

GENOME REPORT

Open Access



# Draft genome sequence of non-shiga toxin-producing *Escherichia coli* O157 NCCP15738

Taeso Kwon<sup>1,2</sup>, Jung-Beom Kim<sup>3</sup>, Young-Seok Bak<sup>4</sup>, Young-Bin Yu<sup>5</sup>, Ki Sung Kwon<sup>6</sup>, Won Kim<sup>1\*</sup> and Seung-Hak Cho<sup>7\*</sup>

## Abstract

**Background:** The non-shiga toxin-producing *Escherichia coli* (non-STEC) O157 is a pathogenic strain that cause diarrhea but does not cause hemolytic-uremic syndrome, or hemorrhagic colitis. Here, we present the 5-Mb draft genome sequence of non-STEC O157 NCCP15738, which was isolated from the feces of a Korean patient with diarrhea, and describe its features and the structural basis for its genome evolution.

**Results:** A total of 565-Mbp paired-end reads were generated using the Illumina-HiSeq 2000 platform. The reads were assembled into 135 scaffolds throughout the de novo assembly. The assembled genome size of NCCP15738 was 5,005,278 bp with an N50 value of 142,450 bp and 50.65 % G+C content. Using Rapid Annotation using Subsystem Technology analysis, we predicted 4780 ORFs and 31 RNA genes. The evolutionary tree was inferred from multiple sequence alignment of 45 *E. coli* species. The most closely related neighbor of NCCP15738 indicated by whole-genome phylogeny was *E. coli* UMNK88, but that indicated by multilocus sequence analysis was *E. coli* DH1(ME8569).

**Conclusions:** A comparison between the NCCP15738 genome and those of reference strains, *E. coli* K-12 substr. MG1655 and EHEC O157:H7 EDL933 by bioinformatics analyses revealed unique genes in NCCP15738 associated with lysis protein S, two-component signal transduction system, conjugation, the flagellum, nucleotide-binding proteins, and metal-ion binding proteins. Notably, NCCP15738 has a dual flagella system like that in *Vibrio parahaemolyticus*, *Aeromonas* spp., and *Rhodospirillum centenum*. The draft genome sequence and the results of bioinformatics analysis of NCCP15738 provide the basis for understanding the genomic evolution of this strain.

**Keywords:** Non-shiga toxin-producing *Escherichia coli* O157, Draft genome, Dual flagella

## Background

*Escherichia coli* is a gram-negative bacterium that colonizes the human gastrointestinal tract. Most *E. coli* serotypes are non-pathogenic but some serotypes cause food poisoning. *E. coli* strains are divided into three subgroups, according to their pathogenicity: nonpathogenic, pathogenic, and extra intestinal pathogenic *E. coli*. There are 190 serotypes [1] of *E. coli*, based on the major surface

antigens (O, H, and K) [2]. The serotype O157:H7 is the major strain in the enterohemorrhagic *E. coli* (EHEC) serotype and since 1982, these strains have been found to be important food-borne pathogens [3]. This type of *E. coli* can cause hemorrhagic colitis and hemolytic uremic syndrome (HUS) [4, 5]. O157:H7 can be identified by a combination of biochemical and immunological markers, such as sorbitol [6] in combination with O antigen [7]. *E. coli* O157:H7 is characterized by the expression of shiga-like toxins even though it produces various other virulence factors [8–10]. Shiga toxins are classified into two major groups, Stx1 and Stx2, which are encoded on a prophage [11]. These genes can be transferred horizontally to *E. coli* and other *Enterobacteriaceae* species [12],

\*Correspondence: wonkim@plaza.snu.ac.kr; skcho38@korea.kr

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

<sup>7</sup> Division of Enteric Diseases, Center for Infectious Diseases, Korea National Institute of Health, Cheongju 363-951, Republic of Korea  
Full list of author information is available at the end of the article

allowing transformation of shiga-like toxin non-producing strains into shiga-like toxin-producing strains [13]. The capillary endothelial cells are the major target sites of shiga toxins released by shiga toxin-producing *E. coli* (STEC). Specifically, the shiga toxins target the globotriaosylceramide receptor on the cells and are transported into the cells by receptor-mediated endocytosis [14]. Shiga toxins halt protein synthesis by cleaving an adenine base from the ribosomes of the intruded cells [15]. This blockage can cause kidney failure, as in HUS [14].

Although serotype STEC O157 strains are prevalent, non-STEC O157 strains have also been reported in children with diarrhea [16]. Very little is known about the symbiosis and pathogenicity of non-STEC O157 strains in the host; therefore, their genomes should be sequenced to assess horizontal gene transfer (HGT) and to understand the evolution of these strains. In this study, we performed genomic sequencing to investigate the genetic background of the evolution of non-STEC O157 NCCP15738 isolated from a patient with diarrhea. We also performed genomic comparison between the genomes of NCCP15738 and two reference strains, *E. coli* K-12 substr. MG1655 [17] and EHEC O157:H7 str. EDL933 [18] to study their evolution and phylogenetic linkage.

## Methods

### Strain, isolation, and serotyping

A fecal sample from a patient with diarrhea was plated on MacConkey agar directly or, occasionally, after enrichment in trypticase soy broth containing vancomycin (Sigma Chemicals Co., St. Louis, MO). Candidate colonies were then plated on trypticase soy agar medium and biochemically characterized using the API20E system (Biomérieux, Marcy l'Etoile, France). For O-antigen determination, we used the method described by Guinee et al. [7] and all available O (O1–O181) antisera. All antisera were absorbed with the corresponding cross-reacting antigens to remove non-specific agglutinins. The O antisera were produced at Laboratorio de Referencia de *E. coli* (Lugo, Spain [<http://www.lugo.usc.es/ecoli>]). This research was approved by the Research Ethics Committee of the Korea Centers for Disease Control and Prevention, and written informed consent was obtained from the patient. The isolated strain was deposited at the National Culture Collection for Pathogens (NCCP) at Korea National Institute of Health under the accession number NCCP15738. *E. coli* K-12 substr. MG1655 and EHEC O157:H7 str. EDL933 were used as the reference strains because these strains represent non-STECs and STECs, respectively.

### Library preparation and whole-genome sequencing

The genomic DNA of NCCP15738 was purified and fragmented randomly. After fragmentation, the overhangs were converted into blunt ends using T4 DNA polymerase, Klenow Fragment, and T4 Polynucleotide Kinase (New England Biolabs, MA, USA). Sequencing adapters were ligated to the ends of the end-repaired DNA fragments. The DNA fragments that met the required length were retained by gel electrophoresis and amplified by PCR. We used the Illumina-HiSeq 2000 (Illumina, San Diego, CA, USA) platform for whole-genome sequencing and produced 565,810,000 bp data with paired end reads of 90-bp length and 500-bp insert size.

### Genome assembly and annotation

For quality control of the sequencing data, the following steps were employed. First, reads with more than 9 % Ns' bases or low complexity reads were discarded. Second, reads with more than 40 bases of low quality ( $\leq Q20$ ) were discarded. Third, adapter sequences with at least 15 bp overlap between adapter and reads that allowed 3 bp mismatches were removed. Fourth, duplicated reads were discarded. After quality control removals, we obtained 504 Mbp of high quality reads. SOAPdenovo (version 1.05) [19] was used for de novo assembly of the genome using the high quality reads. For the purpose of assembly correction, all reads that passed the quality control were aligned against the assembly result using SOAPaligner (version 2.21) [20]. The single base error of the assembly result was corrected using mapping information. Scaffolds over 500 bp in length were considered for downstream analysis. To predict open reading frames (ORFs) and annotate the ORFs, we used the RAST (Rapid Annotation using Subsystem Technology, version 4.0) [21] server pipeline. We compared the predicted CDSs (coding DNA sequences) of NCCP15738 to those of two *E. coli* strains, K-12 substr. MG1655 and *E. coli* O157:H7 str. EDL933, using OrthoMCL software (version 2.0.9) [22]. Orthologous protein sequences were clustered into groups and the orthologous proteins of all three *E. coli* strains in each group were counted. To identify the virulence factor genes in NCCP15738, we performed a BLAST (Basic Local Alignment Search Tool) search of whole NCCP15738 ORFs against the virulence factor genes listed in VFDB [23] with an e-value of  $1e-5$ . Insertion sequences (ISs) were identified by mapping to a sequence database that was downloaded from IS Finder DB (<http://www-is.biotoul.fr>), using RepeatMasker (version 4.0.1) (<http://www.repeatmasker.org>). Phage-associated gene clusters in the scaffolds of NCCP15738 were searched using the PHAST server [24] (data not shown).

### Phylogenetic analysis and comparative genomic analysis

To infer the evolutionary history of NCCP15738, we performed a multiple sequence alignment of the whole genome using Mugsy (version 1.2.3) [25] and approximately-maximum-likelihood phylogenetic trees were inferred using FastTree (version 2.1.7) [26] with a GTR (generalized time-reversible) + CAT model [27]. The tree was visualized using FigTree (version 1.3.1) (<http://tree.bio.ed.ac.uk/software/figtree/>). In order to exclude the effect of HGT in our phylogenetic analysis, we used the multilocus sequence analysis (MLSA) method [28, 29]. Seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*) from 45 *E. coli* strains were retrieved and concatenated. A phylogenetic tree of multi locus sequence typing (MLST) genes was created using the method employed for whole-genome phylogenetic analysis. Mauve (version 2.3.1) [30] was used for comparative genomics using the Move Contig tool. The scaffolds were reordered against the complete genome of the reference *E. coli* strain, K-12 substr. MG1655. From the comparative genomic study, we identified a syntenic region that aligned against the reference genome. Unaligned scaffolds against the reference genome were defined as unique regions of NCCP15738. We also used the progressive alignment algorithm of Mauve for comparative alignment of NCCP15738, *E. coli* strain K-12 substr. MG1655, and *E. coli* strain O157:H7 str. EDL933 genomes. The BLAST algorithm was used to identify syntenic genes between the species and to analyze the genes of interest.

### Quality assurance

The genomic DNA was purified from a pure culture of a single bacterial isolate of NCCP15738. Potential contamination of the genomic library by other microorganisms was assessed using a BLAST search against the non-redundant database. We also checked for contamination by other genomes by confirming coverage distribution.

## Results and discussion

### General features

Whole-genome sequencing by Illumina-HiSeq 2000 showed 565,810,000 bp with paired end reads that were 90 bp in length. After quality control, 504 Mbp of high quality reads were kept for assembly. The average sequencing depth was 86.7-fold coverage and the coverage ratio was 84.77 %. The high quality reads were assembled into 135 scaffolds by de novo assembly with an N50 value of 142,450 bp. The predicted genome size of NCCP15738 was 5,005,278 bp with 50.65 % G+C content. RAST analysis identified 4780 putative ORFs and 31 RNA genes, of which 4181 (80.6 %) could be functionally annotated (Fig. 1). The monosaccharides (212 ORFs) and central carbohydrate metabolism (135 ORFs) subsystems

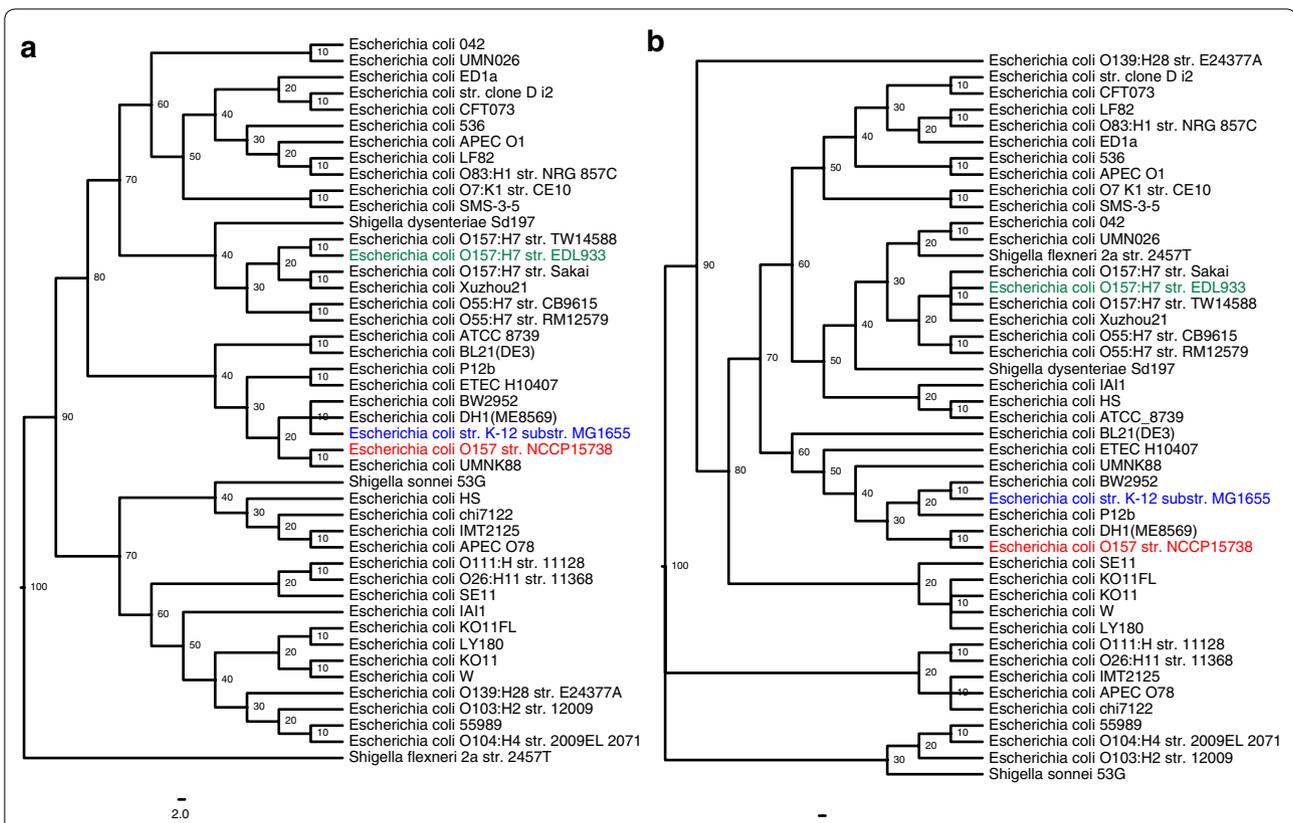
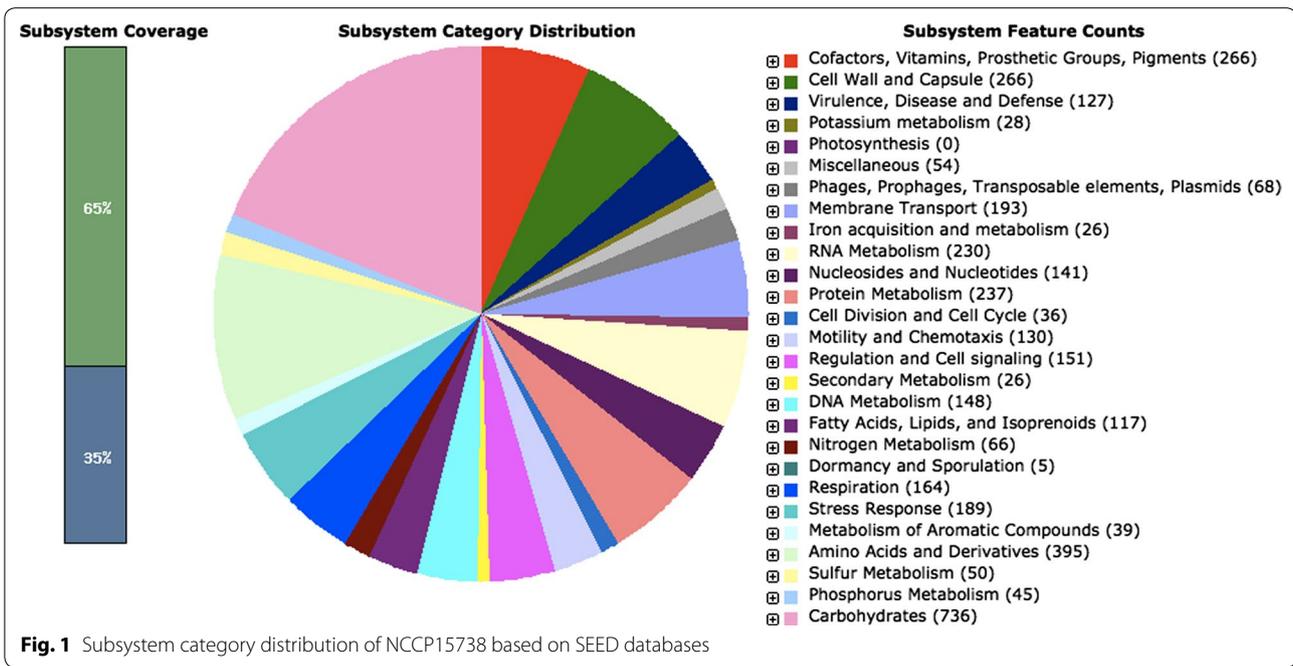
were significantly abundant among the subsystems (18.7 %). According to the subsystem results, we can assume that the NCCP15738 developed systems that can utilize various monosaccharides in addition to glucose to adapt to an extreme environment. A large number of ORFs were also associated with the “Amino acids and derivatives” subsystem (395 ORFs), “cofactors, vitamins, prosthetic groups, pigments” subsystem (266 ORFs) and “cell wall and capsule” subsystem (266 ORFs).

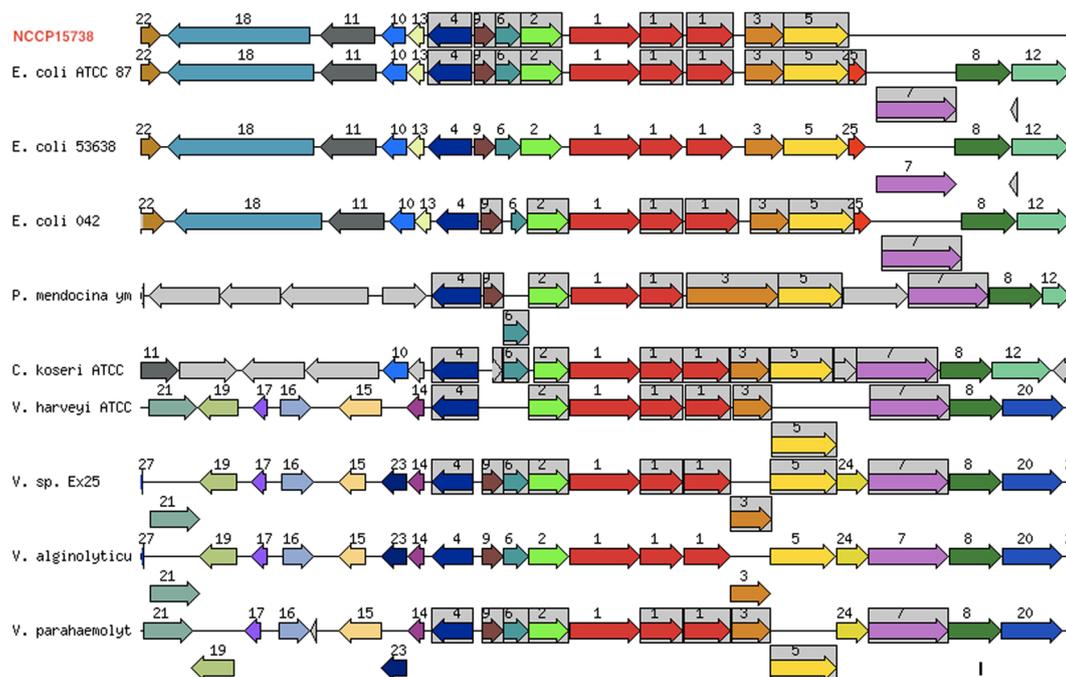
### Comparative genomics of NCCP15738 with other *E. coli* strains

The phylogenetic comparison of gene candidates predicted by SEED [31] revealed *E. coli* O104:H4 GOS1 [32] as the closest neighbor of NCCP15738 (score 513). To investigate the detailed evolutionary history of NCCP15738, we performed a multiple sequence alignment of 45 *E. coli* species including NCCP15738 (Fig. 2, Additional file 1: Table S1). Whole-genome phylogenetic analysis revealed that NCCP15738 did not cluster with *E. coli* strain K-12 substr. MG1655 into a single clade. Moreover, NCCP15738 was not grouped with *E. coli* O157:H7 str. EDL933. The most closely related neighbor of NCCP15738 was the pathogenic *E. coli* UMNK88 [33]. In MLSA, NCCP15738 clustered with *E. coli* DH1 (ME8569) into a single clade. The *E. coli* UMNK88 strain and K-12 substr. MG1655 were farther from NCCP15738 in the MLSA tree than in the whole-genome phylogenetic tree. However, this difference between the whole-genome phylogenetic tree and the MLST phylogenetic tree was not significant, as there was consensus in the topology among trees. It is concordant with previous research with Phylomark [34].

### Comparison of functional genes

A comparison of NCCP15738 genes and those of the two reference strains, *E. coli* K-12 substr. MG1655 and *E. coli* O157:H7 str. EDL933 showed that most of the functional genes of NCCP15738 were conserved in the two reference strains, but 941 genes were unique (Additional file 2: Table S2). Unique genes in NCCP15738 included those encoding lysis protein S, a two-component signal transduction system, conjugation, the flagellum, nucleotide-binding proteins, and metal ion binding proteins, explain the phenotypic differences that result from environmental adaptation. In particular, NCCP15738 has a dual flagella system used for swarming in viscous media. This system resembles those found in *Vibrio parahaemolyticus*, *Aeromonas* spp., and *Rhodospirillum centenum* [35]. Sixty-five genes encoded the flagellar biosynthesis protein or the flagellar structural protein. Seven of the flagella-related proteins (1–6, and 9) were highly conserved in *V. parahaemolyticus* and in nine other strains (Fig. 3,





**Fig. 3** Comparative map of lateral flagella in NCCP15738 genome and other closely related species. Nine genes were highly conserved in 11 strains, but only seven genes (1–6, and 9) were conserved in NCCP15738. Numbers indicate genes encoding for the following proteins: flagellar hook protein FlgE, flagellar basal-body rod protein FlgF and FlgG from left to right (1); FlgD (2); FlgH (3); FlgA (4); FlgI (5); FlgC (6); FlgB (7); FlgK (8); FlgL (9); hypothetical protein (10); hypothetical protein (11); lysine-N-methylase (EC 2.1.1.-) (12); hypothetical protein (13); LfgM (14); membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) (15); MutT/nudix family protein (16); hemolysin (17); Cps2A (18); hypothetical protein (19); putative flagellin (20); hypothetical protein (21); glycerol-3-phosphate cytidyltransferase (EC 2.7.7.39) (22); LfgN (23); FlgJ (24 and 25). Gray background boxes indicate that the genes in the relative position are conserved in at least four species. The comparative map was created with the genome browser of the SEED viewer (version 2.0)

Additional file 3: Table S3). Lateral flagella have no effect on pathogenicity, but the polar flagellum is important in the pathogenesis of *V. parahaemolyticus* [36]. Therefore, we can suppose that the polar flagellum of NCCP15738 is the major machinery for swarming and has a pathogenic effect. In contrast, the lateral flagellum of NCCP15738 is likely related only to locomotion in this strain.

### Virulence factors

Even though NCCP15738 belongs to serotype O157, it causes diarrhea but not HUS in human hosts. Because of this, we were particularly interested in identifying the potential virulence factors within the genome of NCCP15738. The features that we identified through sequence analysis are detailed in Table 1, which includes a variety of pilus and fimbriae genes and their associated operons. However, NCCP15738 produces no shiga toxins, such as Stx1 (*stx1A*, *stx1B*) or Stx2 (*stx2A*, *stx2B*), and has only one locus of enterocyte effacement (LEE) that encodes type three secretion system (TTSS) (*escR*)

[18]. From our comparison of NCCP15738 with *E. coli* K-12 substr. MG1655 and *E. coli* O157:H7 str. EDL933, we found that NCCP15738 has only one unique virulence gene, *papD*. NCCP15738 has 19 virulence genes and 18 of these genes had been previously reported in the other two strains.

### Future directions

This study shows a broad comparative genomics approach to the study of the NCCP15738 genome and describes the features of this type of non-STE C O157. This information will be useful for studying the evolution of the pathogenic mechanisms in this strain and its adaptation to the environment.

### Availability of supporting data

Nucleotide sequence accession numbers: This Whole Genome Shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number ASHB00000000.

**Table 1 Virulence genes of NCCP15738**

Category	Subcategory	Genes	Gene name
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2959	ecpA
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2960	ecpB
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2961	ecpC
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2962	ecpD
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2963	ecpE
Adherence	<i>E. coli</i> common pilus (ECP)	fig 562.1020.peg.2958	ecpR
Adherence	F1C fimbriae	fig 562.1020.peg.2245	focC
Adherence	Type I fimbriae	fig 562.1020.peg.1416	fimA
Adherence	Type I fimbriae	fig 562.1020.peg.2244	fimA
Adherence	Type I fimbriae	fig 562.1020.peg.1417	fimB
Adherence	Type I fimbriae	fig 562.1020.peg.1414	fimC
Adherence	Type I fimbriae	fig 562.1020.peg.944	fimC
Adherence	Type I fimbriae	fig 562.1020.peg.1413	fimD
Adherence	Type I fimbriae	fig 562.1020.peg.2246	fimD
Adherence	Type I fimbriae	fig 562.1020.peg.3956	fimD
Adherence	Type I fimbriae	fig 562.1020.peg.4109	fimD
Adherence	Type I fimbriae	fig 562.1020.peg.1412	fimF
Adherence	Type I fimbriae	fig 562.1020.peg.2247	fimF
Adherence	Type I fimbriae	fig 562.1020.peg.1411	fimG
Adherence	Type I fimbriae	fig 562.1020.peg.2248	fimG
Adherence	Type I fimbriae	fig 562.1020.peg.1410	fimH
Adherence	Type I fimbriae	fig 562.1020.peg.2249	fimH
Adherence	Type I fimbriae	fig 562.1020.peg.1415	fimI
Auto transporter	Adhesin involved in diffuse adherence	fig 562.1020.peg.1313	aida
Iron uptake	Salmochelin siderophore	fig 562.1020.peg.1061	iroN
Secretion system	LEE locus encoded TTSS	fig 562.1020.peg.3365	escR

## Additional files

**Additional files 1: Table S1.** Bacterial strain list used for phylogenetic analysis.

**Additional files 2: Table S2.** Annotated genes of non-STEC O157 NCCP15738 using RAST server.

**Additional files 3: Table S3.** Lateral flagella genes of non-STEC O157 NCCP15738.

## Abbreviations

CDS: coding DNA sequences; EHEC: enterohemorrhagic *Escherichia coli*; GTR: generalized time-reversible; HGT: horizontal gene transfer; HUS: hemolytic-uremic syndrome; IS: insertion sequences; MLSA: multilocus sequence analysis; MLST: multi locus sequence typing; NCCP: National Culture Collection for Pathogens; ORFs: open reading frames; RAST: Rapid Annotation Using Subsystem Technology; STEC: shiga toxin-producing *Escherichia coli*; str.: strain; substr: substrain.

## Authors' contributions

SHC and WK planned and directed the project, and interpreted the results. SHC drafted the manuscript. KSK, YBY, and YSB interpreted the results. JBK characterized the strain and prepared the genomic DNA. 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> Division of Biosafety Evaluation and Control, Korea National Institute of Health, Cheongju 363-951, Republic of Korea. <sup>3</sup> Department of Food Science and Technology, Suncheon National University, Suncheon, Jeonnam 540-950, Republic of Korea. <sup>4</sup> Department of Emergency Medical Service, College of Medical Science, Konyang University, Daejeon 302-832, Republic of Korea. <sup>5</sup> Department of Biomedical Laboratory Science, College of Medical Science, Konyang University, Daejeon 302-832, Republic of Korea. <sup>6</sup> New Hazardous Substances Team, National Institute of Food and Drug Safety Evaluation, Cheongju 363-700, Republic of Korea. <sup>7</sup> Division of Enteric Diseases, Center for Infectious Diseases, Korea National Institute of Health, Cheongju 363-951, Republic of Korea.

## Acknowledgements

This work was supported by a Grant from Marine Biotechnology Program (Genome Analysis of Marine Organisms and Development of Functional Applications) funded by the Ministry of Oceans and Fisheries and the Korea National Institute of Health (NIH 4800-4845-300 and NIH 4800-4847-300 to S.H.C.).

## Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2016 Accepted: 4 March 2016

Published online: 19 April 2016

## References

- Stenutz R, Weintraub A, Widmalm G. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev.* 2006;30(3):382–403.
- Orskov I, Orskov F, Jann B, Jann K. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev.* 1977;41(3):667–710.
- Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, et al. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med.* 1983;308(12):681–5.
- Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev.* 1998;11(1):142–201.
- Caprioli A, Morabito S, Brugere H, Oswald E. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet Res.* 2005;36(3):289–311.
- Ratnam S, March SB, Ahmed R, Bezanson GS, Kasatiya S. Characterization of *Escherichia coli* serotype O157:H7. *J Clin Microbiol.* 1988;26(10):2006–12.
- Guinee PA, Agterberg CM, Jansen WH. *Escherichia coli* O antigen typing by means of a mechanized microtechnique. *Appl Microbiol.* 1972;24(1):127–31.
- Griffin PM, Tauxe RV. The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiol Rev.* 1991;13:60–98.
- Johannes L, Romer W. Shiga toxins—from cell biology to biomedical applications. *Nat Rev Microbiol.* 2010;8(2):105–16.
- Suh JK, Hovde CJ, Robertus JD. Shiga toxin attacks bacterial ribosomes as effectively as eucaryotic ribosomes. *Biochemistry.* 1998;37(26):9394–8.
- Friedman DI, Court DL. Bacteriophage lambda: alive and well and still doing its thing. *Curr Opin Microbiol.* 2001;4(2):201–7.
- Beutin L. Emerging enterohaemorrhagic *Escherichia coli*, causes and effects of the rise of a human pathogen. *J Vet Med B Infect Dis Vet Public Health.* 2006;53(7):299–305.
- O'Brien AD, Newland JW, Miller SF, Holmes RK, Smith HW, Formal SB. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science.* 1984;226(4675):694–6.
- Karmali MA. Infection by Shiga toxin-producing *Escherichia coli*: an overview. *Mol Biotechnol.* 2004;26(2):117–22.
- Sandvig K, Bergan J, Dyve AB, Skotland T, Torgersen ML. Endocytosis and retrograde transport of Shiga toxin. *Toxicon.* 2010;56(7):1181–5.
- Blank TE, Lacher DW, Scaletsky IC, Zhong H, Whittam TS, Donnenberg MS. Enteropathogenic *Escherichia coli* O157 strains from Brazil. *Emerg Infect Dis.* 2003;9(1):113–5.
- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997;277(5331):1453–62.
- Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature.* 2001;409(6819):529–33.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010;20(2):265–72.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics.* 2009;25(15):1966–7.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008;9:75.
- Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13(9):2178–89.
- 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(Database issue):D325–8.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res.* 2011;39(Web Server issue):W347–52.
- Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics.* 2011;27(3):334–42.
- 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.
- Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Parallel and Distributed Processing Symposium, 2006 IPDPS 2006 20th International 2006.
- 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, Kampfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol.* 2015;38(4):237–45.
- Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14(7):1394–403.
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014;42(Database issue):D206–14.
- Brzuszkiewicz E, Thurmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, Boelter J, Petersen H, Gottschalk G, Daniel R. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol.* 2011;193(12):883–91.
- Shepard SM, Danzeisen JL, Isaacson RE, Seemann T, Achtman M, Johnson TJ. Genome sequences and phylogenetic analysis of K88- and F18-positive porcine enterotoxigenic *Escherichia coli*. *J Bacteriol.* 2012;194(2):395–405.
- Sahl JW, Matalka MN, Rasko DA. Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. *Appl Environ Microbiol.* 2012;78(14):4884–92.
- McCarter LL. Dual flagellar systems enable motility under different circumstances. *J Mol Microbiol Biotechnol.* 2004;7(1–2):18–29.
- Lee H-G, Jeong B-G, Park K-S. Role of Dual Flagella in the Pathogenesis of *Vibrio parahaemolyticus*. *Fisheries Aqua Sci.* 2011;14(2):73–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

