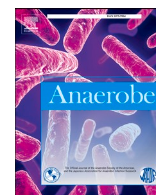
Anaerobe 91 (2025) 102926

Contents lists available at ScienceDirect

Anaerobe

journal homepage: www.elsevier.com/locate/anaerobe



Clostridioides difficile (including epidemiology)

Clostridioides difficile hypervirulent strain ST1 isolated from clinical stool specimens obtained from three Provinces in South Africa

Hlambani Shirinda [a], Anthony M. Smith [a,d], Ben Prinsloo [b], Marleen M. Kock [a,c], Mishalan Moodley [e], Mohamed Said [a,c], Marthie M. Ehlers [a,c,*]

[a] Department of Medical Microbiology, University of Pretoria, Pretoria, South Africa
[b] Lancet laboratories, Arcadia, Pretoria, South Africa
[c] National Health Laboratory Service (NHLS), Academic Division, Pretoria, South Africa
[d] Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa
[e] Sequencing Core Facility, National Institute for Communicable Diseases, Johannesburg, South Africa

ARTICLE INFO

Handling Editor: Dr. M. Rupnik

Keywords:
Clostridioides difficile
EnteroBase
Whole genome sequencing
South Africa

ABSTRACT

Objectives: Clostridioides difficile infection is a serious healthcare-associated infection linked to antimicrobial use. The severity of the disease can be associated with hypervirulent ribotypes such as RT027. The study aimed to investigate the molecular epidemiology and genomic characteristics of *C. difficile* isolates from private and public healthcare settings in South Africa.

Methods: One hundred clinical stool specimens were cultured on cycloserine-cefoxitin-fructose agar. Conventional multiplex polymerase chain reaction (M-PCR) assays were conducted for isolate identification and detection of toxin genes. Genomic characteristics of the isolates were determined using whole genome sequencing (WGS) and data was analysed using pubMLST, EnteroBase, Pathogenwatch and CARD.

Results: One hundred clinically presumptive *C. difficile* positive stool specimens were collected, of which 62 % (62/100) were confirmed as *C. difficile* by M-PCR assay. Among the 62 identified *C. difficile* isolates, 97 % (60/62) were toxigenic, with the most dominant toxin profile being $A^{+} B^{+} CDT^{+}$ according to the M-PCR assay. The results showed that 93 % (40/43) of the WGS analysed *C. difficile* strains clustered into clades 1 to 5. These 40 strains were categorized into 16 sequence types (STs), with ST1 (clade 2) being the most prevalent, representing 45 % (18/40), this strain is an RT027-associated strain previously epidemic hypervirulent strain. One major cluster (n = 18) comprising ST1 strains was identified in Gauteng Province and all the isolates associated with this cluster showed the same resistome (antimicrobial resistance genes and mutations: *CDD-1*, *aac (6′)-Ie-aph (2″)-Ia, PnimB*^{G} and *Thr82Ile*). The study also identified one strain as ST11, this strain is well known for its zoonotic potential, and two strains were identified as ST37 known as an epidemic strain. Strains from public healthcare settings exhibited genetic similarity, while those from private settings showed greater genetic diversity.

Conclusion: The study reported, for the first time, hypervirulent strains ST1 in Africa and ST11 in South Africa, with a minimum spanning tree indicating an ongoing ST1 outbreak.

1. Introduction

Clostridioides difficile (*C. difficile*), is a gram-positive bacterial pathogen that causes antimicrobial-associated diarrhea in high-income and low-income countries [1–3]. *Clostridioides difficile* infections significantly impact quality of life, morbidity and healthcare utilisation especially in older patients and immunocompromised patients [4]. Symptoms range from mild diarrhea to severe pseudomembranous colitis and the major virulence factors are toxin A, toxin B and binary toxins [2,5]. According to the literature, Africa has not implemented routine surveillance and strain typing of *C. difficile;* prospective sampling and comprehensive studies are infrequent and are often limited by resource scarcity and requisite laboratory expertise [6]. Worldwide, PCR-ribotyping is the most used typing method; however, its

* Corresponding author. Department of Medical Microbiology, School of Medicine, University of Pretoria, HW Snyman South Building University of Pretoria, Bophelo Rd, 0084, Pretoria, South Africa.
E-mail address: Marthie.ehlers@up.ac.za (M.M. Ehlers).

https://doi.org/10.1016/j.anaerobe.2024.102926
Received 22 August 2024; Received in revised form 25 November 2024; Accepted 27 November 2024
Available online 28 November 2024
1075-9964/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

standardisation is challenging due to interlaboratory variation in banding patterns and interpretation, causing data comparison issues [7]. Whole genome sequencing (WGS) effectively addresses challenges by offering insights into genetic relatedness, genotyping and public health risks [8]. Ribotyping and toxinotyping studies have identified 446 strains from five African countries, with Egypt and South Africa identifying global strains [e.g., ST3 (RT001), ST2 (RT014/020), ST37 (RT017), ST11 (RT078)]; however, the hypervirulent strain ST1 has not been documented [6,9–11]. This study aimed to investigate the molecular epidemiology and genomic characteristics of *C. difficile* strains from South African private and public healthcare settings to provide additional information on *C. difficile* strain diversity in Africa.

2. Methods

2.1. Study settings, sample collection and processing

This laboratory-based study analysed residual stool samples testing positive for toxigenic *C. difficile* from two major diagnostic laboratories representing private and public healthcare settings. A total of 100 non repeated stool specimens were included, sourced from 20 community hospitals including 11 private and nine public hospitals across Gauteng, Limpopo and Mpumalanga Provinces in South Africa. The study included all specimens confirmed positive for toxigenic *C. difficile* by either the GeneXpert real-time PCR assay (Cepheid, USA) or the TECHLAB C. diff Quick Check Complete assay (TECHLAB, USA), without restriction on patient age or sex. Since the specimens were residual samples collected post-diagnosis, no direct patient contact was involved and patient consent was not required. Furthermore, no clinical data were associated with the specimens, as they were collected solely for diagnostic purposes.

Stool specimens were obtained between June 2022 and December 2022 and stored at −20 °C until analysis. The sample collection included an equal number of specimens from each healthcare setting (50 from private and 50 from public). Private healthcare settings utilized the GeneXpert real-time PCR assay (Cepheid, USA) to detect *C. difficile* toxins A, B, and binary toxins. In contrast, public healthcare settings used the TECHLAB C. diff Quick Check Complete enzyme immunoassay (TECHLAB, USA), which detects both the glutamate dehydrogenase (gdh) enzyme and toxins in a single assay. Discrepancies in the enzyme immunoassay were further analysed using the GeneXpert PCR assay (Cepheid, USA). Fig. 1 presents the methodological framework employed in this study.

2.2. Isolation of Clostridioides difficile

A total of 100 stool specimens from private and public healthcare settings were cultured on cycloserine-cefoxitin-fructose agar (CCFA) (Liofilchem, Italy) [12]. Briefly, a volume of 1mL of stool specimen was mixed with 1 mL 96 % ethanol (Merck, UK) for spore selection, vortexed and incubated at room temperature (25 °C ± 5 °C) for one hour. The suspension was cultured on CCFA (Liofilchem, Italy) supplemented with 5 % horse blood (South African Vaccine Producers) for isolation of *C. difficile* from stool specimens and were incubated in a Mitsubishi™ AnaeroPack (Davies Diagnostics, South Africa) anaerobic jar containing AnaeroPack-Anaero (Davies Diagnostics, RSA) gas-generating sachets at 37 °C for 48 h. The colonies were identified based on their morphology (characteristics: 4 mm–6 mm diameter irregular, raised opaque, grey-white in colour) and were sub-cultured on 5 % Columbia Horse Blood Agar (CHBA) (Thermo Scientific, USA) and incubated at 37 °C for 48 h for purity.

2.3. DNA extraction

The total genomic DNA of 87 presumptive *C. difficile* isolates was extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymogen Fermentas, USA) according to the manufacturer's instructions.

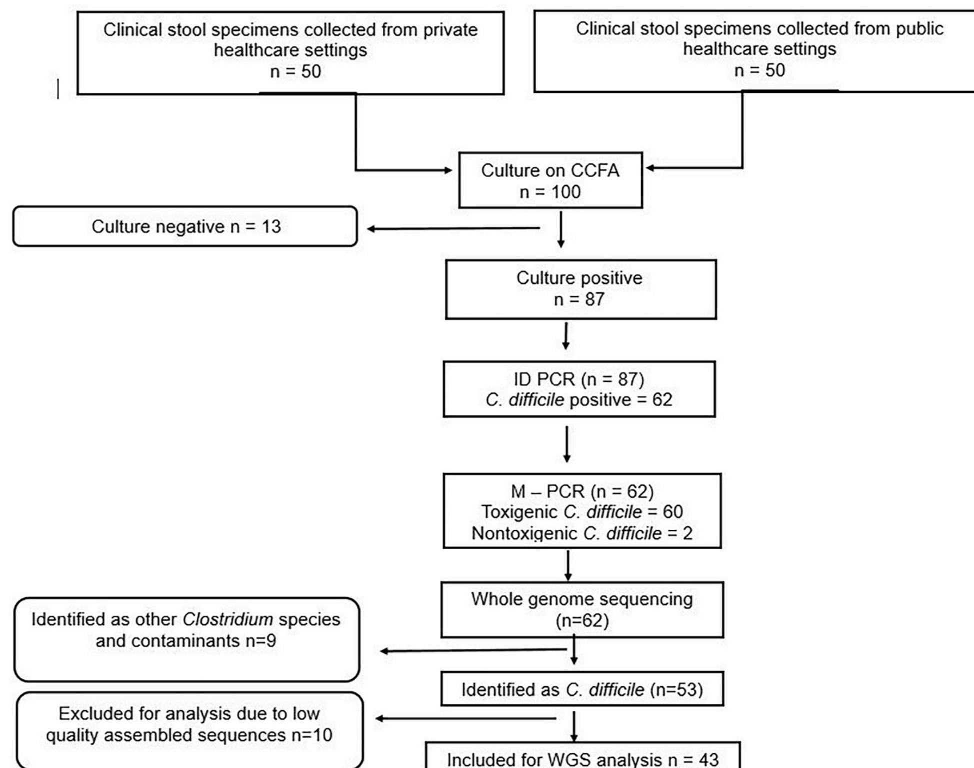


Fig. 1. Flow diagram showing the methodological framework employed in this study.

2.4. *Molecular identification and characterization of Clostridioides difficile using conventional PCR*

Conventional-polymerase chain reaction (PCR) assays were used for the identification of *C. difficile* targeting the genus-specific (*16S rRNA*), species-specific (*gdh*) gene and toxin profiling (*tcdA, tcdB* and *cdtA* and *cdtB*) using CelTaq DNA Polymerase master mix (Celtic Molecular Diagnostics, South Africa) [13–19]. Agarose gel electrophoresis was performed using a 2 % (m/v) SeaKem agarose gel (Lonza, Switzerland) for PCR amplification product detection. The amplicon product bands were visualised under a transilluminator (UV) light (UVP products, USA) using the Bio-Rad gel DocTM EZ (Bio-Rad, USA) system. *Clostridioides difficile* American Type Culture Collection (ATCC) BAA-1870 ($tcdA^+$, $tcdB^+$, and $cdtA/B^+$) and *C. difficile* ATCC 700057 ($tcdA^-$, $tcdB^-$ and $cdtA/B^-$) strains (Davies Diagnostics, South Africa) were used as the positive and negative controls respectively. A no-template control consisting of ultrapure water (ThermoFisher Scientific, USA) was included as a negative control.

2.5. *Whole genome sequencing of Clostridioides difficile isolates*

Sixty-two (toxigenic and non-toxigenic) *C. difficile* DNA extracts were submitted to the Sequencing Core Facility (SCF), NICD, for WGS. The Nextera DNA Flex library prep kit (Illumina, San Diego, CA, USA) was used for library preparation, with the inclusion of a gBlock Gene Fragment (Integrated DNA Technologies, Coralville, IA, USA) as a quality control measure. The Illumina NextSeq 2000 platform (Illumina, USA) was used for sequencing at 100 × coverage, using 2 × 150 base pairs (bp) paired-end sequencing for each flow cell.

2.6. *Bioinformatics analysis*

Sequencing reads were processed using the JEKESA pipeline (v1.0). In summary, FastQC (v.0.11.9) was used for quality control (available online at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trim Galore! (v.0.6.7) (available online at https://github.com/FelixKrueger/TrimGalore) set to a minimum Phred quality score of 30 and minimum read length of 50 bp. Kraken 2 was used to determine whether any contaminating sequences other than *C. difficile* were present. Assembly was performed using SPAdes (v.3.14.1), while Shovil (v.1.1.0) was used for polishing (available online: https://github.com/tseemann/shovill) [18,20] The final read assembly was evaluated using QUAST [21]. Assembled sequence uploaded on Pathogenwatch (https://pathogen.watch/) to confirm identity. Raw sequencing data (FastQ files) were uploaded to the EnteroBase platform (available online: https://enterobase.warwick.ac.uk/species/index/clostridium) where core-genome MLST (cgMLST) and hierarchical clustering (HC) were used to investigate phylogeny amongst clinical *C. difficile* isolates [22]. The genomic relationships were visualised using GrapeTree with the MSTree V2 algorithm based on the cgMLST + HierCC scheme and Microreact [21–23]. Once the GrapeTree was produced, the settings of the tool were set to 'collapse branches' at a value of '6', which resulted in isolates showing allelic differences to collapse together into a 'cluster'. This was how a cluster of isolates was created and visualised. Cluster definition used in the current study was ≥2 isolates showing ≤6 allelic differences, as obtained by the above actions, following cgMLST analysis and generation of a GrapeTree [7]. Assembled sequences were submitted to the PubMLST sequence query page (http://pubmlst.org/cdifficile/) where MLST was used to obtain the sequence type (ST) and clade, looking at seven housekeeping genes (*adk, atpA, dxr, glyA, recA, sodA* and *tpi*). Detection of antimicrobial resistance determinants was performed using Comprehensive antibiotic resistance database (CARD) and EnteroBase.

2.7. *Genome sequence data availability*

All sequencing data was uploaded to the public EnteroBase platform (https://enterobase.warwick.ac.uk/species/index/clostridium) and are freely available at the EnteroBase platform. In addition, sequencing data were deposited in the National Center for Biotechnology Information (NCBI) under the project accession number PRJNA1138394.

3. Results

3.1. *Molecular characterization of Clostridioides difficile*

One hundred clinical stool specimens from private and public healthcare settings were screened for toxigenic *C. difficile* and 87/100 isolates had presumptive colonies that resembled *C. difficile*. Conventional PCR confirmed 71 % (62/87) of isolates as *C. difficile* due to the presence of the *16S rRNA* and *gdh* genes. Among the 62 identified *C. difficile* strains, 97 % (60/62) were confirmed as toxigenic, while 3 % (2/62) were non-toxigenic. The most prevalent *C. difficile* toxin profile was $tcdA^+B^+$, $cdtA^+B^+$ ($A^+B^+CDT^+$) [37 % (23/62)]. Table 1 shows the distribution of toxin profiles for all isolates identified as *C. difficile* by M-PCR.

3.2. *Genetic relatedness and phylogenetic analysis*

Among the 62 isolates submitted for WGS, 86 % (53/62) were identified as *C. difficile*, while 14 % (9/62) were identified as *Clostridium* species (n = 6) and contaminants (n = 3) based on the kraken2 analysis and pathogenwatch. Among the 53 *C. difficile* strains, 19 % (10/53) were excluded because they did not meet the EnteroBase assembly minimum requirements. Therefore, 43 strains were included for WGS analysis and 93 % (40/43) clustered in clades 1 to 5, with most strains belonging to clade 2 [45 % (18/40)]. The 40 *C. difficile* strains were assigned to 16 STs and ST1 clade 2 was the most prevalent clone accounting for 45 % (18/40) with the majority 83 % (15/18) from public health settings, followed by ST3 clade 1 and ST35 clade 1 with 7 % (3/43) strains each. Other STs included: two isolates each belonging to: ST4, ST37 and ST104, while one isolate each belonging to ST2, ST5, ST11, ST23, ST29, ST43, ST54, ST63, ST397and ST558. Amongst the *C. difficile* strains, 48 % (19/40) were classified as hypervirulent strains, comprising 18 ST1 and one ST11 strain. Four STs (ST1, ST3, ST35 and ST104) were detected in both private and public health settings, while ten STs (ST2, ST4, ST5, ST11, ST23, ST29, ST37, ST43, ST63 and ST397) were only detected in private health settings and two STs (ST54 and ST558) in public health settings. The remaining 7 % (3/43) of the strains belonged to unknown clade (s) and were identified as ST122 (n = 2) and ST668, all from private healthcare settings. Table 2 shows distribution of clinical *C. difficile* strains according pubMLST and Fig. 2 is a minimum spanning tree that shows genetic relationship of the *C. difficile* strains.

3.3. *Demographic characteristics of patients with Clostridioides difficile infection*

The majority [58 % (25/43)] of the patients with CDI were younger than 65 years and the median was 61 years (range: 7 months to 89 years). The majority [61 %, (26/43)] were from females. Detailed demographic characteristics are described in Tables 3A–3C.

3.4. *Core-genome MLST and hierarchal clustering of Clostridioides difficile strains*

As illustrated in Fig. 2, one major cluster comprising ST1 strains (n = 18) was identified, the majority of the strains (67 %, 12/18) were from females and the majority (44 %, 8/18) were from patients aged between 26 years and 44 years. All strains from this cluster were from seven different hospitals (private and public) in Gauteng Province. All strains

Table 1
Distribution of *Clostridioides difficile* toxin profiles.

Toxin profile	Number of isolates with toxin profiles by M-PCR n = 62 (%)	Identified as *C. difficile* by whole genome sequencing n = 62 (%)	*C. difficile* isolates that met minimum assembly requirements on EnteroBase and included for further analysis n = 62 (%)
$A^+B^+CDT^+$	30 (48 %)	27 (44 %)	23 (37 %)
A^+B^+	23 (37 %)	19 (31 %)	16 (26 %)
$A\text{-}B^+$	4 (7 %)	4 (7 %)	2 (3 %)
A^+B^-	1 (2 %)	0	0
A-B + CDT+	1 (2 %)	1 (2 %)	0
None toxigenic	2 (3 %)	2 (3 %)	2 (3 %)

Table 2
Distribution of *Clostridioides difficile* strains by clade, sequence type, and healthcare setting from which stool specimens were collected.

Clade	Sequence Type (s)	Count n = 43	Healthcare Setting
Clade 1	ST2	1	Private
	ST3	3	Private n = 1 Public n = 2
	ST4	2	Private
	ST29	1	Private
	ST35	3	Private n = 2 Public n = 1
	ST43	1	Private
	ST54	1	Public
	ST63	1	Private
	ST104	2	Private n = 1 Public n = 1
	ST397	1	Private
	ST558	1	Public
Clade 2	ST1	18	Private n = 15 Public n = 3
Clade 3	ST5	1	Private
Clade 4	ST23	1	Private
	ST37	2	Private
Clade 5	ST11	1	Private
Unknown Clade	ST122	2	Private
	ST668	1	Private

associated with the cluster 1 showed similar genotypic properties. Genotypic properties for the strains are summarised in Table 4 and the genotypic relationships are displayed in Fig. 3A and B and are briefly described as follows.

From cluster 1, four groups of genetically indistinguishable strains were detected at HC level 0 (HC0), namely, HC0:ST11385, HC0:ST26375, HC0:ST26377 and HC0:ST26384. The HC0:ST11385 and HC0:ST26375 comprised four strains each while HC0:ST26377 and HC0:ST26384 comprised two strains each. Three of the four strains in the HC0:ST11385 originated from the same hospital within public healthcare settings, while one strain was from a private healthcare setting. The HC0: ST26375 (n = 4) includes strains distributed as follows: three from public healthcare settings (two from the same hospital, one from a different hospital) and one from a private hospital. The HC0:ST26384 comprised strains from public healthcare settings exclusively, while the HC0: ST26377 cluster comprised strains from private healthcare settings exclusively. All strains from cluster 1 showed a multidrug resistance (MDR) genotype with the presence of intrinsic antimicrobial resistance (AMR) genes (beta-lactamase *CDD-1*, cephalosporin], acquired AMR genes (*aac (6′)-Ie-aph (2″)-Ia,* aminoglycosides), mutations (*PnimB*^{G}, metronidazole and *Thr82Ile*, fluoroquinolones)] and efflux gene (*qacG*, disinfecting agents and antiseptics). An additional group comprising two undistinguishable ST122 strains designated HC0:ST26392 was detected. Both strains from HC0:ST26392 were from a private hospital in Gauteng Province. Fig. 3A shows five groups of undistinguishable *C. difficile* strains in private and public healthcare settings.

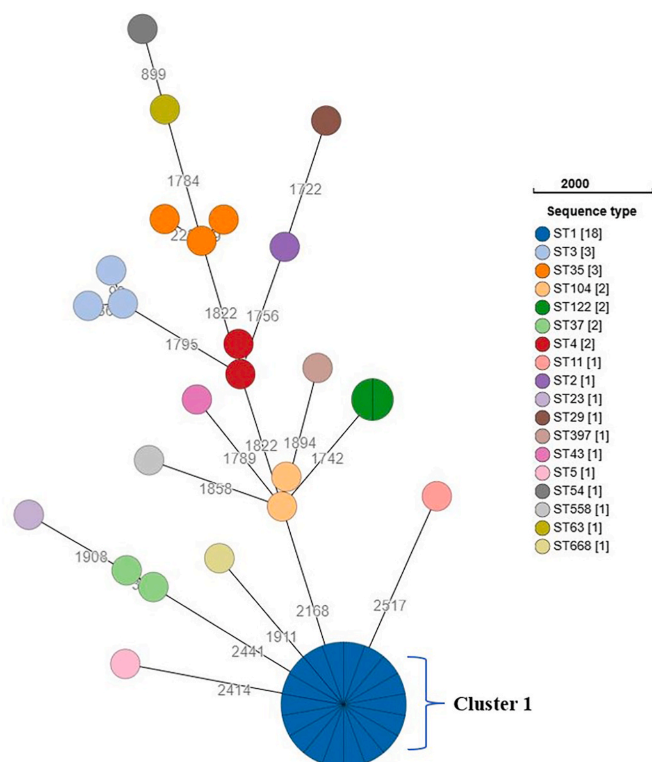


Fig. 2. Minimum spanning tree based on the core-genome MLST and hierarchical clustering of *Clostridioides difficile* strains on EnteroBase, visualised using GrapeTree (n = 43). Different colours represent different sequence types. Isolates showing ≤6 allelic differences, are collapsed together into a single circular node. The larger the circular node, the more isolates which are reflected. The number of segments within a circular node are indicative of the number of isolates. The number values between adjacent nodes indicate the number of allele differences between connecting nodes (isolates). Different colours represent genetically indistinguishable strains. A major cluster of isolates is indicated (n = 18). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

This study reports the circulation of hypervirulent strain ST1 which was found to be the dominant strain, while one strain belonging to ST11 (not previously reported in South Africa) was identified. The primary STs identified were ST1, ST3 and ST35; with all these strains seen in private and public healthcare settings. These findings are similar to findings by Liu et al. [24], conducted in China, where ST3 and ST35 were the most predominant strains. Based on the literature, hypervirulent strain ST1 is the representative ST in clade 2 and is associated with RT027 increased CDI severity [23,24]. Since its first discovery in North America, this strain has been reported in Europe, Australia and Asia, and the current study reports its presence in Africa [3,25–27]. In the current

Table 3A
Demographic characteristics of patients with *Clostridioides difficile* from private and public healthcare settings per sequence type.

Sequence type	Age Group	Private healthcare settings n = 43	Public healthcare settings n = 43	Total Count n = 43
ST1	≥65	2 (5 %)	2 (5 %)	4 (9 %)
	≤65	1 (2 %)	12 (28 %)	13 (30 %)
	Unknown	–	1 (2 %)	1 (2 %)
ST2	≤65	1 (2 %)	–	1 (2 %)
ST3	≥65	1 (2 %)	1 (2 %)	2 (5 %)
	Unknown	–	1 (2 %)	1 (2 %)
ST4	≥65	2 (5 %)	–	2 (5 %)
ST5	≤65	1 (2 %)	–	1 (2 %)
ST11	≥65	1 (2 %)	–	1 (2 %)
ST23	≤65	1 (2 %)	–	1 (2 %)
ST29	≤65	1 (2 %)	–	1 (2 %)
ST35	≤65	2 (5 %)	1 (2 %)	3 (7 %)
ST37	≥65	1 (2 %)	–	1 (2 %)
	≤65	1 (2 %)	–	1 (2 %)
ST43	≤65	1 (2 %)	–	1 (2 %)
ST54	≤65	–	1 (2 %)	1 (2 %)
ST63	≤65	1 (2 %)	–	1 (2 %)
ST104	≥65	1 (2 %)	–	1 (2 %)
	Unknown	–	1 (2 %)	1 (2 %)
ST122	≥65	2 (5 %)	–	2 (5 %)
ST397	≤65	1 (2 %)	–	1 (2 %)
ST558	≤65	–	1 (2 %)	1 (2 %)
ST668	≥65	1 (2 %)	–	1 (2 %)

Table 3B
Age distribution for patients with *Clostridioides difficile*.

Age Group	Count n = 43 (%)
≥65	15 (35 %)
≤65	25 (58 %)
Unknown	3 (7 %)

Table 3C
Gender distribution for patients with *Clostridioides difficile*.

Sex	Count n = 43
Female	26 (61 %)
Male	16 (37 %)
Unknown	1 (2 %)

study, ST1 was the most prevalent strain [42 % (18/43) followed by ST3 [7 %, n = 3/43]. Similar to the results in the current study, ST3 was the second most prevalent strain detected in patients attending tuberculosis hospitals in Cape Town, South Africa in 2016 [9]. The study by Rajabally et al. [10], identified ST37 in 50 % (16/32) of the strains, while ST37 was only detected among 5 % (2/43) of the strains in the current study. In the current study, the ST37 comprised of one epidemic strain RT017 (HC10:ST17).

Sequence type 11, a representative strain of clade 5, is widely recognised as a key causal agent of community-acquired (CA-CDI) and hospital-acquired *C. difficile* infections (HA-CDI) in Europe and North America [25–28]. This strain is a significant infectious source in animal populations and accounts for a substantial number of CDIs in humans. It has been isolated from human clinical samples, veterinary cases and environmental sources across Australia, Asia, Europe and North America [29]. In Africa, ST11 was identified in Egypt and Malawi, however, this is the first study to identify ST11 in South Africa [6].

Two ST122 strains belonging to an unknown clade were also identified, this ST was described for the first time in Europe by Knetsch et al. [30] and it demonstrated a virulence profile similar to that of ST1 and ST11 hypervirulent strains. An Australian study by Hong et al. [27], detected a high prevalence of *CDT* $^{+}$ strains in public health settings

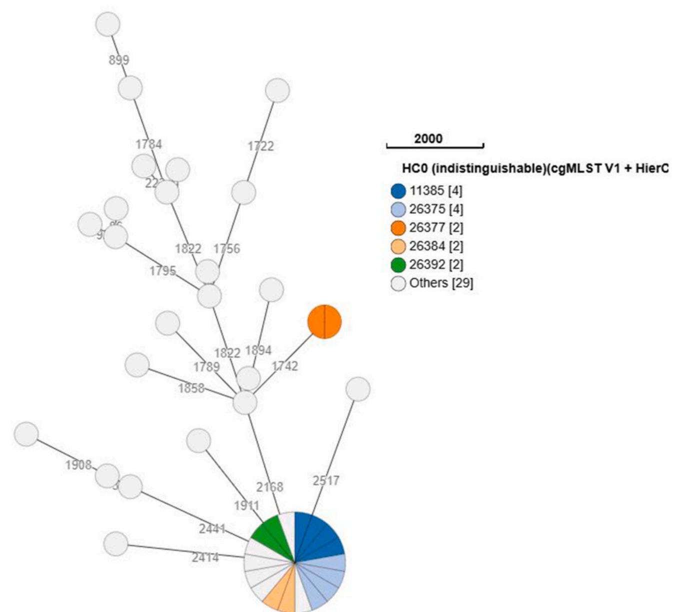


Fig. 3A. Minimum spanning tree based on the core-genome MLST and hierarchical clustering of *Clostridioides difficile* strains on EnteroBase, visualised using GrapeTree (n = 43). Isolates showing ≤6 allelic differences, are collapsed together into a single circular node. Different colours undistinguishable isolates. The larger the circular node, the more isolates which are reflected. The number of segments within a circular node are indicative of the number of isolates. The number values between adjacent nodes indicate the number of allele differences between connecting nodes (isolates). Different colours represent genetically indistinguishable strains. Colourless indicate non-identical strains. A major cluster of isolates is indicated (n = 18). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Genotypic properties of *Clostridioides difficile* isolates associated with the cluster identified in the Gauteng province of South Africa.

Cluster	MLST	cgMLST HC0 (indistinguishable)	cgMLST HC150 (clonal complex)	Presence of intrinsic AMR genes	Presence of acquired AMR genes	Mutations	Plasmid
1	1	11385 n = 4	4711	*CDD-1*	*aac (6′)-Ie-aph (2″)-Ia*	$PnimB^{G}$, *Thr82Ile*	None
		26375 n = 4	4711	*CDD-1*	*aac (6′)-Ie-aph (2″)-Ia*	$PnimB^{G}$, *Thr82Ile*	None
		26377 n = 2	4711	*CDD-1*	*aac (6′)-Ie-aph (2″)-Ia*	$PnimB^{G}$, *Thr82Ile*	None
		26384 n = 2	4711	*CDD-1*	*aac (6′)-Ie-aph (2″)-Ia*	$PnimB^{G}$, *Thr82Ile*	None
		Strains with allele difference > 0 n = 6	4711	*CDD-1*	*aac (6′)-Ie-aph (2″)-Ia*	$PnimB^{G}$, *Thr82Ile*	None

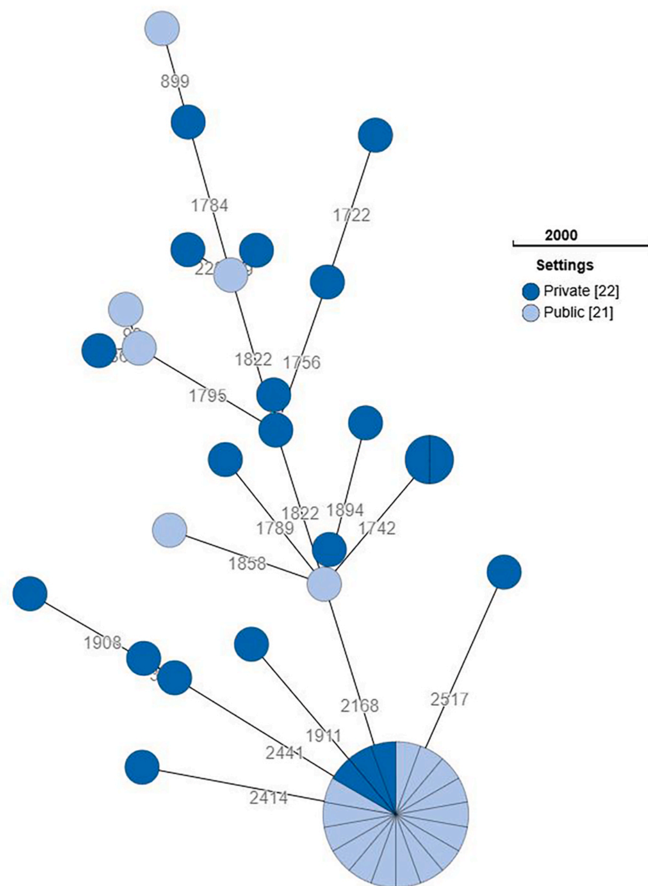


Fig. 3B. Minimum spanning tree based on the core-genome MLST and hierarchical clustering of *Clostridioides difficile* strains on EnteroBase, visualised using GrapeTree (n = 43). The tree is coloured by the settings in which the clinical specimen was collected. Isolates showing ≤6 allelic differences, are collapsed together into a single circular node. The larger the circular node, the more isolates which are reflected. The number of segments within a circular node are indicative of the number of isolates. The number values between adjacent nodes indicate the number of allele differences between connecting nodes (isolates). A major cluster of isolates is indicated (n = 18). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

compared with private healthcare settings, similarly, the current study observed a high prevalence of *CDT* $^{+}$ strains in public health settings. The majority of the *C. difficile* strains were from patients younger than 65 years, especially those with the ST1 strain. This demonstrates that *C. difficile* could spread amongst younger patients, which could be a result of the high prevalence of comorbidities such as human immunodeficiency virus and tuberculosis in South Africa [4]. These findings highlight the importance and significance of continuous epidemiological surveillance of CDI to prevent generalisations. Core-genome MLST and HC showed that strains from public healthcare settings were genetically similar, while strains from private healthcare settings were genetically diverse. This was evidenced by the clustering of strains at levels HC0 - HC 2500. This suggests that there may have been patient-to-patient transmission possibly resulting from shared hospital facilities, transmission from healthcare workers (HCWs) and patient transfers between clinical settings. Additionally, public healthcare settings possibly acquire infections from common environmental sources in the community. The high diversity of strains in private healthcare settings could be due to strict infection control measures compared to public healthcare settings. Moreover, strains (ST1, ST2, ST3, ST11 and ST37) with zoonotic potential were identified in the current study which underscores the importance of One Health approach.

One large cluster comprising 18 ST1 strains with four different groups of undistinguishable clones was detected in samples from Gauteng Province. This demonstrated a possible ongoing outbreak of ST1. Most importantly, all the strains from this cluster proved to be genotypically MDR. This is of public health concern because this strain is known to cause epidemic outbreaks.

Non-toxigenic *C. difficile* isolates may result from patient carriage of both toxigenic and non-toxigenic strains. The GeneXpert PCR assay (Cepheid, USA) has misidentified strains as presumptive RT027 due to genetic similarities such as PaLoc and *cdtB*. For instance, RT244 and RT591 were misidentified as presumptive RT027 in studies by Lim et al. [31] in Australia and Skinner et al. [32] in the USA, respectively. In this study, most strains presumed RT027 by GeneXpert were identified as ST1 through WGS.

Some of the clinical specimens failed to yield colonies on CCFA (Liofilchem, Italy), potentially due to the prolonged storage period of stool samples before testing, as they were processed approximately a year after collection. The CCFA (Liofilchem, Italy) also has limitations in its inability to differentiate between species, which could contribute to colony misidentification, therefore proper verification is recommended when using this media. Additionally, there was a misidentification of other *Clostridium* species as *C. difficile* by M-PCR assay specifically designed for *C. difficile*. However, there is a possibility that the primers used in the M-PCR assay were non-specific, which potentially led to the misidentification of *Clostridium* species as *C. difficile*. In M-PCR, primers are designed to target specific DNA sequences, but if they are not highly specific, they can bind to similar sequences in closely related species, resulting in the amplification of non-target DNA. To reduce such errors, tools such as MALDI-TOF-MS (Bruker Daltonics, Billerica, MA, USA) are recommended for more accurate species identification, minimizing the risk of sending misidentified samples for labour-intensive and expensive tests such as M-PCR and WGS. Additionally, only one colony per sample was isolated and at this point, the toxigenic *C. difficile* could have been missed, thereby isolating non – *C. difficile* isolates with similar morphology to that of *C. difficile*. The WGS has proven invaluable in distinguishing between *C. difficile* and *Clostridium* species.

5. Conclusion

This study is the first to report hypervirulent *C. difficile* strains ST1 in Africa and ST11 in South Africa, linked to epidemics, severe symptoms and outbreaks. Sequence type 1 was dominant, indicating an ongoing outbreak. Whole-genome sequencing and genomic tools (EnteroBase, pathogenwatch and pubMLST) highlighted diverse *C. difficile* genotypes in private healthcare settings, while undistinguishable strains in public healthcare settings suggest possible cross-transmission among patients. This study was limited to two laboratories that collected 100 stool specimens from 20 healthcare facilities across three Provinces. Consequently, the findings may not fully represent South Africa's broader population. Thus, multi-centre studies or surveillance programs in both sectors are recommended to enhance epidemiological insights and early detection of epidemic strains. Additionally, South Africa's *C. difficile* diagnostic algorithm [TECHLAB C. diff Quick Check Complete assay (TECHLAB, USA) and the GeneXpert real-time PCR assay (Cepheid, USA)] should be updated to include culture-based, AST, and molecular methods, as current protocols have missed the spread of ST1, highlighting a critical gap in current diagnostic practices.

CRediT authorship contribution statement

Hlambani Shirinda: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anthony M. Smith:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition. **Ben Prinsloo:** Investigation. **Marleen M. Kock:** Investigation. **Mishalan Moodley:** Investigation.

Mohamed Said: Investigation. **Marthie M. Ehlers:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Ethical approval

Ethics clearance (number 612/2022) was obtained from the Research Ethics Committee of the Faculty of Health Sciences, University of Pretoria. Permission letters to collect stool specimens were obtained from both public and private sector diagnostic laboratories.

Funding

The funds required for this research project were provided by Professor MM Ehlers, through an NHLS trust grant. Whole-genome sequencing of bacterial isolates was made possible by support from the SeqAfrica project which is funded by the Department of Health and Social Care's Fleming Fund using UK aid. The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health and Social Care or its Management Agent, Mott MacDonald. A student bursary was provided by the National Research Foundation.

Declaration of competing interest

All authors declare to have no financial or non-financial conflict of interest regarding this study.

Acknowledgment

The authors thank the University of Pretoria, Lancet laboratories, National Health Laboratory Services, the National Research Foundation, SeqAfrica and the National Institute for Communicable Diseases.

Data availability

Data will be made available on request.

References

[1] A. Majumdar, R. Govind, Regulation of *Clostridioides difficile* toxin production, Curr. Opin. Microbiol. 65 (2022) 95–100, https://doi.org/10.1016/j.mib.2022.08.003.

[2] C.F. Rizek, R.C. Martins, E.S. Girão, B. de M. Tavares, S.A. dos Santos, G. L. Gamarra, A.G. Abreu, *Clostridioides difficile* from Brazilian hospitals: characterization of virulence genes by whole genome sequencing, Microb. Infect. 24 (2022) 104953, https://doi.org/10.1016/j.micinf.2022.104953.

[3] C. Aguayo, R. Flores, S. Lévesque, P. Araya, S. Ulloa, J. Lagos, J. Kuehn, Rapid spread of *Clostridium difficile* NAP1/027/ST1 in Chile confirms the emergence of the epidemic strain in Latin America, Epidemiol. Infect. 143 (2015) 3069–3073, https://doi.org/10.1017/S0950268815000558.

[4] T. Nana, C. Moore, T. Boyles, A.J. Brink, J. Cleghorn, L.M. Devenish, J. Lues, South African Society of Clinical Microbiology *Clostridioides difficile* infection diagnosis, management and infection prevention and control guidelines, South. Afr. J. Infect. Dis. 35 (2020), https://doi.org/10.1080/23115533.2020.1839488.

[5] R.E. Ooijevaar, Y.H. van Beurden, E.M. Terveer, A. Goorhuis, M.P. Bauer, J. J. Keller, C.J.J. Mulder, E.J. Kuijper, Update of treatment algorithms for *Clostridium difficile* infection, Clin. Microbiol. Infection 24 (2018) 452–462, https://doi.org/10.1016/j.cmi.2017.07.015, 2018.

[6] B. Kullin, V.R. Abratt, S.J. Reid, T.V. Riley, *Clostridioides difficile* infection in Africa: a narrative review, Anaerobe 74 (2022), https://doi.org/10.1016/j.anaerobe.2021.102547.

[7] A. Baktash, J. Corver, C. Harmanus, W.K. Smits, W. Fawley, M.H. Wilcox, R.F. de Boer, Comparison of whole-genome sequence-based methods and PCR ribotyping for subtyping of *Clostridioides difficile*, J. Clin. Microbiol. 60 (2022) e02358, https://doi.org/10.1128/JCM.02358-21, 21.

[8] S. Bletz, S. Janezic, D. Harmsen, M. Rupnik, A. Mellmann, Defining and evaluating a core genome multilocus sequence typing scheme for genome-wide typing of *Clostridium difficile*, J. Clin. Microbiol. 56 (2018), https://doi.org/10.1128/JCM.00297-18.

[9] I. Janssen, P. Cooper, K. Gunka, M. Rupnik, D. Wetzel, O. Zimmermann, E. Asamoah, High prevalence of nontoxigenic *Clostridium difficile* isolated from hospitalized and non-hospitalized individuals in rural Ghana, Int. J. Med. Microbiol. 306 (2016) 652–656, https://doi.org/10.1016/j.ijmm.2016.06.001.

[10] N. Rajabally, B. Kullin, K. Ebrahim, T. Brock, A. Weintraub, A. Whitelaw, I. Zulu, et al., A comparison of *Clostridium difficile* diagnostic methods for identification of local strains in a South African centre, J. Med. Microbiol. 65 (2016) 320–327, https://doi.org/10.1099/jmm.0.000217.

[11] L. Legenza, S. Barnett, W. Rose, M. Bianchini, N. Safdar, R. Coetzee, Epidemiology and outcomes of *Clostridium difficile* infection among hospitalised patients: results of a multicentre retrospective study in South Africa, BMJ Glob. Health 3 (2018), https://doi.org/10.1136/bmjgh-2018-000725.

[12] H. Kunishima, H. Ohge, H. Suzuki, A. Nakamura, K. Matsumoto, H. Mikamo, N. Mori, Y. Morinaga, K. Yanagihara, Y. Yamagishi, S. Yoshizawa, Japanese clinical practice guidelines for management of *Clostridioides (Clostridium) difficile* infection, J. Infect. Chemother. 28 (2022) 1045–1083, https://doi.org/10.1016/j.jiac.2021.12.011.

[13] M. Muñoz, D.I. Ríos-Chaparro, G. Herrera, S.D. Leon, C. Sara, C. Birchenall, J. D. Ramírez, New insights into *Clostridium difficile* (CD) infection in Latin America: novel description of toxigenic profiles of diarrhea-associated with CD in Bogotá, Colombia, Front. Microbiol. 9 (2018) 74, https://doi.org/10.3389/fmicb.2018.00074.

[14] P. Naaber, J. Štšepetova, I. Smidt, M. Rätsep, S. Kõljalg, K. Lõivukene, A. Tamm, Quantification of *Clostridium difficile* in antibiotic-associated diarrhea patients, J. Clin. Microbiol. 49 (2011) 3656–3658, https://doi.org/10.1128/JCM.00160-11.

[15] S. Paltansing, R.J. van den Berg, R.A. Guseinova, C.E. Visser, E.R. van der Vorm, E. J. Kuijper, Characteristics and incidence of *Clostridium difficile*-associated disease in The Netherlands, 2005, Clin. Microbiol. Infection 13 (2007) 1058–1064, https://doi.org/10.1111/j.1469-0691.2007.01725.x.

[16] D. Griffiths, W. Fawley, M. Kachrimanidou, R. Bowden, D.W. Crook, R. Fung, T. Golubchik, K. Harding, R.M. Jeffery, K.A. Jolley, A.P. Kirton, T.E. Peto, G. Rees, N. Stoesser, A. Vaughan, A. Walker, B.C. Young, M. Wilcox, Multilocus sequence typing of *Clostridium difficile*, J. Clin. Microbiol. 48 (2010) 770–778, https://doi.org/10.1128/JCM.01454-09.

[17] L. Lemée, A. Dhalluin, S. Testelin, M.A. Mattrat, K. Maillard, J.F. Lemeland, J. L. Pons, Multiplex PCR targeting tpi (triose phosphate isomerase), *tcdA* (toxin A), and *tcdB* (toxin B) genes for toxigenic culture of *Clostridium difficile*, J. Clin. Microbiol. 42 (2004) 5710–5714, https://doi.org/10.1128/JCM.42.12.5710-5714.2004.

[18] S. Stubbs, M. Rupnik, M. Gibert, J. Brazier, B. Duerden, M. Popoff, Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*, FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 186 (2006) 307–312, https://doi.org/10.1111/j.1574-6968.2006.00113.x.

[19] R.A. Stabler, M. He, L. Dawson, M. Martin, E. Valiente, C. Corton, T.D. Lawley, M. Sebaihia, M.A. Quail, G. Rose, D.N. Gerding, M. Gibert, M.R. Popoff, J. Parkhill, G. Dougan, B.W. Wren, Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provide insight into the evolution of a hypervirulent bacterium, Genome Biol. 10 (2009), https://doi.org/10.1186/gb-2009-10-3-r83.

[20] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (2019) 1–3, https://doi.org/10.1186/s13059-019-1891-0.

[21] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.

[22] M. Frentrup, Z. Zhou, M. Steglich, J.P. Meier-Kolthoff, M. Göker, T. Riedel, B. Bunk, C. Spröer, J. Overmann, M. Blaschitz, A. Indraet, A publicly accessible database for *Clostridioides difficile* genome sequences supports tracing of transmission chains and epidemics, Microb. Genom. 6 (2020), https://doi.org/10.1099/mgen.0.000340.

[23] Z. Zhou, N.F. Alikhan, M.J. Sergeant, N. Luhmann, C. Vaz, A.P. Francisco, J. A. Carrico, M. Achtman, GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens, Genome Res. 28 (2018) 1395–1404, https://doi.org/10.1101/gr.228166.117.

[24] X. Liu, W. Li, W. Zhang, Y. Wu, J. Lu, Molecular characterization of *Clostridium difficile* isolates in China from 2010 to 2015, Front. Microbiol. 9 (2018), https://doi.org/10.3389/fmicb.2018.00861.

[25] T. Lv, Y. Chen, L. Guo, Q. Xu, S. Gu, P. Shen, J. Quan, Y. Fang, L. Chen, Q. Gui, G. Ye, L. Li, Whole genome analysis reveals new insights into the molecular characteristics of *Clostridioides difficile* NAP1/BI/027/ST1 clinical isolates in the People's Republic of China, Infect. Drug Resist. 12 (2019) 1783–1794, https://doi.org/10.2147/IDR.S212868.

[26] C. Aguayo, R. Flores, S. Levesque, P. Araya, S. Ulloa, J. Lagos, J.C. Hormazabal, J. Tognarelli, D. Ibáñez, P. Pidal, O. Duery, The rapid spread of Clostridium difficile NAP1/027/ST1 in Chile confirms the emergence of the epidemic strain in Latin America, Epidemiol. Infect. 143 (2015) 3069–3073, https://doi.org/10.1017/S0950268815000558.

[27] S. Hong, P. Putsathit, N. George, C. Hemphill, P.G. Huntington, T.M. Korman, D. Kotsanas, M. Lahra, R. McDougall, C.V. Moore, G.R. Nimmo, Laboratory-based surveillance of *Clostridium difficile* infection in Australian health care and community settings, *2013 to 2018*, J. Clin. Microbiol. 58 (2020) e01028, https://doi.org/10.1128/JCM.01028-20, 20.

[28] C.W. Knetsch, N. Kumar, S.C. Forster, T.R. Connor, H.P. Browne, C. Harmanus, I. M. Sanders, S.R. Harris, L. Turner, T. Morris, M. Perry, Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans, J. Clin. Microbiol. 56 (2018), https://doi.org/10.1128/JCM.01028-20.

[29] K.D.R. Knight, K.B. Kullin, A.G.O. Androga, B.F. Barbut, E.C. Eckert, J.S. Johnson, &T.V. Riley, Evolutionary and genomic insights into *Clostridioides difficile* sequence type 11: a diverse zoonotic and antimicrobial-resistant lineage of global one health importance, mBio 10 (2019), https://doi.org/10.1128/mBio.01011-19.

[30] C.W. Knetsch, E.M. Terveer, C. Lauber, A.E. Gorbalenya, C. Harmanus, E.J. Kuijper, J. Corver, H.C. van Leeuwen, Comparative analysis of an expanded *Clostridium difficile* reference strain collection reveals genetic diversity and evolution through six lineages, Infect. Genet. Evol. 12 (2012) 1577–1585, https://doi.org/10.1016/j.meegid.2012.07.008.

[31] A.M. Skinner, S.T. Phillips, M.M. Merrigan, K.J. O'Leary, S.P. Sambol, F. Siddiqui, L.R. Peterson, D.N. Gerding, S. Johnson, The relative role of toxins A and B in the virulence of *Clostridioides difficile*, J. Clin. Med. 10 (2020) 96, https://doi.org/10.3390/jcm10010096.

[32] S.K. Lim, R.L. Stuart, K.E. Mackin, G.P. Carter, D. Kotsanas, M.J. Francis, M. Easton, K. Dimovski, B. Elliott, T.V. Riley, G. Hogg, Emergence of a ribotype 244 strain of *Clostridium difficile* associated with severe disease and related to the epidemic ribotype 027 strain, Clin. Infect. Dis. 58 (2014) 1723–1730, https://doi.org/10.1093/cid/ciu191.