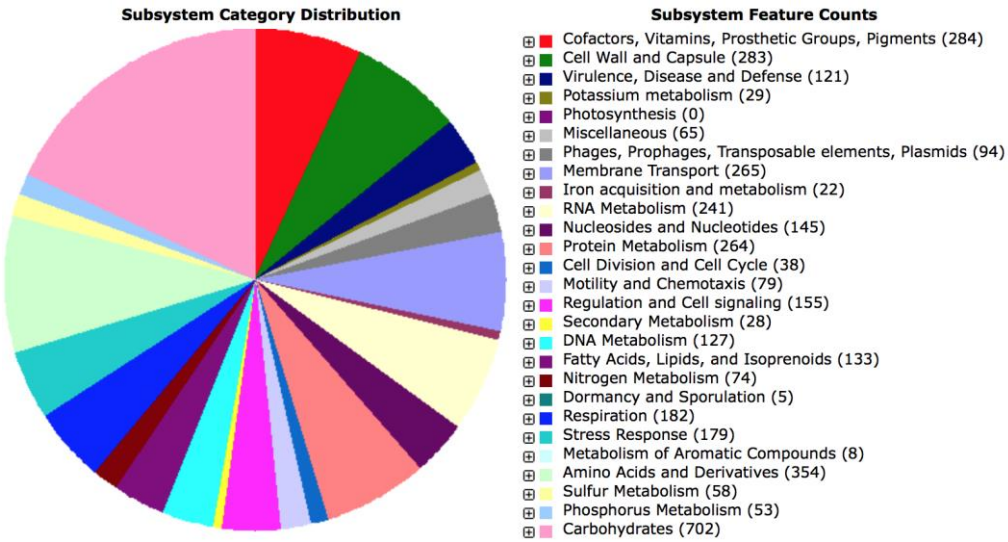

Fig. 16. Subsystem category distribution of NCCP15741. Based on RAST analysis, 4,512 coding sequences were detected in NCCP15741, of which 3,988 (88%) were functional.

Table. 15. Genomic characteristics of NCCP15741.

Strain	**NCCP15741**
Genome size	4,718,388
% GC	50.5
N50	9,158
L50	154
Total contigs	943
Total subsystems	589
Total CDS	4,512
tRNAs	16
Serotype	O25:H16
Toxin type	LT
CF type	CS6
Phylo-group	A

Phylogenetic analysis

To study evolutionary relationships among *E. coli* O25 strains, 10 whole-genome sequences of ETEC strains reported as serotype O25 isolates over a similar time period were selected. Seven *E. coli* O25 strains had been isolated globally from 1980 to 2011 (von Mentzer et al. 2014). Two strains, 2014EL-1343-2 and F5505-C1 were used as reference isolated from patients in the United States and cruise ship outbreaks (Smith, Lindsey et al. 2018). Moreover, ATCC43886 (or called E2539C1) was selected because it was isolated from a shipboard outbreak of human diarrheal illness (Wachsmuth et al. 1979, Smith et al. 2018). Genomic features of 11 O25 strains were summarized in Table. 16.

Whole-genome phylogenetic analysis and MLST-based phylogenetic analysis were performed of sequence of 11 ETEC O25 strains (Fig. 17-18). The *E. coli* O25 strains had simply phylogenetic relationships, represented by three clades, according to both the whole-genome and MLST data. The whole genome phylogeny showed that NCCP15741 belonged to a group of strains isolated from ETEC outbreaks in USA and was distinct from two strains (E1122 and E1119) with same serotype and toxin type. In contrast to the whole-genome phylogeny, three strains isolated from ETEC outbreaks in USA clustered in same group and NCCP15741 clustered two strains with MLST ST3931.

Table. 16. Genomic features of whole genome datasets of ETEC O25 strains.

Strain	Genome (Mb)	Phylo-group	MLST type	Serotype	Toxin type	CF type[1]
E. coli NCCP15741	4.7	A	ST1491	O25:H16	LT	CS6
E. coli ATCC43886	5.0	A	ST1312	O25:H16	LT	CF NEG
E. coli 2014EL-1343-2	4.9	A	ST1312	O25:H16	STh	CF NEG
E. coli F5505-C1	5.0	A	ST1312	O25:H16	LT	CF NEG
E. coli E97	5.2	A	ST96	O25:H42	LT+STp	CS4+CS6
E. coli E106	5.1	A	ST96	O25:H42	LT+STp	CS4+CS6
E. coli E1114	6.6	A	ST3931	O25:H16	LT	CS6+CS8
E. coli E1118	7.0	A	ST3931	O25:H39	LT	CS6+CS8
E. coli E1119	5.5	A	ST5238	O25:H16	LT	CS6+CS21
E. coli E1122	5.3	A	ST5238	O25:H16	LT	CS6
E. coli E1516	5.2	A	ST1312	O25:H16	LT	CS6

[1] CF NEG: CF “negative”

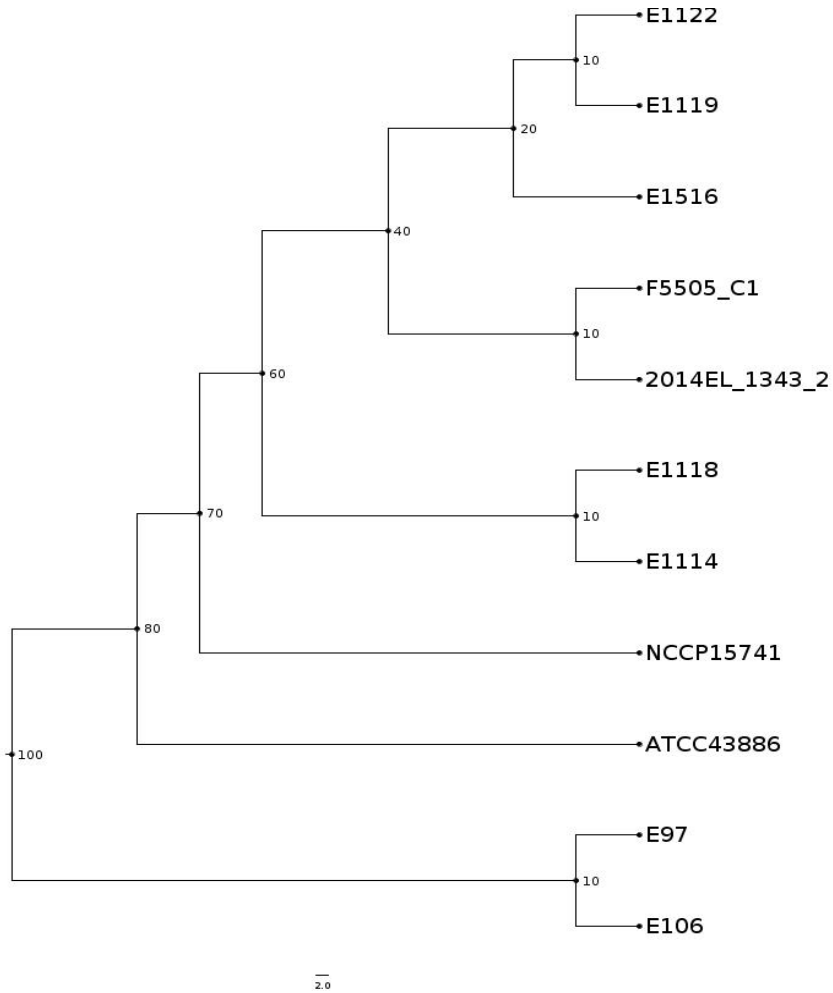

Fig. 17. Maximum likelihood (ML) tree of whole-genome sequences of ETEC O25 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of whole-genomes, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site.

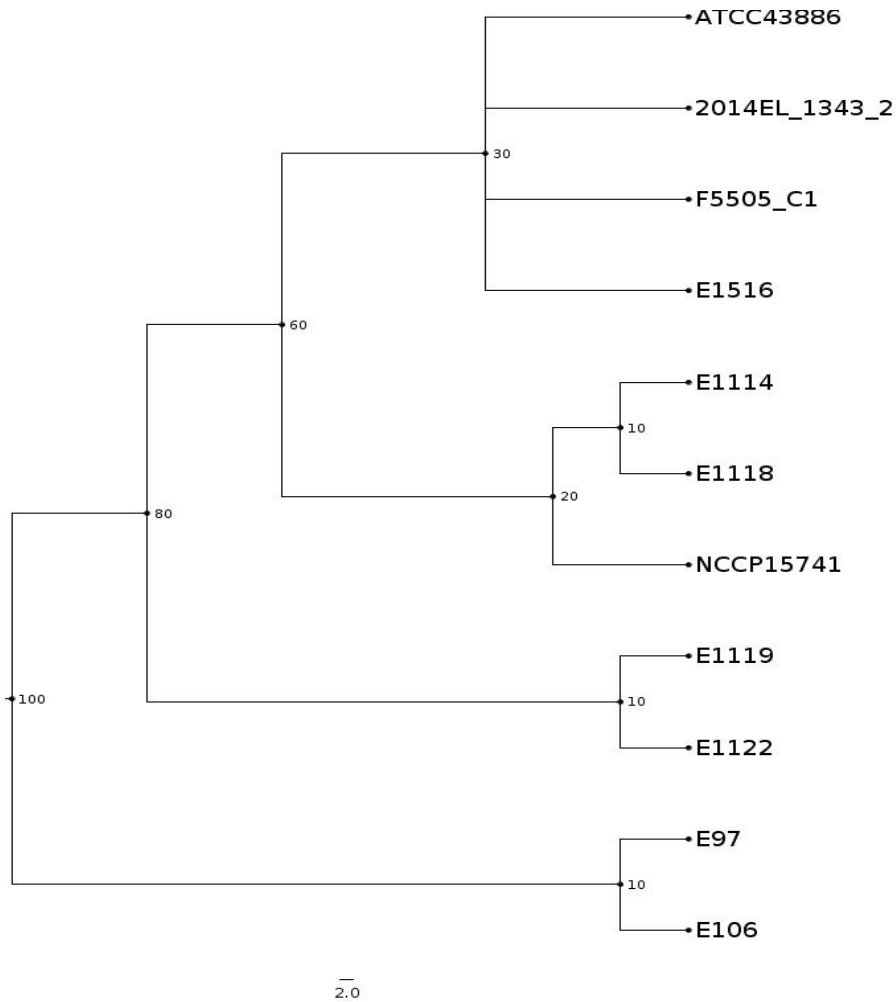

Fig. 18. Maximum likelihood (ML) tree of MLST sequences of ETEC O25 strains. The tree was obtained by approximately-maximum-likelihood analysis of a concatenated alignment of nucleotide sequences of seven MLST genes (*adk, fumC, gyrB, icd, mdh, purA*, and *recA)*, using FastTree (version 2.1.7) with a GTR (generalized time-reversible) +CAT model. Bootstrap values based on 1,000 replications are included. Evolutionary time is adjusted to 100. The scale bar indicates 2.0 substitutions per site.

Virulence factor analysis

To determine the causal mechanisms of pathogenicity, virulence factors in O25 strains were compared. 234 virulence factors were grouped into 30 categories and 78 subcategories. Hierarchical clustering based on the presence or absence of all virulence genes was constructed using R (version 3.4.3) (Fig. 19). These results showed that pathogenicity of NCCP15741 isolated from ETEC outbreaks in Korea was different to those of three strains isolated from ETEC outbreaks in USA.

Among 234 virulence factors, 138 virulence genes were found in all of O25 strains. NCCP15741 harbored 154 of the 234 virulence factors (59%). Most virulence factors detected in all O25 strains were included in adherence, anti-phagocytosis, endotoxin. However, some of virulence factors belonging to auto transporter, secretion and iron uptake systems showed different patterns in each O25 strains and these factors are the main contributors to ETEC diarrhea (Guerra et al. 2014).

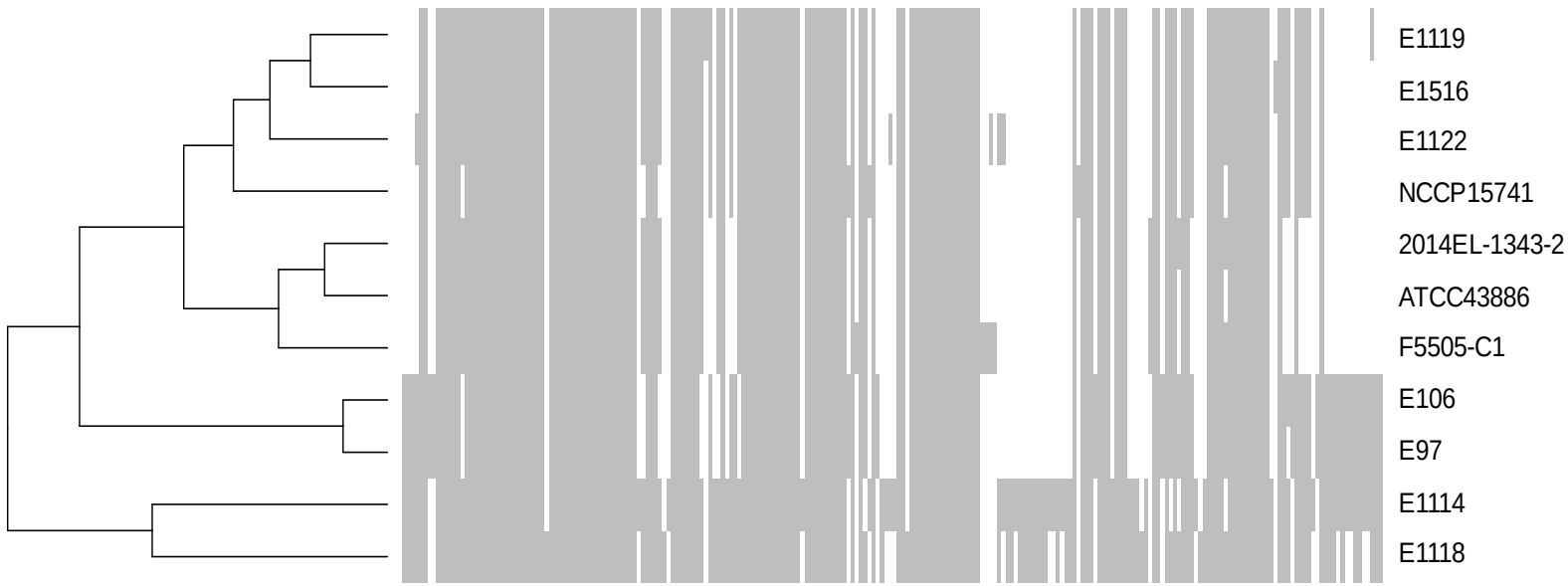

Fig. 19. Hierarchical clustering of 11 ETEC strains according to virulence factors. Dendrogram and associated heatmap were generated on the basis of gene presence/absence considering 234 genes involved in virulence factors. Grey/white color indicates gene presence / absence.

In adherence systems, some of 10 potentially functional systems: CFA/I fimbriae, ECP (*E. coli* Common Pilus), *EaeH, UpaG*, Flagella, Type I fimbriae, Curli fimbriae, Stg fimbriae, P fimbriae and *fimZ* were found in virulence factors of ETEC O25 strains (Table. 17). But, NCCP15741 strain had not harbor *cfaABCD* genes related CFA/I fimbriae and *UpaG* related adhesion system. Interestingly, E1114 with same serotype of NCCP15731 and CS6+CS8 colonization factors, had only unique virulence genes: *papX* and Stg fimbriae systems. These Stg fimbriae systems contribute to the attachment of human epithelial cells (Lymberopoulos et al. 2006).

Table. 17. Adherence system encoded by ETEC O25 strains.

Adherence system	Gene or Genes cluster
Curil fimbriae	*csgDEFG, csgCBA*
ECP (*E. coli* Common Pilus)	*ecpCDER*
Flagella adhesin	*motAB, cheABRWYZ*
UpaG adhesin	*upaG/ehaG*
EaeH adhesin	*eaeH*
CFA/I fimbriae	*ibeBC, cfaABCD*
P fimbriae	*papX*
Fim	*fimZ*
Stg fimbriae	*stgABCD*
Type I fimbriae	*fimACDFGHI*

Table. 18. Anti-phagocytosis system encoded by ETEC O25 strains genome.

Anti-phagocytosis system	Gene or Genes cluster
Alginate regulation	*algU*
Capsular polysaccharide	*rmlABC*
	wbfV/wcvB
	wbjD/wecB
	wecC
	wza

In relation to toxins, enterobactin synthesis (*entABCDEF*) and ferric-enterobactin tansport (*fepABCDG*) were present in all of the O25 strains. Alpha-hemolysin related gene (*hlyA* and *hlyE/clyA*) plays a major pathogen role in ETEC and other pathogenic *E. coli* strains was present in all of the O25 strains (Burgos and Beutin 2010). NCCP15741 had both *eltA* and *eltB* genes encoding LT enterotoxin types. The presence of both *eltA* and *eltB* genes was common in O25 strains (8 out of 11, 73%). Among all of 11 strains, CS6 colonization factor was most prevalent colonization factor type and five strains had one more colonization factor (CS4, CS8 and CS21). Any colonization factors of three strains obtained from ETEC outbreaks in USA were not detected. Interestingly, *E. coli* 2014EL-1343-2 strain that only expressed STh had less virulence factors.

Concerning iron uptake systems, *hemCELM* genes related Heme-biosynthesis were present in all of the O25 strains, including NCCP15741, whereas Ferrous iron transport, Hemin uptake and Yersiniabactin systems-related genes were only present in *E. coli* E1114 and E1118 strains.

In secretion categories, five potentially functional systems were found in ETEC O25 strains (Table. 19). Most of genes were distributed in different aspects in each O25 strains. The ETEC O25 strains harbored EPS type II secretion system, Type II secretion system, *aec15* gene related ACE T6SS system. *IpaH* gene related TTSS effector was found in all ETEC O25 strains, whereas, *ipaH2.5* gene was found in 8 out of 11 strains (NCCP15741 had not harbor these gene). NCCP15741 had *aatBDP* genes related ABC transporter of Type I secretion system (8 out of 11 strains). Although *aec15* gene was found in all O25 strains, additional *aec* genes were found in four strains (E106, E1114, E1118 and E1119), but not in NCCP15741.

Table. 19. Secretion system encoded by ETEC O25 strains.

Secretion system	Gene or Genes cluster
EPS type II secretion	*gspD*
Mxi-Spa TTSS effectors	*ipaH and ipaH2.5*
T2SS Type II secretion system	*gspCEFGHIJKLM*
Type I secretion system	*aatBDP*
Type VI secretion system	*aec15*

DISCUSSION

Enterotoxigenic *Escherichia coli* (ETEC) is a common cause of bacterial infection that leads to diarrhea and is responsible for diarrhea-related deaths per year in young children (Organization 2006). The enterotoxin type, i.e., LT and ST enterotoxin and relative proportions of ETEC colonization factors differ depending on the geo graphical source (Gaastra and Svennerholm 1996, Svennerholm and Lundgren 2012, von Mentzer et al. 2014). In recent studies, results of whole genome sequencing of a representative collection of ETEC isolated between 1980 and 2011 showed that they shared distinct colonization factors and enterotoxin profiles and were divided into phylo-groups (von Mentzer et al. 2014). This suggests that researching long-term ETEC clades can find a combination that optimizes fitness and transmissibility, and affect the tracking and prevention of ETEC disease.

In previous study in Korea, 258 isolates from patients with diarrhea in Korea were analyzed for CFs and subjected to multi-locus sequence typing (MLST). ST171 (24%) was identified as the most prevalent ETEC type in Korea, followed by ST955, ST964, and ST656 (Oh et al. 2014). However, there have been limited study conducted on the pathogenicity of ETEC outbreaks in Korea and comparative analysis with ETEC distributed globally. Therefore, it is required to study pathogen by revealing specific characters of ETEC in Korea because these results will provide important information for the development of an ETEC vaccine candidate and play a key role in increasing public health. In this study, whole-

genome sequencing was performed to reveal the genomic characteristics of five predominant MLST types of ETEC isolates in Korea (Table. 1).

In chapter 1, NCCP15731 and NCCP15733 appeared to belong to the O159:H34 serotype and ST218 MLST-type. Unlike other O159 strains, any colonization factor gene of NCCP15731 and NCCP15733 did not detected. And it suggests that additional colonization factors might exist and research of colonization factors of ETEC from Korea is needed (Isidean et al. 2011, Svennerholm and Lundgren 2012). Two domestic strains had unique genes related capsular polysaccharide, auto transporter, and secretion system. Moreover, both of strains did not contain LT genes. Based on the presence or absence of a total of 108 major virulence factors of NCCP15731, NCCP15733 and other reference strains, results of hierarchical clustering showed that phylo-group A and B1 strains clustered each phylo-group. This result supports hypothesis that phylo-groups are also linked to genetic characteristics and virulence factors of ETEC strains (Escobar-Páramo, Clermont et al. 2004, Steinsland, Valentiner-Branth et al. 2004, Turner, Chaudhuri et al. 2006, Rivera, Ochoa et al. 2010, Steinsland, Lacher et al. 2010, Sahl and Rasko 2012).

In chapter 2, NCCP15732 appeared to belong to the O6:H16 serotype and ST4 MLST-type. Compared with other O6 strains, NCCP15732 had carried both *eltA* and *eltB* enterotoxins. All of the ETEC O6 strains had *eltA* gene and STh toxin type but half of the ETEC O6 strains had *eltB*. Among 16 ETEC O6:H16 strains, a combination of CS2+CS3+CS21 colonization factors with STh+LT enterotoxin type was the most prevalent type of those ETEC. In Korea isolates from ETEC

outbreaks, NCCP15732 had CS1+CS3 with STh+LT combination whereas, NCCP15740 had CS2+CS3 with STh+LT combination. In recent study, three combination type were found in O6 serotype strains, globally and demonstrated that they were not only phylogenetically related but also shared consistent plasmid-encoded virulence profiles (von Mentzer et al. 2014). Comparing NCCP15740 isolated from diarrheal patients and non-ETEC O6 ATCC25922, NCCP15732 had two unique virulence genes: *fimZ* related adherence system and *tia* related invasion system. Interestingly, two genes are non-classical virulence factor known to contribute to diarrhea. As results, the combination type and virulence factors varied with time, even in same area.

In chapter 3, comparative analysis of O169:H41 strains, all of strains harbored CS6 colonization factor with STp enterotoxin type. A combination of CS6 and STp was one of typical ETEC toxin combinations (Gaastra and Svennerholm 1996, Svennerholm and Lundgren 2012, von Mentzer et al. 2014). Most pathogenic *E. coli* strains isolated from outbreaks belonged to phylo-group A and B1 (Escobar-Páramo et al. 2004, Badouei et al. 2015) and about 90% of ETEC outbreaks in Korea belonged to phylo-group A and B1 (Koo et al. 2012). However, NCCP15735 isolated from ETEC outbreaks in Korea and four strains (2014EL-1345-2, F6326-C1, F8111-1SC3, and F9792) isolated from ETEC outbreaks in USA belonged to phylo-group F. Five strains did not cluster in whole-genome phylogeny and hierarchical clustering of virulence factor.

In chapter 4, NCCP15741 appeared to belong to the O25:H16 serotype. Compared with other O25 strains, NCCP15741 had both eltA and eltB genes

encoding LT enterotoxin types. The presence of both eltA and eltB genes was common in O25 strains (8 out of 11, 73%). The CS6 colonization factors was most prevalent type, including NCCP15741. As a result of phylogenetic analysis and hierarchical clustering of total virulence factors, domestic food poisoning ETEC and US food poisoning ETEC were grouped differently. Second, the US food poisoning ETEC were grouped together in MLST-phylogenetic tree and clustering of virulence factors. Compared to ETEC O25 strains, NCCP15741 has additional Type I secretion system.

This study shows a broad comparative genomic analysis of five ETEC isolated from diarrhea patients in Korea and describes the genomic and pathogenic features of each serotype ETEC strains. This study will be a basic study to understand epidemiology of ETEC strains from outbreaks in Korea. This research will be useful for studying the evolution of pathogenic mechanisms in those strains and provide important information for future diagnosis and development of ETEC vaccine candidate.

CONCLUSION

In summary, new knowledges of ETEC strains in Korea were found. First, NCCP 15731 and NCCP157333 had unique genes related capsular polysaccharide, auto transporter, and secretion system. Second, phylo-groups are linked to genetic characteristics and toxic factors of ETEC strains. Third, NCCP15732 had the same sequence type as that of isolates from the USA but had different CF types and had two unique non-classical virulence genes known to contribute to diarrhea. Fourth, a combination of colonization factors with enterotoxin type of Kores isolates were consistent with that of globally distributes ETEC isolates. Fifth, as results of bioinformatics analysis of the five ETEC strains as major MLST types of domestic ETEC, all of their MLST type was changed. However, because the results were obtained from in silico analysis, experimental confirmation of these results is required.

REFERENCES

Angiuoli, S. V. and S. L. Salzberg (2010). "Mugsy: fast multiple alignment of closely related whole genomes." Bioinformatics **27**(3): 334-342.

Antão, E.-M., L. H. Wieler and C. Ewers (2009). "Adhesive threads of extraintestinal pathogenic *Escherichia coli*." Gut pathogens **1**(1): 22.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass and M. Kubal (2008). "The RAST Server: rapid annotations using subsystems technology." BMC genomics **9**(1): 75.

Badouei, M. A., M. Jajarmi and A. Mirsalehian (2015). "Virulence profiling and genetic relatedness of Shiga toxin-producing *Escherichia coli* isolated from humans and ruminants." Comparative immunology, microbiology and infectious diseases **38**: 15-20.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham and A. D. Prjibelski (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." Journal of computational biology **19**(5): 455-477.

Beatty, M. E., C. A. Bopp, J. G. Wells, K. D. Greene, N. D. Puhr and E. D. Mintz (2004). "Enterotoxin-producing *Escherichia coli* O169: H41, United States." Emerging infectious diseases **10**(3): 518.

Bölin, I., G. Wiklund, F. Qadri, O. Torres, A. L. Bourgeois, S. Savarino and A.-M. Svennerholm (2006). "Enterotoxigenic *Escherichia coli* with STh and STp genotypes is associated with diarrhea both in children in areas of

endemicity and in travelers." Journal of clinical microbiology **44**(11): 3872-3877.

Burgos, Y. and L. Beutin (2010). "Common origin of plasmid encoded alpha-hemolysin genes in *Escherichia coli*." BMC microbiology **10**(1): 193.

Chen, L., J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen and Q. Jin (2005). "VFDB: a reference database for bacterial virulence factors." Nucleic acids research **33**(suppl_1): D325-D328.

Clermont, O., J. K. Christenson, E. Denamur and D. M. Gordon (2013). "The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups." Environmental microbiology reports **5**(1): 58-65.

Crossman, L. C., R. R. Chaudhuri, S. A. Beatson, T. J. Wells, M. Desvaux, A. F. Cunningham, N. K. Petty, V. Mahon, C. Brinkley and J. L. Hobman (2010). "A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407." Journal of bacteriology **192**(21): 5822-5831.

Devasia, R. A., T. F. Jones, J. Ward, L. Stafford, H. Hardin, C. Bopp, M. Beatty, E. Mintz and W. Schaffner (2006). "Endemically acquired foodborne outbreak of enterotoxin-producing *Escherichia coli* serotype O169: H41." The American journal of medicine **119**(2): 168. e167-168. e110.

Escobar-Páramo, P., O. Clermont, A.-B. Blanc-Potard, H. Bui, C. Le Bouguénec and E. Denamur (2004). "A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*." Molecular biology and evolution **21**(6): 1085-1094.

Fleckenstein, J. M., P. R. Hardwidge, G. P. Munson, D. A. Rasko, H. Sommerfelt and H. Steinsland (2010). "Molecular mechanisms of enterotoxigenic *Escherichia coli* infection." Microbes and Infection **12**(2): 89-98.

Gaastra, W. and A.-M. Svennerholm (1996). "Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC)." Trends in microbiology **4**(11): 444-452.

Glaeser, S. P. and P. Kämpfer (2015). "Multilocus sequence analysis (MLSA) in prokaryotic taxonomy." Systematic and applied microbiology **38**(4): 237-245.

Gómez-Aldapa, C. A., E. Rangel-Vargas, H. Bautista-De León, M. E. Vázquez-Barrios, A. J. Gordillo-Martínez and J. Castro-Rosas (2013). "Behavior of enteroaggregative *Escherichia coli*, non-O157-shiga toxin-producing *E. coli*, enteroinvasive *E. coli*, enteropathogenic *E. coli* and enterotoxigenic *E. coli* strains on mung bean seeds and sprout." International journal of food microbiology **166**(3): 364-368.

Guerra, J. A., Y. C. Romero-Herazo, O. Arzuza and O. G. Gómez-Duarte (2014). "Phenotypic and genotypic characterization of enterotoxigenic *Escherichia coli* clinical isolates from Northern Colombia, South America." BioMed research international **2014**.

Hazen, T. H., J. Michalski, Q. Luo, A. C. Shetty, S. C. Daugherty, J. M. Fleckenstein and D. A. Rasko (2017). "Comparative genomics and transcriptomics of *Escherichia coli* isolates carrying virulence factors of both enteropathogenic and enterotoxigenic *E. coli*." Scientific Reports **7**(1): 3513.

Heberle, H., G. V. Meirelles, F. R. da Silva, G. P. Telles and R. Minghim (2015). "InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams." BMC bioinformatics **16**(1): 169.

Isidean, S., M. Riddle, S. Savarino and C. Porter (2011). "A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression." Vaccine **29**(37): 6167-6178.

Joensen, K. G., A. M. Tetzschner, A. Iguchi, F. M. Aarestrup and F. Scheutz (2015). "Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data." Journal of clinical microbiology: JCM. 00008-00015.

Khan, N. H., M. Ahsan, S. Yoshizawa, S. Hosoya, A. Yokota and K. Kogure (2008). "Multilocus sequence typing and phylogenetic analyses of *Pseudomonas aeruginosa* isolates from the ocean." Applied and environmental microbiology **74**(20): 6194-6205.

Koo, H.-J., H.-S. Kwak, S.-H. Yoon and G.-J. Woo (2012). "Phylogenetic group distribution and prevalence of virulence genes in *Escherichia coli* isolates from food samples in South Korea." World Journal of Microbiology and Biotechnology **28**(4): 1813-1816.

Kwon, T., S.-y. Chung, Y.-H. Jung, S.-J. Jung, S.-G. Roh, J.-S. Park, C.-H. Kim, W. Kim, Y.-S. Bak and S.-H. Cho (2017). "Comparative genomic analysis and characteristics of NCCP15740, the major type of enterotoxigenic *Escherichia coli* in Korea." Gut pathogens **9**(1): 55.

Lasaro, M., J. Rodrigues, C. Mathias-Santos, B. Guth, A. Balan, M. Sbrogio-Almeida and L. Ferreira (2008). "Genetic diversity of heat-labile toxin

expressed by enterotoxigenic *Escherichia coli* strains isolated from humans." Journal of bacteriology **190**(7): 2400-2410.

Leatham-Jensen, M. P., J. Frimodt-Møller, J. Adediran, M. E. Mokszycki, M. E. Banner, J. E. Caughron, K. A. Krogfelt, T. Conway and P. S. Cohen (2012). "The streptomycin-treated mouse intestine selects *Escherichia coli* envZ missense mutants that interact with dense and diverse intestinal microbiota." Infection and immunity **80**(5): 1716-1727.

Li, R., H. Zhu, J. Ruan, W. Qian, X. Fang, Z. Shi, Y. Li, S. Li, G. Shan and K. Kristiansen (2010). "*De novo* assembly of human genomes with massively parallel short read sequencing." Genome research **20**(2): 265-272.

Liu, J., S. Silapong, P. Jeanwattanalert, P. Lertsehtakarn, L. Bodhidatta, B. Swierczewski, C. Mason, A. L. McVeigh, S. J. Savarino and R. Nshama (2017). "Multiplex real time PCR panels to identify fourteen colonization factors of enterotoxigenic *Escherichia coli* (ETEC)." PloS one **12**(5): e0176882.

Lymberopoulos, M. H., S. Houle, F. Daigle, S. Léveillé, A. Brée, M. Moulin-Schouleur, J. R. Johnson and C. M. Dozois (2006). "Characterization of Stg fimbriae from an avian pathogenic *Escherichia coli* O78: K80 strain and assessment of their contribution to colonization of the chicken respiratory tract." Journal of bacteriology **188**(18): 6449-6459.

Madhavan, T. V., J. A. Steen, P. Hugenholtz and H. Sakellaris (2014). "Genome sequence of enterotoxigenic *Escherichia coli* strain B2C." Genome announcements **2**(2): e00247-00214.

McWilliams, B. and A. Torres Enterohemorrhagic *Escherichia coli* adhesins.

Microbiol Spectrum. 2014; doi: 10.1128/microbiolspec, EHEC-0003-2013.

Minogue, T., H. Daligault, K. Davenport, K. Bishop-Lilly, S. Broomall, D. Bruce, P. Chain, O. Chertkov, S. Coyne and T. Freitas (2014). "Complete genome assembly of *Escherichia coli* ATCC 25922, a serotype O6 reference strain." Genome announcements **2**(5): e00969-00914.

Nataro, J. P. and J. B. Kaper (1998). "Diarrheagenic *escherichia coli*." Clinical microbiology reviews **11**(1): 142-201.

Oh K-H, Chung GT, Yoo C-K. 2015. Waterborne and Foodborne Disease by Pathogenic *Escherichia coli* in Korea, 2014. Public Health Weekly Report, KCDC 8: 1250-54

Oh, K.-H., D. W. Kim, S.-M. Jung and S.-H. Cho (2014). "Molecular characterization of enterotoxigenic *Escherichia coli* strains isolated from diarrheal patients in Korea during 2003–2011." PLoS One **9**(5): e96896.

Organization, W. H. (2006). "Future directions for research on enterotoxigenic *Escherichia coli* vaccines for developing countries= Orientations futures de la recherche sur les vaccins contre *Escherichia coli* entérotoxinogène destinés aux pays en développement." Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire **81**(11): 97-104.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello and M. Shukla (2013). "The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)." Nucleic acids research **42**(D1): D206-D214.

Pattabiraman, V. and C. A. Bopp (2014). "Draft whole-genome sequences of 10 serogroup O6 enterotoxigenic *Escherichia coli* strains." Genome

announcements **2**(6): e01274-01214.

Pattabiraman, V. and C. A. Bopp (2015). "Draft whole-genome sequences of 10 enterotoxigenic *Escherichia coli* serogroup O6 strains." Genome announcements **3**(3): e00564-00515.

Pi, H., S. A. Jones, L. E. Mercer, J. P. Meador, J. E. Caughron, L. Jordan, S. M. Newton, T. Conway and P. E. Klebba (2012). "Role of catecholate siderophores in gram-negative bacterial colonization of the mouse gut." PLoS One **7**(11): e50020.

Price, M. N., P. S. Dehal and A. P. Arkin (2009). "FastTree: computing large minimum evolution trees with profiles instead of a distance matrix." Molecular biology and evolution **26**(7): 1641-1650.

Qadri, F., A.-M. Svennerholm, A. Faruque and R. B. Sack (2005). "Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention." Clinical microbiology reviews **18**(3): 465-483.

Rice, P., I. Longden and A. Bleasby (2000). EMBOSS: the European molecular biology open software suite, Elsevier Current Trends.

Rivera, F., T. Ochoa, R. Maves, M. Bernal, A. Medina, R. Meza, F. Barletta, E. Mercado, L. Ecker and A. Gil (2010). "Genotypic and phenotypic characterization of enterotoxigenic *Escherichia coli* strains isolated from Peruvian children." Journal of clinical microbiology **48**(9): 3198-3203.

Rodas, C., V. Iniguez, F. Qadri, G. Wiklund, A.-M. Svennerholm and Å. Sjöling (2009). "Development of multiplex PCR assays for detection of enterotoxigenic *Escherichia coli* colonization factors and toxins." Journal

of clinical microbiology **47**(4): 1218-1220.

Sahl, J. W. and D. A. Rasko (2012). "Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A." Infection and immunity **80**(3): 1232-1242.

Searle, L. J., G. Méric, I. Porcelli, S. K. Sheppard and S. Lucchini (2015). "Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *Escherichia coli*." PLoS One **10**(3): e0117906.

Shin, J., K.-B. Yoon, D.-Y. Jeon, S.-S. Oh, K.-H. Oh, G. T. Chung, S. W. Kim and S.-H. Cho (2016). "Consecutive outbreaks of enterotoxigenic *Escherichia coli* O6 in schools in South Korea caused by contamination of fermented vegetable kimchi." Foodborne pathogens and disease **13**(10): 535-543.

Simon-Assmann, P., C. Spenle, O. Lefebvre and M. Kedinger (2010). The role of the basement membrane as a modulator of intestinal epithelial mesenchymal interactions. Progress in molecular biology and translational science, Elsevier. **96:** 175-206.

Smith, P., R. L. Lindsey, L. A. Rowe, D. Batra, D. Stripling, L. Garcia-Toledo, D. Drapeau, K. Knipe and N. Strockbine (2018). "High-Quality Whole-Genome Sequences for 21 Enterotoxigenic *Escherichia coli* Strains Generated with PacBio Sequencing." Genome announcements **6**(2): e01311-01317.

Stamatakis, A. (2006). Phylogenetic models of rate heterogeneity: a high performance computing perspective. Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International, IEEE.

Steinsland, H., D. W. Lacher, H. Sommerfelt and T. S. Whittam (2010). "Ancestral lineages of human enterotoxigenic *Escherichia coli*." Journal of clinical microbiology **48**(8): 2916-2924.

Steinsland, H., P. Valentiner-Branth, P. Aaby, K. Mølbak and H. Sommerfelt (2004). "Clonal relatedness of enterotoxigenic *Escherichia coli* strains isolated from a cohort of young children in Guinea-Bissau." Journal of clinical microbiology **42**(7): 3100-3107.

Svennerholm, A.-M. and A. Lundgren (2012). "Recent progress toward an enterotoxigenic *Escherichia coli* vaccine." Expert review of vaccines **11**(4): 495-507.

Team, R. C. (2013). "R: A language and environment for statistical computing."

Trees, E., N. Strockbine, S. Changayil, S. Ranganathan, K. Zhao, R. Weil, D. MacCannell, A. Sabol, A. Schmidtke and H. Martin (2014). "Genome sequences of 228 Shiga toxin-producing *Escherichia coli* isolates and 12 isolates representing other diarrheagenic *E. coli* pathotypes." Genome announcements **2**(4): e00718-00714.

Turner, S. M., R. R. Chaudhuri, Z.-D. Jiang, H. DuPont, C. Gyles, C. W. Penn, M. J. Pallen and I. R. Henderson (2006). "Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages." Journal of clinical microbiology **44**(12): 4528-4536.

von Mentzer, A., T. R. Connor, L. H. Wieler, T. Semmler, A. Iguchi, N. R. Thomson, D. A. Rasko, E. Joffre, J. Corander and D. Pickard (2014). "Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with

long-term global distribution." Nature genetics **46**(12): 1321.

Von Mentzer, A., J. Tobias, G. Wiklund, S. Nordqvist, M. Aslett, G. Dougan, Å. Sjöling and A.-M. Svennerholm (2017). "Identification and characterization of the novel colonization factor CS30 based on whole genome sequencing in enterotoxigenic *Escherichia coli* (ETEC)." Scientific reports **7**(1): 12514.

Wachsmuth, K., J. Wells, P. Shipley and R. Ryder (1979). "Heat-labile enterotoxin production in isolates from a shipboard outbreak of human diarrheal illness." Infection and immunity **24**(3): 793-797.

Wennerås, C. and V. Erling (2004). "Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world." Journal of Health, Population and Nutrition: 370-382.

Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. Maiden and H. Ochman (2006). "Sex and virulence in Escherichia coli: an evolutionary perspective." Molecular microbiology **60**(5): 1136-1151.

Yoon, S.-H., S.-m. Ha, J. Lim, S. Kwon and J. Chun (2017). "A large-scale evaluation of algorithms to calculate average nucleotide identity." Antonie van Leeuwenhoek **110**(10): 1281-1286.

Appendix 1. Accession number of reference strains in this study.

Strain	Serotype	ERS number
E. coli E159	O159:H34	ERS077713
E. coli E927	O159:H20	ERS038952
E. coli E943	O159:H20	ERS044499
E. coli E133	O159:H4	ERS044470
E. coli E1532	O159:H4	ERS077694
E. coli E1573	O159:H4	ERS077704
E. coli E1679sc	O159:H4	ERS163315
E. coli E391	O159:H21	ERS077592
E. coli E833	O159:H34	ERS077628
E. coli E885	O159:H21	ERS077638
E. coli E1640	O159:H34	ERS077734
E. coli E5085	O159:H34	ERS055670
E. coli E5088	O159:H34	ERS055673
E. coli E5051	O159:H18	ERS055664
E. coli E344	O169:H41	ERS038934
E. coli E370	O169:H41	ERS038935
E. coli E520	O169:H41	ERS077602
E. coli E554	O169:H41	ERS038939
E. coli E1245	O169:H41	ERS038963
E. coli E1328	O169:H41	ERS206745
E. coli E1329	O169:H41	ERS206746
E. coli E1373	O169:H41	ERS038966
E. coli E1484	O169:H41	ERS206752
E. coli E2256	O169:H41	ERS055653
E. coli E5084	O169:H41	ERS055669
E. coli E97	O25:H42	ERS044465
E. coli E106	O25:H42	ERS044455
E. coli E1114	O25:H16	ERS206736

E. coli E1118	O25:H39	ERS206740
E. coli E1119	O25:H16	ERS206741
E. coli E1122	O25:H16	ERS206744
E. coli E1516	O25:H16	ERS206754

Strain	Serotype	Accession number
E. coli NCCP15740	O16:H16	MOED01
E. coli CFA/II str. B2C	O16:H16	AUZS01
E. coli 99-3165	O16:H16	JHJW01
E. coli F5656C1	O16:H16	CP024260
E. coli 2011EL-1370-2	O16:H16	JPUT01
E. coli 2011EL-1369-1	O16:H16	JPUS01
E. coli 2011EL-1497-2	O16:H16	JPUU01
E. coli M9803	O16:H16	JPXK01
E. coli 2011EL-1640-5	O16:H16	JPXN01
E. coli 2011EL-1251-4	O16:H16	JPXM01
E. coli K1506-c2	O16:H16	JYIB01
E. coli 2014EL-1181-1	O16:H16	JYIE01
E. coli K1884 sc	O16:H16	JYIC01
E. coli B1020-2	O16:H16	JYIF01
E. coli F5524-c2	O16:H16	JYIG01
E. coli ATCC25922	O16:H1	CP009072
E. coli 2014EL-1345-2	O169:H41	CP024223
E. coli F6326-C1	O169:H41	CP024263
E. coli F8111-1SC3	O169:H41	CP024269
E. coli F9792	O169:H41	CP024273
E. coli ATCC43886	O25:H16	CP024256
E. coli 2014EL-1343-2	O25:H16	CP024228
E. coli F5505-C1	O25:H16	CP024266

Appendix 2. Virulence factors of ETEC O159 strains.

Sample Number	Strain	Serotype
1-1	*E. coli* NCCP15731	O159:H34
1-2	*E. coli* NCCP15733	O159:H34
1-3	*E. coli* E159	O159:H34
2-1	*E. coli* E927	O159:H20
2-2	*E. coli* E943	O159:H20
3-1	*E. coli* E133	O159:H4
3-2	*E. coli* E1532	O159:H4
3-3	*E. coli* E1573	O159:H4
3-4	*E. coli* E1679sc	O159:H4
4-1	*E. coli* E391	O159:H21
4-2	*E. coli* E885	O159:H21
5-1	*E. coli* E833	O159:H34
5-2	*E. coli* E1640	O159:H34
5-3	*E. coli* E5085	O159:H34
5-4	*E. coli* E5088	O159:H34
6-1	*E. coli* E5051	O159:H18

Category	Subcategory	gene	1-1	1-2	1-3	2-1	2-2	3-1	3-2	3-3	3-4	4-1	4-2	5-1	5-2	5-3	5-4	6-1
Adherence	CFA/I fimbriae	cfaA	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	cfaB	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	cfaC	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	cfaD	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	ibeB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	ibeC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpA	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
Adherence	E. coli common pilus (ECP)	ecpB	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpC	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpD	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpE	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpR	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Adherence	EaeH	eaeH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Fimbriae	fimA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Fimbriae	fimB	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0
Adherence	Fimbriae	fimC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Adherence	Fimbriae	fimD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flhA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flhB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Adherence	Flagella	fliF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliL	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Adherence	Flagella	fliM	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliQ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	motA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	motB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Hsp60	htpB	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	LOS synthesis	rfaF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	P fimbriae	papX	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0

Adherence	Type I fimbriae	fimE	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
Adherence	Type I fimbriae	fimF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimI	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
Adherence	Type IV pili	pilP	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Adherence	Type IV pili	pilW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Adhesin	Cell surface hydrophobicity proteins	cshB	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
Anaerobic respiration	Nitrate reductase	narH	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Antiphagocytosis	Alginate regulation	algU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	cpsA	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Antiphagocytosis	Capsular polysaccharide	rmlA	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Antiphagocytosis	Capsular polysaccharide	rmlB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wbfV/wcvB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wbjD/wecB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wecA	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Antiphagocytosis	Capsular polysaccharide	wecC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wza	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsule I	manC	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0

Autotransporter	AatA, AIDA-I type	aatA	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
Autotransporter	Antigen 43, AIDA-I type	agn43	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0
Autotransporter	Cah, AIDA-I type	cah	0	0	1	1	1	1	0	0	1	0	0	1	1	1	1	1
Autotransporter	EhaA, AIDA-I type	ehaA	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Autotransporter	EhaB, AIDA-I type	ehaB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autotransporter	Enteroaggregative immunoglobulin repeat protein	air/eaeX	0	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0
Autotransporter	EspC, SPATE	espC	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Autotransporter	Pic	pic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Autotransporter	UpaG adhesin, trimeric AT	upaG/ehaG	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1
Autotransporter	UpaH, AIDA-I type	upaH	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Chemotaxis and motility	Flagella	cheA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheW	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	galU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	gmhA/l	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

		pcA															
Endotoxin	LOS	kdsA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxB	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Endotoxin	LOS	lpxC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	msbA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	rfaD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	rfaE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Fim	fimZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence	Sta	staB	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1

determinants																		
Fimbrial adherence determinants	Sta	staC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Sta	staD	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
Fimbrial adherence determinants	Stc	stcB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stc	stcC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stf	stfD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stg	stgA	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stg	stgB	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stg	stgC	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stg	stgD	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Host immune evasion	Exopolysaccharide	mrsA/glmM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Host immune evasion	Exopolysaccharide	pgi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Immune evasion	LPS	acpXL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Immune evasion	LPS	gmd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Invasion	Flagella	cheD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Invasion	Tia/Hek	tia	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0
Iron and heme acquisition	Heme biosynthesis	hemC	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Iron and heme acquisition	Heme biosynthesis	hemE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Brucebactin	entD	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entA	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entF	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepD	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Ferrous iron transport	sitA	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Iron uptake	Ferrous iron transport	sitB	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Iron uptake	Ferrous iron transport	sitC	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Iron uptake	Ferrous iron transport	sitD	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Lipid and fatty acid metabolism	Isocitrate lyase	icl	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Metal uptake	Mycobactin	mbtI	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Motility	Flagella	flhC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Motility	Flagella	flhD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Motility	Flagella	fliY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspL1	espL1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspL4	espL4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspR1	espR1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspR4	espR4	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Non-LEE encoded TTSS effectors	EspX1	espX1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspX4	espX4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspX5	espX5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Non-LEE encoded TTSS effectors	EspY1	espY1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
Others	Hemolysin	hlyA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Others	MsbB2	msbB2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Others	VirK	virK	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1
Quorum sensing	Autoinducer-2	luxS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Alternative sigma factor RpoS	rpoS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Carbon storage regulator A	csrA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Regulation	GacS/GacA two-component system	gacA	0	0	1	1	1	0	0	0	0	1	1	0	0	0	0	1
Regulation	PhoP/R	phoP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Sigma A	sigA/rpoV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulator	AggR	aggR	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Bsa T3SS	orgA	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	1
Secretion system	Bsa T3SS	orgB	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1
Secretion system	EPS type II secretion system	gspD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Flagella (cluster I)	flhE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Flagella (cluster I)	fliZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Mxi-Spa TTSS effectors controlled by MxiE	ipaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Mxi-Spa TTSS effectors controlled by MxiE	ipaH2.5	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	1
Secretion system	T2SS (Type II secretion system)	gspC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Secretion system	T2SS (Type II secretion system)	gspJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	TTSS (SPI-1 encode)	iagB	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1
Stress adaptation	Catalase-peroxidase	katG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Stress protein	SodCI	sodCI	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	0
Toxins	Heat-labile enterotoxin	eltA	0	0	1	0	0	1	1	1	1	1	1	1	1	0	0	1
Toxins	Heat-labile enterotoxin	eltB	0	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1
Toxins	Hemolysin/cytolysin A	hlyE/clyA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Two-component system	PhoPQ	phoQ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type I secretion system	ABC transporter for dispersin	aatB	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
Type I secretion system	ABC transporter for dispersin	aatD	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
Type I secretion system	ABC transporter for dispersin	aatP	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0
Type VI secretion system	AAI/SCI-II T6SS	aaiA	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiB	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiD	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1

Type VI secretion system	AAI/SCI-II T6SS	aaiE	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiF	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiH	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiJ	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiK	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiL	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	aaiN	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	clpV/aaiP	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	icmF/aaiO	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Type VI secretion system	AAI/SCI-II T6SS	vgrG	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec15	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec17	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec18	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec19	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec22	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec23	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1

Type VI secretion system	ACE T6SS	aec24	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec25	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec26	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec27/clpV	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec28	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec29	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec30	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Appendix 3. Virulence factors of ETEC O6 strains.

Sample Number	Strain	Serotype
1-1	*E. coli* NCCP15732	O6:H16
1-2	*E. coli* NCCP15740	O6:H16
2-1	*E. coli* O6:H16 str. 99-3165	O6:H16
3-1	*E. coli* strain 2011EL-1251-4	O6:H16
3-2	*E. coli* strain 2011EL-1369-1	O6:H16
3-3	*E. coli* strain 2011EL-1370-2	O6:H16
3-4	*E. coli* strain 2011EL-1497-2	O6:H16
3-5	*E. coli* strain 2011EL-1640-5	O6:H16
4-1	*E. coli* strain 2014EL-1181-1	O6:H16
5-1	*E. coli* strain B1020-2	O6:H16
6-1	*E. coli* O6:H16:CFA/II str. B2C	O6:H16
6-2	*E. coli* strain F5524-c2	O6:H16
6-3	*E. coli* strain F5656C1	O6:H16
7-1	*E. coli* strain K1506-c2	O6:H16
7-2	*E. coli* strain M9803	O6:H16
7-3	*E. coli* strain K1884 sc	O6:H16

Category	Subcategory	Gene	1-1	1-2	2-1	3-1	3-2	3-3	3-4	3-5	4-1	5-1	6-1	6-2	6-3	7-1	7-2	7-3
Adherence	CFA/I fimbriae	ibeB	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	CFA/I fimbriae	ibeC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpB	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpC	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpD	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpE	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Adherence	E. coli common pilus (ECP)	ecpR	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1
Adherence	EaeH	eaeH	0	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Adherence	EtpA	etpA	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Adherence	Fimbriae	fimA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Fimbriae	fimB	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0
Adherence	Fimbriae	fimC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Fimbriae	fimD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Adherence	Flagella	flgE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flgN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flhA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	flhB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Adherence	Flagella	fliK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliQ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	fliT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	motA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Flagella	motB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Hsp60	htpB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	LOS synthesis	rfaF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Toxin-coregulated pilus (type IVB pilus)	tcpC	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
Adherence	Type I fimbriae	fimE	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0
Adherence	Type I fimbriae	fimF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adherence	Type I fimbriae	fimI	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0

Antiphagocytosis	Alginate regulation	algU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	rmlB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wbfV/wcvB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wbjD/wecB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wecC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Antiphagocytosis	Capsular polysaccharide	wza	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autotransporter	AatA, AIDA-I type	aatA	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Autotransporter	Cah, AIDA-I type	cah	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Autotransporter	EhaB, AIDA-I type	ehaB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autotransporter	Pic	pic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheW	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chemotaxis and motility	Flagella	cheZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	galU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Endotoxin	LOS	gmhA/lpcA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	kdsA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	lpxD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	msbA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	rfaD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Endotoxin	LOS	rfaE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Agf/Csg	csgG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Fim	fimZ	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1

Fimbrial adherence determinants	Sta	staC	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stc	stcB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stc	stcC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fimbrial adherence determinants	Stf	stfD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Host immune evasion	Exopolysaccharide	mrsA/glmM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Host immune evasion	Exopolysaccharide	pgi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Immune evasion	LPS	acpXL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Invasion	Flagella	cheD	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Invasion	Tia/Hek	tia	1	0	1	0	1	1	0	0	1	1	0	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron and heme acquisition	Heme biosynthesis	hemN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Brucebactin	entD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Iron uptake	Enterobactin synthesis	entE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin synthesis	entF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Iron uptake	Enterobactin transport	fepG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lipid and fatty acid metabolism	Isocitrate lyase	icl	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Motility	Flagella	flhC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Motility	Flagella	flhD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Motility	Flagella	fliY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspL1	espL1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspL4	espL4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspR1	espR1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspX1	espX1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspX4	espX4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Non-LEE encoded TTSS effectors	EspX5	espX5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Other adhesion-related proteins	EF-Tu	tuf	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

Others	Hemolysin	hlyA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Others	MsbB2	msbB2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Others	VirK	virK	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Quorum sensing	Autoinducer-2	luxS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Alternative sigma factor RpoS	rpoS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Carbon storage regulator A	csrA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	GacS/GacA two-component system	gacA	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	0
Regulation	PhoP/R	phoP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulation	Sigma A	sigA/rpoV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulator	AggR	aggR	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Secretion system	EPS type II secretion system	gspD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Flagella (cluster I)	flhE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Flagella (cluster I)	fliZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Mxi-Spa TTSS effectors controlled by MxiE	ipaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	Mxi-Spa TTSS effectors controlled by MxiE	ipaH2.5	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Secretion system	T2SS (Type II secretion system)	gspG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspH	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspJ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspK	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Secretion system	T2SS (Type II secretion system)	gspM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Stress adaptation	Catalase-peroxidase	katG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toxins	Heat-labile enterotoxin	eltA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toxins	Heat-labile enterotoxin	eltB	1	1	1	0	0	0	1	1	0	0	0	1	0	1	1	0
Toxins	Hemolysin/cytolysin A	hlyE/clyA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Two-component system	PhoPQ	phoQ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type I secretion system	ABC transporter for dispersin	aatB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type I secretion system	ABC transporter for dispersin	aatD	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0
Type I secretion system	ABC transporter for dispersin	aatP	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Type VI secretion system	ACE T6SS	aec30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Type VI secretion system	ACE T6SS	aec32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ABSTRACT (In Korean)

장독소형대장균 (Enterotoxigenic *Escherichia coli*, ETEC)은 주요 설사 및 식중독을 유발하는 박테리아 감염균으로 오염된 음식과 물을 통해 감염을 시킨다. ETEC은 장관독소 (Enterotoxin)인 열에 민감한 독소 (LT) 혹은 열에 안정한 독소 (ST)를 생산하여 인간을 포함하는 포유류에 복통을 수반한 장염을 일으킨다. ETEC는 숙주의 장 표면의 상피 세포에 정착하기 위해 하나 혹은 그 이상의 식민 인자 (colonization factor, CF)로 불리는 세포 부착 인자를 보유하고있다.

이전 연구에서 국내 설사 환자 258 명을 대상으로 CF에 대한 분석을 실시하고 MLST (multilocus sequence typing)를 시행했다. ST171 (24 %)은 한국에서 가장 널리 사용되는 ETEC 유형으로 확인되었으며, ST955, ST964, ST656등이 주요한 MLST유형으로 밝혀졌다. 그러나 한국의 주요 MLST 유형들의 ETEC의 지놈 특성은 아직 조사되지 않았다.

본 연구에서는 2003년에서 2001년까지 한국에서 가장 많이 발견된 5가지 MLST 유형의 ETEC 분리 균의 지놈 및 독성인자의 특성을 밝히기 위해 전체 지놈 서열 분석을 수행했다. 본 논문은 ETEC O159 NCCP15731과 NCCP15733 (1 장), ETEC O6 NCCP15732 (2 장), ETEC

O169 NCCP15735 (3 장) 및 ETEC O25 NCCP15741 (4 장)로 구성되어 있다. 또한 비슷한 기간 동안 세계적으로 보고 된 ETEC 와 비교 유전체 분석을 수행했다.

요약하자면 본 연구에서 한국의 ETEC 에 대하여 몇 가지 새로운 지식이 발견되었다. 국내 ETEC는 다른 나라에서의 ETEC와 비교한 결과, 특이적 유전자를 가지고 있으며, 이들의 장내 독소 (enterotoxin) 유형에 의한 식민지 인자의 조합은 세계적으로 분포하는 ETEC의 독소조합들과 일치한다. 이 논문은 ETEC의 병원성과 한국의 특정 병원체의 생물학적 기능을 이해하기위한 기초 연구로 활용될 것으로 사료된다.

주요어 : 병원균, 전체 지놈 서열 분석, 비교 지놈 분석, 대장균, 장독소형대장균, 독성 인자

학번 : 2016 - 22634