MICROBIAL GENOMICS

RESEARCH ARTICLE

Imwattana *et al.*, *Microbial Genomics* 2022;8:000792

DOI 10.1099/mgen.0.000792

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Global evolutionary dynamics and resistome analysis of *Clostridioides difficile* ribotype 017

Korakrit Imwattana[1,2], Papanin Putsathit[3], Deirdre A. Collins[3], Teera Leepattarakit[2], Pattarachai Kiratisin[2], Thomas V. Riley[1,3,4,5] and Daniel R. Knight[1,4,*]

Abstract

Clostridioides difficile PCR ribotype (RT) 017 ranks among the most successful strains of *C. difficile* in the world. In the past three decades, it has caused outbreaks on four continents, more than other 'epidemic' strains, but our understanding of the genomic epidemiology underpinning the spread of *C. difficile* RT 017 is limited. Here, we performed high-resolution phylogenomic and Bayesian evolutionary analyses on an updated and more representative dataset of 282 non-clonal *C. difficile* RT 017 isolates collected worldwide between 1981 and 2019. These analyses place an estimated time of global dissemination between 1953 and 1983 and identified the acquisition of the *ermB*-positive transposon Tn*6194* as a key factor behind global emergence. This coincided with the introduction of clindamycin, a key inciter of *C. difficile* infection, into clinical practice in the 1960s. Based on the genomic data alone, the origin of *C. difficile* RT 017 could not be determined; however, geographical data and records of population movement suggest that *C. difficile* RT 017 had been moving between Asia and Europe since the Middle Ages and was later transported to North America around 1860 (95% confidence interval: 1622–1954). A focused epidemiological study of 45 clinical *C. difficile* RT 017 genomes from a cluster in a tertiary hospital in Thailand revealed that the population consisted of two groups of multidrug-resistant (MDR) *C. difficile* RT 017 and a group of early, non-MDR *C. difficile* RT 017. The significant genomic diversity within each MDR group suggests that although they were all isolated from hospitalized patients, there was probably a reservoir of *C. difficile* RT 017 in the community that contributed to the spread of this pathogen.

DATA SUMMARY

All new whole-genome sequence data generated in this study, highlighted in the Supplementary Document, have been submitted to the European Nucleotide Archive under BioProject PRJEB44406 (sample accessions ERS6268756–ERS6268798). The complete genome of *C. difficile* MAR286 was submitted to GenBank under BioProject PRJNA679085 (accession CP072118). Details of genomes included in the final analyses (282 genomes in the global analysis and an additional 13 genomes from the smaller analysis), as well as records of the phenotypic analyses are available in the Supplementary Document, available at https://www.doi.org/10.6084/m9.figshare.14544792. An interactive version of the Bayesian phylogenetic tree in Fig. 1 is available at https://microreactorg/project/v89tzQ8rii55PkAGF5Jo2r/64c80194.

Received 21 September 2021; Accepted 03 February 2022; Published 22 March 2022

Author affiliations: [1]School of Biomedical Sciences, The University of Western Australia, Australia; [2]Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand; [3]School of Medical and Health Sciences, Edith Cowan University, Australia; [4]Medical, Molecular and Forensic Sciences, Murdoch University, Australia; [5]Department of Microbiology, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Australia.

***Correspondence:** Daniel R. Knight, daniel.knight@uwa.edu.au

Keywords: *Clostridioides difficile*; ribotype 017; population; evolution; AMR; outbreaks.

Abbreviations: AMR, antimicrobial resistance; ANI, average nucleotide identity; CDI, *C. difficile* infection; c-di-GMP, cyclic dimeric guanosine monophosphate; cgSNP, core genome SNP; CI, confidence interval; E, epidemic lineage; ESS, effective sample size; GWAS, genome-wide association study; MCMC, Markov chain Monte Carlo; MDR, multi-drug resistance; MLST, multi-locus sequence typing; NAD, nicotinamide adenine dinucleotide; NE, non-epidemic lineage; OD, optical density; ONT, Oxford Nanopore Technology; PaLoc, pathogenicity locus; QR, quartile range; RT, ribotype; ST, sequence type; WGS, whole-genome sequencing.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. A supplementary file is available with the online version of this article.

Impact Statement

This study utilizes genomic sequence data from 282 non-clonal *Clostridioides difficile* ribotype (RT) 017 isolates collected from around the world to delineate the origin and spread of this epidemic lineage, as well as explore possible factors that have driven its success. It also reports on a focused epidemiological investigation of a cluster of *C. difficile* RT 017 in a tertiary hospital in Thailand to identify possible sources of transmission in this specific setting.

INTRODUCTION

Clostridioides difficile PCR ribotype (RT) 017, or sequence type (ST) 37, ranks among the most successful strains of *C. difficile*. Despite producing only one functional toxin (toxin B), *C. difficile* RT 017 has spread widely and caused outbreaks globally [1]. The severity of *C. difficile* infection (CDI) caused by RT 017 has been comparable to infection caused by *C. difficile* strains producing two or three toxins [2–4]. One factor that may have contributed to the success of *C. difficile* RT 017 is antimicrobial resistance (AMR) [5].

The evolutionary origins of *C. difficile* RT 017 remain contentious [1]. Possible contributing factors included the early erroneous dismissal of *C. difficile* RT 017 as non-pathogenic due to its lack of toxin A [6], and the use of diagnostic tests that only detected toxin A [7]. By the time that the pathogenicity of *C. difficile* RT 017 was recognized (1995) [8], the strain had already spread across the globe [1].

Based on multi-locus sequence type, *C. difficile* RT 017 is a member of evolutionary clade 4 [1]. This clade comprises several non-toxigenic *C. difficile* and toxigenic *C. difficile* that only produces toxin B (A-B+CDT-) [9]. Epidemiological evidence suggested that *C. difficile* clade 4 originated in Asia and *C. difficile* RT 017 later spread globally. First, *C. difficile* RT 017 has been the dominant strain in Asia for decades [10–14] and has only appeared sporadically in other regions [8, 15–20]. Second, reports of other *C. difficile* clade 4 strains have been exclusively from Asian countries, such as *C. difficile* ST 81 in China [21, 22], *C. difficile* RT 369 in Japan [23] and, most importantly, the high diversity of clade 4 non-toxigenic *C. difficile* in Southeast Asia [24]. However, there have not been many historical *C. difficile* RT 017 strains available from the region to verify this hypothesis [10–14].

In 2017, Cairns *et al.* analysed the whole-genome sequence (WGS) data of 277 *C. difficile* RT 017 strains from around the world. Their results suggested an alternative hypothesis, that *C. difficile* RT 017 instead originated in North America, spread to Europe in the 1990s and later spread to other regions [25]. A more recent study based mainly on the same dataset agreed with this hypothesis but estimated the time of spread to be before the 1970s [26]. Despite the large dataset, this conclusion might have been influenced by a strain selection bias, as the North American strains included in the study were relatively older than strains from other regions [25, 26].

To improve upon the previous analyses, the present study included a larger number of strains, with a few early European strains and a greater diversity of Asian strains. We aimed to explore the origin and spread of *C. difficile* RT 017, as well as the key genetic factors driving its success.

METHODS

C. difficile RT 017 genomes

This study started with 929 *C. difficile* RT 017 strains from three collections: a set of 45 clinical *C. difficile* RT 017 strains from Thailand [32 phenotypically multidrug-resistant (MDR) and 13 non-MDR] some of which have been described previously [27], 97 previously unpublished *C. difficile* RT 017 strains from our laboratory's collection and 787 *C. difficile* RT 017 genomes publicly available at the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/) as of January 2020. These collections included genomes of three *C. difficile* RT 017 isolated in the early 1980s (courtesy of Dr Jon Vernon and Prof. Mark Wilcox in Leeds, UK, but originally part of Prof. S. P. Borriello's collection) [28]. Multilocus sequence typing (MLST) was performed directly from sequence read files using SRST2 v0.2.0 and the PubMLST *C. difficile* database (https://pubmlst.org/organisms/clostridioides-difficile/) as previously described [29, 30]. After excluding clonal and redundant strains (see below), 282 *C. difficile* strains remained in the final dataset.

Assembly of a new complete *C. difficile* RT 017 genome from Southeast Asia

To facilitate phylogenomic analysis of *C. difficile* strains from Thailand, a Thai *C. difficile* strain was selected for hybrid assembly of a closed reference genome. *C. difficile* MAR286 was a non-MDR strain as opposed to the existing MDR reference strain *C. difficile* M68 [31]. Short-read sequencing was performed on an Illumina HiSeq sequencing platform (Illumina) using 150 bp paired-end chemistry to a depth of 39× coverage as previously described [30]. Long-read sequencing was performed on the MinION Mk1C machine (Nanopore). The sequencing libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109) and run on a

FLO-MIN106 (R9.4.1) flow cell, according to the manufacturer's instructions, for 24 h. Hybrid assembly was performed with Unicycler v0.4.8 using a conservative mode [32]. The final assembly graph was visualized and polished with Bandage v0.8.1 [33]. Genome annotation was performed using the NCBI Prokaryotic Genomes Annotation Pipeline [34].

AMR genotyping

AMR genotyping was performed as previously described [24]. Briefly, all read files were interrogated against the ARGannot database (for known accessory AMR genes) with two additional genes recently described in *C. difficile*, *erm*(52) and *mefH* [27], and a customized *gyrA*, *gyrB* and *rpoB* alleles database (for known resistance-conferring point mutations) using SRST2 [29, 35]. Strains that were positive for either *ermB* or *tetM* were interrogated for known transposons using SRST2 as previously described [29].

Evolutionary analysis of *C. difficile* RT 017

To investigate the evolution and spread of *C. difficile* RT 017, core genome SNP (cgSNP) and Bayesian evolutionary analyses were performed. All paired-end reads were trimmed using TrimGalore v0.6.4 to remove low-quality and adapter sequences (https://github.com/FelixKrueger/TrimGalore), mapped to the genome of *C. difficile* M68 and variants identified using Snippy v4.4.5 (https://github.com/tseemann/snippy). The resulting VCF file was then screened to exclude variants occurring in the repetitive region using SnpSift v4.3t [36] and to exclude indels using VCF-annotate v0.1.15 [37]. Gubbins v2.4.1 was used to identify and remove recombination sites [38]. SNP-dists v0.7.0 was used to generate a pairwise cgSNP table (https://github.com/tseemann/snp-dists). Following the approach of Eyre *et al.* [39] and Didelot *et al.* [40], a threshold of 0–2 cgSNPs was used to determine if groups of two or more strains were clonally related.

To facilitate the Bayesian analysis, clonal strains were removed from the dataset, leaving only one representative for each clonal cluster (n=282). Bayesian evolutionary analysis was performed using BactDating v1.0.1 [41]. In our previous study, BactDating allowed the comparison of a longer sequence alignment than the conventional BEAST software, which led to more precise confidence intervals [9]. It is also compatible with Gubbins, which was used in the previous step [41]. BactDating was run using a Gubbins recombination-adjusted phylogenetic tree from the previous analysis (1455 sites) as an input with the following settings: Markov chain Monte Carlo (MCMC) chains of 5×10^8 iterations sampled every 5×10^5 iterations with a 50% burn-in and a strict model with a rate of 1.4 mutations per genome per year as published by Didelot *et al.* [40]. These parameters were first tested on a smaller dataset (n=45, see below) and produced the best model. The molecular clock used in this study is best suited for the microevolutionary analysis and the investigation of transmission [40], which was the main focus of this study. However, this clock estimate does not take into account the quiescence of *C. difficile* spores and the analysis may have underestimated the time of spread (see Discussion). Traces were inspected to ensure convergence and the effective sample size (ESS) values for all estimated continuous variables were >200. The final Bayesian tree was annotated using iTOL v6 [42]. An interactive version of the Bayesian phylogenetic tree in Fig. 1 was uploaded to Microreact [43].

Bayesian analysis was also performed on a subset of 45 Thai *C. difficile* genomes, for which patient data and phenotypic AMR results were available [27, 44]. With this small dataset, several Bayesian evolutionary analyses were performed with different parameters, including the use of different input phylogenetic trees (Gubbins [41] versus PhyML [45]), strict versus relaxed clock, inclusion versus omission of collection dates, as well as different MCMC parameters. The cgSNP analysis was performed using the following reference genomes: MAR286 (this study; accession CP072118.1), M68 (accession FN668375.1), 630 (accession AM180355.1) and M120 (accession FN665653.1), to evaluate whether the choice of reference genome had any effect on downstream analysis. A pairwise whole-genome average nucleotide identity (ANI) between each *C. difficile* strain and the reference strains was generated using FastANI [46], and the results were used to compare strain relatedness with each reference.

Pangenome-wide association study

The cgSNP and Bayesian analyses identified two distinct *C. difficile* RT 017 sublineages. To determine significant genetic loci associated with each lineage, all *C. difficile* genomes were assembled and a pangenome-wide association study (pan-GWAS) was performed as previously described [9]. Briefly, Panaroo v1.1.0 was run with default settings on the annotated *C. difficile* genomes [47], and the results were used as an input for Scoary v1.6.16 to identify the significant genetic loci associated with each lineage [48].

Assessment of motility and cell aggregation

We also evaluated motility and cell aggregation in *C. difficile* RT 017 from the two lineages; *C. difficile* strain 1470 [ATCC 43598, non-epidemic lineage (NE)], MAR006 [epidemic lineage (E)], MAR024 (lineage E) and MAR 286 (lineage NE). First, a motility assay was performed as described by Tasteyre *et al.* in four separate batches [49]. In each batch, the growth diameter was recorded three times at different angles. Second, cell aggregation was assessed by measuring the optical density at 600 nm (OD_{600}) of the undisturbed and disturbed 48-h-old growth in brain heart infusion broth [50]. These tests were performed with at least three biological replicates. *C. difficile* strain IS58 (RT 033, non-motile) was included as a negative control [51].

Statistical analysis

All statistical analyses were performed using online tools by Social Science Statistics available at https://wwwsocscistatisticscom/. A *P*-value of ≤0.05 was considered to be statistically significant.

RESULTS

The epidemic *C. difficile* RT 017 lineage emerged from Asia in the middle of the 20th century

To study the global population structure of *C. difficile* RT 017, cgSNP and Bayesian evolutionary analyses were performed on 282 non-clonal *C. difficile* RT 017 genomes collected worldwide between 1981 and 2019 (Fig. S1, available in the online version of this article). The overall median year of isolation for this dataset was 2011 [quartile range (QR): 2008–2014]. The median years of isolation for the three main continents were as follow: Asia, 2014 (2010–2016), Europe, 2010 (2006–2012) and North America, 2009 (2004–2017). The Bayesian tree of *C. difficile* RT 017 could be divided into two parts based on the tree branching and topology (Fig. 1). The first part had deep temporal branches with a small number of strains, indicating an ancient lineage with limited spreading. This lineage was thus called the non-epidemic lineage (NE) and could be further divided into three sublineages (NE_1, NE_2 and NE_3). The second part stemmed from sublineage NE_3, which had shallow temporal branches with a large number of strains, indicating a rapid expansion of the lineage. This lineage was thus called the epidemic lineage (E). Table 1

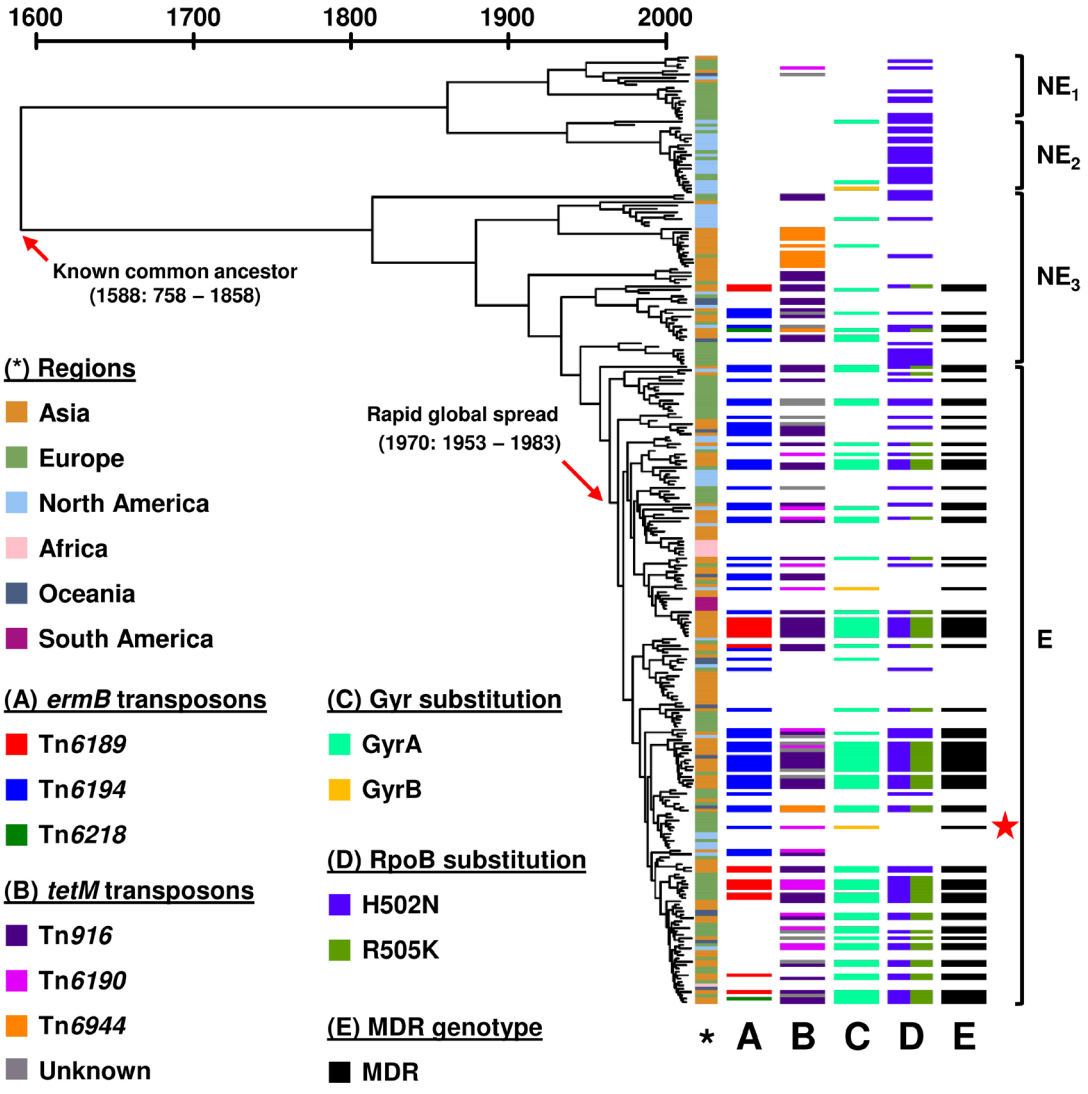

Fig. 1. Bayesian tree of 282 non-clonal *C. difficile* RT 017 genomes from around the world. The *C. difficile* RT 017 population could be divided into non-epidemic (NE; sublineages NE_1–NE_3) and epidemic (E) lineages. *The region of origin for each strain. Important genotypic AMR determinants are displayed on the right (A–E). The red star represents *C. difficile* M68, the reference genome in this analysis.

Table 1. List of lineage-defining cgSNPs

Position*	Strand†	Product	*N/S*‡	Lineages			
				NE_1	NE_2	NE_3	E
Lineage NE vs. lineage E							
867 703	F	Diguanylate kinase signalling protein	*N*	G§	G§	G§	T
Sublineages NE_1 and NE_2 vs. sublineage NE_3							
263 571	F	FlgG	*N*	T§	T§	C	C
480 088	R	UvrA	*S*	A§	A§	G	G
1 486 937	F	Gfo/Idh/MocA family oxidoreductase	*N*	T§	T§	G	G
1 789 300	F	Serine *O*-acetyltransferase	*S*	C§	C§	T	T
3 254 867	R	ABC transporter	*N*	T§	T§	C	C
3 808 791	N/A	Non-coding region	–	G§	G§	A	A
Sublineage NE_1 vs. sublineage NE_2							
1 299 679	F	Penicillin-binding protein 2	*N*	G	T§	G	G
1 486 584	F	Gfo/Idh/MocA family oxidoreductase	*N*	C	T§	C	C
2 928 003	R	ABC transporter	*N*	G	T§	G	G
3 066 957	R	Thioether cross-link-forming SCIFF peptide maturase	*N*	C	T§	C	C

*Position on *C. difficile* M68 genome.
†Coding strand (F, forward; R, reverse).
‡Non-synonymous substitutions (*N*) and synonymous substitutions (*S*).
§Different from the reference genome.

summarizes 11 lineage-defining SNPs identified. None of the mutations was on the pathogenicity locus (PaLoc), the genetic region containing the toxin genes *tcdA* and *tcdB*. All *C. difficile* RT 017 strains carried the same *tcdB* allele (*tcdB*_9 according to the PubMLST database) and were expected to produce variant type toxin B [1]. Sublineages NE_1, NE_2 and NE_3 consisted mainly of strains from Europe, North America and Asia, respectively, and the common ancestor of the three sublineages was estimated to have emerged in 1588 [95% confidence interval (CI): 758–1858]. Sublineages NE_1 and NE_2 split around 1860 (95% CI: 1622–1954). Lineage E was estimated to have split from sublineage NE_3 around 1958 (95% CI: 1920–1977) and later spread globally around 1970 (95% CI: 1953–1983).

The acquisition of *ermB* was probably the driving factor of the epidemic *C. difficile* RT 017 lineage

After incorporating genotypic AMR data, an association between the acquisition of AMR genotype and the spread of *C. difficile* RT 017 was evident. Genotypically MDR *C. difficile* RT 017 strains were in the lower part of sublineage NE_3 and lineage E, and only emerged around 1935 (95% CI: 1851–1969). There had been multiple acquisition events for the two most common accessory AMR determinants: *tetM* and *ermB*. The earliest acquisition of *tetM* was probably through gaining Tn*916,* which occurred around 1914 (95% CI: 1799–1964), while the earliest acquisition of *ermB* was probably through gaining Tn*6194*, which occurred around 1958 (95% CI: 1920–1977), notably the same timeframe as the predicted time of emergence of lineage E.

Non-synonymous substitutions in RpoB (H502N, conferring rifamycin resistance) and in GyrA (T82I, conferring fluoroquinolone resistance) were found scattered throughout the population. In contrast, an R505K substitution in RpoB was found only in strains from sublineage NE_3 and lineage E and was more common among Asian strains (37.2% vs. 8.9%, $P<0.0001$). The only European strains with an R505K substitution in RpoB were from an outbreak in Portugal [18]. Three independent GyrB substitution events were identified in this dataset: two D426N substitution events in North America around 2008 (95% CI: 1998–2011) and 2015 (95% CI: 2012–2016), and one D426V substitution event in Ireland (*C. difficile* M68, the reference strain) around 2004 (95% CI: 2001–2005) (star in Fig. 1). In addition to the important AMR determinants described above, the *aac6-aph2* gene was also common among *C. difficile* RT 017, found in 73 strains in this dataset (25.9%), and more commonly among Asian strains (43.4% vs. 14.2%, $P<0.0001$).

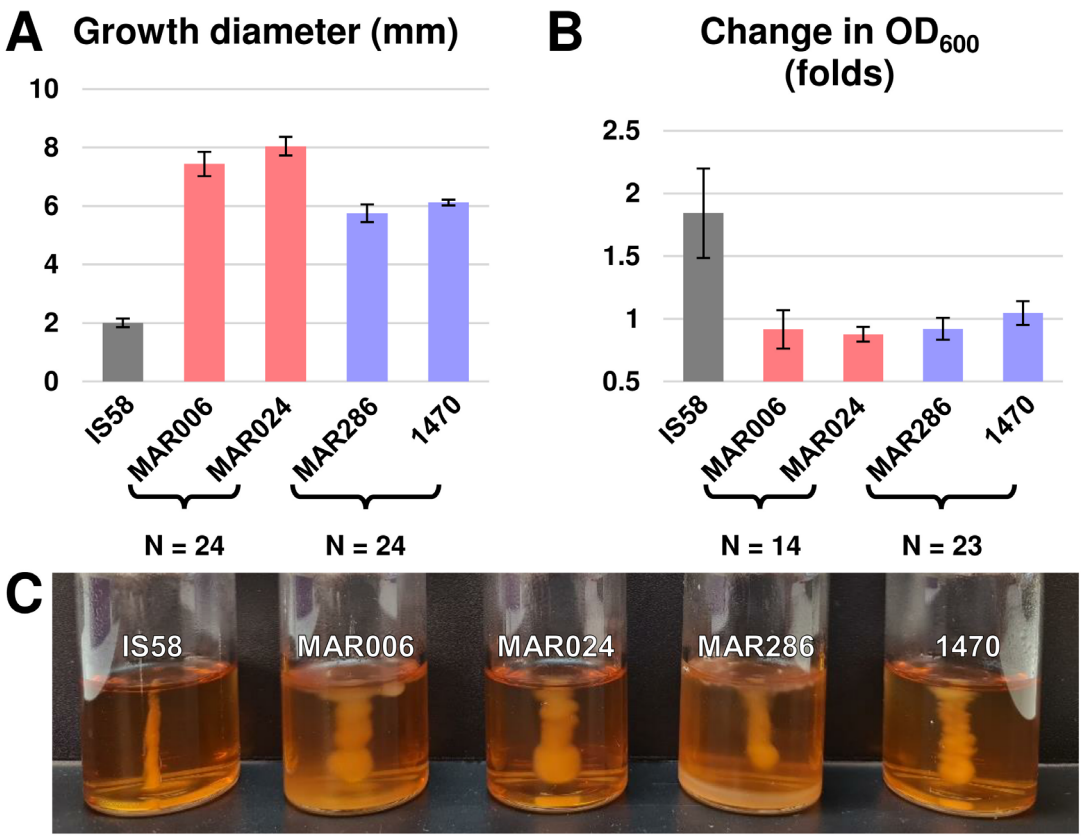

Fig. 2. Comparison of motility and cell aggregation between Lineages E (pink) and NE (lilac). (a) Lineage E had a larger growth diameter in semi-solid media. (b) Lineage E displayed a lower cell aggregation as measured by the difference in OD_{600} between undisturbed and disturbed broths. (c) The semisolid media for all tested strains. *C. difficile* IS58 (RT 033, dark grey) was used as a negative control. All error bars display 95% confidence intervals.

The epidemic *C. difficile* RT 017 lineage expresses higher motility

The cgSNP that differentiated between lineages NE and E resulted in a substitution in a diguanylate kinase signalling protein, which may play a role in motility and biofilm formation in *C. difficile* [50, 52]. Thus, motility and cell aggregation assays were performed (Fig. 2). Strains from lineage E had an increase in growth diameter compared to lineage NE (average diameter 7.7 vs. 5.9 mm, Wilcoxon rank sum $P<0.0001$, Fig. 2a) with a marginally significant change in the level of cell aggregation as shown by the lower change in OD_{600} between undisturbed and disturbed cultures (0.88 vs. 0.99, Wilcoxon rank sum $P=0.031$; for comparison, the non-motile *C. difficile* IS58 had 1.84-fold change in OD_{600}, Fig. 2b).

In addition to the lineage-specific cgSNPs (Table 1) and the difference in the prevalence of genotypic AMR, pan-GWAS was performed to identify other significant lineage-specific genetic loci. A total of 32863 genes was identified in the dataset, 3560 (10.8%) of which were found in more than 95% of strains and classified as core genes. Based on the GWAS, the locus most significantly associated with lineage E was the aminoglycoside resistance locus [containing *aac6-aph2* and a gene resembling *ant6*(Ib) (72% identity, E-value=5.01e-157); sensitivity 85.3%, specificity 97.8%]. Apart from AMR-related loci, lineage E was associated with a truncation of the formate dehydrogenase FdhF protein (sensitivity 75.3%, specificity 97.8%). A comparison of the FdhF protein is shown in Fig. S2 [53]. In an analysis of 260 representative *C. difficile* genomes across eight evolutionary clades [9], this truncated FdhF protein was not found in other *C. difficile* strains.

C. difficile RT 017 strains in Thailand were probably acquired outside of the hospital

In this study, a smaller subset of *C. difficile* RT 017 genomes from a single hospital in Thailand (n=45) was analysed to determine the best parameters to be used in the analyses above (Gubbins tree model, strict model with a rate of 1.4 mutations per genome per year, the inclusion of collection dates and the MCMC parameters described in the Methods). First, a local reference genome (MAR286; GenBank accession CP072118.1) was generated to evaluate the effect of the different reference genomes on the downstream analysis. Comparison of the *C. difficile* MAR286 genome with that of *C. difficile* M68, a commonly used reference genome of *C. difficile* RT 017, is shown in Table 2. Pairwise whole-genome ANI and cgSNP analyses were performed on Thai *C.*

Table 2. Comparison of two *C. difficile* RT 017 reference genomes

Parameter	M68	MAR286
Accession	FN668375.1	CP072118.1
Genome size (bp)	4 308 325	4 242 261
Genes	3983	3892
Coding sequences	3830	3761
rRNAs	40	35
tRNAs	109	92
Non-coding RNAs	4	4
CRISPR arrays	4	6
GC content	28.9%	28.8%
AMR loci	*ermB* (Tn*6194*) [MLS_B], *tetM* (Tn*6190*) [tetracyclines], D426V (GyrB) [fluoroquinolones]	*ermB* (Tn*6194*) [MLS_B], *tetM* (Tn*916*) [tetracyclines]
Pairwise ANI	99.92%	

difficile RT 017 genomes against different reference genomes and the results are summarized in Table 3. Thai *C. difficile* strains were closest to *C. difficile* M68. Using *C. difficile* M68 as a reference resulted in the longest average mapped length, significantly longer than *C. difficile* MAR286, the second closest reference genome ($P<0.0001$). Accordingly, *C. difficile* M68 was chosen as a reference for subsequent analysis. The average number of pairwise cgSNP differences based on *C. difficile* M68 and *C. difficile* MAR286 was 0.49 SNPs (95% CI: 0.44–0.54). The difference between *C. difficile* strains in this study and the other two reference genomes was more pronounced, resulting in a greater number of pairwise cgSNP differences compared to *C. difficile* M68: 5.42 SNPs (95% CI: 5.15–5.69) for *C. difficile* 630 and 9.39 SNPs (95% CI: 9.05–9.72) for *C. difficile* M120.

Using *C. difficile* M68 as a reference, 308 high-quality cgSNPs were identified across 45 *C. difficile* strains. The final Bayesian phylogenetic tree is shown in Fig. 3. Based on this phylogeny, 44 *C. difficile* RT 017 strains, excluding the outlier, could be classified roughly into three groups: the oldest group (G1, n=13), most of which were non-MDR *C. difficile* RT 017; a group of early MDR *C. difficile* RT 017 (G2, n=15); and the most recent and rapidly expanding clade of MDR *C. difficile* RT 017 (G3, n=16). The common ancestor of all Thai *C. difficile* RT 017 was estimated to have arisen around 1988 (95% CI: 1949–2000). The common ancestors of the three groups were estimated to have arisen around 1999 (1993–2004), 2003 (1995–2007) and 2012 (2009–2013), respectively.

Seven small clonal groups (CGs) were identified across the tree (CG1–CG7 in Fig. 3), three of which (CG2, CG5 and CG7) were from different patients who were in the hospital during the same period, suggesting possible direct patient–patient transmission (red boxes). Two CGs (CG1 and CG3), and two small CGs in CG5, included strains that were isolated from the same patients within 2 months, suggesting recurrence of CDI (blue boxes). The other two CGs (CG4 and CG6) included strains isolated from different patients without an obvious epidemiological link, one of which included strains from two specimens collected 3 years apart, suggesting contaminations in the hospital environment (red asterisks in Fig. 3). The remaining *C. difficile* strains were non-clonal.

Table 3. Effect of the choice of reference genome on cgSNP analysis

Reference	ST (clade)	Accession	Average mapped length (bp)	No. of SNPs	ANI (%)
MAR286	37 (4)	CP072118.1	4134703.82	311	99.88
M68	37 (4)	FN668375.1	4176850.73	308	99.93
630	54 (1)	AM180355.1	3836370.82	267	97.98
M120	11 (5)	FN665653.1	3579796.21	235	96.11

ANI, average nucleotide identity; SNPs, single nucleotide polymorphisms; ST, sequence type.

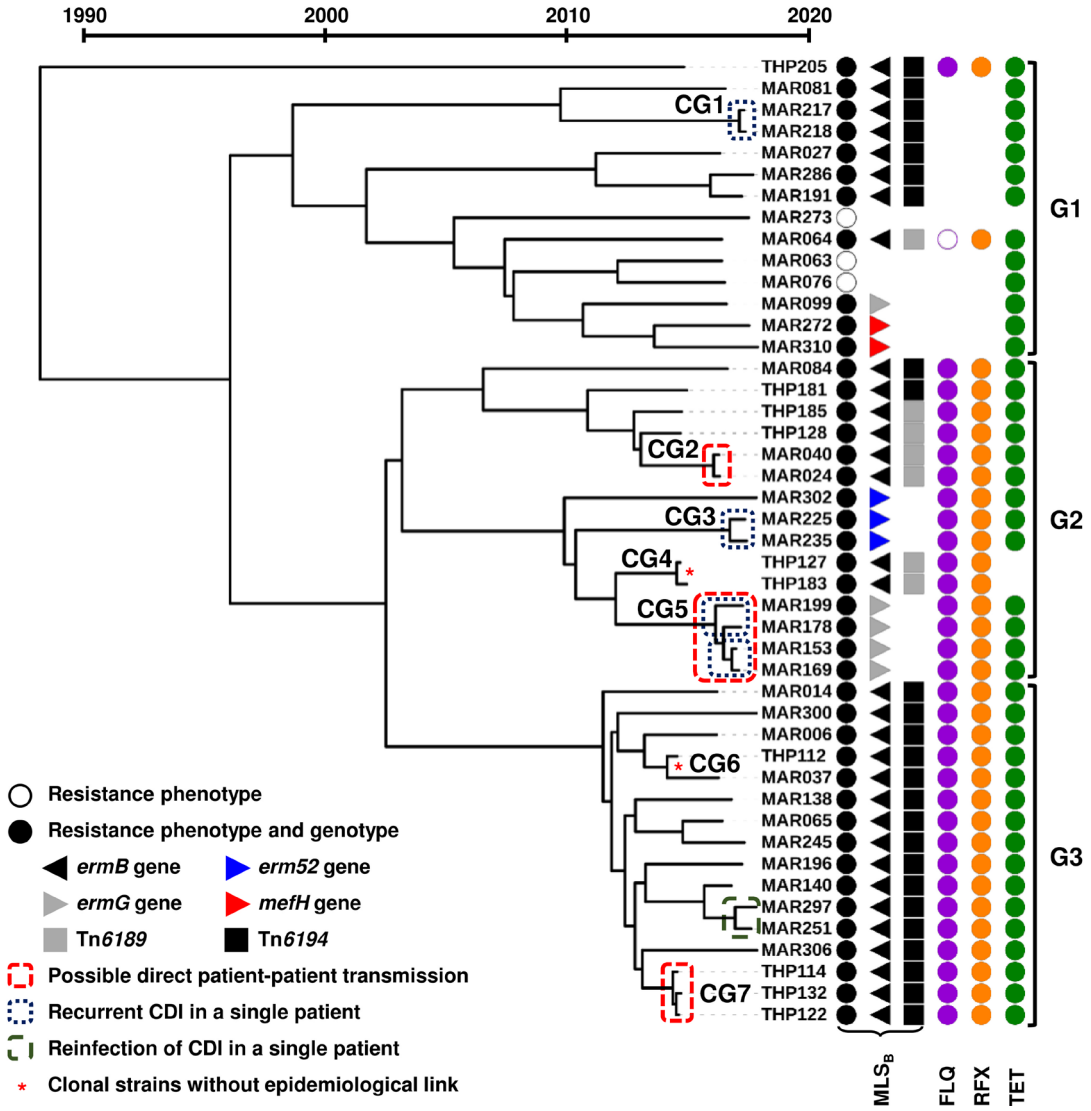

Fig. 3. Bayesian tree of 45 Thai *C. difficile* RT 017 strains. 'THP' refers to strains isolated in 2015 and 'MAR' to strains isolated in 2017–2018. Red boxes indicate that the patients were in the same department when the strains were isolated. Blue boxes indicate that the strains were isolated from the same patient within 2–8 weeks.

DISCUSSION

Despite being one of the most successful strains of *C. difficile*, very little is known about the evolution and spread of *C. difficile* RT 017. This study addresses this knowledge gap using high-resolution phylogenomic analyses on a comprehensive and diverse dataset of 282 global *C. difficile* RT 017 genomes. We found that the population of *C. difficile* RT 017 can be divided into two lineages, agreeing with the previous study by Cairns *et al.* [25]. However, the data disagree on the geographical origin of *C. difficile* RT 017. Our study suggests that *C. difficile* RT 017 may have originated in Asia, supporting the epidemiological studies [1], then spread to Europe and North America. This probably resulted from the inclusion of a few older European strains (isolated between 1981 and 1985) to reduce the gap in collection years between the two continents (P=0.6745 in this dataset) and a large diversity of Asian strains from 11 countries and administrative regions.

Based on the difference in structure, the two lineages of *C. difficile* RT 017 were classified as non-epidemic (NE, a small number of strains with little population expansion) and epidemic (E, a larger number of strains with rapid population expansion) lineages. Although not exclusively containing strains from one continent, the NE lineage could be divided into three sublineages predominantly containing strains from Asia, Europe and North America. This suggests that the spread of *C. difficile* RT 017 between these continents had occurred since the end of the 16th century. This roughly coincides with the estimated time of PaLoc acquisition ~500 years ago [54]. Sublineages NE_1 (Europe) and NE_2 (North America) were more closely related to one another than to sublineage NE_3 (Asia). In

turn, sublineage NE_3 was more closely related to sublineage NE_1 than sublineage NE_2, as demonstrated by fewer cgSNP differences (Table 1). Thus, the spread of *C. difficile* RT 017 probably began with population movement between Asia and Europe (1588, 95% CI: 758–1858) before spreading from Europe to North America (1860, 95% CI: 1622–1954). The direction of the spread between Asia and Europe cannot be determined from this analysis; however, based on the high prevalence and diversity of clade 4 strains in Asia [10–13, 24], it is likely that *C. difficile* RT 017, as well as other strains in clade 4, originated in Asia, travelled to Europe and subsequently crossed the Atlantic to North America.

Even though *C. difficile* RT 017 could be found in at least three continents by the end of the 19th century, the Bayesian analysis suggests that the epidemic lineage E emerged solely from Asia (sublineage NE_3) following the acquisition of *ermB*-positive Tn*6194* in 1958 (95% CI: 1920–1977), before spreading globally in 1970 (95% CI: 1953–1983). The time of acquisition of the *ermB* element coincides with the introduction of clindamycin into clinical practice in the 1960s [55]. This pattern of spread is similar to *C. difficile* RT 027, another epidemic strain that spread in and from North America in the early 2000s [56] driven by the acquisition of fluoroquinolone resistance in 1993/94 [56], following the widespread use of levofloxacin for community-acquired pneumonia [57]. This provides supporting evidence that the use of antimicrobials and the acquisition of AMR determinants are significant in the spread of *C. difficile*. Although the prevalence of fluoroquinolone and rifamycin resistance was also high in *C. difficile*, the widespread resistance across all lineages suggests the independent acquisition of resistance after the spread of the strain.

The analyses were first performed on a small dataset of Thai clinical *C. difficile* RT 017 isolates (n=45) with complete metadata to evaluate the performance of the pipeline. These analyses accurately identified four pairs of *C. difficile* strains isolated from the same patients, provided good correlations between AMR phenotypes and genotypes [27], as well as AMR genotypes and cgSNP population structure. When performed on the global dataset (n=282), the analyses accurately predicted the emergence of *C. difficile* M68 (2001–2005), a strain from a 2003 outbreak in Ireland [31]. Also, appropriate timelines for the emergence of Argentinian (1996–2000) and Portuguese (2003–2011) clusters [18, 20] were estimated, supporting the accuracy of the analyses.

Besides the aforementioned AMR genes, the epidemic lineage E was also associated with the presence of an aminoglycoside resistance locus and a truncated FdhF protein. Being a strictly anaerobic bacterium, *C. difficile* is intrinsically resistant to aminoglycosides, and the presence of an additional aminoglycoside-resistance locus is unlikely to have provided any advantage to the bacterium [58]. However, it may suggest that the epidemic strains were from an area with a high prevalence of aminoglycoside-resistant enteric bacteria, especially enterococci [59]. Formate dehydrogenase is an enzyme involved in the reoxidation of NAD [60]. Based on the prediction by the UniProt database [61], the truncated region is the coiled-coil domain that probably serves as a binding site for NAD. Thus the truncated protein is probably non-functional, although *C. difficile* has several pathways for oxidizing NAD and the truncated FdhF may not ultimately have any effect on growth or virulence [60]. Another significant genetic variant associated with lineage E was a point substitution (W366L) on the diguanylate kinase signalling protein (Table 1). This protein is involved in the regulation of cyclic dimeric guanosine monophosphate (c-di-GMP), which plays a role in motility and biofilm formation [50, 52]. In our preliminary assessment, strains from lineage E had increased motility *in vitro*. This provides a foundation for further *in vivo* studies to determine the effect of these phenotypes on the virulence and transmissibility of the epidemic strains.

Analyses of the Thai clinical *C. difficile* strains provided information on disease transmission in the country that differs from a previous report from the UK [16]. The UK study reported a cluster of closely related *C. difficile* RT 017 strains in a single hospital in London that was different to strains from other parts of the city, suggesting an intra-hospital outbreak [16]. In the current study, all Thai strains were isolated in a single tertiary hospital over 4 years (2015–2018), but most of them were not closely related. Overall, these strains were more closely related to *C. difficile* M68, a strain isolated in Ireland in a different decade [31], than to a non-epidemic strain from the same hospital. This suggests that the high prevalence of *C. difficile* RT 017 in the hospital was not due to an ongoing outbreak. Indeed, evidence of direct patient–patient transmission could be identified in only a few cases. The remaining cases acquired *C. difficile* RT 017 elsewhere, probably from the community [62, 63].

Although evidence of *C. difficile* RT 017, or other clade four strains, in the environment in Asia has yet to be provided [24, 64], it may be inferred by the persistence of *C. difficile* RT 017 in the human population [10–14]. This mimics the situation in North America, where successful public health interventions have led to a significant decrease in the burden of the epidemic *C. difficile* RT 027, although complete eradication has not been achieved [65]. By contrast, *C. difficile* RT 027 has almost completely disappeared in Europe [66]. The persistence of *C. difficile* RT 027 in North America was linked strongly to continuous spillover from several environmental sources, including household environments and companion animals [67, 68]. Similar studies in Asia are needed to verify the presence of *C. difficile* RT 017 in the environment, which would further suggest that *C. difficile* RT 017 has long been integrated into the Asian community.

This study also demonstrates the effect of reference genome selection on downstream analysis (Table 3). The results were comparable when a reference from the same ST was used (an average difference of 0.49 SNPs, clonality cut-off point of 2 SNPs) [40]. Differences became more pronounced as the reference strain became less related, suggesting that a reference genome from the same ST should be used to ensure accurate cgSNP results. With the introduction of ONT, it is now possible to assemble a complete genome of a local reference strain to maximize the accuracy of cgSNP analysis using a combination of short- and long-read sequences.

A limitation of this study remained the relatively low number of early *C. difficile* RT 017 strains in general and the lack of older strains from Asia. This probably led to some uncertainty in the estimations, as reflected by wide 95% CIs, especially around the root of the Bayesian tree. Also, the biological clock of 1.4 mutations per genome per year used in this study did not account for the presence of *C. difficile* spores, the genomes of which may remain unchanged for decades or centuries [40]. This may affect the estimated time at the root of the tree, which could be earlier than the current estimate. The inclusion of more early strains will help adjust the model leading to a more accurate estimate. Although it may be difficult to acquire old clinical strains, it may be possible to get historical strains from other sources. Soil is one promising source for ancient *C. difficile*, as it is a reservoir for *C. difficile* spores and several methods have been developed to measure the age of the soil [69], which can be used as a substitution for the collection date in a Bayesian evolutionary analysis.

In conclusion, *C. difficile* RT 017 had been circulating between Asia and Europe for centuries before spreading to North America. The epidemic lineage of *C. difficile* RT 017 emerged from Asia in the middle of the 20th century following the acquisition of *ermB*. A focused investigation of contemporary *C. difficile* RT 017 in Thailand revealed that the population was highly diverse and community reservoirs/sources may have played an important role in the transmission of disease in this country.

Additional information

The Supplementary Data Sheet 1 is available at DOI: 10.6084/m9.figshare.14544792.

Funding information
This work was supported, in part, by funding from The Raine Medical Research Foundation (RPG002-19) and a Fellowship from the National Health and Medical Research Council (APP1138257) awarded to D.R.K. and Edith Cowan University School of Medical and Health Sciences Research Grant awarded to D.A.C. K.I. is a recipient of the Mahidol Scholarship from Mahidol University, Thailand. This research used the facilities and services of the Pawsey Supercomputing Centre (Perth, Western Australia).

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
This study involved the use of de-identified patient data. It was approved by the Human Research Ethics Committee of The University of Western Australia (reference file RA/4/20/4704) and the Siriraj Institutional Review Board (protocol number 061/2558 [EC1]).

References

1. Imwattana K, Knight DR, Kullin B, Collins DA, Putsathit P, *et al*. *Clostridium difficile* ribotype 017 - characterization, evolution and epidemiology of the dominant strain in Asia. *Emerg Microbes Infect* 2019;8:796–807.
2. Kim J, Kim Y, Pai H. Clinical characteristics and treatment outcomes of *Clostridium difficile* infections by PCR Ribotype 017 and 018 Strains. *PLoS One* 2016;11:e0168849.
3. Kim J, Pai H, Seo MR, Kang JO. Clinical and microbiologic characteristics of tcdA-negative variant *Clostridium difficile* infections. *BMC Infect Dis* 2012;12:109.
4. Goorhuis A, Debast SB, Dutilh JC, van Kinschot CM, Harmanus C, *et al*. Type-specific risk factors and outcome in an outbreak with 2 different *Clostridium difficile* types simultaneously in 1 hospital. *Clin Infect Dis* 2011;53:860–869.
5. Imwattana K, Knight DR, Kullin B, Collins DA, Putsathit P, *et al*. Antimicrobial resistance in *Clostridium difficile* ribotype 017. *Expert Rev Anti Infect Ther* 2020;18:17–25.
6. Lyerly DM, Saum KE, MacDonald DK, Wilkins TD. Effects of *Clostridium difficile* toxins given intragastrically to animals. *Infect Immun* 1985;47:349–352.
7. Lyerly DM, Sullivan NM, Wilkins TD. Enzyme-linked immunosorbent assay for *Clostridium difficile* toxin A. *J Clin Microbiol* 1983;17:72–78.
8. al-Barrak A, Embil J, Dyck B, Olekson K, Nicoll D, *et al*. An outbreak of toxin A negative, toxin B positive *Clostridium difficile*-associated diarrhea in a Canadian tertiary-care hospital. *Can Commun Dis Rep* 1999;25:65–69.
9. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, *et al*. Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. *elife* 2021;10:e64325.
10. Putsathit P, Maneerattanaporn M, Piewngam P, Kiratisin P, Riley TV. Prevalence and molecular epidemiology of *Clostridium difficile* infection in Thailand. *New Microbes New Infect* 2017;15:27–32.
11. Imwattana K, Wangroongsarb P, Riley TV. High prevalence and diversity of tcdA-negative and tcdB-positive, and non-toxigenic, *Clostridium difficile* in Thailand. *Anaerobe* 2019;57:4–10.
12. Riley TV, Collins DA, Karunakaran R, Kahar MA, Adnan A, *et al*. High prevalence of toxigenic and nontoxigenic *Clostridium difficile* strains in Malaysia. *J Clin Microbiol* 2018;56:e00170-18.
13. Collins DA, Gasem MH, Habibie TH, Arinton IG, Hendriyanto P, *et al*. Prevalence and molecular epidemiology of *Clostridium difficile* infection in Indonesia. *New Microbes New Infect* 2017;18:34–37.
14. Rupnik M, Kato N, Grabnar M, Kato H. New types of toxin A-negative, toxin B-positive strains among *Clostridium difficile* isolates from Asia. *J Clin Microbiol* 2003;41:1118–1125.
15. Arvand M, Hauri AM, Zaiss NH, Witte W, Bettge-Weller G. Clostridium difficile ribotypes 001, 017, and 027 are associated with lethal *C. difficile* infection in Hesse, Germany. *Euro Surveill* 2009;14:45.
16. Cairns MD, Preston MD, Lawley TD, Clark TG, Stabler RA, *et al*. Genomic epidemiology of a protracted hospital outbreak caused by a toxin a-negative *Clostridium difficile* sublineage PCR ribotype 017 strain in London, England. *J Clin Microbiol* 2015;53:3141–3147.
17. Åkerlund T, Alefjord I, Dohnhammar U, Struwe J, Norén T, *et al*. Geographical clustering of cases of infection with moxifloxacin-resistant *Clostridium difficile* PCR-ribotypes 012, 017 and 046 in Sweden, 2008 and 2009. *Euro Surveill* 2011;16:10.
18. Isidro J, Santos A, Nunes A, Borges V, Silva C, *et al*. Imipenem resistance in *Clostridium difficile* ribotype 017, Portugal. *Emerg Infect Dis* 2018;24:741–745.
19. Drudy D, Harnedy N, Fanning S, O'Mahony R, Kyne L. Isolation and characterisation of toxin A-negative, toxin B-positive *Clostridium difficile* in Dublin, Ireland. *Clin Microbiol Infect* 2007;13:298–304.

20. Goorhuis A, Legaria MC, van den Berg RJ, Harmanus C, Klaassen CHW, *et al*. Application of multiple-locus variable-number tandem-repeat analysis to determine clonal spread of toxin A-negative *Clostridium difficile* in a general hospital in Buenos Aires, Argentina. *Clin Microbiol Infect* 2009;15:1080–1086.
21. 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.
22. Qin J, Dai Y, Ma X, Wang Y, Gao Q, *et al*. Nosocomial transmission of *Clostridium difficile* genotype ST81 in a general teaching hospital in China traced by whole genome sequencing. *Sci Rep* 2017;7:9627.
23. Senoh M, Kato H, Fukuda T, Niikawa A, Hori Y, *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;64:1226–1236.
24. 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.
25. Cairns MD, Preston MD, Hall CL, Gerding DN, Hawkey PM, *et al*. Comparative genome analysis and global phylogeny of the toxin variant *Clostridium difficile* PCR ribotype 017 reveals the evolution of two independent sublineages. *J Clin Microbiol* 2017;55:865–876.
26. Xu X, Luo Y, Chen H, Song X, Bian Q, *et al*. Genomic evolution and virulence association of *Clostridioides difficile* sequence type 37 (ribotype 017) in China. *Emerg Microbes Infect* 2021;10:1331–1345.
27. 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;27:1505–1512.
28. Vernon JJ. *Multidrug resistant Clostridioides difficile: the Presence of Antimicrobial Resistance Determinants in Historical and Contemporaneous Isolates, and the Impact of Fluoroquinolone Resistance Development on PCR Ribotype 027 Fitness*. University of Leeds, 2019.
29. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, *et al*. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;6:11.
30. 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.
31. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, *et al*. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci U S A* 2010;107:7527–7532.
32. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
33. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 2015;31:3350–3352.
34. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
35. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, *et al*. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014;58:212–220.
36. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 2014;6:80–92.
37. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al*. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–2158.
38. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
39. Eyre DW, Walker AS. *Clostridium difficile* surveillance: harnessing new technologies to control transmission. *Expert Rev Anti Infect Ther* 2013;11:1193–1205.
40. Didelot X, Eyre DW, Cule M, Ip CL, Ansari M, *et al*. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol* 2012;13:R118.
41. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* 2018;46:22.
42. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
43. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, *et al*. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016;2:e000093.
44. Imwattana K, Putsathit P, Leepattarakit T, Kiratisin P, Riley TV. Mild or malign: clinical characteristics and outcomes of *Clostridium difficile* infection in Thailand. *J Clin Microbiol* 2020;58:e01217-20.
45. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704.
46. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
47. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al*. Producing polished prokaryotic pangenomes with the panaroo pipeline. *Genome Biol* 2020;21:180.
48. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Erratum to: Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 2016;17:262.
49. Tasteyre A, Barc M-C, Karjalainen T, Dodson P, Hyde S, *et al*. A *Clostridium difficile* gene encoding flagellin. *Microbiology* 2000;146:957–966.
50. Purcell EB, McKee RW, McBride SM, Waters CM, Tamayo R. Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *J Bacteriol* 2012;194:3307–3316.
51. Geric B, Carman RJ, Rupnik M, Genheimer CW, Sambol SP, *et al*. Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters. *J Infect Dis* 2006;193:1143–1150.
52. Bordeleau E, Fortier LC, Malouin F, Burrus V. c-di-GMP turnover in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. *PLoS Genet* 2011;7:e1002039.
53. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
54. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, *et al*. Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol Evol* 2014;6:36–52.
55. Meyers BR, Kaplan K, Weinstein L. Microbiological and pharmacological behavior of 7-chlorolincomycin. *Appl Microbiol* 1969;17:653–657.
56. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, *et al*. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet* 2013;45:109–113.
57. Lynch JP 3rd, File TM Jr, Zhanel GG. Levofloxacin for the treatment of community-acquired pneumonia. *Expert Rev Anti Infect Ther* 2006;4:725–742.
58. Khanafer N, Daneman N, Greene T, Simor A, Vanhems P, *et al*. Susceptibilities of clinical *Clostridium difficile* isolates to antimicrobials: a systematic review and meta-analysis of studies since 1970. *Clin Microbiol Infect* 2018;24:110–117.
59. Daigle DM, Hughes DW, Wright GD. Prodigious substrate specificity of AAC(6')-APH(2"), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci. *Chem Biol* 1999;6:99–110.
60. Neumann-Schaal M, Jahn D, Schmidt-Hohagen K. Metabolism the difficile way: the key to the success of the pathogen *Clostridioides difficile*. *Front Microbiol* 2019;10:219.

61. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, *et al*. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2021;49:D480–D489.
62. Rodriguez-Palacios A, Stämpfli HR, Duffield T, Peregrine AS, Trotz-Williams LA, *et al*. *Clostridium difficile* PCR ribotypes in calves, Canada. *Emerg Infect Dis* 2006;12:1730–1736.
63. Lim SC, Knight DR, Riley TV. *Clostridium difficile* and one health. *Clin Microbiol Infect* 2020;26:857–863.
64. Putsathit P, Neela VK, Joseph NMS, Ooi PT, Ngamwongsatit B, *et al*. Molecular epidemiology of *Clostridium difficile* isolated from piglets. *Vet Microbiol* 2019;237:108408.
65. Guh AY, Mu Y, Winston LG, Johnston H, Olson D, *et al*. Trends in U.S. burden of *Clostridioides difficile* infection and outcomes. *N Engl J Med* 2020;382:1320–1330.
66. Nagy E. What do we know about the diagnostics, treatment and epidemiology of *Clostridioides (Clostridium) difficile* infection in Europe? *J Infect Chemother* 2018;24:164–170.
67. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB, *et al*. More than 50% of *Clostridium difficile Isolates* from pet dogs in Flagstaff, USA, carry toxigenic genotypes. *PLoS One* 2016;11:e0164504.
68. Weese JS, Finley R, Reid-Smith RR, Janecko N, Rousseau J. Evaluation of *Clostridium difficile* in dogs and the household environment. *Epidemiol Infect* 2010;138:1100–1104.
69. Bradley RS. Chapter 3 - dating methods I. In: Bradley RS (eds). *Paleoclimatology*, 3rd ed. Academic Press; 2015. pp. 55–101.