GENOME REPORT

Whole-genome sequencing and comparative genomic analysis of *Escherichia coli* O91 strains isolated from symptomatic and asymptomatic human carriers

Taesoo Kwon[1†], Young-Seok Bak[2†], Young-Hee Jung[3], Young-Bin Yu[4], Jong Tae Choi[5], Cheorl-Ho Kim[6], Jung-Beom Kim[7], Won Kim[1‡] and Seung-Hak Cho[8*‡]

Abstract

Background: The Shiga toxin–producing *Escherichia coli* (STEC) O91:H21 strains NCCP15736 and NCCP15737 were isolated during a single outbreak in Korea, NCCP15736 from a symptomatic carrier and NCCP15737 from an asymptomatic carrier. To investigate genomic differences between the two strains, we performed whole-genome sequencing of both strains and conducted a comparative genomic analysis.

Results: Using the Illumina HiSeq 2000 platform and Rapid Annotation using the Subsystem Technology (RAST) server, whole-genome sequences of NCCP15736 and NCCP15737 were obtained and annotated. Phylogenetic analysis of ten *E. coli* strains showed that NCCP15736 and NCCP15737 are evolutionarily close. The two strains were found to be most close to *E. coli* O91:NM str. 2009C-3745. The genomic comparison showed that the *fimD* gene of NCCP15737 is truncated and that the truncation could underlie the defects in infection and pathogenicity of NCCP15737. The two strains showed the same virulence factor profiles, and we identified 25 virulence factors from NCCP15736 and NCCP15737, respectively. We identified ten and nine phage-associated regions in the NCCP15736 and NCCP15737 genomes, respectively; the two strains share five of these.

Conclusions: NCCP15736 and NCCP15737 differ at the genomic level, even though they share features such as virulence-related genes. NCCP15737 has a deletion in *fimD,* which may underlie its asymptomatic character. We conclude that complete genome sequencing and integration of other types of omics data are needed to fully reveal the mechanism underlying the asymptomatic character of NCCP15737.

Keywords: Shiga-like toxin-producing *Escherichia coli* O91, Draft genome, Type I fimbriae, *fimD*, Truncated protein

Background

Escherichia coli is a typical member of the normal microflora of the human gastrointestinal tract [1].

However, some *E. coli* isolates cause serious disease. They can be divided into three major subgroups: commensal or nonpathogenic strains, pathogenic strains that cause intestinal infection, and extraintestinal pathogenic strains [2]. Intestinal pathogenic *E. coli* include enteroaggregative *E. coli*, enterohemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli*, and enterotoxigenic *E. coli* (ETEC). Shiga toxin-producing *E. coli* (STEC) O157:H7 in humans

*Correspondence: skcho38@korea.kr
†Taesoo Kwon and Young-Seok Bak contributed equally to the work
‡Won Kim and Seung-Hak Cho contributed equally to the work
[8] 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

Kwon *et al. Gut Pathog* (2016) 8:57

Page 2 of 8

was first reported in 1983 [3–5]. STEC causes a variety of diarrheal diseases and hemolytic uremic syndrome (HUS) [6]. EHEC belongs to the STEC group but it is associated with a distinctive clinical syndrome, namely hemorrhagic colitis (HC), mainly caused by *E. coli* O157:H7 [7, 8]. Shiga toxin (Stx) inhibits protein synthesis by disrupting the 28S RNA of the 60S ribosomal subunit [9]. Shiga toxins can be classified into two groups: Stx1 and Stx2 [7]. Stx1 originates from *Shigella dysenteriae* and there are three subtypes: Stx1a, Stx1c and Stx1d; these genes are highly conserved in STECs. Stx2 shows a lower degree of conservation and includes several variants: Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2f, and Stx2g [10]. Most outbreaks involve STEC O157:H7, but outbreaks caused by non-O157 STEC have shown a recent increase [11]. Thus, a better understanding of the causes of the asymptomatic character of STEC strains is required. Non-O157 STEC includes the O8:H, O26:H, I26:H11, O91:H21, O103:H2, O111:H, O113:H21, O128:H2, and O145:H [7] serotypes.

Two STEC O91:H21 isolates were used in this study, one from a symptomatic carrier and one from an asymptomatic carrier, both isolated during a recent outbreak in Korea [12]. Molecular and cellular analyses to investigate differences in pathogenicity between the isolates were performed in a previous study. A reduced adherence phenotype and transcriptional repression of type I fimbriae genes were identified in the isolates from the asymptomatic carrier; these two factors may explain why the isolates cause no symptoms. However, the mechanism underlying the transcriptional repression of type I fimbriae is not yet understood at the genomic level. To investigate the differences between the O91:H21 isolates from symptomatic and asymptomatic carriers and to explore the genetic basis underlying these differences, whole-genome sequencing and comparative genomic analyses were performed.

Methods

Strain, isolation, and serotyping

An outbreak of STEC at an elementary school was reported in Gwangju, Korea on July 2004 [12]. A total of 1643 stool samples were obtained from asymptomatic individuals and all isolates were biochemically characterized using the API20E system (Biomerieux, Marcy l'Etoile, France). A total of 74 STEC isolates were characterized as positive for STEC but caused no symptoms. Apart from the isolates from asymptomatic carriers, one STEC isolate from a symptomatic carrier was characterized. The isolated strains were deposited in the National Culture Collection for Pathogens (NCCP) at the Korean National Institute of Health under accession numbers NCCP15736 and NCCP15737. For the present study,

NCCP15736 and NCCP15737 were obtained from the NCCP for whole-genome sequencing. This research has been reviewed and approved by the Institutional Review Board of the Korean Centers for Disease Control and Prevention.

Library preparation and whole-genome sequencing

A sequencing library was constructed using the TruSeq Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Genomic DNA was end repaired and ligated with paired-end sequencing adapters. DNA fragments with the desired length of ~500 bp were selected by gel electrophoresis. A sequencing library was produced by PCR amplification. The Illumina HiSeq 2000 platform was used for whole-genome sequencing.

Genome assembly and annotation

Low-complexity reads, reads with quality scores <Q20, adapter sequences, and duplicate reads were discarded. De novo assembly of high-quality reads was performed with SOAPdenovo (version 1.05) [13]. The de novo assembly results were corrected based on alignment of all reads that passed the quality control threshold against the assembly results using SOAPaligner (version 2.21) [14]. After correction, scaffolds >500 bp in length were considered for downstream analysis.

Open reading frames and annotated open reading frames were identified using the Rapid Annotation using Subsystem Technology (RAST version 4.0) [15] server pipeline. The coding sequences (CDSs) of NCCP15736 and NCCP15737 were compared using the sequence base comparison functionality of the RAST server. For comparison of type I fimbriae gene clusters between the two strains, the sequence base comparison functionality of the RAST server was also used. To investigate the virulence factor genes, a BLAST search of the total open reading frames (ORFs) of NCCP15736 and NCCP15737 against the virulence factor genes of *E. coli* listed in VFDB [16] was performed with an e-value threshold of 1e − 5. To select homologous virulence factor genes, the BLAST Score Ratio (BSR) was calculated and only genes with a BSR score ≥0.4 were used in further analyses. The BSR score was calculated using our in-house scripts. We excluded genes with coverage lower than 60%, even if they showed high sequence identity. Phage-associated gene clusters in the genome sequences of NCCP15736 and NCCP15737 were identified using the PHAST server [17]. Three scenarios for the completeness of the predicted phage-associated regions were defined according to how many genes/proteins of a known phage the region contained: intact (≥90%), questionable (90–60%), and incomplete (≤60%).

Kwon *et al. Gut Pathog* (2016) 8:57

Page 3 of 8

Phylogenetic analysis and genomic structure comparison

To infer the evolutionary relationships among *E. coli* O91, including NCCP15736 and NCCP15737, multiple sequence alignments of the whole genome were performed with Mugsy (version 1.2.3) [18]. The generalized time-reversible [19] + CAT model [20] was used to infer the structure of maximum-likelihood phylogenetic trees using FastTree (version 2.1.7) [21]. FigTree (version 1.3.1) (http://tree.bio.ed.ac.uk/software/figtree/) was employed for tree visualization. For comparison of genomic structures between the two strains, the progressive alignment algorithm in Mauve (version 2.3.1) [22] was used. The BLAST algorithm was used to compare phage-associated regions.

Quality assurance

The genomic DNA was purified from a pure culture of a single bacterial isolate of NCCP15736 and NCCP15737, respectively. 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

A total of 569,860,000 bp and 576,270,000 bp of paired-end reads were generated with the Illumina HiSeq 2000 platform from genomic DNA of NCCP15736 and NCCP15737. We used 517 Mbp and 477 Mbp of high-quality reads for assembly after quality control. After de novo assembly, a total of 151 scaffolds with a scaffold N50 of 133,815 bp were obtained for NCCP15736 and 156 scaffolds with a scaffold N50 of 140,358 bp were assembled for NCCP15737. The draft genome size of NCCP15736 was 5079,147 bp and that of NCCP15737 was 5,126,930 bp. The genomic features of NCCP15736 and NCCP15737 are summarized in Table 1. Based on a RAST analysis, 4823 putative CDSs and 14 tRNA genes were identified in the NCCP15736 genome. A total of 4924 putative CDSs and 26 RNAs were identified in the NCCP15737 genome (Fig. 1; Additional file 1: Table S1).

Table 1 Genomic features of NCCP15736 and NCCP15737 strains of *Escherichia coli*

Strain	NCCP15736	NCCP15737
Genome (Mb)	5.08	5.13
% GC	50.59	50.62
Total open reading frames	4823	4924
tRNAs	14	22
rRNAs	0	4

Comparison of genome structure

In the comparative analysis of genomic structure performed using the progressive alignment function of Mauve, we detected structures that were highly conserved between NCCP15736 and NCCP15737 (Fig. 2). Several unaligned scaffolds were also detected.
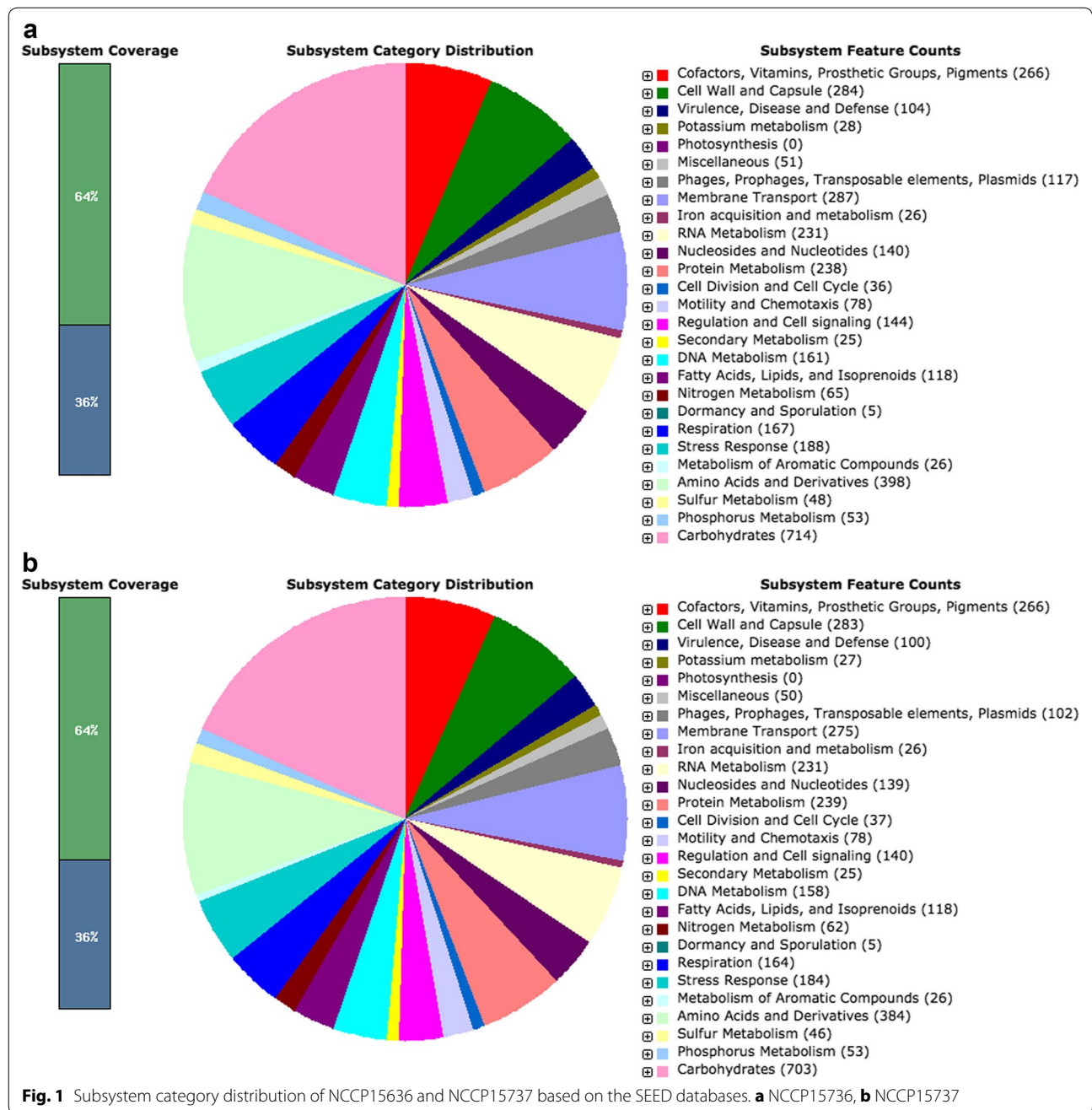
Phylogenetic analysis

Phylogenetic comparison of candidate genes, implemented in SEED [23], showed that NCCP15736 and NCCP15737 are most close to *E. coli* O104:H4 str. GOS1 (score 516). A whole-genome phylogenetic tree showed that NCCP15736 is close to NCCP15737 and that both strains are closer to *E. coli* O91:NM str. 2009C-3745 (Fig. 3).

Type I fimbriae operon

In a previous study, it was reported that the cell surface of NCCP15737 is completely bald; the lack of type I fimbriae was concluded to be the main cause of the asymptomatic character of NCCP15737 [12]. From the comparison of NCCP15736 and NCCP15737 using the sequence-based comparison functionality of the RAST server, we determined that *fimD* is truncated in NCCP15737, and its product is 591 amino acids instead of the full 852 amino acids (Fig. 4). The product of *fimD* is also known as fimbrial usher protein, which anchors the type I pilus to the cell surface [24]. Type I fimbriae are important for the virulence and survival of *E. coli* [25]. To investigate the role of *fimD* in the infection and pathogenicity of *E. coli* O91:H21, further experiments such as a *fimD* deletion study and microarray analysis of gene expression in NCCP15736 and NCCP15737 are required.
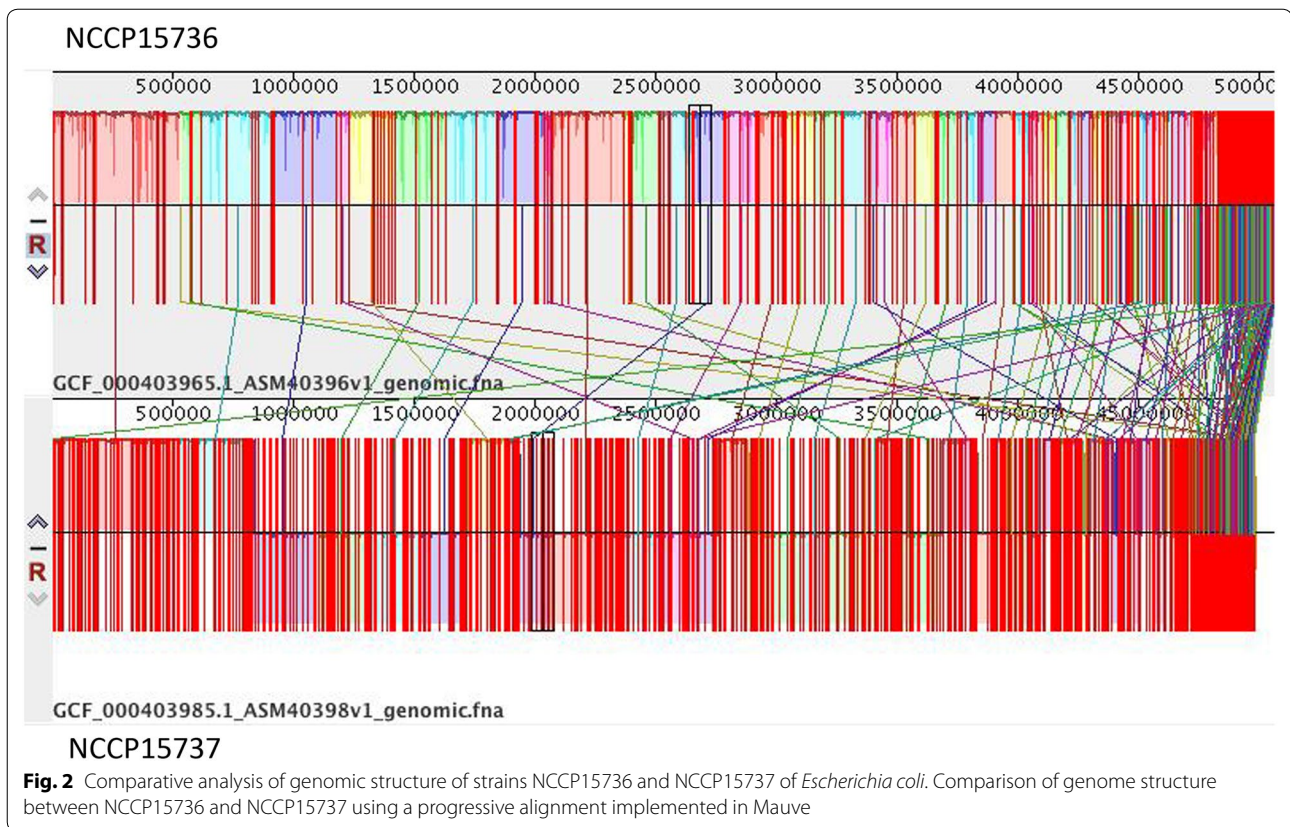
Virulence factors

NCCP15736 was isolated from a symptomatic human carrier but NCCP15737 was isolated from an asymptomatic human carrier. To determine the causal mechanisms underlying the observed pathogenicity, we investigated the virulence factors of NCCP15736 and compared these factors with those of NCCP15737. Using a BLAST search against VFDB, we identified the same number, 25, of virulence factors from NCCP15736 and NCCP15737, respectively (Additional file 2: Table S2). The 25 virulence genes present in NCCP15736 were also present in NCCP15737. The virulence genes of NCCP15736 and NCCP15737 can be classified into five categories: adherence, invasion, iron uptake, secretion system, and toxins. In the adherence category, *E. coli* common pilus (ECP)-related genes (*ecpA, B, C, D, E*, and *ecpR*) F1C fimbriae (*focC*), and type I fimbriae genes (*fimA, B, C, D, E, F, G, H*, and *I*) were identified.

Kwon *et al. Gut Pathog* (2016) 8:57

Page 4 of 8

Fig. 1 Subsystem category distribution of NCCP15636 and NCCP15737 based on the SEED databases. **a** NCCP15736, **b** NCCP15737

Tia invasion determinant (*tia*) [26], which belongs to the invasion category and originates from *E. coli* O1:K1, was identified in both strains. In the iron uptake category, iron-regulated element gene (*ireA*) and salmochelin siderophore-related gene (*iroN*) were identified in NCCP15736 and NCCP15737. Neither strain contained all of the genes in the LEE-encoded TTSS effectors category, harboring only one secretion gene, *escR* [27]. In the toxins category, alpha-hemolysin–related genes (*hlyA, B* and *D*) [28] were identified. Alpha-hemolysin is a major

Kwon *et al. Gut Pathog* (2016) 8:57

Page 5 of 8

Fig. 2 Comparative analysis of genomic structure of strains NCCP15736 and NCCP15737 of *Escherichia coli*. Comparison of genome structure between NCCP15736 and NCCP15737 using a progressive alignment implemented in Mauve
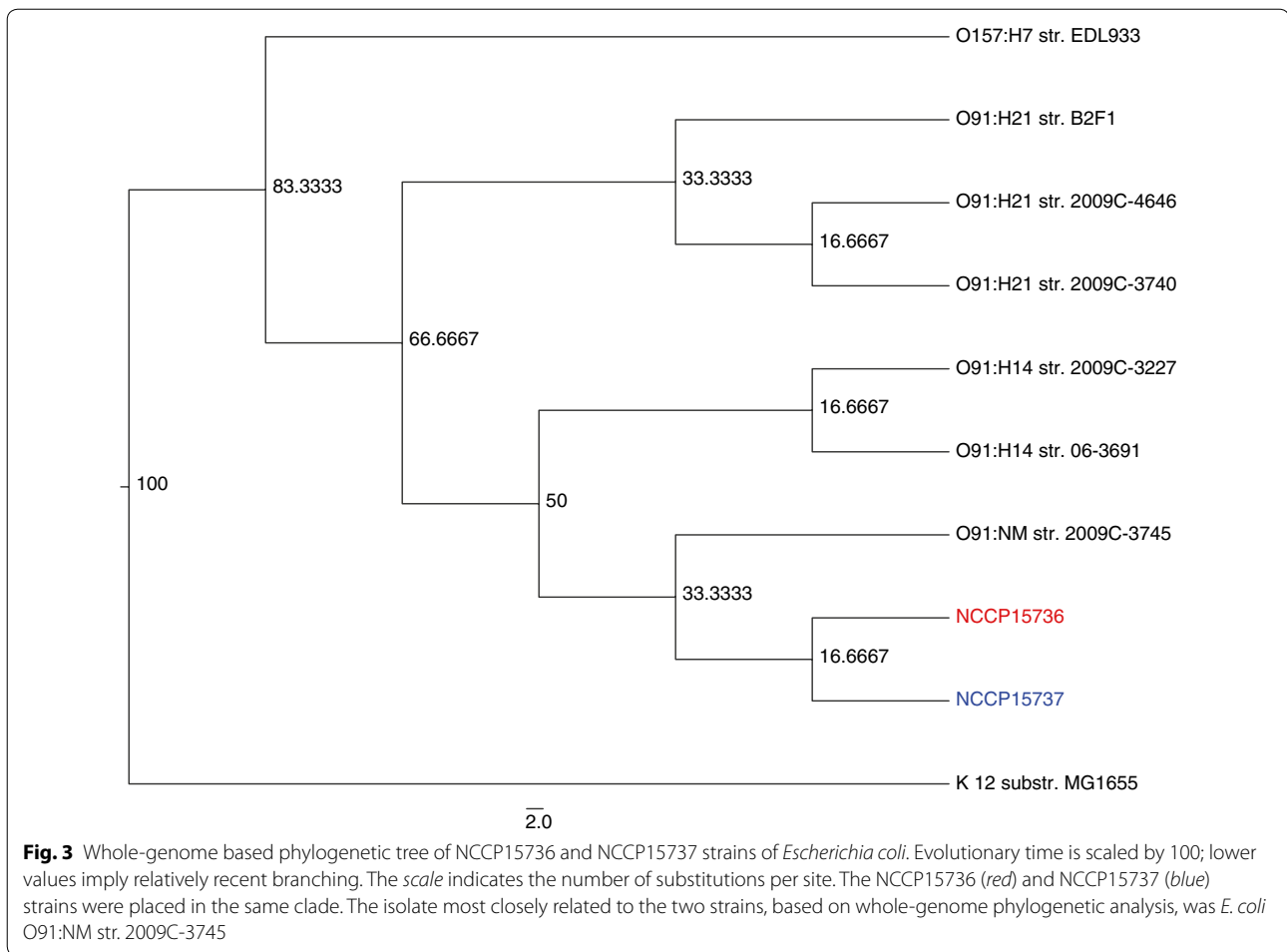
virulence factor present in ETEC, STEC, and EPEC strains. It is acquired by horizontal gene transfer via conjugative plasmids [29]. Shiga-like toxin-related genes (*stx1A* and *1B*) [30] were present in both of the strains and exhibited 100% sequence conservation. In summary, the NCCP15736 and NCCP15737 strains showed the same virulence factors, although NCCP15736 was isolated from a symptomatic carrier and NCCP15737 was isolated from an asymptomatic carrier. In a previous report [12], the expression of type I fimbriae genes was found to be significantly repressed, and the repression was hypothesized to be the main cause of the asymptomatic nature of NCCP15737.

Phage-associated regions
Prophages are mobile genetic elements that can deliver antimicrobial-resistance genes [31] or virulence factors
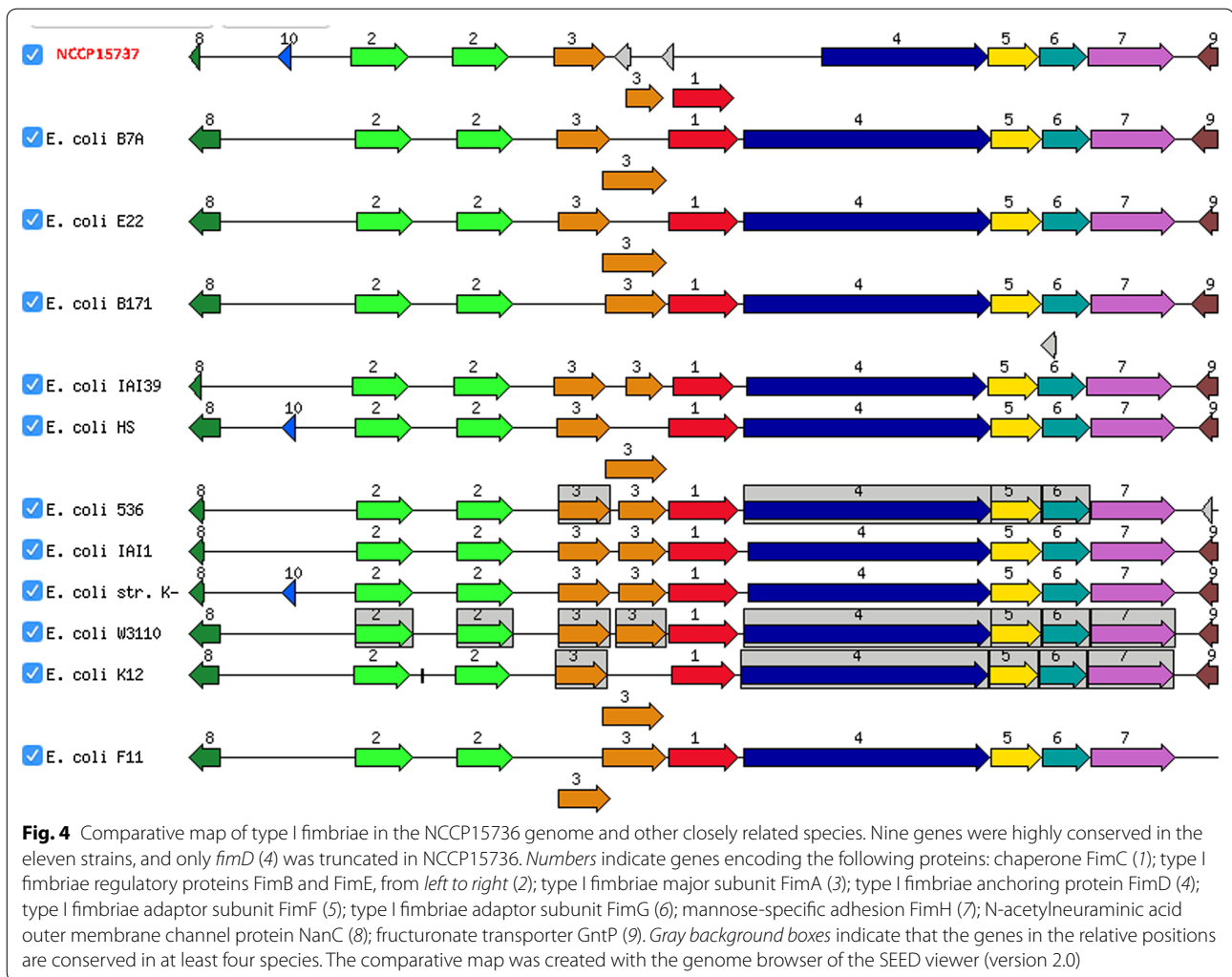
[32] to bacterial hosts and contribute to the diversity of host genomes [33]. We identified ten phage-associated regions (S1–S10) in the NCCP15736 genome and nine phage-associated regions (A1–A9) in the NCCP15737 genome using the PHAST algorithm (Additional file 3: Table S3). Seven of the ten phages in NCCP1576 were intact, and seven of the nine phages in NCCP15737 were intact. NCCP15736 and NCCP15737 each contain two incomplete prophages. Only one questionable prophage, in the S6 region (Stx2-converting phage 1717), was identified in the NCCP15736 genome. Five of the identical phage-associated regions, as determined via a BLAST search, were shared by the two strains. The prophage-associated regions S2, S6, S8, S9, and S10 were unique to NCCP15736, and the A5, A7, A8, and A9 regions were unique to the NCCP15737 genome.

Kwon *et al. Gut Pathog (2016) 8:57*

Page 6 of 8

Fig. 3 Whole-genome based phylogenetic tree of NCCP15736 and NCCP15737 strains of *Escherichia coli*. Evolutionary time is scaled by 100; lower values imply relatively recent branching. The *scale* indicates the number of substitutions per site. The NCCP15736 (*red*) and NCCP15737 (*blue*) strains were placed in the same clade. The isolate most closely related to the two strains, based on whole-genome phylogenetic analysis, was *E. coli* O91:NM str. 2009C-3745

Future directions

The number of outbreaks caused by non-O157 STEC has increased recently and is causing growing concern. In this study, we performed whole-genome sequencing and comparative genomic analysis of two strains, NCCP15736 and NCCP15737. Our whole-genome sequencing and bioinformatics analyses revealed that NCCP15736 and NCCP15737 have the same virulence gene profiles, but NCCP15737 *fimD* shows a deletion. Even though our results did not reveal the genomic basis of the transcriptional repression of type I fimbriae genes in NCCP15737, we provided a structural basis for the relationship between the deficiency in the gene encoding type I fimbriae and the asymptomatic character of NCCP15737. We suggest that complete genome

Kwon *et al. Gut Pathog* (2016) 8:57

Page 7 of 8

Fig. 4 Comparative map of type I fimbriae in the NCCP15736 genome and other closely related species. Nine genes were highly conserved in the eleven strains, and only *fimD* (*4*) was truncated in NCCP15736. *Numbers* indicate genes encoding the following proteins: chaperone FimC (*1*); type I fimbriae regulatory proteins FimB and FimE, from *left to right* (*2*); type I fimbriae major subunit FimA (*3*); type I fimbriae anchoring protein FimD (*4*); type I fimbriae adaptor subunit FimF (*5*); type I fimbriae adaptor subunit FimG (*6*); mannose-specific adhesion FimH (*7*); N-acetylneuraminic acid outer membrane channel protein NanC (*8*); fructuronate transporter GntP (*9*). *Gray background boxes* indicate that the genes in the relative positions are conserved in at least four species. The comparative map was created with the genome browser of the SEED viewer (version 2.0)

sequencing and integration of other types of omics data are required to fully reveal the mechanism underlying the asymptomatic character of NCCP15737.

Additional files

Additional file 1. Annotated genes of NCCP15736 and NCCP15737 using the RAST server.

Additional file 2. Virulence genes of NCCP15736 and NCCP15737.

Additional file 3. Phage-associated regions in NCCP15736 and NCCP15737.

Abbreviations

CDS: coding DNA sequences; EHEC: enterohemorrhagic *Escherichia coli*; NCCP: National Culture Collection for Pathogens; 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. YSB, YBY, JBK, JTC, CHK and YHJ interpreted the

results. YHJ performed the MLST database search. YSB characterized the strain and prepared the genomic DNA. TK performed the gene annotation and comparative genomic analysis and wrote the manuscript. All authors read and approved the final manuscript before submission.

Author details
[1] Interdisciplinary Program in Bioinformatics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. [2] Department of Emergency Medical Service, Sun Moon University, Asan-si, Chungcheongnam-do 31460, Republic of Korea. [3] Division of Antimicrobial Resistance, Center for Infectious Diseases, Korea National Institute of Health, Cheongju 363-951, Republic of Korea. [4] Department of Biomedical Laboratory Science, College of Medical Science, Konyang University, Daejeon 302-832, Republic of Korea. [5] Department of Biomedical Laboratory Science, Kyungdong University, 815 Gyeonhwon-ro, Munmak-eup, Wonju-si, Gangwon-do 26495, Republic of Korea. [6] 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. [7] Department of Food Science and Technology, Sunchon National University, Sunchon, Jeonnam 540-950, Republic of Korea. [8] Division of Enteric Diseases, Center for Infectious Diseases, Korea National Institute of Health, Cheongju 363-951, Republic of Korea.

Kwon *et al. Gut Pathog* (2016) 8:57

Page 8 of 8

References
1. Gill SR, Pop M, Deboy 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.
2. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004;2(2):123–40.
3. Karmali MA, Steele BT, Petric M, Lim C. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet. 1983;1(8325):619–20.
4. 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.
5. Wells JG, Davis BR, Wachsmuth IK, Riley LW, Remis RS, Sokolow R, Morris GK. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. J Clin Microbiol. 1983;18(3):512–20.
6. Corrigan JJ Jr, Boineau FG. Hemolytic-uremic syndrome. Pediatr Rev. 2001;22(11):365–9.
7. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998;11(1):142–201.
8. 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.
9. Sandvig K, Bergan J, Dyve AB, Skotland T, Torgersen ML. Endocytosis and retrograde transport of Shiga toxin. Toxicon. 2010;56(7):1181–5.
10. Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, Karch H, Mellmann A, Caprioli A, Tozzoli R, Morabito S, et al. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. J Clin Microbiol. 2012;50(9):2951–63.
11. Luna-Gierke RE, Griffin PM, Gould LH, Herman K, Bopp CA, Strockbine N, Mody RK. Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. Epidemiol Infect. 2014;142(11):2270–80.
12. Kim JB, Oh KH, Park MS, Cho SH. Repression of type-1 fimbriae in Shiga toxin-producing *Escherichia coli* O91:H21 isolated from asymptomatic human carriers in Korea. J Microbiol Biotechnol. 2013;23(5):731–7.
13. 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.
14. 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.
15. 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 Genom. 2008;9:75.
16. 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(1):325–8.
17. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:347–52.
18. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics. 2011;27(3):334–42.
19. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences. 1986;17:57–86.
20. Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Parallel and distributed processing symposium 2006. 20th International 2006 IPDPS.
21. 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.
22. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–403.
23. Overbeek 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. 2014;42:206–14.
24. Nishiyama M, Horst R, Eidam O, Herrmann T, Ignatov O, Vetsch M, Bettendorff P, Jelesarov I, Grutter MG, Wuthrich K, et al. Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. EMBO J. 2005;24(12):2075–86.
25. Connell I, Agace W, Klemm P, Schembri M, Marild S, Svanborg C. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proc Natl Acad Sci USA. 1996;93(18):9827–32.
26. Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, Skyberg JA, Lynne AM, Johnson JR, Nolan LK. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. J Bacteriol. 2007;189(8):3228–36.
27. Pallen MJ, Beatson SA, Bailey CM. Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion. BMC Microbiol. 2005;5:9.
28. Burgos Y, Beutin L. Common origin of plasmid encoded alpha-hemolysin genes in Escherichia coli. BMC Microbiol. 2010;10:193.
29. Lim JY, Yoon J, Hovde CJ. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. J Microbiol Biotechnol. 2010;20(1):5–14.
30. Lee JE, Reed J, Shields MS, Spiegel KM, Farrell LD, Sheridan PP. Phylogenetic analysis of Shiga toxin 1 and Shiga toxin 2 genes associated with disease outbreaks. BMC Microbiol. 2007;7:109.
31. Colomer-Lluch M, Imamovic L, Jofre J, Muniesa M. Bacteriophages carrying antibiotic resistance genes in fecal waste from cattle, pigs, and poultry. Antimicrob Agents Chemother. 2011;55(10):4908–11.
32. 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.
33. Ventura M, Canchaya C, Bernini V, Altermann E, Barrangou R, McGrath S, Claesson MJ, Li Y, Leahy S, Walker CD, et al. Comparative genomics and transcriptional analysis of prophages identified in the genomes of *Lactobacillus gasseri*, *Lactobacillus salivarius*, and *Lactobacillus casei*. Appl Environ Microbiol. 2006;72(5):3130–46.