

GENOME REPORT

Open Access



Genome sequence of *Escherichia coli* NCCP15653, a group D strain isolated from a diarrhea patient

Min-Jung Kwak^{1†}, Myung-Soo Kim^{1†}, Soon-Kyeong Kwon¹, Seung-Hak Cho² and Jihyun F. Kim^{1*} 

Abstract

Background: Pathogenic strains in *Escherichia coli* can be divided into several pathotypes according to their virulence features. Among them, uropathogenic *E. coli* causes most of the urinary tract infections and has a genotype distinct from other virulent strains of *E. coli*. In this study, we sequenced and analyzed the genome of *E. coli* NCCP15653 isolated from the feces of a diarrhea patient in 2007 in South Korea.

Results: A phylogenetic tree based on MLST showed that NCCP15653 belongs to the D group of *E. coli* and located in the lineage containing strains ST2747, UMN026 and 042. In the genome of NCCP15653, genes encoding major virulence factors of uropathogenic *E. coli* were detected. They include type I fimbriae, hemin uptake proteins, iron/manganese transport proteins, yersiniabactin siderophore proteins, type VI secretion proteins, and hemolysin. On the other hand, genes encoding AslA, OmpA, and the K1 capsule, which are virulence factors associated with invasion of neonatal meningitis-causing *E. coli*, were also present, while a gene encoding CNF-1 protein, which is a cytotoxic necrotizing factor 1, was not detected.

Conclusions: Through the genome analysis of NCCP15653, we report an example of a genome of chimeric pathogenic properties. The gene content of NCCP15653, a group D strain, demonstrates that it could be both uropathogenic *E. coli* and neonatal meningitis-causing *E. coli*. Our results suggest the dynamic nature of plastic genomes in pathogenic strains of *E. coli*.

Keywords: Extraintestinal *E. coli*, UPEC, NMEC, Cystitis, Pyelonephritis

Background

Escherichia coli can be divided into commensal and pathogenic strains. Commensal *E. coli* is a member of the normal flora of animal intestine and other body sites, but pathogenic strains of *E. coli* cause several health problems. Many *E. coli* strains can cause diarrhea, but not serious [1]. However, some pathogenic strains such as *E. coli* O104:H4 that caused the German outbreak in 2011 may be fatal [2]. According to the virulence factors and phenotypes, pathogenic *E. coli* strains can be classified into enteroaggregative *E. coli* (EAEC),

enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), uropathogenic *E. coli* (UPEC), and *E. coli* that causes neonatal meningitis (NMEC) [3–8]. Among them, UPEC and NMEC are extraintestinal pathogenic *E. coli* (ExPEC), and most of the urinary tract infections (UTIs) are caused by UPEC strains [9]. The urinary tract is a harsh environment because of continuous urine excretion, antibacterial factors, and strong immune system, and these features of urinary tract can make UPEC possible to have genotypes distinct to other pathogenic strains [10]. In the urinary tract, it needs adhesion to urinary epithelial cells, several resistance factors against the antibacterial factors and host immune systems, and iron-acquisition systems to obtain iron, which is limited in the urinary tract. UPEC causes the infection in the bladder

*Correspondence: jfk1@yonsei.ac.kr

[†]Min-Jung Kwak and Myung-Soo Kim contributed equally to this work

¹ Department of Systems Biology and Division of Life Sciences, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Republic of Korea
Full list of author information is available at the end of the article

and sometimes in the kidneys through entering the ureters from the bladder to trigger symptoms such as cystitis and pyelonephritis, and even bacteremia and sepsis through entering the bloodstream [11]. In this study, we sequenced and analyzed the genome of pathogenic *E. coli* strain NCCP15653 isolated from the feces of a patient suffering from diarrhea.

Methods

Bacteria and DNA isolation

E. coli strain NCCP15653 was isolated from the feces of a Korean patient with the diarrhea symptom in 2007. This strain was deposited at the National Culture Collection for Pathogens in Korea National Institute of Health (KNIH) and its accession number is NCCP15653. Genomic DNA was extracted using chemical and enzymatic methods as described in Molecular cloning, a laboratory manual [12].

Genome sequencing, de novo assembly and annotation

For the genome sequencing of NCCP15653, Genome Analyzer IIX of the Illumina platform at the Biomedical Genomics Research Center of the Korea Research Institute of Bioscience and Biotechnology was used and 18,521,148 of raw sequencing reads with 76-bp of average read length were generated from a 500-bp paired-end library. The sequencing reads were imported into CLC Genomics Workbench version 5.1 (CLC bio, Qiagen, Netherlands) with the parameters of 400–700 of paired-end distance and 1.5–1.7 version of Illumina quality score. Trimming of the imported reads was performed with the parameters of 0.01 quality score, none of the ambiguous nucleotide, and 70-bp of minimum read length. De novo assembly of 13,864,337 high-quality reads were conducted using CLC Genomics Workbench with the parameters of similarity fraction of 1.0, length fraction of 0.5, and minimum contig length of 500 bp. SSPACE [13] was used for scaffolding and IMAGE [14] was used for automatic gap filling. Manual contig extension and gap filling were performed with CLC Genomics Workbench. Structural gene prediction was accomplished with Glimmer3 [15], and functional annotation of predicted genes was performed using the MicroScope database [16].

Genome analysis

A phylogenetic tree based on multilocus sequence typing (MLST) was constructed with MEGA5 [17]. Nucleotide sequences of seven MLST genes (*adk* adenylate kinase, *fumC* fumarate hydratase, *gyrB* DNA gyrase, *icd* isocitrate/isopropylmalate dehydrogenase, *mdh* malate dehydrogenase, *purA* adenylosuccinate dehydrogenase, *recA* ATP/GTP binding motif) [18] and Jukes-Cantor model

were used for tree construction. To determine the serotype of NCCP15653 in silico, amino-acid sequences of the *wzx* and *wzy* gene for O-antigen and the *fliC* gene for H-antigen were used and the neighbor-joining trees were constructed with MEGA5. SerotypeFinder program [19] was also used for the analysis. Average nucleotide identity based on blast (ANIb) value was calculated using JSpecies [20]. Calculation of the core genome was conducted with OrthoMCL (ver. 2.0.3) [21] with parameters of e -value $\leq 1e-5$, identity ≥ 85 %, and coverage ≥ 80 % [22]. Functional classification of the genes was conducted by BLASTP with the COG and subsystem databases. Prediction of phage sequences and clustered regularly interspaced short palindromic repeats (CRISPRs) was performed with PHAST [23] and CRISPRfinder [24], respectively. Detection of the virulence genes was conducted using BLAST software. Typing of the specific virulence genes were referenced to the virulence factors of pathogenic bacteria database (<http://www.mgc.ac.cn/VFs/main.htm>) [25].

Quality assurance

E. coli NCCP15653 was maintained in pure culture at KNIH and genomic DNA was isolated from a single isolate. Possibilities for the contamination of other genomes and misassembly were checked through mapping reads to the contigs. The read mapping of the draft genome of NCCP15653 indicated that the distance between paired-end reads is in the range of expected size distribution and the coverage of the reads was consistent throughout the genome.

Results and discussion

General features

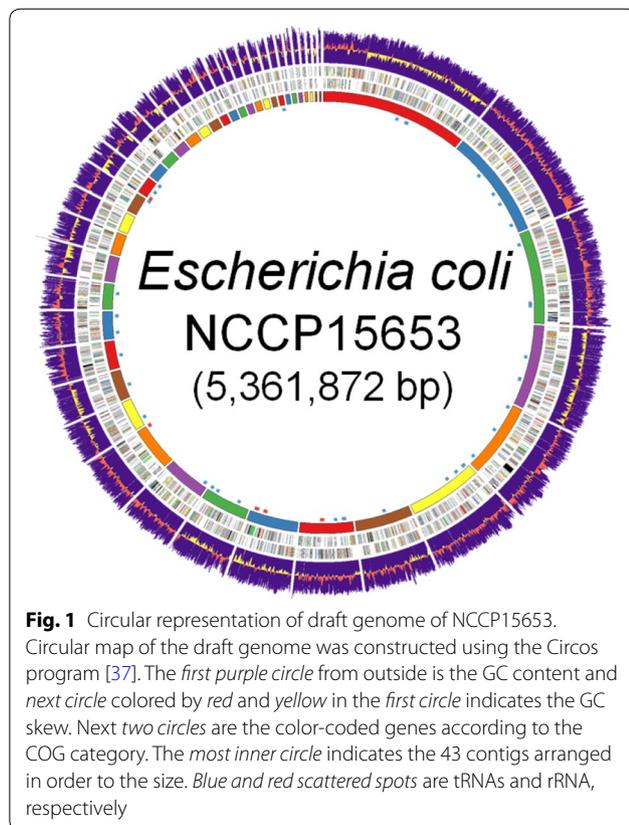
The draft genome of *E. coli* NCCP15653 consists of 43 contigs and the sum of the length of the contigs is 5,361,872 bp with 50.56 % of GC content (Table 1 and Fig. 1). The number of predicted protein coding sequences (CDSs) is 5203 and the percentages of subsystem and COG assigned proteins were 76.40 and 76.42 % respectively. The numbers of predicted transfer RNA and ribosomal RNA are 73 and 21, respectively. In the genome of NCCP15653, six intact phages and four CRISPR candidates were detected. Among the four CRISPR candidates, one has the *cas* genes next to the repeat array and nine spacers.

Phylogenetic relationships

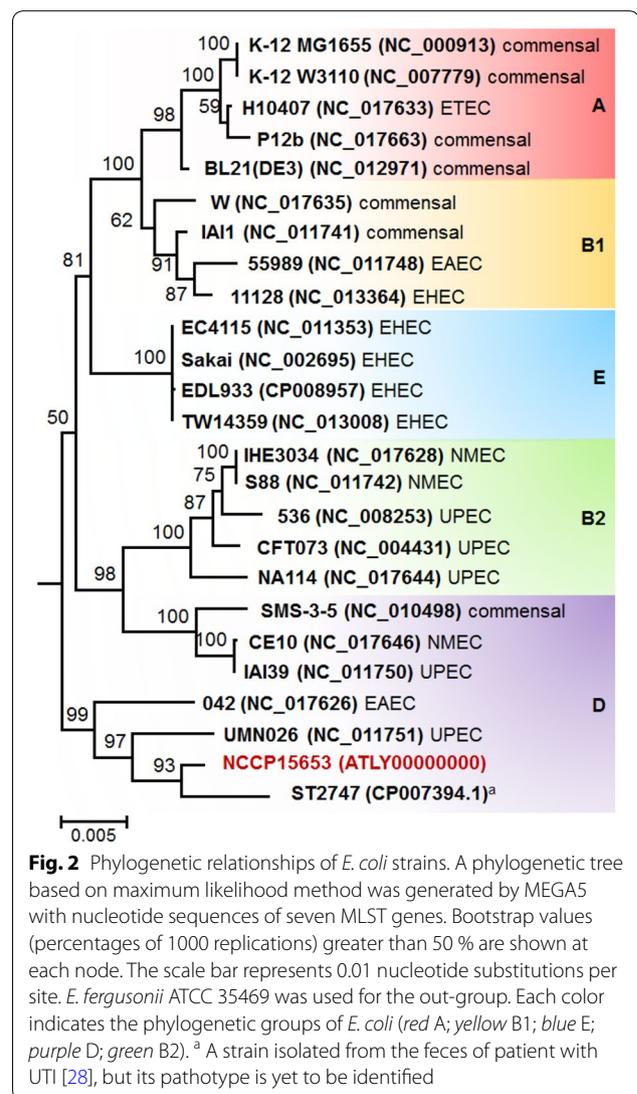
The phylogenetic tree based on MLST showed that NCCP15653 belong to the D group of *E. coli* (Fig. 2). In accordance with previous reports [26, 27], strains belonging to group D are contained in two distinct phylogenetic lineages and may have a polyphyletic origin.

Table 1 General features of the *E. coli* NCCP15653 genome

Item	Value
Number of contigs	43
Total contig length (bp)	5,361,872
Fold coverage (x)	171.19
N50 (bp)	242,877
G + C content (%)	50.56
Number of protein coding genes	5297
Number of predicted transfer RNAs	73
Number of predicted ribosomal RNAs	21
GenBank accession number	ATLY000000000



One is located in the outermost branches of *E. coli* outside the groups A, B1, B2, and E, and the other forms a sister clade of group B2. NCCP15653 is placed in the former clade. The *E. coli* group D includes several pathogenic strains such as ST2747 (isolated from feces of patient with UTI, but pathotype not identified) [28], UMN026 (UPEC), 042 (EAEC), IAI39 (UPEC), and CE10 (NMEC) as well as commensal strains SMS-3-5 [29–32]. NCCP15653 is placed next to strain ST2747. Calculation of ANIb between the group D strains also indicated



that NCCP15653 is most similar to ST2747; average ANIb value and genome coverage are 98.18 and 85.54 %, respectively (Table 2). A serotype analysis using the genes encoding O-antigen and H-antigen indicated that O-antigen of NCCP15653 is untypable but H-antigen can be clustered with those of the H18 serotype.

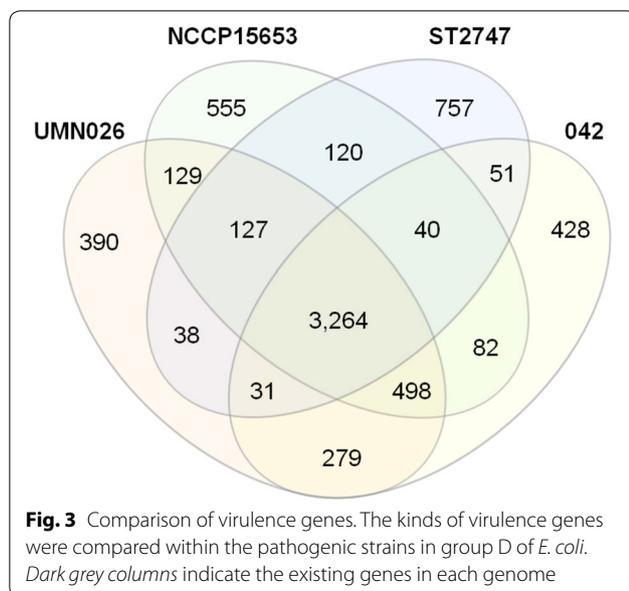
Virulence genes

Dr adhesins, F1C fimbriae, P fimbriae, S fimbriae, type 1 fimbriae, immuno-evasion protein, aerobactin, enterobactin, Chu proteins, siderophore receptor, proteases, CNF-1 toxin, and hemolysin are the major virulence factors of UPEC [25]. In the genome of NCCP15653, several virulence factors for UTI were detected and shown in Fig. 3. They include genes encoding type I fimbriae, hemin uptake proteins, iron/manganese transport

Table 2 Average nucleotide identity values based on blast

	NCCP15653	ST2747	UMN026	0 42	SMS-3-5	IAI39	CE10
NCCP15653	---	98.16 (82.75%)	97.95 (85.21%)	97.64 (83.39%)	97.49 (82.41%)	97.47 (79.80%)	97.44 (82.94%)
ST2747	98.20 (88.32%)	---	98.16 (86.34%)	97.85 (85.91%)	97.51 (84.15%)	97.34 (81.23%)	97.34 (83.79%)
UMN026	98.06 (84.35%)	98.09 (80.64%)	---	98.32 (87.58%)	97.32 (80.66%)	97.53 (79.00%)	97.49 (82.59%)
0 42	97.74 (82.78%)	97.80 (79.76%)	98.36 (87.74%)	---	97.23 (81.15%)	97.33 (77.90%)	97.27 (81.12%)
SMS-3-5	97.46 (83.93%)	97.38 (80.30%)	97.25 (83.45%)	97.06 (83.29%)	---	98.13 (82.96%)	98.10 (85.05%)
IAI39	97.40 (84.06%)	97.14 (79.65%)	97.49 (84.69%)	97.21 (82.36%)	98.09 (85.71%)	---	99.85 (97.95%)
CE10	97.40 (82.48%)	97.09 (78.27%)	97.39 (83.55%)	97.14 (81.62%)	98.04 (83.40%)	99.82 (92.65%)	---

Numbers in parentheses indicate the genome coverage



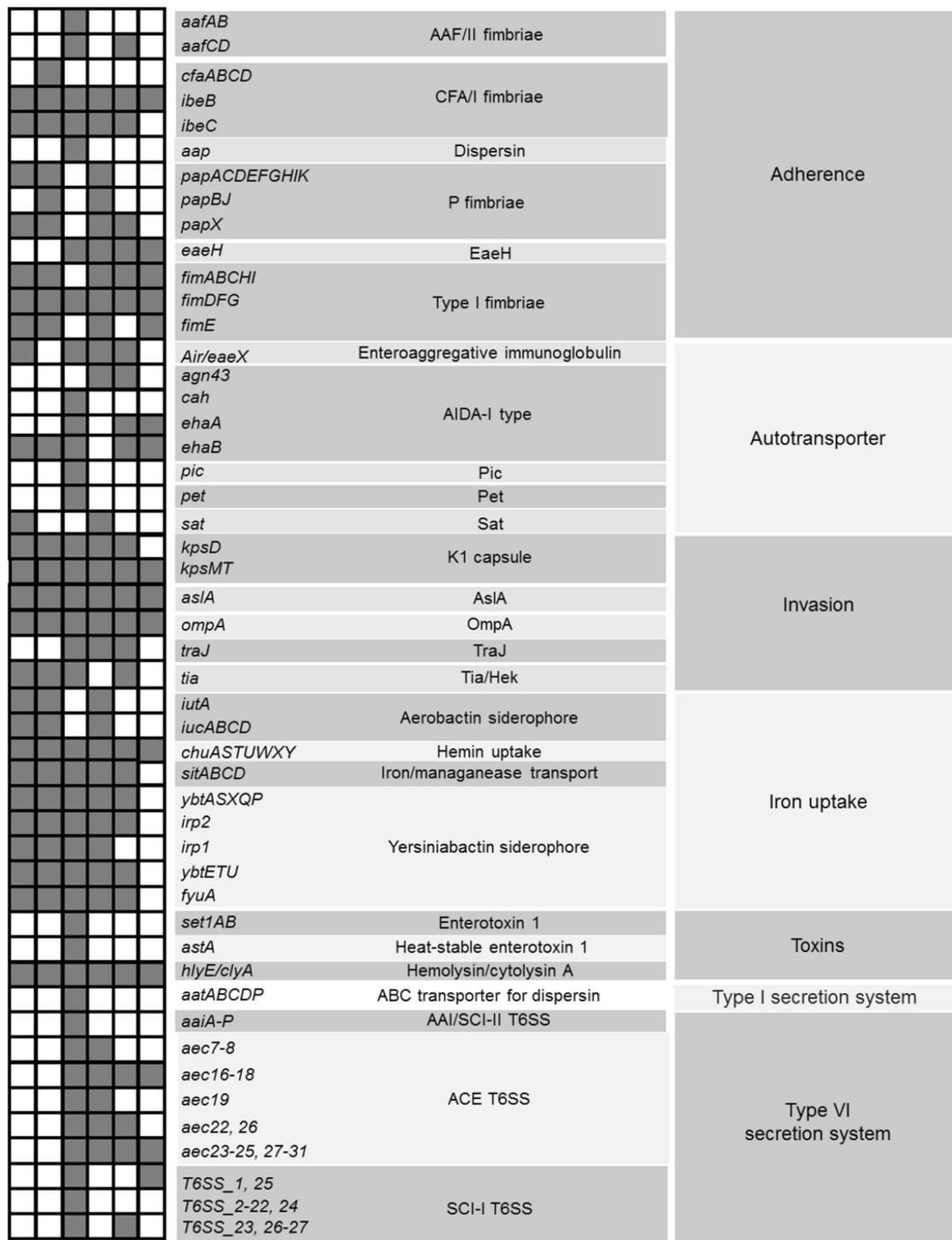
proteins, yersiniabactin siderophore proteins, type VI secretion proteins, and hemolysin (Fig. 3). Type I fimbriae are known to promote intracellular invasion and persistence [11], and hemolysin is known to kill the host cell by making pores to the surface [33]. Genes associated with iron-uptake are expected to make *E. coli* possible to survive in iron-deprived environments like the urinary tract [34]. In the genome of NCCP15653, genes encoding AslA and OmpA, which are virulence factors associated with invasion of NMEC, were discovered. Moreover, *kps* genes encoding proteins that form the K1 capsule were also identified in the genome of NCCP15653. The K1 capsule is known as a predominant capsular polysaccharide detected in approximately 80 % of the NMEC strains

[35] and known to play important roles in invasion and survival in the host cell [36]. On the other hand, the *cnf1* gene encoding cytotoxic necrotizing factor 1, which is a toxin of NMEC, was not present.

Comparison with other *E. coli* strains in group D

An analysis of the core genome of four strains in group D, which were located in the same lineage with NCCP15653, inferred that they share 3264 core genes (Fig. 4). The core gene set contains genes encoding CFA/I fimbrial proteins, hemolysin E, and flagella-biosynthetic proteins and proteins as well as proteins related to general cell metabolism. Genes conserved in NCCP15653, ST2747, and UMN026, three strains that share the same common ancestor, as compared with 042, an EAEC strain outside of them, include genes encoding adhesin for cattle intestine colonization. Genes conserved in NCCP15653 and ST2747 compared with UMN026 and 042 include genes encoding entericidin and toxin-antitoxin system proteins RelB and RelE.

Comparison of the virulence genes in the pathogenic strains in group D suggests that they may be divided into three groups (Fig. 3). The first group includes CE10 and IAI39, which are NMEC and UPEC, respectively. The second has UMN026, a UPEC strain, NCCP15653, and ST2747, which was isolated from a patient with UTI, but its pathotype is not yet determined [28]. The third group contains the EAEC strain 042 alone, which has a quite different gene content compared to those in the first and second groups. Strains in the first and second groups show similar gene contents. However, in the genomes of CE10 and IAI39, *aec (tss)* genes encoding the type VI secretion system were not detected, and in the genome of NCCP15653, biosynthetic genes for aerobactin siderophore and P fimbriae were not present.



CE10
IAN39
042
UMN026
NCCPP15653
ST2747

Fig. 4 Numbers of core gene set and shared genes among strains NCCP15653, ST2747, UMN026, and 042. Gene designations are according to ref. [25]

Conclusions

NCCP15653 was isolated from the feces of a diarrhea patient. However, an MLST-based phylogenetic tree and ANiB values indicated that NCCP15653 belongs to the D group of *E. coli* and is a sister strain of ST2747. In addition, in the genome of NCCP15653, genes encoding UPEC-type virulence factors of were detected, and those included type I fimbriae, hemin uptake proteins, iron/manganese transport proteins, yersiniabactin siderophore proteins, type VI secretion proteins, and hemolysin. Moreover, NCCP15653 has genes associated with the invasion of NMEC, which include those for the K1 capsule and putative arylsulfatase. Genome analysis results of NCCP15653 will be useful for further research of genome dynamics in the pathogenic *E. coli* strains causing UTI.

Availability of supporting data

This whole genome shotgun project of NCCP15653 has been deposited at GenBank under the accession ATLY000000000.

Authors' contributions

JFK conceived, organized and supervised the project, interpreted the results, and edited the manuscript. SHC characterized the strains and maintained it in pure cultures. SKK prepared the high-quality genomic DNA and arranged the acquisition of sequence data. MJK and MSK performed the sequence assembly, gene prediction, gene annotation, analyzed the genome information, and drafted the manuscript. All authors read and approved the final manuscript.

Author details

¹ Department of Systems Biology and Division of Life Sciences, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Republic of Korea. ² Division of Enteric Diseases, Center for Infectious Diseases, Korea National Institute of Health, Heungdeok-Gu, Cheongju 363-951, Republic of Korea.

Acknowledgements

The authors are thankful to Byung Kwon Kim, Ju Yeon Song, Seon-Young Kim, and the KRIBB sequencing team for technical assistance. This work was financially supported by the National Research Foundation of the Ministry of Science, ICT and Future Planning (NRF-2011-0017670 to J.F.K.) and Korea National Institute of Health (KNIH 4800-4845-300 to S.H.C.), Republic of Korea.

Competing interests

The authors declare that they have no competing interests.

Received: 18 November 2015 Accepted: 4 January 2016

Published online: 23 February 2016

References

- Dobrindt U. (Patho-)Genomics of *Escherichia coli*. *Int J Med Microbiol.* 2005;295:357–71.
- Muniesa M, Hammerl JA, Hertwig S, Appel B, Brussow H. Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Appl Environ Microbiol.* 2012;78:4065–73.
- Jeong H, Zhao F, Iguri D, Oh KH, Kim SY, Kang SG, Kim BK, Kwon SK, Lee CH, Song JY, et al. Genome sequence of the hemolytic-uremic syndrome-causing strain *Escherichia coli* NCCP15647. *J Bacteriol.* 2012;194:3747–8.
- Song JY, Yoo RH, Jang SY, Seong WK, Kim SY, Jeong H, Kang SG, Kim BK, Kwon SK, Lee CH, et al. Genome sequence of enterohemorrhagic *Escherichia coli* NCCP15658. *J Bacteriol.* 2012;194:3749–50.
- Kim BK, Song GC, Hong GH, Seong WK, Kim SY, Jeong H, Kang SG, Kwon SK, Lee CH, Song JY, et al. Genome sequence of the Shiga toxin-producing *Escherichia coli* strain NCCP15657. *J Bacteriol.* 2012;194:3751–2.
- Leonard SR, Lacher DW, Lampel KA. Draft genome sequences of the enteroinvasive *Escherichia coli* strains M4163 and 4608-58. *Genome Announc.* 2015;3:e01395.
- Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, et al. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J Bacteriol.* 2009;191:347–54.
- Kaper JB, Nataro JP, Mobley HLT. Pathogenic *Escherichia coli*. *Nat Rev Microbiol.* 2004;2:123–40.
- Kohler CD, Dobrindt U. What defines extraintestinal pathogenic *Escherichia coli*? *Int J Med Microbiol.* 2011;301:642–7.
- Johnson JR, Russo TA. Extraintestinal pathogenic *Escherichia coli*: "the other bad *E. coli*". *J Lab Clin Med.* 2002;139:155–62.
- Bower JM, Eto DS, Mulvey MA. Covert operations of uropathogenic *Escherichia coli* within the urinary tract. *Traffic.* 2005;6:18–31.
- Green MR, Sambrook J. *Molecular cloning: a laboratory manual.* 4th ed. New York: Cold Spring Harbor Laboratory Press; 2012.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27:578–9.
- Tsai JJ, Otto TD, Berriman M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 2010;11:R41.
- Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 1998;26:544–8.
- Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Medigue C. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* 2006;34:53–65.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28:2731–9.
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol.* 2006;60:1136–51.
- Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheut F. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol.* 2015;53:2410–26.
- Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA.* 2009;106:19126–31.
- Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–89.
- Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, et al. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol.* 2009;394:644–52.
- Zhou Y, Liang YJ, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res.* 2011;39:W347–52.
- Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007;35:W52–7.
- Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* 2012;40:D641–5.
- Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol.* 2010;8:207–17.
- Skippington E, Ragan MA. Within-species lateral genetic transfer and the evolution of transcriptional regulation in *Escherichia coli* and *Shigella*. *BMC Genomics.* 2011;12:532.
- Xavier BB, Vervoort J, Stewardson A, Adriaenssens N, Coenen S, Harbarth S, Goossens H, Malhotra-Kumar S. Complete genome sequences of nitrofurantoin-sensitive and -resistant *Escherichia coli* ST540 and ST2747 strains. *Genome Announc.* 2014;2:e00239.
- Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 2009;5:e100034.

30. Zhou ZM, Li XM, Liu B, Beutin L, Xu JG, Ren Y, Feng L, Lan RT, Reeves PR, Wang L. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. *PLoS One*. 2010;5:e8700.
31. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, Petty NK, Mahon V, Brinkley C, Hobman JL, et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol*. 2010;192:5822–31.
32. Lu ST, Zhang XB, Zhu YF, Kim KS, Yang J, Jin Q. Complete genome sequence of the neonatal-meningitis-associated *Escherichia coli* strain CE10. *J Bacteriol*. 2011;193:7005.
33. Kerenyi M, Allison HE, Batai I, Sonnevend A, Emody L, Plaveczyk N, Pal T. Occurrence of *hlyA* and *sheA* genes in extraintestinal *Escherichia coli* strains. *J Clin Microbiol*. 2005;43:2965–8.
34. Tree JJ, Ulett GC, Ong CLY, Trott DJ, McEwan AG, Schembri MA. Trade-off between iron uptake and protection against oxidative stress: Deletion of *cueO* promotes uropathogenic *Escherichia coli* virulence in a mouse model of urinary tract infection. *J Bacteriol*. 2008;190:6909–12.
35. Kim KS, Itabashi H, Gemski P, Sadoff J, Warren RL, Cross AS. The K1-capsule is the critical determinant in the development of *Escherichia coli* meningitis in the rat. *J Clin Invest*. 1992;90:897–905.
36. Hoffman JA, Wass C, Stins MF, Kim KS. The capsule supports survival but not traversal of *Escherichia coli* K1 across the blood-brain barrier. *Infect Immun*. 1999;67:3566–70.
37. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.

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

