Gut Pathogens

GENOME REPORT Open Access

Genomic characterization of nine *Clostridioides difficile* strains isolated from Korean patients with *Clostridioides difficile* infection

Seung Woo Ahn[1†], Se Hee Lee[1†], Uh Jin Kim[2], Hee-Chang Jang[2], Hak-Jong Choi[1], Hyon E. Choy[3], Seung Ji Kang[2*] and Seong Woon Roh[1*]

Abstract

Background: *Clostridioides difficile* infection (CDI) is an infectious nosocomial disease caused by *Clostridioides difficile*, an opportunistic pathogen that occurs in the intestine after extensive antibiotic regimens.

Results: Nine *C. difficile* strains (CBA7201–CBA7209) were isolated from nine patients diagnosed with CDI at the national university hospital in Korea, and the whole genomes of these strains were sequenced to identify their genomic characteristics. Comparative genomic analysis was performed using 51 reference strains and the nine isolated herein. Phylogenetic analysis based on 16S rRNA gene sequences confirmed that all 60 *C. difficile* strains belong to the genus *Clostridioides*, while core-genome tree indicated that they were divided into five groups, which was consistent with the results of MLST clade analysis. All strains were confirmed to have a clindamycin antibiotic resistance gene, but the other antibiotic resistance genes differ depending on the MLST clade. Interestingly, the six strains belonging to the sequence type 17 among the nine *C. difficile* strains isolated here exhibited unique genomic characteristics for PaLoc and CdtLoc, the two toxin gene loci identified in this study, and harbored similar antibiotic resistance genes.

Conclusion: In this study, we identified the specific genomic characteristics of Korean *C. difficile* strains, which could serve as basic information for CDI prevention and treatment in Korea.

Keywords: *Clostridioides difficile*, *Clostridioides difficile* infection, Comparative genomics, Pathogenic features, Toxin A and toxin B, Antibiotic resistance gene

Background

Clostridioides difficile is a major nosocomial pathogenic bacterium that poses a threat to public health worldwide [1]. An estimated 453,000 infections occur annually due to this organism; 15,000 deaths are directly attributable to infections caused by *C. difficile* in the United States [2]. In Europe and the UK, about 124,000 and 18,005 people are infected annually, respectively [3]. Recently, studies have revealed a high prevalence of CDI in East Asia, similar to that

*Correspondence: sseungi@chonnam.ac.kr; seong18@gmail.com
†Seung Woo Ahn and Se Hee Lee contributed equally to this work
[1] Microbiology and Functionality Research Group, World Institute of Kimchi, 86, Kimchi-ro, Nam-gu, 61755 Gwangju, Republic of Korea
[2] Department of Infectious Diseases, Chonnam National University Hospital, 61469 Gwangju, Republic of Korea
Full list of author information is available at the end of the article

seen in Europe and North America [4]. In the Republic of Korea, the number of CDI patients per 100,000 people in a population increased from 1.43 in 2008 to 5.06 in 2011, while CDI-associated mortality increased from 0.14 to 0.35 [5]. These numbers are increasing yearly, especially in patients over 65 years old [6]. CDI is an opportunistic infectious disease caused by *C. difficile*, which grows and secretes toxins in the intestinal tract of the patient, resulting in a variety of symptoms including diarrhea and pseudomembranous colitis; it can be life-threatening [7]. CDI is treated with antibiotics such as cephalosporin, clindamycin, quinolone, metronidazole, and vancomycin [8, 9]. However, in some cases, diarrhea may recur or death may occur because these antibiotics can trigger infection [10].

C. difficile is a gram-positive, spore-forming, anaerobic, intestinal bacterium [2]. Recently, 16S rRNA gene sequence analysis of the *Clostridium* genus indicated that the similarity between *C. difficile* and *Clostridium hiranonis* is less than 97%; this has led to the reclassification of *Clostridium difficile* as *Clostridioides difficile* [11]. Secretion of toxin A (enterotoxin) and toxin B (cytotoxin) by *C. difficile* is mainly responsible for the intestinal inflammation resulting from CDI [12]. These two toxins inactivate host cell GTP-binding proteins and destroy the cytoskeleton, inducing apoptosis, severe inflammation, and intestinal cell damage [13–15]. Moreover, some hypervirulent strains (e.g. NAP1/ribotype 027) synthesize the actin-ADP-ribosylating toxin known as binary toxin *C. difficile* transferase (CDT), which leads to the depolymerization of actin filaments, disrupting the actin cytoskeleton in the cytosol [16–19]. Therefore, the toxins A, B, and CDT can cause severe CDI symptoms [20].

The first complete genome sequence of *C. difficile* reported was strain 630 [21]. Subsequently, the genomic information of a variety of *C. difficile* strains has been reported and deposited (https://www.ncbi.nlm.nih.gov/genome/genomes/535). In addition to uncovering the genetic and evolutionary diversity of *C. difficile* strains [22, 23], virulence factors of these strains, such as toxins, antibiotic resistance, mobility, and adhesion have been also investigated through a comparative genomic analysis [24]. Although there have been many studies on *C. difficile* strains isolated from various patients with CDI worldwide, there have been few genomic studies conducted on *C. difficile* strains in the Republic of Korea. Thus, this study aimed to investigate the unique genomic characteristics of nine *C. difficile* strains isolated from South Korean patients through a comparative genomic analysis with previously characterized strains.

Methods

Ethical statement and sample collection

Stool samples were collected from nine patients diagnosed with CDI who visited the Department of infectious disease, Chonnam national university hospital in Gwangju or Hwasun, Republic of Korea. The study protocol was approved by the institutional review boards of the Republic of Korea centers for disease control and Prevention [IRB file no. CNUH-2017-161 and CNUHH-2017-076]. Written informed consent was obtained from all participants. The characteristics of all patients who agreed to fecal sampling after confirmation of CDI and the list of strains isolated from each fecal sample, as well as the prescribed antibiotics for each patient before fecal sampling, are summarized in Table 1.

Culture conditions and identification of *C. difficile* isolated from CDI participants

Collected stool samples were treated with chloroform for efficient isolation of *C. difficile* [25] as chloroform selectively isolates *C. difficile* by removing non-spore forming bacteria in the stool samples. Chloroform (60 µL; concentration, 3%) was added to filtered PBS (1740 µL), after which 200 µL of the stool sample was added. The mixed samples were suspended in a shaking incubator for 1 h at 37 °C, after which chloroform was evaporated with N_2 gas (Automated gas distribution workstation; Raontech, Gwangju, Republic of Korea), followed by culturing in Cycloserine-Cefoxitin Fructose Agar (CCFA) medium, which is an enriched selective and differential medium for the isolation and presumptive identification of *C. difficile*. CCFA medium consists of 40.0 g proteose peptone, 5.0 g sodium phosphate dibasic, 1.0 g potassium phosphate monobasic, 2.0 g sodium chloride, 6.0 g fructose, 15.0 g agar, 9.0 mg neutral red, 500.0 mg cycloserine (10.0% solution), and 15.6 mg cefoxitin (1.56% solution). Cycloserine inhibits the growth of Gram-negative bacteria, while cefoxitin inhibits the growth of both Gram-positive and -negative organisms [26]. *C. difficile* can be resistant to cefoxitin, and CCFA with cefoxitin is an initial formulation that can be used to isolate *C. difficile* strains [26–28]. The samples were cultured under anaerobic conditions in an anaerobic chamber (BACTRON anaerobic chamber; Sheldon Manufacturing, Inc., Cornelius, OR) containing an atmosphere of 90% N_2, 5% H_2, and 5% CO_2, at 37 °C. After incubation for more than 48 h, single colonies were obtained and transferred at least three times until considered pure. The 16S rRNA gene of the pure cultures was amplified using colony PCR with the bacterial universal primers 27F (5′-GTTTGATCCTGGCTCAG-3′) and 1492R (5′-TACGGYTACCTTGTTACGACTT-3′) [29]; identification of the cultures

Table 1 Patients with CDI who provided fecal samples for *C. difficile* isolation and the prescribed antibiotics

Subject no	Subject name	Age (y)	Sex	*C. difficile* strain isolated	Prescribed antibiotics class (product name)
1	Hwasun03	85	M	CBA7201	Nitroimidazole (Furacinil, Trizel, Furacinil), cephalosporin (Zenocef)
2	Gwangju02	79	M	CBA7202	Nitroimidazole (Furacinil), cephalosporin (Cetrazol), glycopeptide (Targocid, Vancozin, IV Vancomycin), fluoroquinolone (Levofexin), linezolid (Zyvox), penicillin (Tazoperan), polymyxin (Colis), glycylcycline (Tygacil), sulfonamide (Ceptrin), azoles (Diflucan), polyene antimycotic (PMS-Nystatin)
3	Hwasun11	63	F	CBA7203	Nitroimidazole (Furacinil), cephalosporin (Cetrazole)
4	Gwangju06	79	F	CBA7204	Nitroimidazole (Furacinil), glycopeptide (Vancocin, Teiconin, Vancomycin), penicillin (Tazoperan)
5	Hwasun12	61	F	CBA7205	Nitroimidazole (Furacinil), cephalosporin (Cetrazol), glycopeptide (Teiconin, IV Vancomycin), fluoroquinolone (Cravit), penicillin (Tazoperan), sulfonamide (Sevatrim, Septrin), carbapenem (Meropen), macrolide (Zithromax)
6	Hwasun13	77	F	CBA7206	Nitroimidazole (Furacinil), fluoroquinolone (Cravit), penicillin (Tazoperan), macrolide (Klaricid)
7	Hwasun15	65	M	CBA7207	Nitroimidazole (Furacinil), cephalosporin (Pacetin), glycopeptide (IV Vancomycin)
8	Hwasun16	62	F	CBA7208	Nitroimidazole (Trizel), glycopeptide (Vancomycin), penicillin (Augmentin)
9	Hwasun18	79	F	CBA7209	Cephalosporin (Cefazolin, Maxipime), glycopeptide (Teiconin, IV Vancomycin)

was performed based on the 16S rRNA gene sequences obtained from Sanger sequencing, with the EzBioCloud Database [30]. For comparative genomic analysis, nine *C. difficile* strains (designated CBA7201–CBA7209) were selected from each patient with CDI.

Genomic DNA extraction and whole-genome sequencing analysis

For genome sequencing of the selected nine *C. difficile* strains, cells were cultivated to the stationary phase in brain heart infusion (BD Biosciences, Franklin Lakes, NJ) broth medium at 37 °C and harvested by centrifugation. Genomic DNA was extracted and purified using MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) and MG Genomic DNA Purification Kit (MGmed, Seoul, Republic of Korea), followed by quantification with PicoGreen (Invitrogen, Carlsbad, CA). The genomes were then sequenced with the PacBio RS II System using single-molecule real-time (SMRT) sequencing technology based on a 20 kb library (Pacific Biosciences, Menlo Park, CA). Assembly was performed using the hierarchical genome assembly process 2 protocol in PacBio SMRT analysis v2.3.0 [31]. The whole-genome sequences of strains CBA7201–CBA7209 were deposited in GenBank (accession numbers QKRF00000000, QLNX00000000, QKRE00000000, CP029566, QLNY00000000, QLNZ00000000, QLOA00000000, QKRD00000000, and QLOB00000000, respectively) and automatically annotated by the NCBI prokaryotic genome annotation pipeline [32]. A total of 60 *C. difficile* strains including the nine strains isolated herein and 51 strains from the NCBI GenBank, were used for comparative genomic analysis. Data were obtained using the *Clostridium difficile* MLST Databases of PubMLST for sequence types (STs) and multi-locus sequence typing (MLST) clades (Table 2) [33].

Phylogenetic analyses of *C. difficile* genomes based on 16S rRNA gene and whole genome sequences

A phylogenetic tree based on the 16S rRNA gene sequences was constructed to infer the phylogenetic relationships among the 60 strains. The 16S rRNA gene sequences were aligned using the fast secondary-structure aware infernal aligner in the ribosomal database project [34]. For pan-genome and core-genome analysis between the 60 strains, the bacterial pan-genome analysis pipeline ver. 1.3 was used [35]. The core-genome of all *C. difficile* strains was extracted through all-against-all comparisons using the USEARCH (ver. 9.0) with a 50% sequence identity cut-off and their concatenated nucleotide sequences were aligned using the MAFFT program (ver. 7.407) available in the Roary pipeline [36]. The phylogenetic trees based on the aligned 16S rRNA gene sequences and the concatenated common gene sequences were constructed using the neighbor-joining (NJ) algorithm in the MEGA7 software [37]. Average nucleotide identity (ANI) and in silico DNA-DNA hybridization (DDH) analysis were used to assess the relatedness among the 60 *C. difficile* strains and two reference strains (*Clostridioides mangenotii* DSM 1289^T and *Clostridium hiranonis* TO-931). The pair-wise ANI value among the genomes was calculated using a stand-alone OrthoANI software [38]. Pair-wise in silico DDH was calculated using the genome-to-genome distance calculator

Table 2 General features of the 60 *C. difficile* genomes employed in this study[a]

Strain name (accession no.)	Genome status[a] (no. of contigs)	Total size[a] (Mb)	No. of genes[a]	G + C content[a](%)	MLST clade[b]	STs[b]	Sampling country
C. difficile CBA7201 (QKRF00000000)[c]	**C (3)**	**4.34**	**4,107**	**28.8**	**1**	**17**	**Korea**
C. difficile CBA7202 (QLNX00000000)[c]	**D (3)**	**4.40**	**4,204**	**28.8**	**1**	**17**	**Korea**
C. difficile CBA7203 (QKRE00000000)[c]	**D (5)**	**4.39**	**4,189**	**28.8**	**1**	**17**	**Korea**
C. difficile CBA7204 (CP029566)[c]	**C (1)**	**4.04**	**3,744**	**28.5**	**1**	**203**	**Korea**
C. difficile CBA7205 (QLNY00000000)[c]	**D (7)**	**4.40**	**4,176**	**28.9**	**1**	**17**	**Korea**
C. difficile CBA7206 (QLNZ00000000)[c]	**D (7)**	**4.16**	**3,911**	**28.7**	**1**	**8**	**Korea**
C. difficile CBA7207 (QLOA00000000)[c]	**D (2)**	**4.33**	**4,101**	**28.8**	**1**	**17**	**Korea**
C. difficile CBA7208 (QKRD00000000)[c]	**C (2)**	**4.08**	**3,786**	**28.5**	**1**	**4**	**Korea**
C. difficile CBA7209 (QLOB00000000)[c]	**D (2)**	**4.40**	**4,191**	**28.8**	**1**	**17**	**Korea**
C. difficile 630 (AM180355-6)	C (2)	4.29	3981	29.0	1	54	Switzerland
C. difficile DSM 27639 (CP011847)	C (1)	4.26	4026	29.1	1	54	Germany
C. difficile DSM 29745 (CP019857)	C (1)	4.24	3975	29.0	1	3	Germany
C. difficile DSM 29688 (CP019858)	C (1)	4.22	4041	28.9	1	15	Germany
C. difficile W0022a (CP025046)	C (1)	4.18	3942	28.9	1	2	USA
C. difficile DSM 29632 (CP019860)	C (1)	4.17	3891	28.6	1	103	Indonesia
C. difficile 08ACD0030 (CP010888)	C (1)	4.16	3907	28.8	1	2	Canada
C. difficile BR81 (CP019870)	C (1)	4.12	3813	28.7	1	42	Korea
C. difficile Mta-79 (CP042267)	C (1)	4.12	3844	28.7	1	34	USA
C. difficile DSM 28666 (CP012321)	C (1)	4.12	3829	29.0	1	48	Ghana
C. difficile DSM 29637 (CP016106)	C (1)	4.11	3844	28.6	1	83	Indonesia
C. difficile W0023a (CP025045)	C (1)	4.11	3804	28.7	1	42	USA
C. difficile FDAARGOS_267 (CP020424-6)	C (3)	4.28	4049	28.7	1	3	USA
C. difficile DH/NAP11/106/ST-42 (CP022524)	C (1)	4.08	3743	28.6	1	42	USA
C. difficile W0003a (CP025047)	C (1)	4.07	3765	28.6	1	8	USA
C. difficile 020477 (CP028524)	C (1)	4.14	3873	28.8	1	110	USA
C. difficile 020709 (CP028529)	C (1)	4.09	3792	28.5	1	21	USA
C. difficile QCD-63q42 (CM000637)	D (28)	4.44	4213	28.6	1	3	Canada
C. difficile DSM 1296^{T} (CP011968-9)	C (2)	4.28	3984	28.7	1	3	UK
C. difficile DSM 27638 (CP011846)	C (1)	4.22	3966	29.0	2	1	Germany
C. difficile DSM 27640 (CP011848)	C (1)	4.22	3964	29.0	2	1	Germany
C. difficile CD-17-01474 (CP026591)	C (1)	4.20	3939	28.9	2	1	Germany
C. difficile R0104a (CP025044)	C (1)	4.19	3924	28.7	2	1	USA
C. difficile 08-00495 (CP026594)	C (1)	4.17	3902	28.7	2	1	Germany
C. difficile 10-00253 (CP026598)	C (1)	4.12	3,844	28.6	2	1	Germany
C. difficile 12-00011 (CP026595)	C (1)	4.11	3,845	28.6	2	1	Germany
C. difficile 09-00072 (CP026599)	C (1)	4.11	3,842	28.6	2	1	Germany
C. difficile 10-00,078 (CP026597)	C (1)	4.11	3,845	28.6	2	1	Germany
C. difficile 10-00071 (CP026596)	C (1)	4.11	3843	28.6	2	1	Germany
C. difficile 12-00008 (CP026593)	C (1)	4.11	3842	28.6	2	1	Germany
C. difficile CD-10-00484 (CP026592)	C (1)	4.11	3843	28.6	2	1	Germany
C. difficile CD196 (FN538970)	C (1)	4.11	3807	28.6	2	1	France
C. difficile QCD-66c26 (CM000441)	D (15)	4.12	3774	28.5	2	1	Canada
C. difficile DSM 102860 (CP020379)	C (1)	4.25	4071	29.0	3	5	Germany
C. difficile DSM 102859 (CP020378)	C (1)	4.24	4050	29.0	3	5	Germany
C. difficile VL_0104 (FAAJ00000000)	D (261)	4.06	3,859	28.7	3	201	Canada
C. difficile VL_0391 (FALK00000000)	D (1,092)	4.16	3939	29.0	3	201	Canada
C. difficile ZJCDC-S82 (JYNK00000000)	D (20)	4.22	3934	29.1	3	5	China
C. difficile CDT4 (CP029152-3)	C (2)	4.28	4035	28.8	4	37	China
C. difficile DSM 29627 (CP016102)	C (1)	4.20	3916	28.8	4	37	Indonesia
C. difficile DSM 28670 (CP012312)	C (1)	4.19	3962	28.8	4	38	Ghana
C. difficile CD161 (CP029154-6)	C (3)	4.47	4277	28.8	4	37	China
C. difficile DSM 28669 (CP012323)	C (1)	4.13	3908	28.8	4	109	Ghana
C. difficile DSM 29629 (CP016104)	C (1)	4.11	3802	28.6	4	39	Indonesia
C. difficile M68 (FN668375)	C (1)	4.3	4025	28.9	4	37	Ireland
C. difficile DSM 29747 (CP019864)	C (1)	4.07	3809	29.1	5	11	Germany
C. difficile 12038 (CP033214-5)	C (2)	4.07	3822	28.8	5	11	China
C. difficile CD10010 (CP033213)	C (1)	4.04	3785	28.7	5	11	China
C. difficile M120 (FN665653)	C (1)	4.04	3756	28.7	5	11	UK
C. difficile CD21062 (CP033216-7)	C (2)	4.10	3898	28.9	5	11	China
C. difficile DSM 29,020 (CP012325)	C (1)	4.13	3905	29.2	5	11	Indonesia

Table 2 (continued)

[a] Bioinformatic genome analysis was carried out using the NCBI prokaryotic genome annotation pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/)

[b] MLST clade and ST information was obtained using PubMLST analysis (https://pubmlst.org/cdifficile/)

[c] Genomes sequenced in this study are highlighted in bold

[d] Genome status: *D* draft genome sequence, *C* complete genome sequence

2.1 [39]. In silico DDH values among the *C. difficile* strains were calculated and visualized using the GENE-E software (https://software.broadinstitute.org/GENE-E/).

Functional and pathogen-associated gene analysis of *C. difficile* strains

The amino acid sequences of 60 *C. difficile* strains were analyzed using GhostKOALA based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to obtain predicted protein annotation information [40]. The resulting KEGG Orthology (KO) numbers were summarized and visualized on the KEGG pathway using iPath2.0 [41]. Flagella assembly, pathogenicity locus (PaLoc) (*tcdRBEAC*), and binary toxin (*cdtAB*) genes in *C. difficile* strains were confirmed through BLASTP analyses using the reference protein sequences available in closely related *C. difficile* strains. Antibiotic resistance genes were identified using the Comprehensive Antibiotic Resistance Database (CARD) [42]. Nucleotide sequence similarity was calculated using EMBOSS Water, the pairwise sequence alignment tool provided by EMBL-EBI (https://www.ebi.ac.uk/Tools/psa/), against the nucleotide sequence of *C. difficile* 630 strain [43].

Quality assurance

Before genomic DNA extraction, the single colonies of each of strain CBA7201–CBA7209 were transferred three times in CCFA medium to obtain pure single colony. After obtaining the whole genome sequence of strain CBA7201–CBA7209, the sequences of the 16S rRNA genes, extracted using RNAmmer 1.21 server, were confirmed using the EzBioCloud database.

Results and discussion

Isolation and phylogenetic relatedness of *C. difficile* strains

A total of nine *C. difficile* strains (CBA7201–CBA7209) were isolated and selected for genomic analysis, as considering different isolation source and 16S rRNA gene sequence homology (Table1). After incubation for 24 h under anaerobic conditions on CCFA medium at 37 °C, the colony morphology of *C. difficile* strains appeared white or grayish-white and had an irregular radial shape. Although *C. difficile* is an anaerobic bacterium, strains CBA7201–CBA7209 were viable when exposed to aerobic conditions for 24 h. *C. difficile* can resist environmental stressors, such as exposure to oxygen, through spore formation. This stress-resistant feature may aid the spread of *C. difficile* in various environments [44, 45].

To assess the phylogenetic relationship between the *C. difficile* strains, a phylogenetic tree based on 16S rRNA gene sequences was constructed using the nine *C. difficile* isolates (CBA7201–CBA7209), 51 strains from GenBank, and two other closely related species, *C. mangenotii* DSM 1289^T and *Clostridium hiranonis* TO-931 (Fig. 1). For strains CBA7201–CBA7209, none were sufficiently different to be classified as strains from other species, as all 60 strains of *C. difficile* tested were grouped into one lineage that was distinct from the two outgroup strains [46, 47]. The 60 strains showed 99.9% 16S rRNA gene sequence similarity with the type strain *C. difficile* DSM 1296^T, and were thus, classified as *C. difficile.* However, the *C. difficile* strains and *Clostridium hiranonis* TO-931 were distinct, supporting the reclassification of *C. difficile* under the genus *Clostridioides* [11]. *C. difficile* was first classified as *Clostridium* because its characteristics (anaerobic, Gram-positive, and spore-forming) were similar to those of other *Clostridium* species. However, further studies using molecular methods indicated a diversity of organisms in the genus *Clostridium*, and 16S rRNA phylogenetic analysis confirmed that *C. difficile* had less than 97% similarity with other species from the genus *Clostridium*. Currently, the genus *Clostridioides* includes two species, *C. difficile* and *C. mangenotii* [11, 48].

Phylogenetic analysis based on the 16S rRNA gene, a molecular ecological marker, showed no differences among the 60 *C. difficile* strains, suggesting a limitation to the information this marker gene can provide [49]. To overcome this, comparative genome analysis with their entire genomes is conducted to identify the characteristics of various *C. difficile* strains [50, 51].

General features of the *C. difficile* genomes

The complete genomes of the nine *C. difficile* strains isolated in this study were obtained by performing whole-genome sequencing using the PacBio RS II System. The *C. difficile* CBA7201–CBA7209 genomes and 51 additional *C. difficile* genomes available in GenBank were compared and their general characteristics described in Table 2. The average genome size and gene numbers were 4.18 ± 0.1 Mb and 3927 ± 131, respectively. The genome of *C. difficile* CBA7204 was the smallest (4.04 Mb), whereas that of *C. difficile* CD161, which

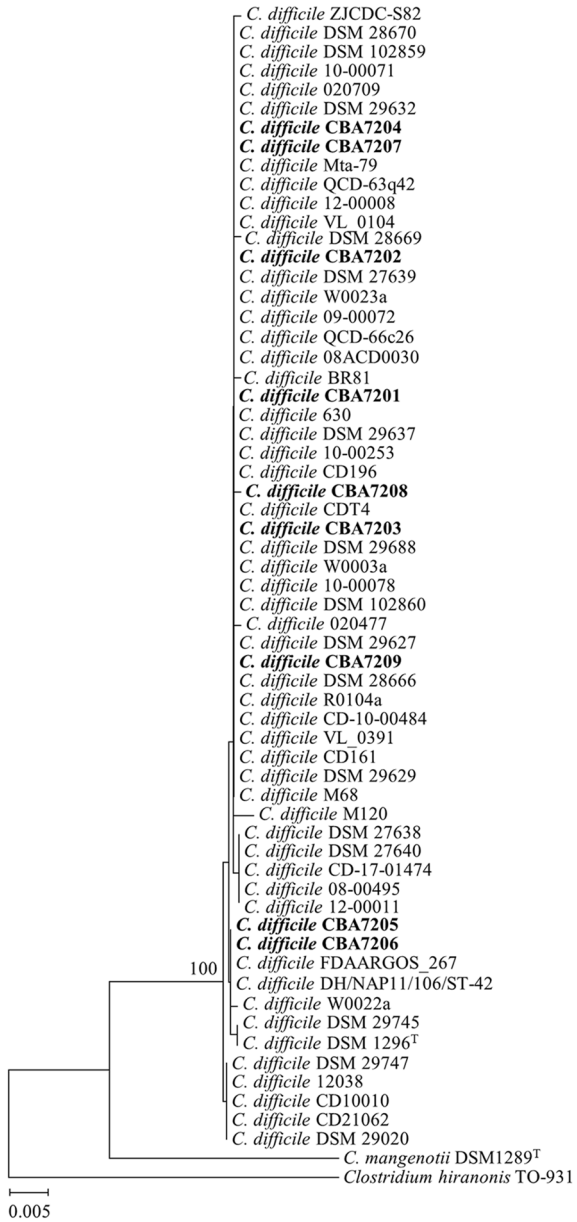

Fig. 1 Phylogenetic tree based on 16S rRNA gene sequences. 16S rRNA gene-based phylogenetic tree constructed using NJ algorithm showing phylogenetic relationships between 60 *C. difficile* strains with *C. mangenotii* DSM 1289^T and *Clostridium hiranonis* TO-931. Strain CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

was isolated in China and is known as a hypervirulent strain, was the largest (4.47 Mb) [52]. The G + C content of the *C. difficile* genomes ranged from 28.5% to 29.2%. The MLST scheme for *C. difficile* is based on the following seven highly conserved housekeeping genes: adenylate kinase (*adk*), ATP synthase subunit alpha (*atpA*), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*), serine hydroxymethyltransferase (*glyA*), recombinase A (*recA*), superoxide dismutase (*sodA*), and triosephosphate isomerase (*tpi*). STs are determined according to a combination of these seven housekeeping genes and are classified into five MLST clades (clade 1–5) [53]. To investigate genomic diversity, MLST clades and STs of the 60 strains of *C. difficile* were assigned using PubMLST and are listed in Table 2. A number of *C. difficile* strains, including CBA7201–CBA7209 and 19 reference strains, were assigned to MLST clade 1, which is consistent with the most frequently identified *C. difficile* strains worldwide [54]. Here, the MLST clade 1 belonging to the 28 strains, included 15 kinds of STs (ST2, ST3, ST4, ST8, ST15, ST17, ST21, ST24, ST42, ST48, ST54, ST83, ST103, ST110, and ST203) and had the most types of STs, making it the most diverse in terms of PCR ribotype (RT), which consists of a combination of STs and toxic genes encoding toxins A, B, and CDT. Among them, ST17 (*C. difficile* CBA7201, CBA7202, CBA7203, CBA7205, CBA7207, CBA7209) is associated with RT018 [55] which is the most prominent ribotype reported in hospitals in the Republic of Korea [56] and Japan [57–59]. RT018 is highly contagious and has been found to account for more than 95% CDI relapse cases. A study also revealed that patients with the RT018 were older than those with other RTs, and there was an association between the infectious RT018 and age [60]. Here, the most identified ST was ST1, which was identified in 14 out of the 60 strains and belongs to MLST clade 2; ST1 has been reported to be associated with an increased mortality rate of communicable diseases in North America and Europe [61–63].

Phylogenetic relatedness of *C. difficile* strains based on pan- and core-genome analysis

Pan-genome analysis is a useful tool for effectively analyzing and expressing the genomic characteristics of bacteria. Through this analysis, we found 5,814 genes in pan-genome and 1,660 genes in core-genome across the 60 strains (Additional file 1: Figure S1), with the number of total unique genes being 643. The curve analysis based on the Heaps' law regression model showed that the pan-genome was open ($B_{pan} = 0.14$), indicating that more sequenced strains are needed to capture the complete gene complement [64]. The number of accessory and unique genes in the 60 strains are listed in Additional file 2: Table S1. OrthoANI values showed the pairwise relatedness of the nine *C. difficile* strains isolated in this study with the reference strains (Additional file 2: Table S2). This result suggests that the nine Korean strains are not part of a common clone. Moreover, the unique genes and different OrthoANI values reflect the existence of evolutionary differences and ecological niches for each

strain that help it adapt to varying environmental stressors. To further investigate the phylogenic relationships among the *C. difficile* strains, we constructed a phylogenetic tree based on the amino acid sequences of 1,660 core genes. Despite being the same species, the core gene-based phylogenetic tree indicated that the strains were divided into five groups (Fig. 2). Unlike the 16S rRNA-based phylogenetic tree, the difference between the strains in the core gene-based phylogenetic tree was more clearly distinguishable. This classification is consistent with the hierarchical classification based on the in silico DDH (Additional file 1: Fig. S2), as well as clustering based on the MLST clade (Fig. 2; vertical bar on the right side). These results indicate that MLST clade-based classification using the seven conserved housekeeping genes is a very efficient method for distinguishing *C. difficile* strains. According to the core gene-based phylogenetic tree analysis, a 28-genomes group containing *C. difficile* strains CBA7201–CBA7209, and strain 630 was consistent with MLST clade 1. Among them, *C. difficile* CBA7201, CBA7202, CBA7203, CBA7205, CBA7207, and CBA7209 formed a distinct group, indicating similar genomic features. The other groups were clustered into MLST clades 2, 3, 4, and 5. Strains belonging to clade 5 in the core gene phylogenetic tree were significantly divergent from the other clade 1–4 strains, indicating that clade 5 strains, which were all ST11 (Table 2), may have undergone different evolutionary processes [50, 65].

Functional category analysis of *C. difficile* genome

To identify the general metabolic diversity, functional features, and virulence factors of the *C. difficile* strains, KEGG analysis was carried out using the core and accessory genes (Fig. 3). All core and accessory genes were most frequently classified under amino acid metabolism, carbohydrate metabolism, and membrane transport. The terms amino acid metabolism and carbohydrate metabolism were more abundant in the core genes than in the accessory genes, which is a common feature of *C. difficile*. In contrast, membrane transport (phosphor transferase system, ABC transporter, and bacterial secretion system) was relatively abundant in the accessory genes, indicating that the bacteria can absorb or secrete various substances depending on the strain. Among the genes assigned to the human disease category, drug resistance genes were relatively abundant in the accessory genes, suggesting that *C. difficile* strains have been exposed to various antibiotics or antibiotic-resistant genes and that their antibiotic resistance was obtained differently depending on the strains [66].

Using CARD, antibiotic resistance-related genes of the 60 *C. difficile* strains are summarized in Table 3 and Additional file 2: Table S3. The 60 *C. difficile* strains contained at least one resistance gene among the 20 Antibiotic Resistance Ontology (ARO) terms; this gene is associated with resistance to erythromycin and clindamycin, which belong to the macrolide and lincosamide class of antibiotics, respectively (ARO term: *C. difficile* 23S rRNA with mutation conferring resistance to erythromycin and clindamycin). Most cases of resistance to these antibiotics can be associated with alterations in nucleotides of 23S rRNA, a component of the large ribosomal subunit [67]. The resistance of *C. difficile* strains to erythromycin and clindamycin has been confirmed in previous CDI-related studies [27]. Moreover, clindamycin has been reported to pose a risk as it promotes CDI; thus, care must be taken when prescribing it [66].

Interestingly, *C. difficile* has different antibiotic resistance genes depending on the MLST clade. Most strains in MLST clade 1–3 possessed resistance genes against vancomycin ($vanR_G$, $vanXY_G$), a class of glycopeptide antibiotics. The $vanR_G$ is a *vanR* variant and $vanXY_G$ is a variant of *vanXY* found in the *vanG* gene cluster. Resistance of enterococci to glycopeptides was reported first [68], after which nine genotypes associated with this resistance were identified. The *vanG* is one of the nine genotypes (*vanA*, *vanB*, *vanC*, *vanD*, *vanE*, *vanG*, *vanL*, *vanM*, and *vanN*) that are involved in glycopeptide antibiotics resistance. It has been reported that similar *vanG* gene clusters also exist in *C. difficile*. The *vanG* phenotype is known to correspond to low-level resistance to vancomycin, which results from the acquisition of two *vanG* operons, *vanG1* and *vanG2*. The $vanR_G$ gene is one of three regulatory genes of *vanG1*, while the $vanXY_G$ gene is one of five effector genes of *vanG1* [69]. However, the similarity of the $vanR_G$ and $vanXY_G$ genes was 77.45%–77.87% and 58.82%–59.22% in CARD, respectively (Additional file 2: Table S3), suggesting that glycopeptide resistance by these two genes is not expected to function properly. Some of the strains belonging to the MLST clade 4 and 5 did not possess resistance genes of glycopeptide antibiotics, but had other resistance genes [ARO term: AAC(6')-Ie-APH(2")-Ia, aad(6), ANT(6)-Ia, ANT(6)-Ib, APH(3')-IIIa, catI, SAT-4, tet(40), tet(44), tet(W/N/W), tetB(P), tetM, tetO)] specific for antibiotics that inhibit protein synthesis by 30S ribosomal subunits, such as aminoglycoside and tetracycline antibiotics [70, 71]. These findings indicate that different antibiotics can be prescribed depending on the MLST clade of the strain causing the infection. Compared with other strains, the nine *C. difficile* strains isolated from Korean patients with CDI possessed a higher number of antibiotic-resistance genes. In some cases, up to 11 antibiotics were prescribed for a patient, Gwangju02 (Table 1), and *C. difficile* strains with the genes resistant to the prescribed antibiotics were isolated from a total of seven patients. Therefore, more

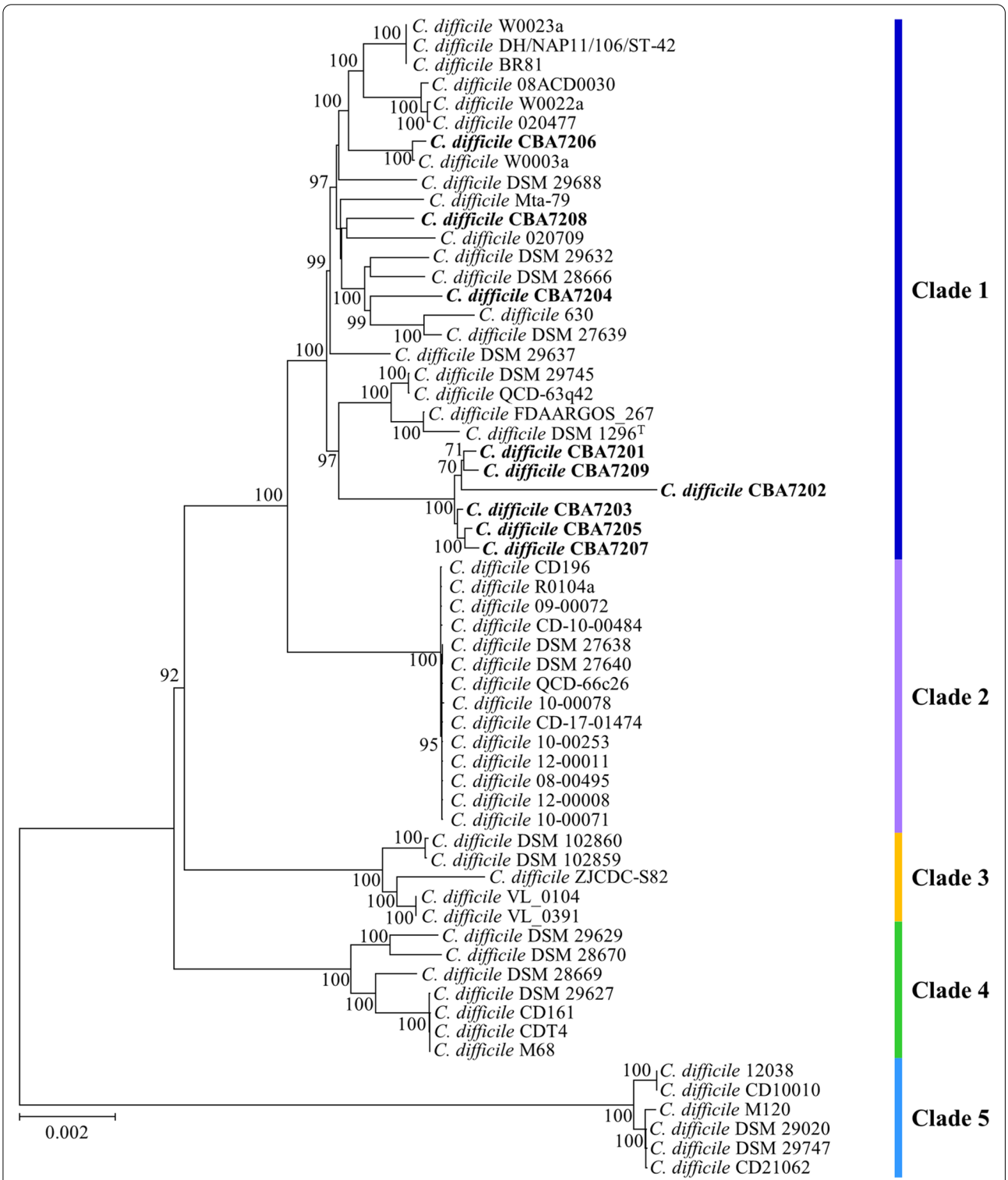

Fig. 2 Phylogenetic tree based on core-gene sequences. Phylogenetic tree generated from the amino acid sequences of the 60 *C. difficile* core genomes showing the relationship between each strain. The MLST clade and ST data were obtained from the PubMLST analysis. The vertical bar on the right indicates the MLST clade to which each *C. difficile* strain belongs. Strain CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

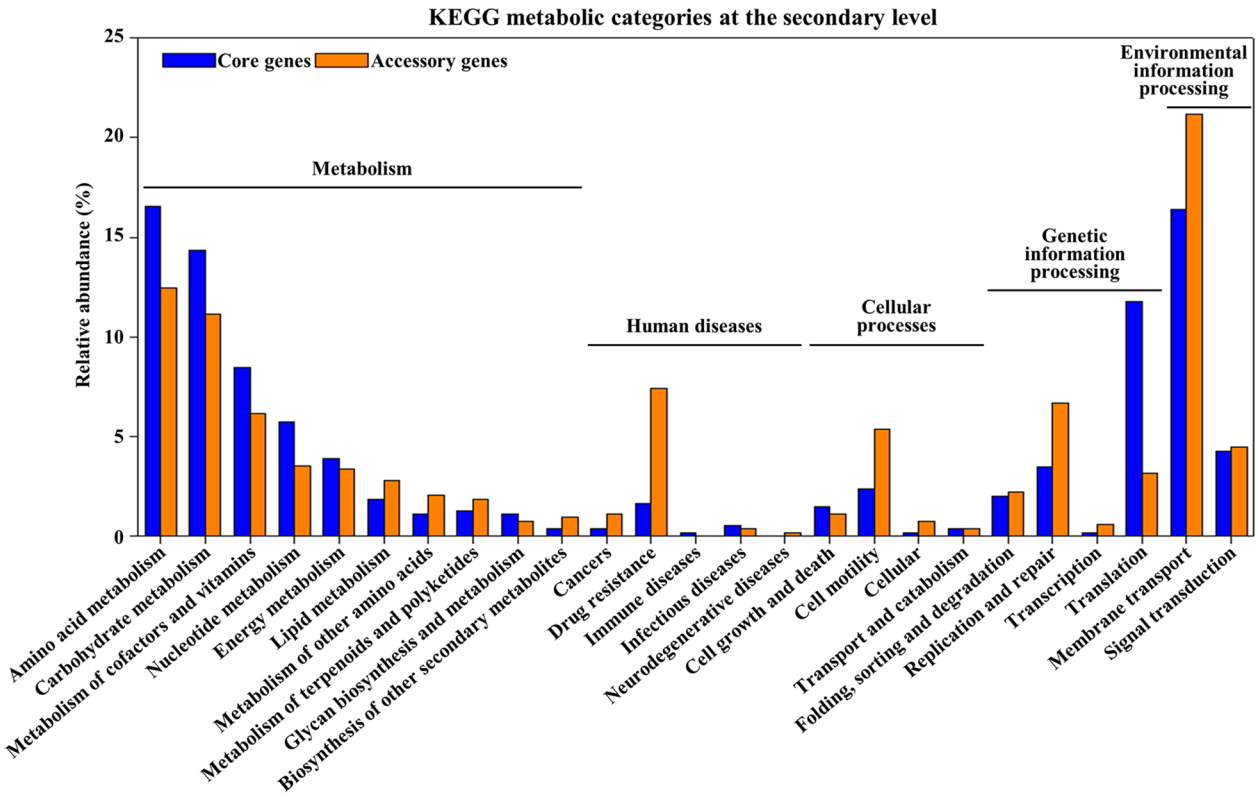

Fig. 3 KEGG functional categories at the secondary levels. Comparison of KEGG functional categories at the secondary levels in core and accessory genes of the 60 *C. difficile* genomes

care should be taken when prescribing antibiotics to prevent persistent CDI and emergence of multidrug-resistant pathogens.

The cell motility category genes were relatively abundant among the accessory genes, indicating that genes associated with cell motility are different depending on the *C. difficile* strain. Most genes assigned to the flagellar assembly in the cell motility category were classified into the core, accessory genes (Additional file 1: Figure S3). The core, soft-core and accessory genomes refer to the set of genes for all 60 strains, 57–59 strains and the remaining 2–56 strains, respectively. There was no soft-core genome identified in the data set. In the flagellar assembly pathway, the genes encoding flagellar motor rotation proteins (*MotA* and *MotB*), flagellin filament structural proteins (*FliC*), flagellar cap proteins (*FliD*), flagellar hook-associated proteins (*FlgL* and *FlgK*), and flagellar secretion chaperone proteins (*FliS*), were found in the core genome of *C. difficile* strains. On the other hand, genes encoding flagellar hook protein (*FlgE*), flagellar basal-body rod modification protein (*FlgD*), flagellar basal-body rod protein (*FlgB*, *FlaC*, and *FlaG*), flagellar hook-basal body complex protein (*FliE*), flagellar M-ring protein (*FliF*), flagellar motor switch protein (*FliG*, *FliM* and *FliN*), flagella biosynthesis protein (*FlhA* and *FlhB*), and flagella assembly protein (*FliH*) and flagellar biosynthetic protein (*FliQ*), were found in the accessory genome of *C. difficile* strains. These results indicate that these genes are well conserved among strains and that flagellar construction as well as attachment and invasion of intestinal epithelial cells, are essential for *C. difficile* infection [72, 73]. Given that flagella motility can affect adhesion and colonization of intestinal epithelial cells, *C. difficile* flagella contribute to pathogenicity and result in mucosal damage and inflammatory responses in the host [74].

PaLoc and CDT locus (CdtLoc) of *C. difficile*

Toxins A and B are encoded by the *tcdA* (enterotoxin) and *tcdB* (cytotoxin) genes located on a chromosomal region called the PaLoc (19.6 kb) [75]. The PaLoc structure consists of the *tcdA* and *tcdB* genes sandwiched between the *tcdR* (positive regulator) and *tcdC* (negative regulator) genes, with the *tcdE* gene (toxin secretion) located between the two toxin genes (Fig. 4). Among the 60 strains, all possessed these toxin genes except for the seven non-toxigenic strains (CBA7204, DSM 29688, DSM 28666, DSM 29637, DSM 28670, DSM 28669, and DSM 29629). However, the sequence similarity and

Table 3 Antibiotics resistance-related genes of *C. difficile* strains based on the CARD

ARO term	AMR gene family	Drug class	Resistance mechanism	*C. difficile* strain
C. difficile 23S rRNA with mutation conferring resistance to erythromycin and clindamycin	23S rRNA with mutation conferring resistance to macrolide antibiotics	Macrolide, lincosamide	Antibiotic target alteration	All 60 strains
ErmB	Erm 23S ribosomal RNA methyltransferase	Macrolide, lincosamide, streptogramin	Antibiotic target alteration	**CBA7201**–**CBA7203**, **CBA7205**, **CBA7207**, **CBA7209**, 630, DSM 27639, DSM 29745, DSM 29688, DSM 28666, W0023a, R0104a, CD161, DSM 28669, M68
vanRG	Glycopeptide resistance gene cluster, *vanR*	Glycopeptide	Antibiotic target alteration	**CBA7201**–**CBA7209**, 630, DSM 27639, DSM 29745, DSM 2,688, W0022a, DSM 29632, 08ACD0030, BR81, Mta-79, DSM 28666, DSM 29637, W0023a, FDAAR-GOS267, DH/NAP11/106/ST-42, W0003a, 20477, 20709, QCD-63q42, DSM 1296^{T}, DSM 27638, DSM 27640, CD-17-01474, R0104a, 08-00495, 10–00253, 12-00011, 09-00072, 10-00078, 10-00071, 12-00008, CD-10-–00484, CD196, QCD-66c26
vanXYG	Glycopeptide resistance gene cluster, *vanXY*	Glycopeptide	Antibiotic target alteration	**CBA7201**, **CBA7203**–**CBA7209**, 630, DSM 27639, DSM 29745, DSM 29688, W0022a, DSM 29632, 08ACD0030, BR81, Mta-79, DSM 28666, DSM 29637, W0023a, FDAAR-GOS267, DH/NAP11/106/ST-42, W0003a, 20477, 20709, QCD-63q42, DSM 1296^{T}, DSM 27,638, DSM 27,640, CD-17–01,474, R0104a, 08-00495, 10-00253, 12-00011, 09-00072, 10-00078, 10-00071, 12-00008, CD-10-00484, CD196, QCD-66c26
Clostridioides difficile gyrA conferring resistance to fluoroquinolones	Fluoroquinolone resistant *gyrA*	Fluoroquinolone	Antibiotic target alteration	**CBA7201–CBA7203**, **CBA7205–CBA7207**, **CBA7209**, DSM 29745, QCD-63q42, DSM 27638, DSM 27,640, CD-17-01474, R0104a, 08-00495, 10-00253, 12-00011, 10-00078, 10–00,071, 12-00008, CD-10-00484, QCD-66c26, VL 0104, VL 0391, CDT4, CD161, DSM 29747
cdeA	Multidrug and toxic compound extrusion (MATE) transporter	Fluoroquinolone, acridine dye	Antibiotic efflux	**CBA7209**, 630, DSM 29745
APH (2″)-If	APH(2″)	Aminoglycoside	Antibiotic inactivation	**CBA7201**–**CBA7203**, **CBA7205**, **CBA7207**, **CBA7209**, DSM 27638, DSM 27640
AAC(6′)-Ie-APH(2″)-Ia	APH(2″), AAC(6′)	Aminoglycoside	Antibiotic inactivation	630, ZJCDC-S82, CD161, M68
aad(6)	ANT(6)	Aminoglycoside	Antibiotic inactivation	DSM 29020
ANT(6)-Ia	ANT(6)	Aminoglycoside	Antibiotic inactivation	M120
ANT(6)-Ib	ANT(6)	Aminoglycoside	Antibiotic inactivation	M120
APH(3′)-IIIa	APH(3′)	Aminoglycoside	Antibiotic inactivation	CD21062

Table 3 (continued)

ARO term	AMR gene family	Drug class	Resistance mechanism	*C. difficile* strain
catI	Chloramphenicol acetyltransferase (CAT)	Phenicol	Antibiotic inactivation	CD21062
SAT-4	Streptothricin acetyltransferase (SAT)	Nucleoside	Antibiotic inactivation	DSM 29020
tet(40)	Major facilitator superfamily (MFS) antibiotic efflux pump	Tetracycline	Antibiotic efflux	DSM 29747, DSM 29020
tet(44)	Tetracycline-resistant ribosomal protection protein	Tetracycline	Antibiotic target protection	M120
tet(W/N/W)	Tetracycline-resistant ribosomal protection protein	Tetracycline	Antibiotic target protection	630
tetB(P)	Tetracycline-resistant ribosomal protection protein	Tetracycline	Antibiotic target protection	DSM 28669
tetM	Tetracycline-resistant ribosomal protection protein	Tetracycline	Antibiotic target protection	DSM 28666, CDT4, DSM 29627, CD161, M68, DSM 29747, M120, CD21062, DSM 29020
tetO	Tetracycline-resistant ribosomal protection protein	Tetracycline	Antibiotic target protection	CD21062

ARO antibiotic resistance ontology, *AMR* antimicrobial resistance

Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

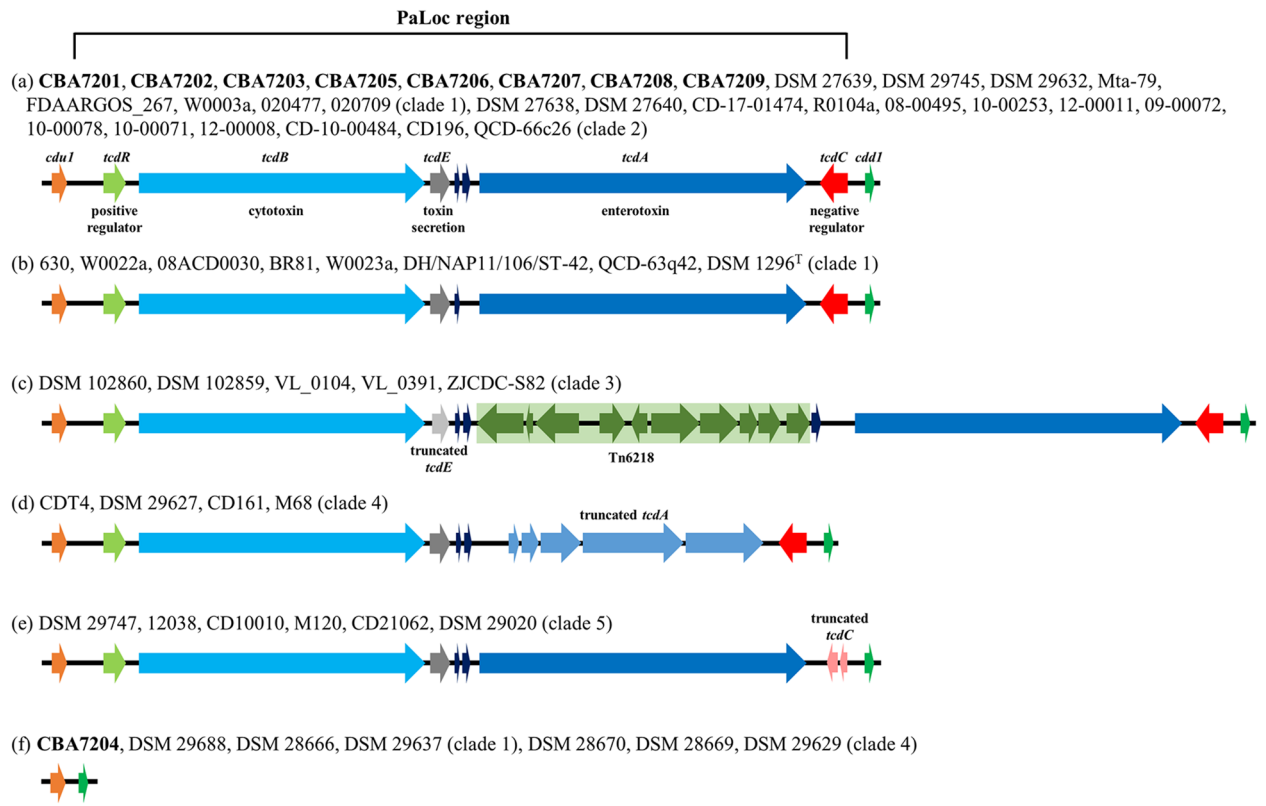

Fig. 4 Schematic representation of the PaLoc region and lateral genes. The PaLoc consisting of the *tcdR*, *tcdB*, *tcdE*, *tcdA*, and *tcdC* genes in this order. **a**–**f** PaLoc with two hypothetical protein-coding genes between the *tcdE* and *tcdA* genes (**a**); one hypothetical protein-coding gene between *tcdE* and *tcdA* genes (**b**); one or two hypothetical protein-coding genes and the mobile genetic elements (MGEs) Tn6218 between *tcdE* and *tcdA* gene (**c**); five truncated *tcdA* genes (**d**); two truncated *tcdC* genes (**e**); and non-toxigenic strains (**f**). Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

structural differences were strain-dependent. All Korean *C. difficile* strains except for strain CBA7204 were found to possess the structure including all of the genes of PaLoc region with two hypothetical protein-coding genes between *tcdE* and *tcdA* genes, as shown in Fig. 4a; 30 strains including 16 strains belonging to MLST clade 1 and 14 strains belonging to clade 2, had the same PaLoc region structure. The remaining strains of clade 1 including strain 630, had only one hypothetical protein-coding gene (Fig. 4b). The insertion of the mobile genetic elements (MGEs) Tn6218, which contains an macrolide, lincosamide and streptogramin-associated antibiotic resistance gene [51, 66] between the truncated *tcdE* and *tcdA* genes, is a common characteristic of clade 3 *C. difficile* strains (Fig. 4c) [76]. Some clade 4 strains face issues with enterotoxin expression due to truncated *tcdA* genes (Fig. 4d) [1]. Meanwhile, clade 5 *C. difficile* strains have a truncated *tcdC* gene, indicating difficulties in suppressing toxin production; thus these strains may become hypervirulent *C. difficile* strains (Fig. 4e) [24]. Four strains including strain CBA7204 belonging to clade 1 and three strains belonging to clade 4 were identified not to have the PaLoc genes (Fig. 4f).

Sequence similarities of PaLoc genes between strain 630 that is the most used reference strain in *C. difficile* genomic analysis and other strains, were compared [21]. Compared with *C. difficile* 630, strains CBA7201, CBA7202, CBA7203, CBA7205, CBA7207, and CBA7209 exhibited similar PaLoc sequence similarities and formed a unique group (Fig. 5). Among the strains belonging to clade 1, except for the non-toxigenic strains and strain QCD-63q42, each PaLoc gene showed a similarity of 99.3% or more. Strain QCD-63q42 had 84.7% similarity of *tcdA* gene with strain 630. As shown in Fig. 4, the strains in clade 2 had the same gene size as the strains in clade 1, but the similarity values of *tcdB* (93.5%), *tcdA* (98.5%), and *tcdC* (95.7%) genes were slightly lower than those in clade 1. In the case of the strains belonging to clade 3, the sequence similarities of *tcdR* (97.7–97.8%), *tcdB* (98.6–98.7%), *tcdA* (98.5–98.8%), and *tcdC* (90.7–95.7%) genes were found. In the case of *tcdE* gene of the strains in clade 3, the gene size was 424 bp (Fig. 4c), which was

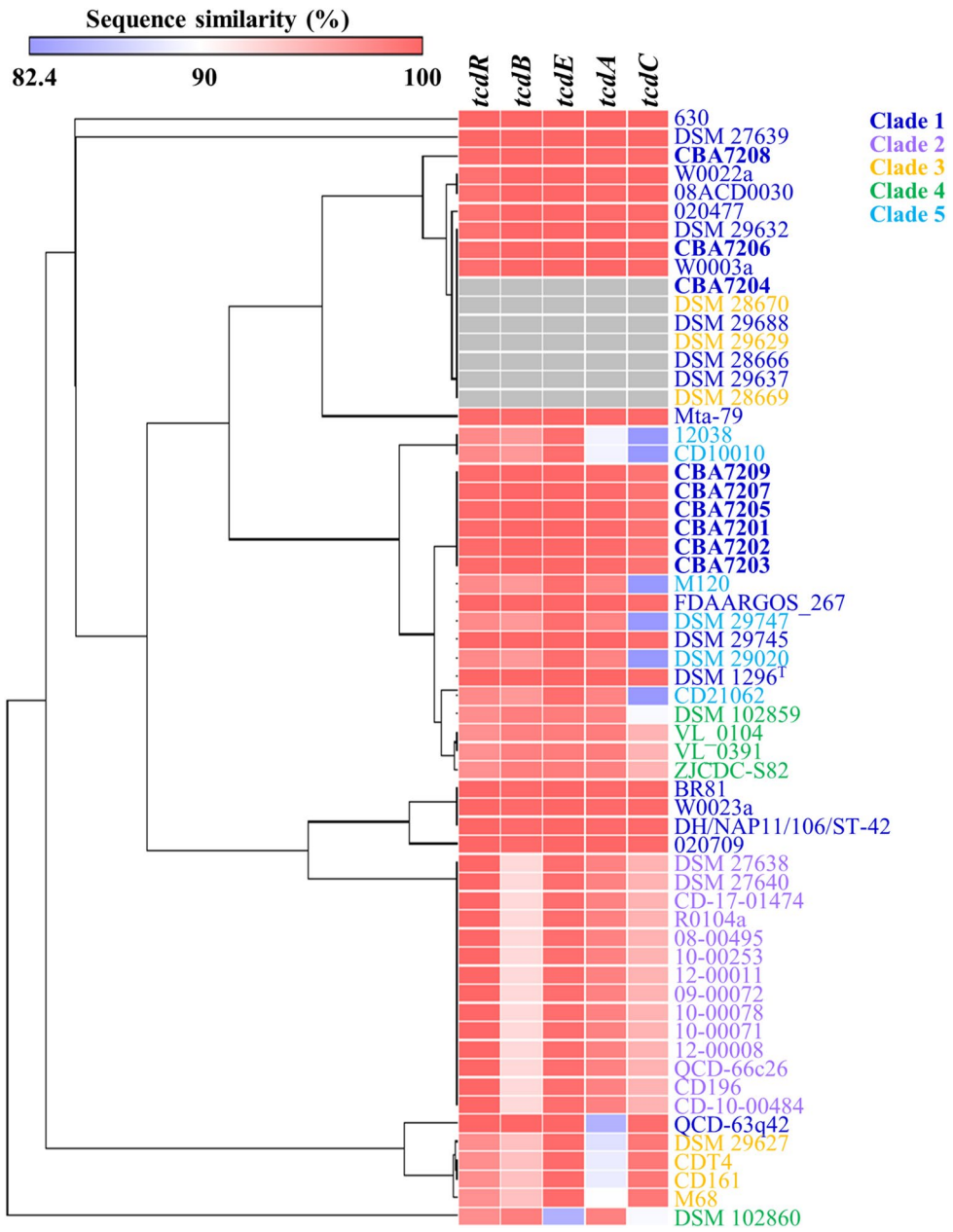

Fig. 5 Heat map and hierarchical cluster analysis of PaLoc region genes. A heat map illustrating the sequence similarity. Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold. Strain names of *C. difficile* strains are indicated differently depending on the MLST clade: navy blue (clade 1), purple (clade 2), yellow (clade 3), green (clade 4), and sky blue (clade 5). Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

shorter than the gene (501 bp) of strain 630, but the similarity of the gene was 98.8%, which was relatively high compared to the rest of the PaLoc genes. In the case of clade 4, 3 of the 7 strains were non-toxigenic (Fig. 4f) and the remaining strains showed low similarity value of *tcdR* (97.8%) and *tcdB* (94.9%) genes. The truncated *tcdA* gene of clade 4 had a similarity value of 88.6–89.6% (Fig. 4d). In the case of clade 5, it was confirmed that the truncated *tcdC* gene had a particularly low similarity with a value of 82.4% (Fig. 4e).

The actin-ADP-ribosylating toxin is located on a chromosomal region CdtLoc. *C. difficile* strains carrying the CDT gene may exhibit stronger virulence due to interaction with the existing toxins A and B [16–19]. The

structure of the CdtLoc region consists of a binary toxin regulatory gene (*cdtR*) and binary toxin genes (*cdtA* and *cdtB*). These three genes differ in sequence similarity and structure depending on the *C. difficile* strains (Fig. 6). Some clade 1 strains and all clade 4 strains were not found to possess a CdtLoc region (Fig. 6a), while other clade 1 strains may show difficulty in producing the toxin due to gene truncation (Fig. 6b–f); for instance, strain DSM 27639 has an inserted gene (approximately 30 kb) between the *cdtA* and *cdtB* genes (Fig. 6d) [77]. However, all strains belonging to clade 2, 3, and 5 possessed an intact CdtLoc region (6.2 kb), indicating the normal expression of toxin-producing genes (Fig. 6g).

Common features of the six *C. difficile* strains isolated from Korean patients with CDI

Finally, we identified several common features among *C. difficile* CBA7201, CBA7202, CBA7203, CBA7205, CBA7207, and CBA7209 isolated from Korean patients with CDI. These strains were not single clones, as evidence by differences between the strains confirmed via phylogenetic analysis; core, accessory, and unique gene analysis; and OrthoANI values (Fig. 3 and Additional file 2: Table S1 and S2). Nevertheless, they share several common features. All the six strains belong to clade 1 and ST17 (Table 2) and possess similar antibiotic-resistance genes (Table 3). In addition, they exhibited similar toxin gene expression in terms of PaLoc and CdtLoc structure (Figs. 4, 5, 6). Interestingly, the cephalosporins such as Zenocef, Cetrazol, Pacetin, and Cefazolin, were commonly prescribed to the six patients with CDI, from whom these six strains were isolated (Table 1). However, the cephalosporin resistance gene was not detected in all six strains isolated from cephalosporin-containing medium (Table 3). In a previous case study, the cephalosporins presented a risk factor to patients with CDI, and the decrease in cephalosporin prescription rate was related to a decrease in diarrhea cases associated with *C. difficile* [9, 78–81]. Therefore, further studies are needed to elucidate the association

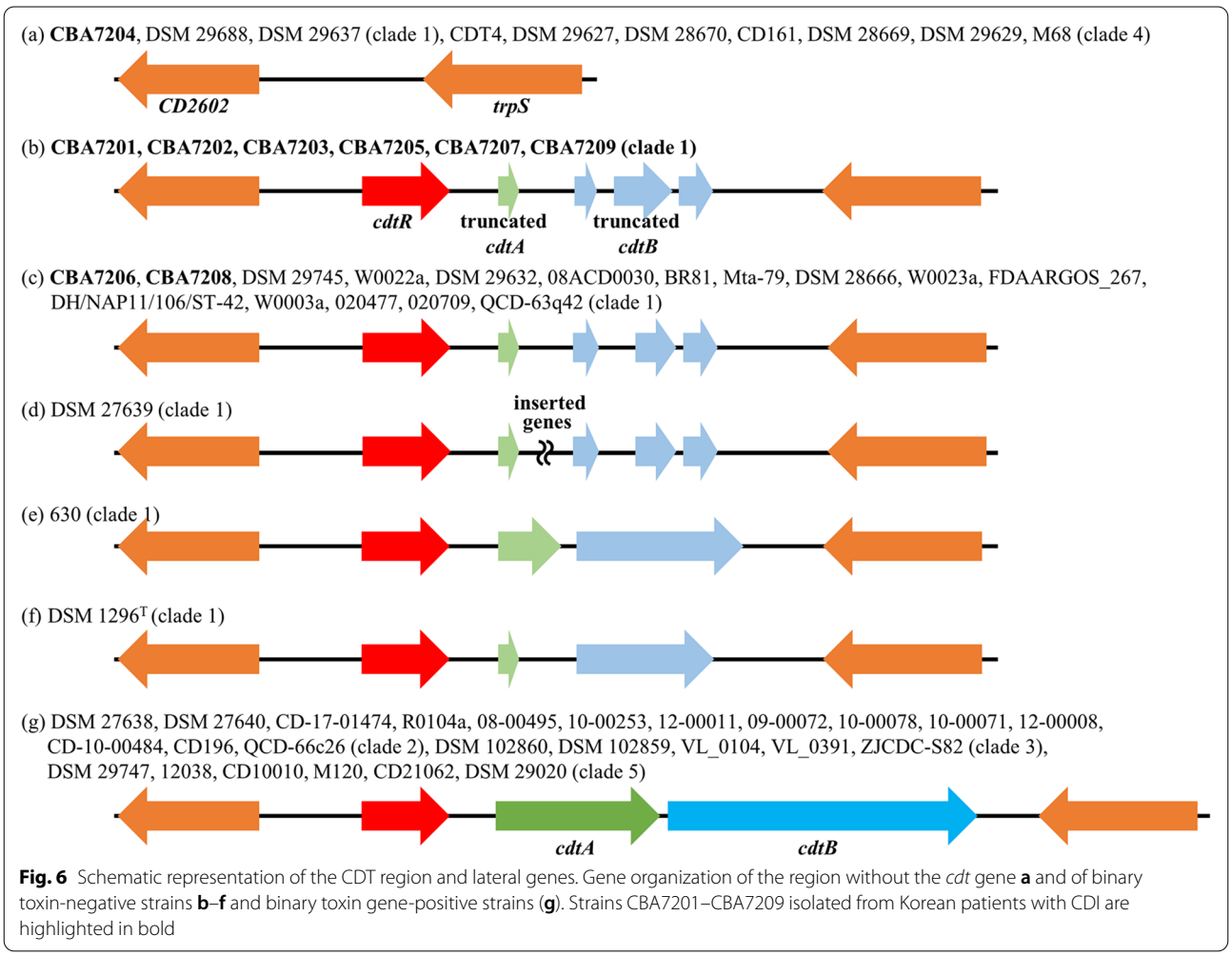

Fig. 6 Schematic representation of the CDT region and lateral genes. Gene organization of the region without the *cdt* gene **a** and of binary toxin-negative strains **b**–**f** and binary toxin gene-positive strains (**g**). Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold

among antibiotics, *C. difficile* strains, and patients with CDI.

Conclusion

In this study, we investigated the genomic, phylogenetic, functional, and pathogenic features of nine *C. difficile* strains isolated from Korean patients and performed a comparative genomic analysis with other strains isolated from various countries. Along with the identified genomic features of Korean *C. difficile* isolates, accumulation of more whole-genome sequence information of diverse *C. difficile* strains could serve as basic information for CDI prevention and treatment in Korea.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13099-021-00451-3.

Additional file 1: Fig. S1. Pan- and core-genome box plot of 60 *C. difficile* strains with standard deviations. The pan-genome represents the total set of genes of the 60 *C. difficile* strains, while the core-genome represents the common genes across all genomes. **Fig. S2.** *In silico* DNA-DNA hybridization (DDH) analyses showing the pair-wise relatedness of 60 *C. difficile* strains and two reference strains (*C. mangenotii* DSM 1289^T and *Clostridium hiranonis* TO-931). Strains CBA7201–CBA7209 isolated from Korean patients with CDI are highlighted in bold. The hierarchical clusters represented by dendrograms were constructed by simple linkage of the *in silico* DDH values. The vertical bar on the right side of the figure indicates the MLST clade to which each *C. difficile* strain belongs. **Fig. S3.** Diagram of the structural components involved in *C. difficile* flagella assembly defined by the number of KEGG orthology genes identified from the genomes of 60 *C. difficile* strains. Flagella assembly genes belonging to the core genome of the 60 *C. difficile* strains are indicated in red; flagella assembly genes belonging to the accessory genome identified from 2–56 genomes are indicated in blue.

Additional file 2: Table S1. The number of the core-genes, accessory-genes (present in more than two strains), and unique-genes present in 60 *C. difficile* strains. **Table S2.** OrthoANI (average nucleotide identity) analyses showing the pair-wise relatedness of 51 *C. difficile* strains and two reference strains (*C. mangenotii* DSM 1289T and *Clostridium hiranonis* TO-931) for nine *C. difficile* strains isolated from Korea. **Table S3.** Lists of antibiotics resistance gene present in 60 *C. difficile* strains.

Acknowledgements
Not applicable.

Authors' contributions
SWA and SHL contributed to conducting experiments, performing genome analysis, and writing manuscript. HJC, HEC, SJK, and SWR contributed to the conception and design of the study. UJK and HCJ contributed to sampling and collecting medical metadata. SWR contributed to the revision of the manuscript and approved the submitted version. All authors contributed to the article and approved the submitted version. All authors read and approved the final manuscript.

Funding
This research was supported by a grant from the World Institute of Kimchi (KE2102-1-2) funded by the Ministry of Science and ICT; a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (HI20C0079040020); a grant of Bio & Medical Technology Development Program of the NRF funded by the Korean government (NRF-2019M3E5D1A02067959), Republic of Korea.

Availability of data and materials
The complete genome data of strain CBA7201–CBA7209 has been deposited in DDBJ/EMBL/GenBank, with accession numbers QKRF00000000, QLNX00000000, QKRE00000000, CP029566, QLNY00000000, QLNZ00000000, QLOA00000000, QKRD00000000, and QLOB00000000, respectively.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the institutional review boards of the Republic of Korea centers for disease control and prevention [IRB file no. CNUH-2017-161 and CNUHH-2017-076]. Written informed consent was obtained from all participants.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Microbiology and Functionality Research Group, World Institute of Kimchi, 86, Kimchi-ro, Nam-gu, 61755 Gwangju, Republic of Korea. [2]Department of Infectious Diseases, Chonnam National University Hospital, 61469 Gwangju, Republic of Korea. [3]Department of Microbiology, Chonnam National University Medical School, 61469 Gwangju, Republic of Korea.

Received: 6 July 2021 Accepted: 8 September 2021
Published online: 16 September 2021

References

1. Chandrasekaran R, Lacy DB. The role of toxins in *Clostridium difficile* infection. FEMS Microbiol Rev. 2017;41(6):723–50.
2. Lessa FC, Winston LG, McDonald LC. Emerging infections program CdST: burden of *Clostridium difficile* infection in the United States. N Engl J Med. 2015;372(24):2369–70.
3. van Dorp SM, Notermans DW, Alblas J, Gastmeier P, Mentula S, Nagy E, Spigaglia P, Ivanova K, Fitzpatrick F, Barbut F, et al. Survey of diagnostic and typing capacity for *Clostridium difficile* infection in Europe, 2011 and 2014. Euro Surveill. 2016. https://doi.org/10.2807/1560-7917.ES.2016.21.29.30292.
4. Borren NZ, Ghadermarzi S, Hutfless S, Ananthakrishnan AN. The emergence of *Clostridium difficile* infection in Asia: a systematic review and meta-analysis of incidence and impact. PLoS ONE. 2017;12(5):e0176797.
5. Lee YJ, Choi MG, Lim CH, Jung WR, Cho HS, Sung HY, Nam KW, Chang JH, Cho YK, Park JM, et al. Change of *Clostridium difficile* colitis during recent 10 years in Korea. Korean J Gastroenterol. 2010;55(3):169–74.
6. Choi HY, Park SY, Kim YA, Yoon TY, Choi JM, Choe BK, Ahn SH, Yoon SJ, Lee YR, Oh IH. The epidemiology and economic burden of *Clostridium difficile* infection in Korea. Biomed Res Int. 2015;2015:510386.
7. Bartlett JG, Gerding DN. Clinical recognition and diagnosis of *Clostridium difficile* infection. Clin Infect Dis. 2008;46(Suppl 1):S12-18.
8. Wu MA, Leidi F. Gruppo di Autoformazione M: Vancomycin vs Metronidazole for *Clostridium difficile* infection: focus on recurrence and mortality. Intern Emerg Med. 2017;12(6):871–2.
9. Nelson DE, Auerbach SB, Baltch AL, Desjardin E, Beck-Sague C, Rheal C, Smith RP, Jarvis WR. Epidemic *Clostridium difficile*-associated diarrhea: role of second- and third-generation cephalosporins. Infect Control Hosp Epidemiol. 1994;15(2):88–94.
10. Kuijper EJ, de Weerdt J, Kato H, Kato N, van Dam AP, van der Vorm ER, Weel J, van Rheenen C, Dankert J. Nosocomial outbreak of *Clostridium difficile*-associated diarrhoea due to a clindamycin-resistant enterotoxin A-negative strain. Eur J Clin Microbiol Infect Dis. 2001;20(8):528–34.
11. Lawson PA, Citron DM, Tyrrell KL, Finegold SM. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. Anaerobe. 2016;40:95–9.

12. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of toxin A and toxin B in *Clostridium difficile* infection. Nature. 2010;467(7316):711–3.
13. Brito GA, Fujji J, Carneiro-Filho BA, Lima AA, Obrig T, Guerrant RL. Mechanism of *Clostridium difficile* toxin A-induced apoptosis in T84 cells. J Infect Dis. 2002;186(10):1438–47.
14. Brito GA, Carneiro-Filho B, Oria RB, Destura RV, Lima AA, Guerrant RL. *Clostridium difficile* toxin A induces intestinal epithelial cell apoptosis and damage: role of Gln and Ala-Gln in toxin A effects. Dig Dis Sci. 2005;50(7):1271–8.
15. Chumbler NM, Farrow MA, Lapierre LA, Franklin JL, Lacy DB. *Clostridium difficile* toxins TcdA and TcdB cause colonic tissue damage by distinct mechanisms. Infect Immun. 2016;84(10):2871–7.
16. Aktories K, Papatheodorou P, Schwan C. Binary *Clostridium difficile* toxin (CDT)—a virulence factor disturbing the cytoskeleton. Anaerobe. 2018;53:21–9.
17. Gerding DN, Johnson S, Rupnik M, Aktories K. *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. Gut Microbes. 2014;5(1):15–27.
18. Beer LA, Tatge H, Schneider C, Ruschig M, Hust M, Barton J, Thiemann S, Fuhner V, Russo G, Gerhard R. The binary toxin CDT of *Clostridium difficile* as a tool for intracellular delivery of bacterial glucosyltransferase domains. Toxins (Basel). 2018;10(6):225.
19. Cowardin CA, Buonomo EL, Saleh MM, Wilson MG, Burgess SL, Kuehne SA, Schwan C, Eichhoff AM, Koch-Nolte F, Lyras D, et al. The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. Nat Microbiol. 2016;1(8):16108.
20. Talbot GH, Kleinman L, Davies E, Hunsche E, Revicki D, Roberts L, Rosenberg D, Nord CE. *Clostridium difficile* infection-daily symptoms (CDI-DaySyms) questionnaire: psychometric characteristics and responder thresholds. Health Qual Life Outcomes. 2019;17(1):77.
21. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet. 2006;38(7):779–86.
22. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, et al. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. 2009;10(9):R102.
23. Citron DM, Tyrrell KL, Merriam CV, Goldstein EJ. Comparative in vitro activities of LFF571 against *Clostridium difficile* and 630 other intestinal strains of aerobic and anaerobic bacteria. Antimicrob Agents Chemother. 2012;56(5):2493–503.
24. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, Dubberke ER, Xavier J, Pamer EG. Pathogenicity Locus, core genome, and accessory gene contributions to *Clostridium difficile* virulence. MBio. 2017. https://doi.org/10.1128/mBio.00885-17.
25. Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, Fukuda S, Saito T, Narushima S, Hase K, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature. 2013;500(7461):232–6.
26. George WL, Sutter VL, Citron D, Finegold SM. Selective and differential medium for isolation of *Clostridium difficile*. J Clin Microbiol. 1979;9(2):214–9.
27. Peng Z, Jin D, Kim HB, Stratton CW, Wu B, Tang YW, Sun X. Update on antimicrobial resistance in *Clostridium difficile*: resistance mechanisms and antimicrobial susceptibility testing. J Clin Microbiol. 2017;55(7):1998–2008.
28. Arroyo LG, Rousseau J, Willey BM, Low DE, Staempfli H, McGeer A, Weese JS. Use of a selective enrichment broth to recover *Clostridium difficile* from stool swabs stored under different conditions. J Clin Microbiol. 2005;43(10):5341–3.
29. Ismaeil M, Yoshida N, Katayama A. *Bacteroides sedimenti* sp. Nov., isolated from a chloroethenes-dechlorinating consortium enriched from river sediment. J Microbiol. 2018;56(9):619–27.
30. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613–7.
31. Jang JY, Oh YJ, Lim SK, Park HK, Lee C, Kim JY, Lee MA, Choi HJ. *Salicibibacter kimchii* gen. nov., sp. Nov., a moderately halophilic and alkalitolerant bacterium in the family Bacillaceae, isolated from kimchi. J Microbiol. 2018;56(12):880–5.
32. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614–24.
33. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.
34. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014. https://doi.org/10.1093/nar/gkt1244.
35. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep. 2016;6:24373.
36. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47(W1):W5–10.
37. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
38. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66(2):1100–3.
39. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60.
40. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–31.
41. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. iPath2.0: interactive pathway explorer. Nucleic Acids Res. 2011. https://doi.org/10.1093/nar/gkr313.
42. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517–25.
43. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636–41.
44. Holy O, Chmelar D. Oxygen tolerance in anaerobic pathogenic bacteria. Folia Microbiol (Praha). 2012;57(5):443–6.
45. Edwards AN, Karim ST, Pascual RA, Jowhar LM, Anderson SE, McBride SM. Chemical and stress resistances of *Clostridium difficile* spores and vegetative cells. Front Microbiol. 2016;7:1698.
46. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Evol Microbiol. 1980;30(1):225–420.
47. Wells JE, Hylemon PB. Identification and characterization of a bile acid 7alpha-dehydroxylation operon in *Clostridium* sp. strain TO-931, a highly active 7alpha-dehydroxylating strain isolated from human feces. Appl Environ Microbiol. 2000;66(3):1107–13.
48. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, Didelot X, Dingle KE, Eyre DW, Rodriguez C, Riley TV. Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. Elife. 2021;10:e64325.
49. Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol Rev. 1994;15(2–3):155–73.
50. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of *Clostridium difficile*. Clin Microbiol Rev. 2015;28(3):721–41.
51. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, Vaughan A, Golubchik T, Fawley WN, Wilcox MH, et al. Evolutionary history of the *Clostridium difficile* pathogenicity locus. Genome Biol Evol. 2014;6(1):36–52.
52. Liu J, Peng L, Su H, Tang H, Chen D, Xu Z, Wu A. Chromosome and plasmid features of two ST37 *Clostridioides difficile* strains isolated in China reveal distinct multidrug resistance and virulence determinants. Microb Drug Resist. 2020;26(12):1503–8.
53. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJ, Jolley KA, et al. Multilocus sequence typing of *Clostridium difficile*. J Clin Microbiol. 2010;48(3):770–8.

54. Munoz M, Rios-Chaparro DI, Patarroyo MA, Ramirez JD. Determining *Clostridium difficile* intra-taxa diversity by mining multilocus sequence typing databases. BMC Microbiol. 2017;17(1):62.
55. Riccobono E, Di Pilato V, Della Malva N, Meini S, Ciraolo F, Torricelli F, Rossolini GM. Draft genome sequence of *Clostridium difficile* belonging to ribotype 018 and sequence type 17. Genome Announc. 2016. https://doi.org/10.1128/genomeA.00907-16.
56. Han SH, Kim H, Lee K, Jeong SJ, Park KH, Song JY, Seo YB, Choi JY, Woo JH, Kim WJ, et al. Epidemiology and clinical features of toxigenic culture-confirmed hospital-onset *Clostridium difficile* infection: a multicentre prospective study in tertiary hospitals of South Korea. J Med Microbiol. 2014;63(Pt 11):1542–51.
57. Kuwata Y, Tanimoto S, Sawabe E, Shima M, Takahashi Y, Ushizawa H, Fujie T, Koike R, Tojo N, Kubota T. Molecular epidemiology and antimicrobial susceptibility of *Clostridium difficile* isolated from a university teaching hospital in Japan. Eur J Clin Microbiol Infect Dis. 2015;34(4):763–72.
58. Mori N, Yoshizawa S, Saga T, Ishii Y, Murakami H, Iwata M, Collins DA, Riley TV, Tateda K. Incorrect diagnosis of *Clostridium difficile* infection in a university hospital in Japan. J Infect Chemother. 2015;21(10):718–22.
59. Senoh M, Kato H, Fukuda T, Niikawa A, Hori Y, Hagiya H, Ito Y, Miki H, Abe Y, Furuta K. Predominance of PCR-ribotypes, 018 (smz) and 369 (trf) of *Clostridium difficile* in Japan: a potential relationship with other global circulating strains? J Med Microbiol. 2015;64(10):1226–36.
60. Baldan R, Trovato A, Bianchini V, Biancardi A, Cichero P, Mazzotti M, Nizzero P, Moro M, Ossi C, Scarpellini P, et al. *Clostridium difficile* PCR Ribotype 018, a Successful epidemic genotype. J Clin Microbiol. 2015;53(8):2575–80.
61. Garey KW, Sethi S, Yadav Y, DuPont HL. Meta-analysis to assess risk factors for recurrent *Clostridium difficile* infection. J Hosp Infect. 2008;70(4):298–304.
62. Jamal W, Rotimi VO, Brazier J, Duerden BI. Analysis of prevalence, risk factors and molecular epidemiology of *Clostridium difficile* infection in Kuwait over a 3-year period. Anaerobe. 2010;16(6):560–5.
63. Abou Chakra CN, McGeer A, Labbe AC, Simor AE, Gold WL, Muller MP, Powis J, Katz K, Garneau JR, Fortier LC, et al. Factors associated with complications of *Clostridium difficile* infection in a multicenter prospective cohort. Clin Infect Dis. 2015;61(12):1781–8.
64. Park SC, Lee K, Kim YO, Won S, Chun J. Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size. Front Microbiol. 2019;10:834.
65. Knight DR, Kullin B, Androga GO, Barbut F, Eckert C, Johnson S, Spigaglia P, Tateda K, Tsai PJ, Riley TV. Evolutionary and genomic insights into *Clostridioides difficile* sequence type 11: a diverse zoonotic and antimicrobial-resistant lineage of global one health importance. MBio. 2019. https://doi.org/10.1128/mBio.00446-19.
66. Spigaglia P. Recent advances in the understanding of antibiotic resistance in *Clostridium difficile* infection. Ther Adv Infect Dis. 2016;3(1):23–42.
67. Vester B, Douthwaite S. Macrolide resistance conferred by base substitutions in 23S rRNA. Antimicrob Agents Chemother. 2001;45(1):1–12.
68. Uttley AH, Collins CH, Naidoo J, George RC. Vancomycin-resistant enterococci. Lancet. 1988;1(8575–6):57–8.
69. Peltier J, Courtin P, El Meouche I, Catel-Ferreira M, Chapot-Chartier MP, Lemee L, Pons JL. Genomic and expression analysis of the vanG-like gene cluster of *Clostridium difficile*. Microbiology. 2013;159(Pt 7):1510–20.
70. Leung V, Vincent C, Edens TJ, Miller M, Manges AR. Antimicrobial resistance gene acquisition and depletion following fecal microbiota transplantation for recurrent *Clostridium difficile* Infection. Clin Infect Dis. 2018;66(3):456–7.
71. Isidro J, Menezes J, Serrano M, Borges V, Paixao P, Mimoso M, Martins F, Toscano C, Santos A, Henriques AO, et al. Genomic study of a *Clostridium difficile* multidrug resistant outbreak-related clone reveals novel determinants of resistance. Front Microbiol. 2018;9:2994.
72. Stevenson E, Minton NP, Kuehne SA. The role of flagella in *Clostridium difficile* pathogenicity. Trends Microbiol. 2015;23(5):275–82.
73. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, Kansau I, Collignon A, Minton NP. The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. PLoS ONE. 2013;8(9):e73026.
74. Batah J, Kobeissy H, Bui Pham PT, Deneve-Larrazet C, Kuehne S, Collignon A, Janoir-Jouveshomme C, Marvaud JC, Kansau I. *Clostridium difficile* flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins. Sci Rep. 2017;7(1):3256.
75. Braun V, Hundsberger T, Leukel P, Sauerborn M, von Eichel-Streiber C. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene. 1996;181(1–2):29–38.
76. Chen R, Feng Y, Wang X, Yang J, Zhang X, Lu X, Zong Z. Whole genome sequences of three Clade 3 *Clostridium difficile* strains carrying binary toxin genes in China. Sci Rep. 2017;7:43555.
77. Lyon SA, Hutton ML, Rood JI, Cheung JK, Lyras D. CdtR regulates TcdA and TcdB production in *Clostridium difficile*. PLoS Pathog. 2016;12(7):e1005758.
78. Settle CD, Wilcox MH, Fawley WN, Corrado OJ, Hawkey PM. Prospective study of the risk of *Clostridium difficile* diarrhoea in elderly patients following treatment with cefotaxime or piperacillin-tazobactam. Aliment Pharmacol Ther. 1998;12(12):1217–23.
79. Ludlam H, Brown N, Sule O, Redpath C, Coni N, Owen G. An antibiotic policy associated with reduced risk of *Clostridium difficile*-associated diarrhoea. Age Ageing. 1999;28(6):578–80.
80. Lai KK, Melvin ZS, Menard MJ, Kotilainen HR, Baker S. *Clostridium difficile*-associated diarrhea: epidemiology, risk factors, and infection control. Infect Control Hosp Epidemiol. 1997;18(9):628–32.
81. Dancer SJ. The problem with cephalosporins. J Antimicrob Chemother. 2001;48(4):463–78.

Publisher's Note