

GENOME REPORT

Open Access



# Comparative genomic analysis of enterotoxigenic *Escherichia coli* O159 strains isolated from diarrheal patients in Korea

Si-yun Chung<sup>1†</sup>, Taesoo Kwon<sup>2†</sup>, Young-Seok Bak<sup>3</sup>, Joung Je Park<sup>4</sup>, Cheorl-Ho Kim<sup>5</sup>, Seung-Hak Cho<sup>6‡</sup> and Won Kim<sup>1\*‡</sup>

## Abstract

**Background:** Enterotoxigenic *Escherichia coli* (ETEC) is a common cause of bacterial infection that leads to diarrhea. Although some studies have proposed a potential association between the toxic profile and genetic background, association between toxin of ETEC and phylo-group has not been reported yet. The objective of this study was to examine genomic and phylogenetic characteristics of ETEC strain NCCP15731 and NCCP15733 by whole genome sequencing and comparative genomic analysis of two phylo-groups of O159 reference strains.

**Results:** Whole genome sequencing showed that genome size of NCCP15731 strain was 4,663,459 bp, containing 4435 CDS and 19 RNAs. The genome size of NCCP15733 was 4,645,336 bp, containing 4369 CDS and 23 RNAs. Both NCCP15731 and NCCP15733 were classified in the phylo-group A, which is one of major *E. coli* phylogenetic groups. Their serotype was O159:H34. They possessed the virulence factor such as adherence systems, auto transporter systems, and flagella segments of major driving force for ETEC pathogenicity. They also harbored STh enterotoxin. Hierarchical clustering result based on the presence or absence of a total of 108 major virulence factors of 14 O159 ETEC strains showed that seven strains in phylo-group A and seven strains in phylo-group B1 were clustered each other, respectively. Colonization factors (CFs) of NCCP15731 or NCCP15733 were not detected.

**Conclusions:** Serotype of NCCP15731 and NCCP15733, representing major types of ETEC in Korea, was O159:H34 and their MLST type was ST218. Comparison with other O159 strains revealed that NCCP15731 was specialized for transporter system and secretion system whereas NCCP15733 had unique genes related to capsular polysaccharide. Compared with E159, the most recent common ancestor, these two strains had different toxin type and virulence factors. These results will improve our understanding of ETEC O159 strains to prevent ETEC disease.

**Keywords:** Enterotoxigenic *Escherichia coli* O159, Whole genome sequencing, Virulence factors, Colonization factors, Phylo-groups

\*Correspondence: wonkim@snu.ac.kr

†Si-yun Chung and Taesoo Kwon contributed equally to the work

‡Won Kim and Seung-Hak Cho contributed equally to this work

<sup>1</sup> School of Biological Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

Full list of author information is available at the end of the article



## Background

*Escherichia coli* is a Gram-negative bacterium belonging to family *Enterobacteriaceae*. Most of *E. coli* are typical members of the normal microflora of the humans and animals [1]. However, they could be classified as pathogenic strains that cause serious disease like diarrhea. Pathogenic *E. coli* can be divided to intestinal pathogenic *E. coli* (IPEC) and extraintestinal pathogenic *E. coli* (ExPEC) [2]. Intestinal pathogenic *E. coli* include enteroaggregative *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC) and Shiga toxin-producing *E. coli* (STEC), while extraintestinal pathogenic *E. coli* include uropathogenic *E. coli* (UPEC) depending on the various virulence genes and phenotypes that play important roles in the pathogenesis [3–8]. ETEC is a common cause of bacterial infection that leads to diarrhea in infants, young children, and travelers in developing countries [9, 10]. Because ETEC outbreaks are gained by consuming contaminated food or water, food contamination is an important concern for public health [11, 12]. ETEC can be defined by the ability to produce 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 [13–15]. While STh is produced by ETEC isolated from humans, STp originally found in pig ETEC is also associated with disease in humans [13, 16]. ETEC harbors one or more cell adhesion factors called colonization factors (CFs) to colonize epithelial cells of intestinal surfaces of hosts. ETEC CFs are named as coli surface antigens (CS) with a number except CFA/I. Currently, more than 30 colonization factors (CFs) have been identified in human ETEC. They are co-expressed with one, two, or three CFs and/or toxic factors, such as CS1 + CS3 ( $\pm$  CS21) with LT + STh, CS2 + CS3 ( $\pm$  CS21) with LT + STh, CS5 + CS6 with LT + STh, CS6 with STp, CFA/I ( $\pm$  CS21) with STh and CS7 with LT [17–19]. However, between 30 and 50% of ETEC have undetectable CFs, suggesting that there are still unknown CFs [20, 21]. Phylogenetic analysis is important for investigating the evolution and diversity of *E. coli* and evaluating bacterial toxicity [22]. Most commensal and pathogenic *E. coli* strains belong to phylo-groups A and B1 [23, 24]. About 90% of foodborne *E. coli* isolates in Korea belong to phylo-groups A and B1 [25].

In a previous study [26], 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. NCCP15740 [27], representing the major MLST type ST171 of ETEC in Korea, has been investigated about its genomic features, CF genes, and virulence factors. However, other

MLST types have not been investigated. In this study, we selected one of ST964 strains and one of ST656 strains identified in the previous work namely NCCP15731 and NCCP15733 [26], because these strains are the most representative strains of ST964 and ST656, and performed whole-genome sequencing. We compared whole-genome sequences of these two strains with NCCP15732 of O6 isolate and those of other ETEC strains reported as serotype O159 isolates.

## Methods

### Bacteria and strain isolation

*Escherichia Coli* were isolated from patients with diarrhea outbreak and identified as third and fourth highest prevalent MLST type (ST964 and ST656) of ETEC in Korea based on 7 isolates obtained from 2003 to 2011, respectively [26]. Candidate colonies of *E. coli* were identified based on phenotypes and biochemical properties using the API20E system (Biomerieux, Marcy l'Etoile, France). These isolated strains were deposited at National Culture Collection for Pathogens (NCCP) under the registration numbers NCCP15731 and NCCP15733. We selected 12 *E. coli* O159 strains as reference strains with the same O serotype as NCCP15731 and NCCP15733 listed in Table 2. These 12 *E. coli* O159 strains had been isolated globally from 1980 to 2011 [18]. Illumina short reads of reference strains were obtained at NCBI SRA (Sequence Read Archive) under the accession numbers listed in Table 2. De novo assemblies were performed with Spades (version 3.5.0) [28]. *E. coli* NCCP15732 was used as the reference strain because this strain represents the second highest prevalent ST955 in Korea. In comparison with 12 strains of *E. coli* O159 and NCCP15732 in Korea, we anticipated that we could identify genomic and pathogenic characteristics of NCCP15731 and NCCP15733.

### Whole genome sequencing, assembly and annotation

Genomic DNAs of a single bacterial isolate of NCCP15731 or NCCP15733 were extracted from a pure culture. Potential contamination of other microorganisms was checked using a BLAST search against non-redundant database. A sequencing library was created using TruSeq sample preparation kit (Illumina, San Diego, CA, USA). Whole genome sequencing of NCCP15731 and NCCP15733 was performed using the Illumina HiSeq 2000 platform (Theragen Etx Bio Institute, Suwon, Republic of Korea). High-quality reads were assembled by discarding low-reads, quality scores < Q20, and duplicated reads using SOAP de novo (version 1.05) [29]. Assembled contigs of NCCP15731 and NCCP15733 were annotated-using the

Rapid Annotation using Subsystem Technology (RAST version 4.0) [30] server pipeline.

**Genomic analysis**

The same method as described by Clermont et al. [22] was performed in silico to identify *E. coli* phylogenetic groups (A, B1, B2, D, E, and F) using Primersearch program from the European Molecular Biology Open Software Suite(EMBOSS) [31]. In silico SerotypeFinder (version 1.1) [32] was used to identify serotype of 14 O159 strains including NCCP15731 and NCCP15733 and NCCP15732. Colonization factors and MLST type were identified using CF primers [33] and *E. coli* MLST database [34]. To identify the genes encoding virulence factors, the total CDSs of NCCP15731, NCCP15732, NCCP15733 and 12 O159 strains were analyzed using BLASTP [35] against the virulence factor genes of *E. coli* listed in VFDB with an e-value of  $1e-5$  [36]. We selected genes with coverage of at least 60%. Resistance genes in Whole-genome sequence of all isolates were identified by ResFinder [37] Default thresholds of coverage of at least 60% and identity of at least 90% were employed.

**Phylogenetic analysis**

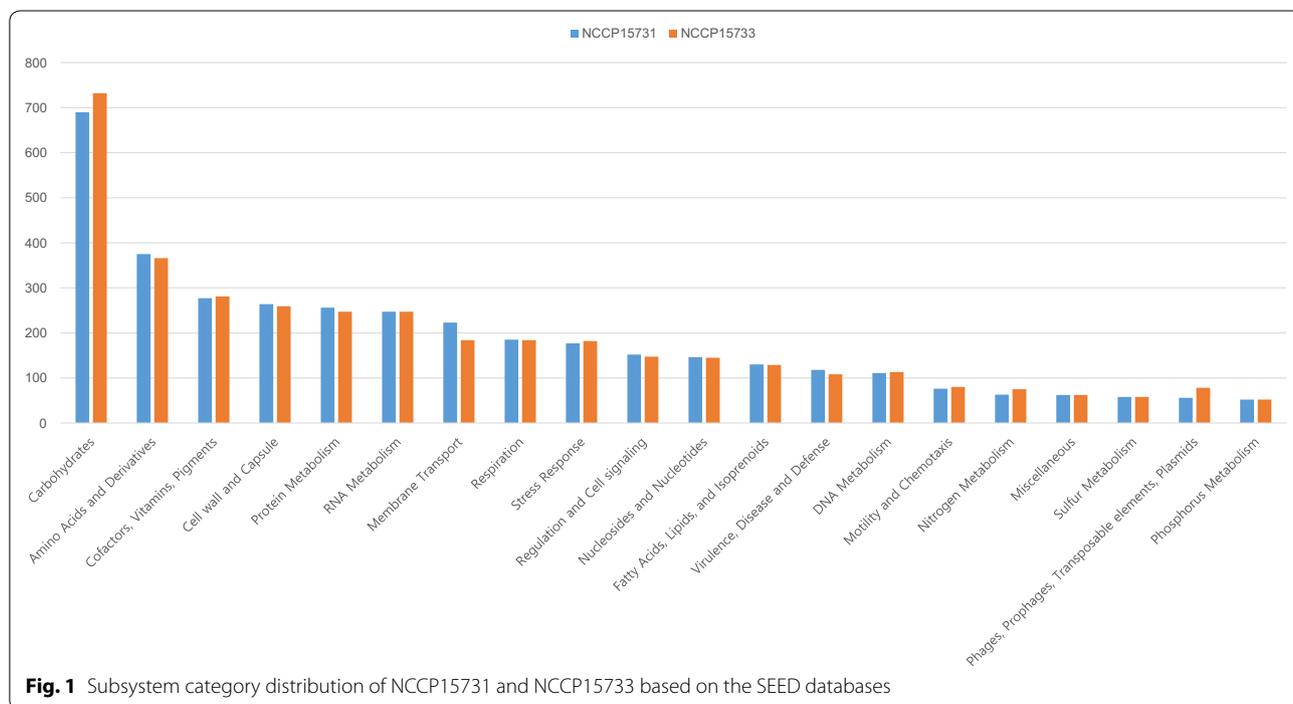
To identify evolutionary relationship of 14 O159 strains and NCCP15732, phylogenetic analysis was performed. Multiple sequence alignments obtained from concatenated whole genome sequences of each of all strains were performed with Mugsy (version 1.2.3) [38]. The

generalized time reversible +CAT model [39] was used for maximum-likelihood phylogenetic tree construction using FastTree (version 2.1.7) [40]. 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 (Horizontal gene transfer) in phylogenetic analysis, multi-locus sequence analysis method was used [41, 42]. Seven housekeeping genes (*adhA*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*) from each of all ETEC sequences were obtained according to the protocol described in *E. coli* MLST database ([http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi\\_html](http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi_html)). A phylogenetic tree based on MLST genes was created using the method employed for whole genome phylogenetic analysis.

**Results and discussion**

**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, 4435 coding sequences and 19 tRNA genes were detected in the genome of NCCP15731, of which 3855 (87%) were functional. A total number of 4369 coding sequences and 23 tRNA genes were detected in the genome of NCCP15733, of which 3852 (88%) were functional (Fig. 1). In silico analysis revealed that serotype of both NCCP15731 and NCCP15733 was O159:H34 belonging to phylo-group A, one of major *E. coli* phylogenetic groups [23]. Genomic and phylogenetic



characteristics of NCCP15731 and NCCP15733 were shown in Table 1. We investigated toxin types and colonization factors of NCCP15731 and NCCP15733. Both NCCP15731 and NCCP15733 possessed STh enterotoxin, but not LT enterotoxin. As shown in Table 2, two strains, E133 and E1679sc, contained both STp and LT. Four strains contained STh or STp, while seven strains contained LT. CFs were not found in eight strains of all isolates, including NCCP15731 and NCCP15733. NCCP15732 had STh and LT enterotoxin and harbored CS1 and CS3 of CFs.

### Phylogenetic analysis

Phylogenetic comparison of candidate genes implemented in SEED [43] showed that NCCP15731 and NCCP15733 were most close to *E. coli* O157:H7 str. 88.1467 (score: 531 and 530, respectively). Whole

genome phylogenetic analysis and MLST phylogenetic analysis were performed based on multiple sequence alignments of whole genomes, and seven MLST genes (*adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) of 15 *E. coli* isolates including NCCP15731, NCCP15732 and NCCP15733, respectively (Fig. 2). Whole genome phylogenetic tree showed that NCCP15732 and 14 *E. coli* O159 strains were clustered into two phylo-groups (A and B1) like previous study [18]. NCCP15731 and NCCP15733 in this study belonged to phylo-group A and were clustered with *E. coli* O159:H34 str. E159 with sequence type of ST218 (Fig. 2a). MLST phylogenetic analysis also showed that NCCP15731 and NCCP15733 belonged to phylo-group A (Fig. 2b). *E. coli* E159 strain was also placed with NCCP15731 and NCCP15733 in MLST phylogenetic tree.

### Virulence factors

The acquisition of virulence factors has been suggested to be a major driving force for ETEC pathogenicity [44, 45]. 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 14 O159 strains and NCCP15732. We identified 222 virulence factors grouped into 28 categories and 74 sub-categories (Additional file 1: Table S1).

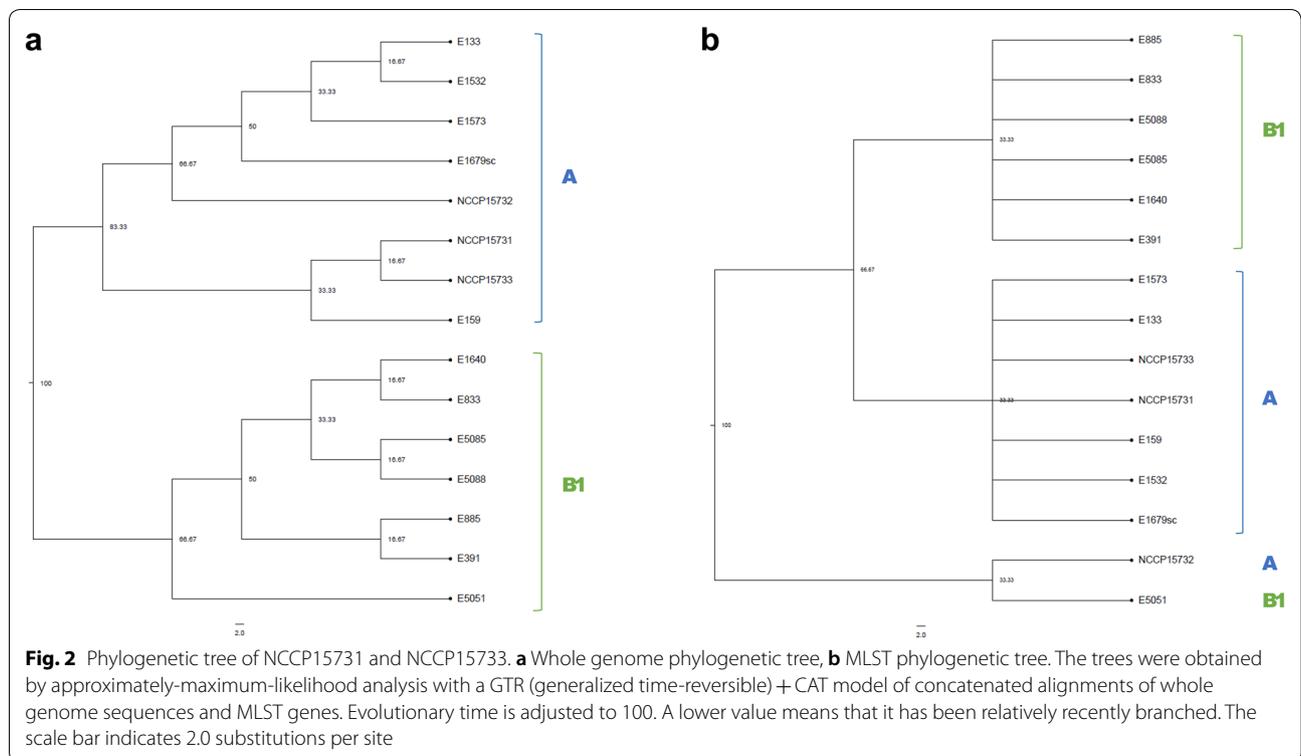
**Table 1 Genomic and phylogenetic characteristics of NCCP15731 and NCCP15733**

Strain	NCCP15731	NCCP15733
Genome size	4,663,459	4,645,336
% GC	50.7	50.6
N50	9457	14,126
L50	147	105
Total contigs	986	694
Total subsystems	587	588
Total CDS	4435	4369
tRNAs	19	23
Serotype	O159:H34	O159:H34
Phylo-group	A	A

**Table 2 Genomic features of whole genome datasets of Enterotoxin *Escherichia coli* strains used in this study**

Strain	Genome (Mb)	Phylo-group	MLST type	Serotype	Toxin type	CF type <sup>1-2</sup>	Accession number
<i>E. coli</i> NCCP15731	4.66	A	218	O159:H34	STh	CF NEG	QICG00000000
<i>E. coli</i> NCCP15733	4.64	A	218	O159:H34	STh	CF NEG	QICF00000000
<i>E. coli</i> E159	4.95	A	218	O159:H34	LT	CF NEG	ERS077713
<i>E. coli</i> E133	4.89	A	10	O159:H4	LT+STp	CS12	ERS044470
<i>E. coli</i> E1532	4.83	A	10	O159:H4	LT	CF NEG	ERS077694
<i>E. coli</i> E1573	4.95	A	10	O159:H4	LT	CF NEG	ERS077704
<i>E. coli</i> E1679sc	4.95	A	10	O159:H4	LT+STp	CS12	ERS163315
<i>E. coli</i> E391	4.91	B1	641	O159:H21	LT	CF NEG	ERS077592
<i>E. coli</i> E833	5.09	B1	641	O159:H34	LT	CS28a	ERS077628
<i>E. coli</i> E885	4.96	B1	641	O159:H21	LT	CF NEG	ERS077638
<i>E. coli</i> E1640	5.01	B1	641	O159:H34	LT	CF NEG	ERS077734
<i>E. coli</i> E5085	5.06	B1	1490	O159:H34	STp	CS6	ERS055670
<i>E. coli</i> E5088	5.07	B1	1490	O159:H34	STp	CS6	ERS055673
<i>E. coli</i> E5051	5.25	B1	295	O159:H18	LT	CS6	ERS055664
<i>E. coli</i> NCCP15732	4.80	A	4	O6:H16	LT+STh	CS1+CS3	–

All strains except NCCP15731 and NCCP15733 were referred to this paper [17]



Among O159 strains, a total of 125 virulence genes 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 of 14 O159 strains was constructed using R (version 3.4.3) (Fig. 3) [46]. 14 O159 strains of phylo-group A and B1 shown in Fig. 2a again completely clustered each other, respectively. These results suggested that the virulence factors were related with the phylo-group [23, 44, 47–50].

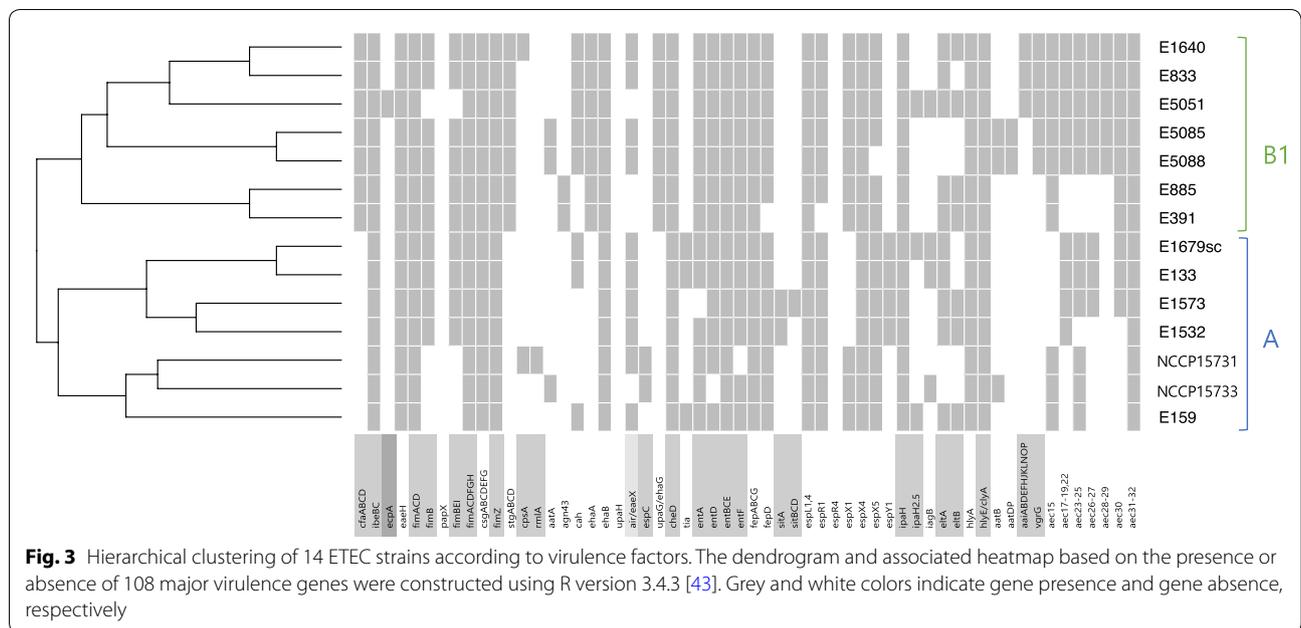
ECP (*E. coli* Common Pilus), *EaeH*, Type I fimbriae, Curli fimbriae, Stg fimbriae, AIDA-I type and *UpaG*, which belong to 13 potentially functional systems in adherence and auto transporter categories, are known to be produced by pathogenic *E. coli* [51]. These functional systems suggest that the ability of ETEC to attach to the host surface is the most important step in successful colonization [52]. Compared with NCCP15732 strain, all of O159 strains had *EaeH* system in adherence category. A virulence factor, *espC*, was only identified in NCCP5731 and NCCP15733. Conversely, *ehaA* auto transporter gene

and stg fimbriae (*stgABCD*) systems were only found in phylo-group B1. It is known that stg fimbriae (*stgABCD*) systems contribute to the attachment of human epithelial cells. They are associated with phylogenetic group B1 [53].

Regarding toxins, enterobactin synthesis (*entABCE*) and ferric-enterobactin transport (*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 [54]. 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 [55]. However, none of these heat-labile genes were found in NCCP15731 or NCCP15733, while NCCP15732, another strain in Korea, had two heat-labile genes.

**Resistance factors**

All isolates in the study were analyzed for antimicrobial susceptibility. We identified 11 resistance factors of six phenotypes (Table 3). One resistance factor, *mdf(A)* of MLS (Macrolide Lincosamide and Streptogramin) phenotype was found in all of isolates. We identified 11 resistance factors of six phenotypes (Table 3). NCCP15731 had six resistance factors of six phenotypes, but NCCP15733 had only one resistance factor. In phylo-group A, except NCCP15731 and E1573, five strains had only a *mdf(A)* resistance factor



**Table 3 Genotypic and phenotypic antimicrobial resistance factors of Enterotoxin *Escherichia coli* strains used in this study**

Strain	Aminoglycoside			Beta-lactam	MLS		Sulphonamide		Tetracycline	Trimethoprim	
	<i>aadA1</i>	<i>aph(6)-I<sub>d</sub></i>	<i>aph(3'')-I<sub>b</sub></i>	<i>blaTEM-1B</i>	<i>mdf(A)</i>	<i>mph(A)</i>	<i>sul1</i>	<i>sul2</i>	<i>tet(A)</i>	<i>dfrA1</i>	<i>dfrA8</i>
NCCP15731	+	-	-	+	+	-	+	-	+	+	-
NCCP15733	-	-	-	-	+	-	-	-	-	-	-
E159	-	-	-	-	+	-	-	-	-	-	-
E1532	-	-	-	-	+	-	-	-	-	-	-
E1679sc	-	-	-	-	+	-	-	-	-	-	-
E133	-	-	-	-	+	-	-	-	-	-	-
E1573	-	+	+	+	+	-	-	+	-	-	+
NCCP15732	-	-	-	-	+	-	-	-	-	-	-
E5051	+	-	-	+	+	-	+	+	+	+	-
E5085	-	-	-	+	+	+	-	-	-	-	-
E5088	-	-	-	+	+	+	-	-	-	-	-
E391	-	+	+	+	+	-	-	+	-	-	+
E833	-	+	+	+	+	-	-	+	+	-	+
E1640	-	-	-	-	+	-	-	-	-	-	-
E885	-	+	+	+	+	-	-	+	+	-	+

of MLS phenotype, while seven strains in phylo-group B had at least three to a maximum of seven resistance factors. NCCP15732 of O6 serotype had only a *mdf(A)* resistance factor of MLS phenotype and any of resistance factors of this strain were not found.

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

To calculate Average Nucleotide identity (ANI), five genomes of O159 of phylo-group A including

NCCP15731 and NCCP15733 were compared with each other using ANIu calculator [56]. *E. coli* E1573, and E1679sc were excluded due to the presence of strains with the same MLST type and toxin type. Results indicated that NCCP15731 was most similar to NCCP15733, with OrthoANIu value of 99.97% (Table 4). Both NCCP15731 and NCCP15733 were most similar to *E. coli* strain E159 among reference

strains in phylo-group A, with OrthoANIu values of 99.90% and 99.91%, respectively.

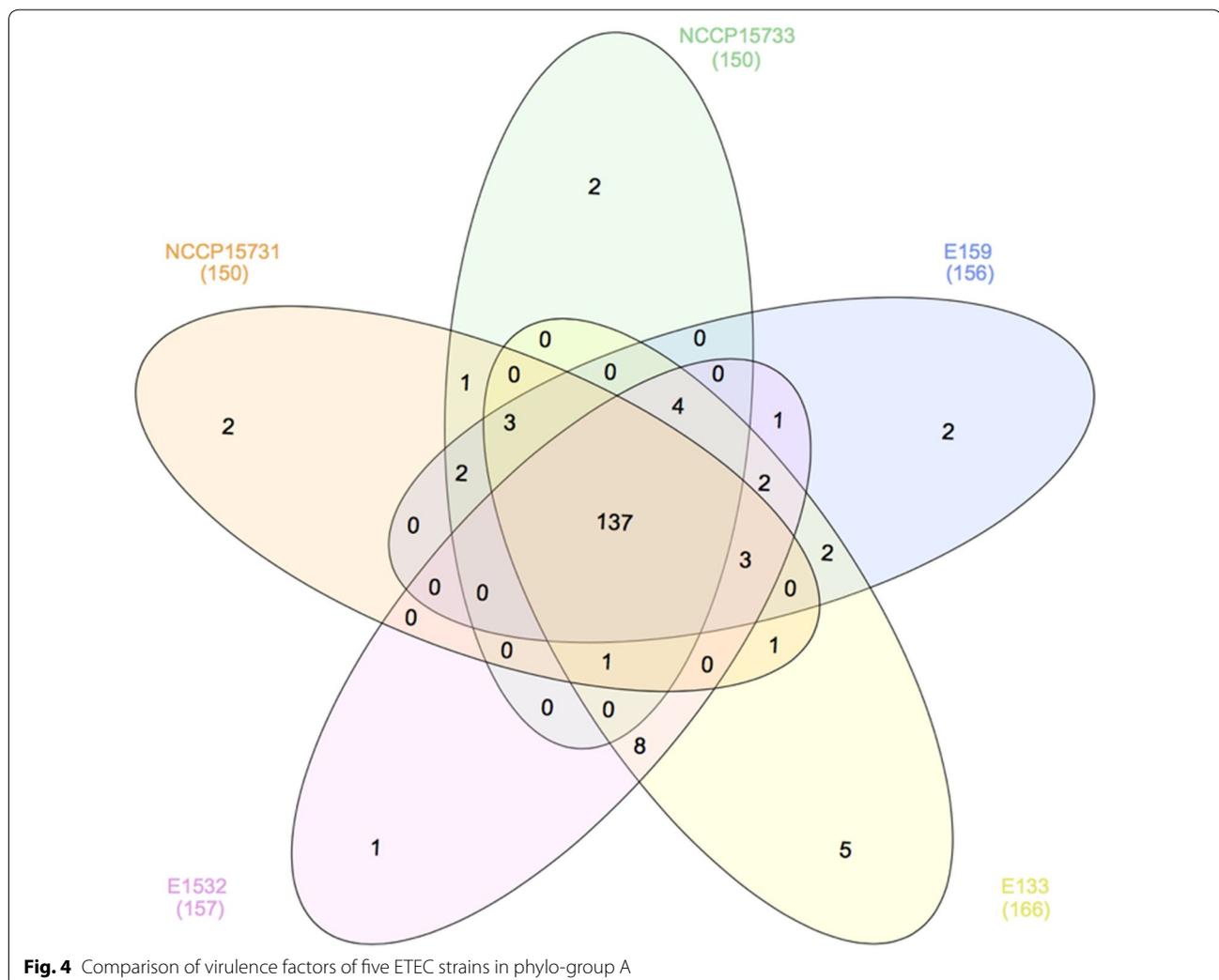
Venn diagram of the virulence factors of the five strains including NCCP15731 and NCCP15733 in phylo-group A was obtained using InteractiVenn [57]. They shared 137 genes encoding *E. coli* common pilus (ECP) proteins,

flagella-biosynthetic proteins, enterobactin transport proteins, and proteins related to Type II secretion system (Fig. 4). Comparison of NCCP15731, NCCP15733, and E159 (the most recent common ancestor of NCCP15731 and NCCP15733) showed that these three strains shared 142 genes. Plus, NCCP15731 and NCCP15733 shared

**Table 4 Average nucleotide identity values based on USEARCH**

	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)	–

Numbers indicate OrthoANIu value (%) and numbers in parentheses indicate the genome coverage (%)



*espC* gene related to 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 to secretion system. Compared to NCCP15731 and E159, NCCP15733 had two unique virulence genes: *cpsA* and *rmlA* encoding capsular polysaccharide. Moreover, 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.

## Conclusions

*Escherichia coli* NCCP15731 and NCCP15733, previously identified as MLST types ST964 and ST656, respectively, were isolated from diarrheal patients. However, multi-locus sequence typing (MLST) profiles showed that the MLST type of each of two strains was ST218. NCCP15731 and NCCP15733 belonged to phylo-group A and their serotype was O159:H34. In both whole genome and MLST phylogenetic analyses, NCCP15731 and NCCP15733 also belonged to phylo-group A. Hierarchical clustering based on the presence or absence of major virulence factors suggest that the virulence factors are associated with the phylogenetic group. In comparison of virulence genes of 14 strains, NCCP15733 has unique genes related to capsular polysaccharide. NCCP15731 has no unique virulence gene. However, in comparison with NCCP15731 and E159, it showed differences in the auto transporter, ABC transporter, and secretion system. Genomic analysis of NCCP15731 and NCCP15733 will be useful for further study on the development of ETEC vaccines.

## Future directions

In summary, serotype and MLST type of NCCP15731 and NCCP15733 were O159:H34 and ST218, respectively. Unlike other O159 strains, CF gene of NCCP15731 and NCCP15733 was not detected. These strains have unique genes related to capsular polysaccharide, auto transporter, and secretion system. Moreover, both strains do not contain LT genes. These results will improve our understanding of ETEC O159 strains to prevent ETEC disease. However, because the results were obtained from in silico analysis, experimental confirmation of these results is required.

## Additional file

**Additional file 1.** List of virulence factors of 16 ETEC O159 strains.

## Abbreviations

ETEC: enterotoxigenic *Escherichia coli*; LT: heat-labile toxin; STh: heat-stable toxin; MLST: multi-locus sequence typing; RAST: Rapid Annotation using Subsystem Technology; CDS: coding DNA sequences; CF: colonization factor; NCCP: National Culture Collection for Pathogens.

## Authors' contributions

SHC and WK planned and directed the project, and interpreted the results. SHC drafted the manuscript. SYC and TK performed the gene annotation, comparative genomic analysis and wrote the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> School of Biological Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. <sup>2</sup> Cloud9, 133, Yeonje-gil, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28164, Republic of Korea. <sup>3</sup> Department of Emergency Medical Services, Sun Moon University, Asan-si, Chungcheongnam-do 31460, Republic of Korea. <sup>4</sup> Department of Emergency Medical Service, Pohang University, Heunghae-eup, Sindeok-ro, Pohang-si, Gyeongsangbuk-do, Republic of Korea. <sup>5</sup> Glycobiology Unit, Department of Biological Science, Sungkyunkwan University and Samsung Advanced Institute for Health Sciences and Technology (SAIHST), 2066 Seobu-ro, Suwon 16419, Republic of Korea. <sup>6</sup> Division of Bacterial Disease Research, Center for Infectious Disease Research, Korea National Institute of Health, Cheongju 363-951, Republic of Korea.

## Acknowledgements

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

Nucleotide sequence accession numbers: The whole-genome shotgun sequencing data for NCCP15731 and NCCP15733 have been deposited in DDBJ/EMBL/GenBank under the accession number QICG000000000 and QICF000000000, respectively.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

This research has been reviewed and approved by the Institutional Review Board of the Korea Centers for Disease Control and Prevention (Reference No.: 2013-12-04-P).

Written informed consent was obtained from all patients with diarrhea to participate the research.

## Funding

This work was supported by a grant from the Marine Biotechnology Program (Genome Analysis of Marine Organisms and Development of Functional Applications) funded by the Ministry of Oceans and Fisheries.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 2 August 2018 Accepted: 11 February 2019

Published online: 21 February 2019

## References

- Gill SR, Pop M, DeBooy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006;312(5778):1355–9.
- Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *J Infect Dis*. 2000;181(5):1753–4.
- Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J Bacteriol*. 2009;191(1):347–54.
- Jeong H, Zhao F, Igori D, Oh K-H, Kim S-Y, Kang SG, Kim BK, Kwon S-K, Lee CH, Song JY. Genome sequence of the hemolytic-uremic syndrome-causing strain *Escherichia coli* NCCP15647. *J Bacteriol*. 2012;194(14):3747–8.
- Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2004;2(2):123.
- Kim BK, Song GC, Hong GH, Seong W-K, Kim S-Y, Jeong H, Kang SG, Kwon S-K, Lee CH, Song JY. Genome sequence of the Shiga toxin-producing *Escherichia coli* strain NCCP15657. *J Bacteriol*. 2012;194(14):3751–2.
- Leonard SR, Lacher DW, Lampel KA. Draft genome sequences of the enteroinvasive *Escherichia coli* strains M4163 and 4608-58. *Genome Announcements*. 2015;3(1):e01395.
- Song JY, Yoo RH, Jang SY, Seong W-K, Kim S-Y, Jeong H, Kang SG, Kim BK, Kwon S-K, Lee CH. Genome sequence of enterohemorrhagic *Escherichia coli* NCCP15658. *J Bacteriol*. 2012;194(14):3749–50.
- Organization WH. Future directions for research on enterotoxigenic *Escherichia coli* vaccines for developing countries—Orientations futures de la recherche sur les vaccins contre *Escherichia coli* entérotogène destinés aux pays en développement. *Wkly Epidemiol Record*. 2006;81(11):97–104.
- Wennerås C, Erling V. Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world. *J Health Populat Nutr*. 2004;1:370–82.
- Shin J, Yoon K-B, Jeon D-Y, Oh S-S, Oh K-H, Chung GT, Kim SW, Cho S-H. Consecutive outbreaks of enterotoxigenic *Escherichia coli* O6 in schools in South Korea caused by contamination of fermented vegetable kimchi. *Foodborne Pathogens Dis*. 2016;13(10):535–43.
- Gómez-Aldapa CA, Rangel-Vargas E, Bautista-De León H, Vázquez-Barrios ME, Gordillo-Martínez AJ, Castro-Rosas J. 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. *Int J Food Microbiol*. 2013;166(3):364–8.
- Qadri F, Svennerholm A-M, Faruque A, Sack RB. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev*. 2005;18(3):465–83.
- Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev*. 1998;11(1):142–201.
- Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. *Microbes Infect*. 2010;12(2):89–98.
- Bölin I, Wiklund G, Qadri F, Torres O, Bourgeois AL, Savarino S, Svennerholm A-M. Enterotoxigenic *Escherichia coli* with STh and STp genotypes is associated with diarrhea both in children in areas of endemicity and in travelers. *J Clin Microbiol*. 2006;44(11):3872–7.
- Gaastra W, Svennerholm A-M. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol*. 1996;4(11):444–52.
- von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, Joffe E, Corander J, Pickard D. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet*. 2014;46(12):1321.
- Svennerholm A-M, Lundgren A. Recent progress toward an enterotoxigenic *Escherichia coli* vaccine. *Expert Rev Vaccines*. 2012;11(4):495–507.
- Rivera F, Ochoa T, Maves R, Bernal M, Medina A, Meza R, Barletta F, Mercado E, Ecker L, Gil A. Genotypic and phenotypic characterization of enterotoxigenic *Escherichia coli* strains isolated from Peruvian children. *J Clin Microbiol*. 2010;48(9):3198–203.
- Isidean S, Riddle M, Savarino S, Porter C. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine*. 2011;29(37):6167–78.
- Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep*. 2013;5(1):58–65.
- Escobar-Páramo P, Clermont O, Blanc-Potard A-B, Bui H, Le Bouguéneq C, Denamur E. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol Biol Evol*. 2004;21(6):1085–94.
- Badouei MA, Jajarmi M, Mirsalehian A. Virulence profiling and genetic relatedness of Shiga toxin-producing *Escherichia coli* isolated from humans and ruminants. *Comp Immunol Microbiol Infect Dis*. 2015;38:15–20.
- Koo H-J, Kwak H-S, Yoon S-H, Woo G-J. Phylogenetic group distribution and prevalence of virulence genes in *Escherichia coli* isolates from food samples in South Korea. *World J Microbiol Biotechnol*. 2012;28(4):1813–6.
- Oh K-H, Kim DW, Jung S-M, Cho S-H. Molecular characterization of enterotoxigenic *Escherichia coli* strains isolated from diarrheal patients in Korea during 2003–2011. *PLoS ONE*. 2014;9(5):e96896.
- Kwon T, Chung S-Y, Jung Y-H, Jung S-J, Roh S-G, Park J-S, Kim C-H, Kim W, Bak Y-S, Cho S-H. Comparative genomic analysis and characteristics of NCCP15740, the major type of enterotoxigenic *Escherichia coli* in Korea. *Gut Pathogens*. 2017;9(1):55.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20(2):265–72.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9(1):75.
- Rice P, Longden I, Bleasby A. EMBOS: the European molecular biology open software suite. *New York: Elsevier Current Trends*; 2000.
- Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol*. 2015;2:00008–15.
- Rodas C, Iniguez V, Qadri F, Wiklund G, Svennerholm A-M, Sjöling Å. Development of multiplex PCR assays for detection of enterotoxigenic *Escherichia coli* colonization factors and toxins. *J Clin Microbiol*. 2009;47(4):1218–20.
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol*. 2006;60(5):1136–51.
- Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(Suppl\_2):W5–9.
- Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. 2005;33(Suppl\_1):D325–8.
- Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*. 2012;67(11):2640–4.
- Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*. 2010;27(3):334–42.
- Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: *Parallel and distributed processing symposium, 2006 IPDPS 2006 20th International*: 2006: IEEE; 2006.
- Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 2009;26(7):1641–50.
- Khan NH, Ahsan M, Yoshizawa S, Hosoya S, Yokota A, Kogure K. Multilocus sequence typing and phylogenetic analyses of *Pseudomonas aeruginosa* isolates from the ocean. *Appl Environ Microbiol*. 2008;74(20):6194–205.
- Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol*. 2015;38(4):237–45.
- Oeverbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res*. 2013;42(D1):D206–14.
- Turner SM, Chaudhuri RR, Jiang Z-D, DuPont H, Gyles C, Penn CW, Pallen MJ, Henderson IR. Phylogenetic comparisons reveal multiple acquisitions

- of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. *J Clin Microbiol.* 2006;44(12):4528–36.
45. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, Petty NK, Mahon V, Brinkley C, Hobman JL. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol.* 2010;192(21):5822–31.
  46. Team RC. R: A language and environment for statistical computing. 2013.
  47. Steinsland H, Lacher DW, Sommerfelt H, Whittam TS. Ancestral lineages of human enterotoxigenic *Escherichia coli*. *J Clin Microbiol.* 2010;48(8):2916–24.
  48. Steinsland H, Valentiner-Branth P, Aaby P, Mølbak K, Sommerfelt H. Clonal relatedness of enterotoxigenic *Escherichia coli* strains isolated from a cohort of young children in Guinea-Bissau. *J Clin Microbiol.* 2004;42(7):3100–7.
  49. Sahl JW, Rasko DA. Analysis of global transcriptional profiles of enterotoxigenic *Escherichia coli* isolate E24377A. *Infect Immun.* 2012;80(3):1232–42.
  50. Regua-Mangia AH, Guth BC, CostaAndrade RJ, Irino K, Pacheco ABF, Ferreira LCS, Zahner V, Teixeira LM. Genotypic and phenotypic characterization of enterotoxigenic *Escherichia coli* (EPEC) strains isolated in Rio de Janeiro city, Brazil. *FEMS Immunol Med Microbiol.* 2004;40(2):155–62.
  51. McWilliams B, Torres A. Enterohemorrhagic *Escherichia coli* adhesins. *Microbiol Spectrum.* 2014. <https://doi.org/10.1128/microbiolspec>.
  52. Antão E-M, Wieler LH, Ewers C. Adhesive threads of extraintestinal pathogenic *Escherichia coli*. *Gut Pathogens.* 2009;1(1):22.
  53. Lymberopoulos MH, Houle S, Daigle F, Léveillé S, Brée A, Moulin-Schouleur M, Johnson JR, Dozois CM. 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. *J Bacteriol.* 2006;188(18):6449–59.
  54. Burgos Y, Beutin L. Common origin of plasmid encoded alpha-hemolysin genes in *Escherichia coli*. *BMC Microbiol.* 2010;10(1):193.
  55. Lasaro M, Rodrigues J, Mathias-Santos C, Guth B, Balan A, Sbrogio-Almeida M, Ferreira L. Genetic diversity of heat-labile toxin expressed by enterotoxigenic *Escherichia coli* strains isolated from humans. *J Bacteriol.* 2008;190(7):2400–10.
  56. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek.* 2017;110(10):1281–6.
  57. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinform.* 2015;16(1):169.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

