Infection and Drug Resistance Dovepress

Open Access Full Text Article ORIGINAL RESEARCH

Genomic Epidemiology of *Clostridioides difficile* ST81 in Multiple Hospitals in China

Xufen Xia[1,*], Tao Lv[2,*], Lisi Zheng[2], Yuhong Zhao[3], Ping Shen[2], Danhua Zhu[2], Yunbo Chen[2]

[1]Department of Laboratory Medicine, Tongde Hospital of Zhejiang Province, Hangzhou, Zhejiang, People's Republic of China; [2]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; [3]Department of Blood Transfusion, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yunbo Chen, Email chenyunbo@aliyun.com

Background: *Clostridioides difficile* sequence type (ST) 81, mainly associated with ribotype (RT) 369, is a TcdA-negative and TcdB-positive genotype and a common ST found in China. Furthermore, ST81 strains are reported with highest resistance rates to many antimicrobial agents. However, given the potential for *C. difficile* ST81 transmission, research into the epidemiological characteristics of this type of ST remain limited.

Methods: We conducted a genomic epidemiology study addressing the genetic characteristics of *C. difficile* ST81 in five tertiary hospitals covering different regions in China between January 2010 and January 2021. Clinical toxigenic *C. difficile* strains were identified, typed by multi-locus sequence typing (MLST), and phylogenetic analysis, antimicrobial resistant gene (AMR) identification were performed after all these strains were conducted by whole genome sequencing (WGS).

Results: In total, 108 clinical *C. difficile* strains of ST81 were isolated and successfully analyzed by WGS, which showed that the percentage of isolates with AMRs was common in this type of ST. Furthermore, two types of transposons, Tn916 and Tn6189, were also detected. We found that all *C. difficile* ST81 genomes were closely related as pairwise core-genomic SNP (cgSNP) distance between the strains was on average 13 cgSNPs (range, 0–425 cgSNPs). Notably, these isolates were split into two sub-lineages (SL I and SL II) by Bayesian analysis, which suggested that both sub-lineages emerged independently. It is noted that some AMRs (such as *clb*A, *dfr*F, and *cfr*B) and Tn916 were only detected in SL I.

Conclusion: *C. difficile* ST81 is among the common STs in this study. Two independent sub-lineages of *C. difficile* ST81 strains are found. Furthermore, the presence of a high number of AMR genes and multiple mobile elements indicate a potential risk for transmission of *C. difficile* ST81. Based on these results, a robust surveillance system is crucial for identifying outbreaks, tracking infection trends, and implementing timely interventions.

Keywords: *Clostridioides difficile* infection, *C. difficile* ST81, antimicrobial resistance, whole genome sequencing, transmission

Introduction

Clostridioides difficile is an important pathogen of antimicrobial-associated diarrhea and nosocomial diarrhea in humans and has been ranked as the highest level of "Urgent threats" by the Centers for Disease Control and Prevention (CDC) in the USA (https://www.cdc.gov/).[1] The clinical outcomes of *C. difficile* infection (CDI) range from self-resolving diarrhea to life-threatening pseudomembranous colitis and toxic megacolon, and even death.[2] Two structurally similar toxins, toxin A (TcdA, an enterotoxin) and toxin B (TcdB, a cytotoxin), are the main virulence determinants associated with CDI.[3] Furthermore, a third toxin, the binary toxin (CDT), has been linked to an increased severity of human infections, as evidenced by the hypervirulent *C. difficile* NAP1/BI/027 strain.[4] Nevertheless, in practice, not all toxin-producing *C. difficile* strains produce the toxins mentioned above.

As a TcdA-negative and TcdB-positive strain, *C. difficile* sequence type (ST) 37 is regarded as one of the most prevailing *C. difficile* STs, which has resulted in widespread CDI and global outbreaks.[5] In Asia, alongside *C. difficile*

Received: 8 October 2024
Accepted: 23 November 2024
Published: 11 December 2024

ST37, ST81 (also a TcdA-negative and TcdB-positive strain) emerges as another common genotype, notably associated with ribotype (RT) 369.[6,7] In initial molecular epidemiological investigations, *C. difficile* ST81 was only found at a low prevalence and it was a predominantly endemic strain.[8] However, recent studies have indicated that *C. difficile* ST81 has gradually replaced ST35, ST3, and ST37, becoming a prevalent strain in some regions throughout mainland China.[6,9] For example, ST81 accounted for 26.4% of all isolated clinical toxigenic *C. difficile* in four tertiary hospitals in Beijing, China.[6] Furthermore, using whole genome sequencing (WGS) to trace the origins of infection, *C. difficile* ST81 strains have been reported to cause outbreak in a tertiary hospital in Shanghai.[10] It is interesting to note a report by Liu et al., in which the authors found that patients infected with *C. difficile* ST81 were more likely to have lower survival rates than those infected with non-ST81 strains.[11] In order to elucidate the epidemiological characteristics of *C. difficile* ST81 isolates, several genomic studies have been carried out. One genome analysis study has revealed that *C. difficile* ST81 has a high-frequency amino acid mutation in the *gyrA* (Thr82Ile), which confers fluoroquinolone resistance.[6] Compared with *C. difficile* ST37, *C. difficile* ST81 isolates exhibit an enhanced ability to transmit between hosts and survive in harsh environments with robust colonization, enhanced spore production, and slightly increased motility.[12]

However, given the potential for *C. difficile* ST81 transmission, high antimicrobial resistance and poor prognosis, research into the epidemiological characteristics of this strain remains limited. Compared with other STs, the molecular characteristics, such as the genetic diversity, the population structure, genome evolution, and pathogenicity in *C. difficile* ST81 in China need to be addressed. Here, we performed WGS and high-resolution phylogenomic analysis on *C. difficile* ST81 isolates in China with the aims to: 1) characterize the genetic diversity and population structure; 2) identify resistance mechanisms; and 3) analyze the phylogenetic relationships of *C. difficile* ST81 isolates. Overall, it is hoped that our data might provide a solid foundation for future in-depth investigations of functional determinants in *C. difficile* ST81 to inform novel strategies for improving infection prevention and control, as well as patient management.

Materials and Methods

Collection of *C. difficile* Isolates

Between January 2010 and January 2021, stool samples were collected from adult- patients with diarrhea required for toxigenic *C. difficile* detection at five tertiary hospitals across different regions of China (HA, HB, HC, HD, HE). The stool sample was cultured anaerobically on cycloserine-cefoxitin-taurocholate agar (CCFA-TA; Oxoid) supplemented with 7% (v/v) sheep serum at 35°C for 48 hours before isolation. The suspected *C. difficile* colonies were identified using Brooke Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF) (Bruker Daltonik GmbH, Bremen, Germany).

Detection of Toxin Genes by Polymerase Chain Reaction (PCR)

Following the 48 h of anaerobic blood agar culture, *C. difficile* isolates were suspended in distilled water in a micro-centrifuge tube. Genomic DNA was then extracted using the simplified alkaline lysis method. All isolated strains were tested for *tcdA, tcdB*, and binary toxin genes by PCR, as previously described.[13] Briefly, the primer pairs were NK9/NK11 for *tcdA*, NK104/NK105 for *tcdB*, cdtApos/cdtArev for *cdtA*, and cdtBpos/cdtBrev for *cdtB*. PCR amplification with primer pair NK9/NK11 was performed for 35 cycles, consisting of 95°C for 20s and 62°C for 120 s. The thermal profile for primer pairs NK104/NK105 was 35 cycles comprising 95°C for 20 s and 55°C for 120 s. Reactions for *cdtA* and *cdtB* genes were subjected to 30 cycles of 94°C for 45 s, 52°C for 60s and 72°C for 80 s. *C. difficile* strain ATCC BAA-1870 (ribotyping 027) was used as the positive template control of toxin genes.

Multi-Locus Sequence Typing (MLST)

To characterize genetic diversity and determining STs of *C. difficile* isolates in the study, MLST of seven housekeeping genes (*adk, atpA, dxr, glyA, recA, soda*, and *tpi*) was used to genotype all toxigenic isolates according to previously described protocols.[14] Briefly, the amplification conditions of each housekeeping gene were 95°C for 15 min, followed by 35 cycles of 94°C for 30 s, 50°C for 40 s, and 72°C for 70 s. Allele designations were obtained through the *C. difficile* PubMLST batch profile query page to determine the ST, as previously described.[8]

Genome Sequencing and Assembly

Sequencing-quality genomic DNA of all *C. difficile* strains identified as ST81 were extracted using a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and quantitated using a Qubit 2.0 system. WGS was performed on an Illumina NovaSeq 6000 instrument (Illumina, San Diego, CA, USA), reaching a sequencing coverage of 200X. Before assembly, quality control of raw sequenced reads was performed using FastQC and adapter regions were trimmed using Trimmomatic.[15] Trimmed reads were assembled *de novo* using SPAdes.[16] The total assembly size of genome is approximately 4.2 Mb, with a scaffold N50 length of approximately 0.5Mb and the GC content of approximately 28.93%. Genomes were annotated using the Prokka.[17]

Phylogenetic and Evolutionary Analyses

To investigate population structure, a maximum-likelihood tree was constructed based on core-genome alignments obtained in Roary using MEGA 11 with 1000 bootstrap replicates and visualized using the Interactive Tree of Life web server.[18,19] Further, to investigate the transmission events of *C. difficile* ST81, single nucleotide polymorphisms (SNPs) variants were called using Snippy (http://github.com/tseemann/snippy) with default parameters. The *C. difficile* M68 (NC_017175.1) chromosome served as the reference. The alignment file was filtered from variants with elevated densities of base substitutions as putative repetitive regions, mobile genetic elements (MGEs) and recombination events by Gubbins and used to calculate pairwise SNPs by SNP-dists (https://github.com/tseemann/snp-dists).[20] The minimum spanning trees were constructed in PHYLOViZ based on the generated pairwise SNP tables.[21]

To explore the evolution of *C. difficile* ST81, Bayesian evolutionary analysis was performed using BEAST, which employs three clock models, two population models, and two site models, as described previously by Xu et al.[22]

Identification of Antimicrobial Resistance (AMR) Genes, Virulence Genes, and Transposons

To investigate the genetic characteristics of *C. difficile* ST81, AMR genes were detected based on the Comprehensive Antibiotic Resistance Database (CARD) (https://card.mcmaster.ca). Further, virulence genes were identified based on the local Virulence Factor Database (VFDB) (https://www.mgc.ac.cn/VFs/). For mutation analysis, *gyrA* and *gyrB* sequences were extracted and compared with reference sequences downloaded from the CARD using the Basic Local Alignment Search Tool (BLAST). Transposable elements were identified using Vrprofile2.[23]

Pangenome-Wide Association Study (Pan-GWAS) and Clusters of Orthologous Groups of Proteins (COG) Analysis

The Bayesian analyses identified two distinct *C. difficile* ST81 sub-lineages, namely sub-lineage I (SL I) and SL II. To determine significant genetic loci associated with each sub-lineage, a pan-GWAS of all *C. difficile* ST81 genomes was performed. Briefly, all *C. difficile* ST81 genome annotations were performed with Prokka.[17] Roary was employed to estimate the size of the core and accessory genomes, and the results were used as an input for Scoary to identify the significant genetic loci associated with each sub-lineage.[19,24] Moreover, genes were extracted from all genomes using an in-house Python script and uploaded to eggNOG-mapper (http://eggnog-mapper.embl.de/) to explore gene function.

Data Analysis

Statistical analyses were performed using SPSS version 23.0 (SPSS, Chicago, IL, USA).

Results

Epidemiologic Analysis and Characterization of *C. difficile* Isolates

In total, 871 (9.0%) non-duplicate toxigenic *C. difficile* isolates were identified from 9675 patients suffering from diarrhea during the period of study. Among these isolates, a total of 59 STs were identified (data not shown), with ST81 accounting for 12.4% (108/871). Most *C. difficile* ST81 strains were isolated from inpatients (94.4%,102/108) while 5.6% (6/108) were isolated from outpatients. During this period, strains of *C. difficile* ST81 were isolated each year from the five participating hospitals, but the

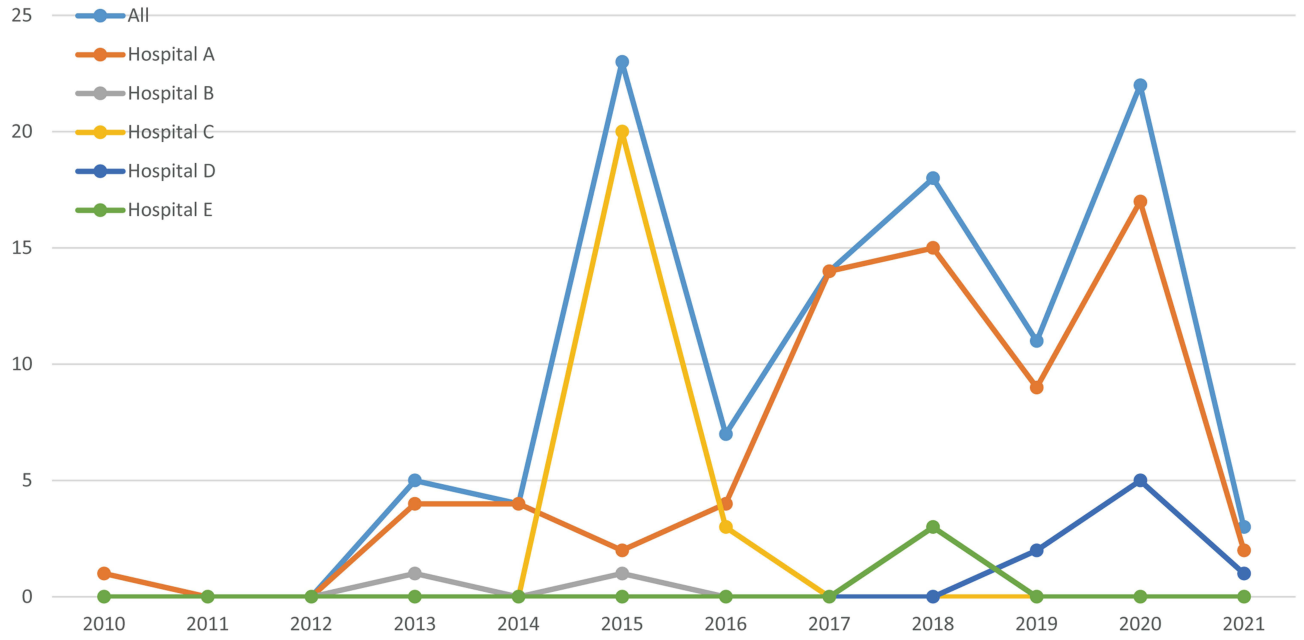

Figure 1 The distribution of 108 isolates in different hospitals and years.

distribution of the *C. difficile* strains varied between the hospitals over time (Figure 1). In Hospital A (HA), the number of *C. difficile* ST81 isolates showed an increasing trend from 2016, peaking in 2020, while Hospital C (HC) had a significantly high number of isolates in 2015.

Phylogenetic Analyses and Genomic Features of *C. difficile* ST81

All 108 WGS data were analyzed after sequence quality control and mapping to *C. difficile* M68 (NC_017175.1). A total of 11,950 genes were predicted to comprise the pangenome of *C. difficile* ST81. The core genes (ie, genes present in more than 99% of the genomes), which is usually used to evaluate the genomic diversity within species, represented 29.9% (3568/11,950) of the pangenome, while cloud genes, shell genes, and soft-core genes comprised 63.8% (7630/11,950), 5.8% (695/11,950), and 0.5% (57/11,950), respectively.

Seven AMR genes were identified while *erm*B and *tet*M were detected in all but one strain each (s15013004 and s13071105, respectively) (Figure 2A). The *clb*A, *dfr*F, and *cfr*B resistance genes were identified in a small number of strains. Furthermore, the majority of *C. difficile* ST81/RT369 isolates were found to have *gyrA* mutations (91.7%, 99/108), which manifested as a Thr82Ile amino acid mutation conferring fluoroquinolone resistance.

As essential mobile genetic elements, integrative and conjugative elements are responsible for horizontal gene transfer, driving increased genetic diversity, and the acquisition of exogenous genes.[25] In these 108 isolates, only two kinds of transposons, Tn*916*, and Tn*6189*, were identified. With the exception of one, we identified Tn*916* in most *C. difficile* ST81/RT369 isolates while 89.8% (97/107) of isolates contained Tn6189. Tn*916*, which encodes resistance to tetracycline and minocycline via Tet (M), is a major family of transposon reported in *C. difficile*, while Tn*6189*, described for the first time in 2019, is a carrier of *erm*B gene.[26,27]

To establish in-depth phylogeography of *C. difficile* ST81 in China, cgSNP calling was performed. The cgSNP alignment consisted of 1,589,277 bp. Pairwise cgSNP distance between the strains was on average 13 cgSNPs (range, 0–425 cgSNPs). For improved resolution of *C. difficile* ST81 phylogeny, we constructed a maximum-likelihood phylogenetic tree and a minimum spanning tree based on these cgSNPs (Figure 2B). Both methods indicated the occurrence of highly similar strains in very different regions in China.

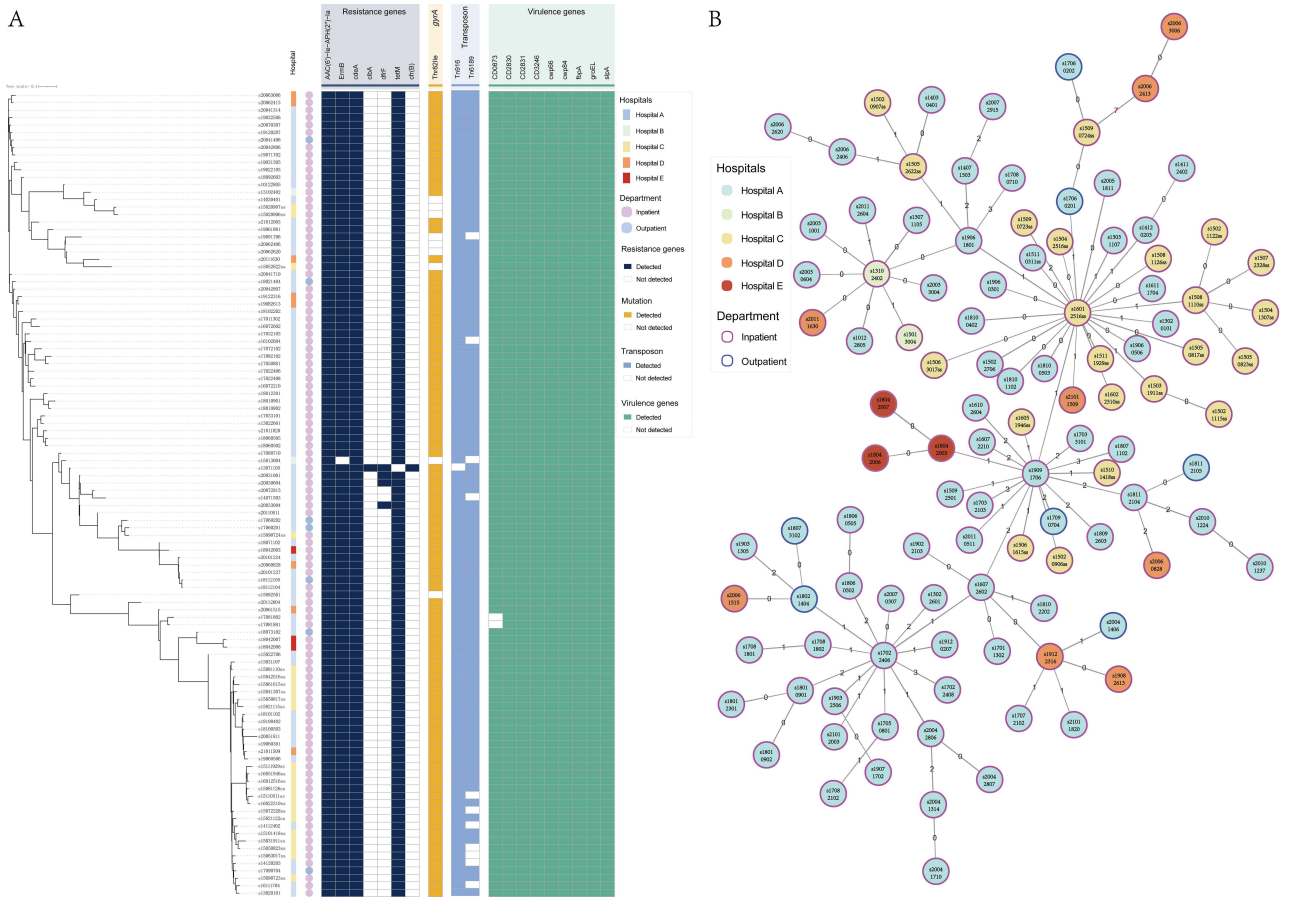

Figure 2 Phylogenetic analysis of 108 *C. difficile* ST81 strains based on single nucleotide polymorphisms (SNPs) in the core genome. (**A**) A phylogenetic tree was constructed, illustrating the relationships between strains and describing the strain number, antimicrobial resistance, transposons, and virulence genes in each hospital. (**B**) A Minimum spanning tree of *C. difficile* ST81 reveals formation of developmental clusters of clonal transmission based on core-genome single nucleotide polymorphisms.

Molecular Evolution and Transmission of *C. difficile* ST81/RT369 in China

To explore evolutionary patterns of *C. difficile* ST81, all genomes with details of collection dates were used to construct the Bayesian phylogeny tree. Based on the topology of the tree, the constructed Bayesian evolutionary tree indicated the presence of two genetically diverse sub-lineages, designated SL I and SL II (Figure 3A), and estimated that SL I emerged with the most recent common ancestors in ~2009 while SL II emerged in ~2010 (median estimates of 95% highest posterior density intervals were 2003 to 2009 for all strains of *C. difficile* ST81). Moreover, the Bayesian evolutionary tree also suggested that both sub-lineages emerged independently. The distribution of strains in SL I was non-hospital and non-time dependent while the strains in SL II were mostly isolated from Hospital A. Interestingly, all of the strains from SL II carried resistant genes *erm*B and *tet*M, the *gyrA* mutation, and Tn916 while the *clb*A, *dfr*F, and *cfr*B resistance genes were only detected in SL I.

Pangenome Analyses

To investigate the variability between the two different *C. difficile* ST81 sub-lineages identified by Bayesian analysis, we conducted pangenome analysis of these isolates. Comparison of the two different *C. difficile* ST81 sub-lineages led to the identification of 2 and 15 unique genes in SL I and SL II isolates, respectively (Figure 3B). To gain an understanding of the function of these unique genes in *C. difficile* ST81 isolates, significant unique genes in both sub-lineages were assigned to COG functional categories. More unique genes in SL II were enriched in various COG functional categories compared with SL I (Figure 3C and Table 1). The largest proportion of unique genes belonged to function unknown, followed by transcription and defense mechanisms. It is important to note that signal transduction mechanisms, cell wall/

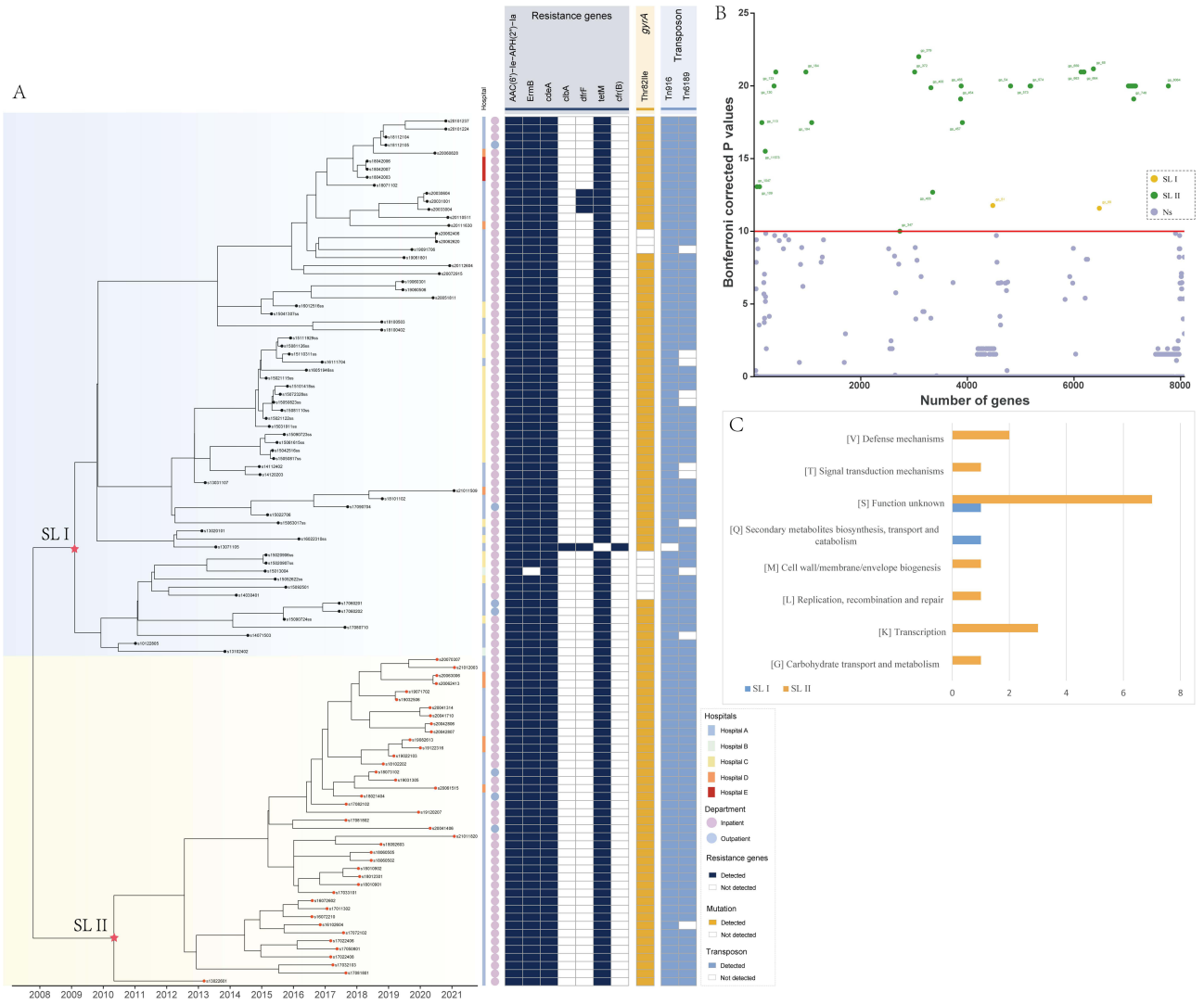

Figure 3 Evolution of 108 *C. difficile* ST81 genomes over time. (**A**) A Bayesian phylogenetic tree was constructed, describing the strain number, antimicrobial resistance, transposons, and virulence genes in each hospital. Two major sub-lineages were defined: SL I and SL II. (**B**) A Manhattan plot of the associations of genes in the two sub-lineages based on genome-wide association analysis. Association of genes with phenotypes was determined using the multiple Kruskal–Wallis test. (**C**) Distribution of differentially expressed proteins among two *C. difficile* ST81 sub-lineages according to COG functional categories.

membrane/envelope biogenesis, replication, recombination, and repair, as well as carbohydrate transport and metabolism genes in SL II isolates were much more abundant than those in SL I isolates, which may endow these isolates with the ability to defend against adverse stimuli, allowing them to use any extra energy to better adapt to the environment.

Table 1 Functional Annotation of Differentially Expressed Genes Between SL I to SL II

Gene	Description	Sub-lineage	Odds_Ratio	Bonferroni_p	Empirical_p	COG_Category
Group_69	Phage Mu protein F like protein	VL I	0.01	2.62E-12	0.01980198	S
Group_51	PFAM Collagen triple helix	VL I	0.01	1.69E-12	0.01980198	Q
Group_347	VanW like protein	VL II	47.09	9.88E-11	0.02970297	V
Group_109	GHKL domain	VL II	183.33	8.63E-14	0.01980198	T
Group_400	Phage-related minor tail protein	VL II	633.75	1.35E-20	0.01980198	S
Group_68	Pfam:Terminase_3C	VL II	1287.00	6.78E-22	0.04950495	S
Group_113	C-5 cytosine-specific DNA methylase	VL II	inf	3.35E-18	0.00990099	S

(*Continued*)

Table 1 (Continued).

Gene	Description	Sub-lineage	Odds_Ratio	Bonferroni_p	Empirical_p	COG_Category
Group_130	NlpC/P60 family	VL II	inf	1.02E-20	0.00990099	GM
Group_184	Phage antirepressor protein	VL II	inf	1.11E-21	0.01980198	K
Group_372	Acetyltransferase	VL II	inf	1.11E-21	0.01980198	K
Group_403	DNA binding	VL II	inf	2.07E-13	0.04950495	S
Group_454	GrpB protein	VL II	inf	7.98E-20	0.04950495	S
Group_573	Uncharacterized protein conserved in bacteria (DUF2313)	VL II	inf	1.02E-20	0.03960396	S
Group_663	Domain of unknown function (DUF4355)	VL II	inf	1.11E-21	0.04950495	S
Group_740	DnaD domain protein	VL II	inf	1.02E-20	0.03960396	L
Group_746	Sigma factor activity	VL II	inf	1.02E-20	0.01980198	K
Group_750	Endonuclease activity	VL II	inf	1.02E-20	0.02970297	V

Discussion

It has been reported that the prevalence of CDI in mainland China was 14% while ST81 is among the common STs of *C. difficile*.[28,29] In this longitudinal and systematic surveillance on the status of CDI in China, a multi-center study was undertaken with five large tertiary hospitals to improve our understanding of the molecular epidemiology of *C. difficile* ST81. Herein, we found that *C. difficile* ST81 accounted for 12.4% of the total toxigenic *C. difficile* isolates. Phylogenetic tree analyses indicated the occurrence of highly similar strains in very different regions in China. Two distinct sub-lineages of *C. difficile* ST81 were identified based on differences in genomic structure.

Several studies of *C. difficile* have shown that *C. difficile* ST37/RT17, a TcdA-negative and TcdB-positive (A-B+) clone, is the most prevalent clone in Asia.[5] However, other studies have also identified A-B+ genotype ST81 as the prevalent clone and causative agent of multiple outbreaks in China, since its identification in 2010.[30] In the present study, although *C. difficile* ST81 was not the most common genotype, this ST accounted for 12.4% of the analyzed isolates, which was similar to data obtained in a Japanese multi-center surveillance study between April 2012 and March 2013.[7] However, in recent years, several studies in China have shown that ST81 was the main common genotype responsible for *C. difficile* outbreaks.[6,9] Although the exact cause remains unclear, some studies have suggested that this may be due to the increased resistance rates along with a higher production of treatment-resistant spores in this genotype.[9,31] However, the composition of *C. difficile* ST81 strains varied in the studied hospitals.[6] Similar trends were observed in this study, which suggests the need to establish a surveillance network as an important strategy for epidemiological monitoring of *C. difficile*, as well as for CDI transmission control.

AMR plays a significant role in the pathogenesis and spread of CDI, as it allows *C. difficile* to survive antimicrobial exposure in the host, while selective pressure allows the emergence and spread of AMR strains.[32] Notably, fluoroquinolone resistance resulting from a *gyr*A (Thr82Ile) mutation in *C. difficile* RT027/ST1 is thought to have significantly facilitated its rapid expansion and dissemination in North America and Europe in the early 2000s.[33] *C. difficile* ST81 isolates have been shown to have the highest in vitro resistance rates for many antimicrobial classes, such as fluoroquinolones (ciprofloxacin, levofloxacin, and moxifloxacin), tetracyclines, macrolide-lincosamide-streptogramin B (MLS_B).[6] Although antimicrobial susceptibility tests were not performed for these *C. difficile* ST81 isolates in the present study, previous research has confirmed that the resistant phenotype and genotype showed good correlation.[34] In this work, we found that most of the isolates carried the resistant genes (*tet*M, *erm*B) and mutations in *gyr*A were detected in almost all strains. It is interesting to note that the *cde*A gene, described as a multidrug efflux transporter belonging to the multidrug and toxic compound extrusion family (MATE-family), has been shown to confer resistance not only to ethidium bromide and acriflavin but also to antibiotics, particularly fluoroquinolones and, in some cases, aminoglycosides. Remarkably, this gene was present in all isolates.[35] However, in another study, this gene was found both in levofloxacin-resistant and susceptible strains.[36] Further research into the role of the *cde*A gene is required to elucidate the mechanisms behind this. Intriguingly, while resistance to rifamycin is characteristic of *C. difficile* ST37/RT17, none of rifamycin resistance genes were identified in isolates analyzed in this study.[13] This phenomenon underscores the possibility that while *C. difficile* ST81 and *C. difficile* ST37/RT17 are the main genotypes of

C. difficile clade 4, these two genotypes may have different molecular features.[12] Another interesting note is that Tn*916* and Tn*6189* are found in most *C. difficile* ST81 isolates. These two transposons carry resistance genes (*tet*M and *erm*B, respectively), which may help explain why *C. difficile* ST81 harbors the aforementioned resistance genes. This also highlights the potential role of transposons in the horizontal transfer of resistance genes.

C. difficile ST81 has been reported to cause nosocomial transmission in a general hospital in China.[10] The present study followed the approach defined by Eyre et al., in which a threshold of 0–2 cgSNPs was used to determine whether groups of two or more strains were clonally related.[37] The constructed phylogenetic tree based on the cgSNPs provided a clear interpretation of the evolutionary relationship between all of the *C. difficile* ST81 isolates in this study, which showed that strains in the same clade are more closely related to each other and may have a common evolutionary ancestor. As the isolates were collected over time across China, and because no intersection of time and space was found after reviewing the clinical information, we could not conclude that this genotype caused spatial-temporal transmission in China. However, we have reason to believe that this genotype evolved more slowly while *C. difficile* ST81 showed low intra-ST cgSNP difference with a mean value of 13 cgSNPs. We also could not conclude that they likely had connections with the present outbreak or were exposed to a common source. These data indicate that isolates of *C. difficile* ST81 are genetically more closely related to each other compared with other ST lineages, and that related isolates can be identified regardless of geographical origin. The pangenome analyses showed an open pangenome, with the accessory genome comprising 70.1%. In comparison, other *C. difficile* lineages, like RT014 and ST11, have accessory genomes of 69.7% and 80.2%, respectively, which is similar to ST81.[38,39] This finding necessitates additional research to understand the evolutionary causes and epidemiological implications of STs with open pangenomes. The low level of diversity found in this study suggested that ST81 was a persistent multi-province clone that was present in different provinces. Another study by Thiel et al. demonstrated that bacteria found in manure can escape into the atmosphere during agricultural land fertilization and be transported thousands of kilometers away.[40,41] However, the true reasons for long-distance transmission or conservative core genome of *C. difficile* ST81 remain unclear and require further investigation.

Bayesian evolutionary analysis showed that *C. difficile* ST81 isolated from the present study formed two independent and stably transmitted sub-lineages based on the differences in structure. Although not exclusively containing strains from one region, the spread of *C. difficile* ST81 had occurred in ~2008, which was derived from the same most recent common ancestor. From this analysis, we inferred that the spread of *C. difficile* ST81 probably began with population movement in Hospital A before spreading to other hospitals in different regions. Although the real mechanism is unclear, given that the *C. difficile* ST81 genotype was first found in Japan, we hypothesized that as Hospital A is located in a major city with frequent exchanges involving Japanese people and international travelers, these interactions could have facilitated the transmission.[7] Furthermore, there were significant difference in genomic characteristics in the two sub-lineages. Using GWAS analysis, we found that SL II isolates containing more unique genes belonged to signal transduction mechanisms and cell wall/membrane/envelope biogenesis, which may allow these isolates to better defend themselves against adverse stimuli and use any extra energy to better adapt to their environment. This further indicates that the SL II strain is isolated in only two regions.

It is important to acknowledge the limitations of this study. Although we collected isolates from patients with diarrhea in five different tertiary hospitals across mainland China, specific areas are needed for further research, such as expanding surveillance to more hospitals, and investigating environmental reservoirs. Secondly, we did not perform in vitro antimicrobial susceptibility tests for these *C. difficile* ST81 isolates, which meant that the lack of in vitro antimicrobial susceptibility testing limits the ability to correlate genetic resistance markers with actual resistance phenotypes. Thirdly, *C. difficile* ST81 is currently not a common genotype outside of China, and there is a lack of WGS data globally. Additionally, more in-depth functional studies on the unique genes found in SL II strains are necessary. Therefore, the representativeness in this work needs to be further studied by involving more isolates.

Conclusions

In this study, we investigated the genetic diversity of 108 *C. difficile* ST81 isolates with temporal and geographical variation. Our data suggests that *C. difficile* ST81 has high resistant rates to different antibiotics and is circulating in mainland China. Phylogeographic analyses of the cgSNPs identified through WGS of the isolates suggests that there are two main sub-lineages (SL I and SL II), which share ancestry, although SL II shows increased adaptability to the

environment with relative genes. Considering the unusual antibiotic resistance and the conservation in core genes of *C. difficile* ST81, surveillance of this genotype needs to be further strengthened to improve disease control and reduce the risk of future outbreaks.

Data Sharing Statement

The genomic sequences of *C. difficile* ST81 isolates in this study were deposited in GenBank under accession number PRJNA432876. The data of this study are available by contacting the corresponding author upon reasonable request.

Ethics Approval

For this observational study, the need for patient consent was not required. Data were not identifiable back to the patients from whom they originated; an ethics approval was waived.

Funding

This study was supported by Basic Public Welfare Research Program of Zhejiang Province (No. LGF20H030004), National Nature Science Foundation of China (No. 82073609).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Sandhu BK, McBride SM. Clostridioides difficile. *Trends Microbiol*. 2018;26(12):1049–1050. doi:10.1016/j.tim.2018.09.004
2. Guh AY, Kutty PK. Clostridioides difficile Infection. *Ann Intern Med*. 2018;169(7):ITC49–ITC64. doi:10.7326/AITC201810020
3. Rupnik M. Heterogeneity of large clostridial toxins: importance of Clostridium difficile toxinotypes. *FEMS Microbiol Rev*. 2008;32(3):541–555. doi:10.1111/j.1574-6976.2008.00110.x
4. Stabler RA, He M, Dawson L, 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. doi:10.1186/gb-2009-10-9-r102
5. Imwattana K, Knight DR, Kullin B, et al. Clostridium difficile ribotype 017 - characterization, evolution and epidemiology of the dominant strain in Asia. *Emerg Microbes Infect*. 2019;8(1):796–807. doi:10.1080/22221751.2019.1621670
6. Cheng JW, Liu C, Kudinha T, et al. The tcdA-negative and tcdB-positive Clostridium difficile ST81 clone exhibits a high level of resistance to fluoroquinolones: a multi-centre study in Beijing, China. *Int J Antimicrob Agents*. 2020;56(1):105981. doi:10.1016/j.ijantimicag.2020.105981
7. Senoh M, Kato H, Fukuda T, et al. 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):1226–1236. doi:10.1099/jmm.0.000149
8. Chen YB, Gu SL, Wei ZQ, et al. Molecular epidemiology of Clostridium difficile in a tertiary hospital of China. *J Med Microbiol*. 2014;63(Pt 4):562–569. doi:10.1099/jmm.0.068668-0
9. Yang Z, Huang Q, Qin J, et al. Molecular epidemiology and risk factors of clostridium difficile ST81 infection in a teaching hospital in Eastern China. *Front Cell Infect Microbiol*. 2020;10:578098. doi:10.3389/fcimb.2020.578098
10. Qin J, Dai Y, Ma X, et al. Nosocomial transmission of Clostridium difficile genotype ST81 in a general teaching hospital in China traced by whole genome sequencing. *Sci Rep*. 2017;7(1):9627. doi:10.1038/s41598-017-09878-8
11. Liu Y, Ma L, Cheng J, Su J. Effects of omeprazole on recurrent clostridioides difficile infection caused by ST81 strains and their potential mechanisms. *Antimicrob Agents Chemother*. 2023;67(6):e0022123. doi:10.1128/aac.00221-23
12. Su T, Chen W, Wang D, et al. Complete genome sequencing and comparative phenotypic analysis reveal the discrepancy between Clostridioides difficile ST81 and ST37 Isolates. *Front Microbiol*. 2021;12:776892. doi:10.3389/fmicb.2021.776892
13. Chen YB, Gu SL, Shen P, et al. Molecular epidemiology and antimicrobial susceptibility of Clostridium difficile isolated from hospitals during a 4-year period in China. *J Med Microbiol*. 2018;67(1):52–59. doi:10.1099/jmm.0.000646
14. Griffiths D, Fawley W, Kachrimanidou M, et al. Multilocus sequence typing of Clostridium difficile. *J Clin Microbiol*. 2010;48(3):770–778. doi:10.1128/JCM.01796-09
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170
16. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–477. doi:10.1089/cmb.2012.0021
17. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–2069. doi:10.1093/bioinformatics/btu153
18. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–3027. doi:10.1093/molbev/msab120
19. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691–3693. doi:10.1093/bioinformatics/btv421
20. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*. 2015;43(3):e15. doi:10.1093/nar/gku1196

21. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrico JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinf*. 2012;13:87. doi:10.1186/1471-2105-13-87
22. Xu X, Luo Y, Chen H, et al. Genomic evolution and virulence association of Clostridioides difficile sequence type 37 (ribotype 017) in China. *Emerg Microbes Infect*. 2021;10(1):1331–1345. doi:10.1080/22221751.2021.1943538
23. Wang M, Goh YX, Tai C, Wang H, Deng Z, Ou HY. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res*. 2022;50(W1):W768–W773. doi:10.1093/nar/gkac321
24. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with scoary. *Genome Biol*. 2016;17(1):238. doi:10.1186/s13059-016-1108-8
25. He M, Sebaihia M, Lawley TD, et al. Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc Natl Acad Sci U S A*. 2010;107(16):7527–7532. doi:10.1073/pnas.0914322107
26. Jon JV, Mark HW, Jane F. Antimicrobial resistance progression in the United Kingdom: a temporal comparison of Clostridioides difficile antimicrobial susceptibilities. *Anaerobe*. 2021;70:102385. doi:10.1016/j.anaerobe.2021.102385
27. Imwattana K, Kiratisin P, Riley TV, Knight DR. Genomic basis of antimicrobial resistance in non-toxigenic Clostridium difficile in Southeast Asia. *Anaerobe*. 2020;66:102290. doi:10.1016/j.anaerobe.2020.102290
28. Tang C, Cui L, Xu Y, et al. The incidence and drug resistance of Clostridium difficile infection in Mainland China: a systematic review and meta-analysis. *Sci Rep*. 2016;6(1):37865. doi:10.1038/srep37865
29. Wen BJ, Dong N, Ouyang ZR, et al. Prevalence and molecular characterization of Clostridioides difficile infection in China over the past 5 years: a systematic review and meta-analysis. *Int J Infect Dis*. 2023;130:86–93. doi:10.1016/j.ijid.2023.03.009
30. Byun JH, Kim H, Kim JL, et al. A nationwide study of molecular epidemiology and antimicrobial susceptibility of Clostridioides difficile in South Korea. *Anaerobe*. 2019;60:102106. doi:10.1016/j.anaerobe.2019.102106
31. Wang B, Peng W, Zhang P, Su J. The characteristics of Clostridium difficile ST81, a new PCR ribotype of toxin A- B+ strain with high-level fluoroquinolones resistance and higher sporulation ability than ST37/PCR ribotype 017. *FEMS Microbiol Lett*. 2018;365(17). doi:10.1093/femsle/fny168
32. Imwattana K, Knight DR, Kullin B, et al. Antimicrobial resistance in Clostridium difficile ribotype 017. *Expert Rev Anti Infect Ther*. 2020;18 (1):17–25. doi:10.1080/14787210.2020.1701436
33. He M, Miyajima F, Roberts P, et al. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet*. 2013;45 (1):109–113. doi:10.1038/ng.2478
34. Imwattana K, Putsathit P, Knight DR, Kiratisin P, Riley TV. Molecular characterization of, and antimicrobial resistance in, clostridioides difficile from Thailand, 2017-2018. *Microb Drug Resist*. 2021;27(11):1505–1512. doi:10.1089/mdr.2020.0603
35. Dridi L, Tankovic J, Petit JC. CdeA of Clostridium difficile, a new multidrug efflux transporter of the MATE family. *Microb Drug Resist*. 2004;10 (3):191–196. doi:10.1089/mdr.2004.10.191
36. Li X, Wang Y, Cao R, et al. Antimicrobial resistance of clostridioides difficile in children from a tertiary pediatric hospital in Shanghai, China. *Infect Drug Resist*. 2024;17:329–339. doi:10.2147/IDR.S441312
37. Eyre DW, Cule ML, Wilson DJ, et al. Diverse sources of C. difficile infection identified on whole-genome sequencing. *N Engl J Med*. 2013;369 (13):1195–1205. doi:10.1056/NEJMoa1216064
38. Knight DR, Squire MM, Collins DA, Riley TV. Genome analysis of Clostridium difficile PCR ribotype 014 lineage in Australian pigs and humans reveals a diverse genetic repertoire and signatures of long-range interspecies transmission. *Front Microbiol*. 2016;7:2138. doi:10.3389/fmicb.2016.02138
39. Knight DR, Kullin B, Androga GO, et al. Evolutionary and genomic insights into clostridioides difficile sequence type 11: a diverse zoonotic and antimicrobial-resistant lineage of global one health importance. *mBio*. 2019;10(2). doi:10.1128/mBio.00446-19
40. Thiel N, Munch S, Behrens W, et al. Airborne bacterial emission fluxes from manure-fertilized agricultural soil. *Microb Biotechnol*. 2020;13 (5):1631–1647. doi:10.1111/1751-7915.13632
41. O'Grady K, Hong S, Putsathit P, et al. Defining the phylogenetics and resistome of the major Clostridioides difficile ribotypes circulating in Australia. *Microb Genom*. 2024;10(5). doi:10.1099/mgen.0.001232