

RESEARCH

Open Access



A molecular epidemiological and transmission analysis of *Clostridioides difficile* using draft whole-genome sequencing in a single hospital

Taito Miyazaki^{1,2}, Kotaro Aoki^{3*}, Tadashi Maeda², Kohji Komori⁴, Sadako Yoshizawa^{3,5}, Yoshikazu Ishii^{3,6}, Yoshihisa Urita² and Kazuhiro Tateda³

Abstract

Background The nosocomial transmission of toxin-producing *Clostridioides difficile* is a significant concern in infection control. *C. difficile*, which resides in human intestines, poses a risk of transmission, especially when patients are in close contact with medical staff.

Methods To investigate the nosocomial transmission of *C. difficile* in a single center, we analyzed the genetic relationships of the bacteria. This was done using draft whole-genome sequencing (WGS) and examining single nucleotide polymorphisms (SNPs) in core-genome, alongside data regarding the patient's hospital wards and room changes. Our retrospective analysis covered 38 strains, each isolated from a different patient, between April 2014 and January 2015.

Results We identified 38 strains that were divided into 11 sequence types (STs). ST81 was the most prevalent ($n = 11$), followed by ST183 ($n = 10$) and ST17 ($n = 7$). A cluster of strains that indicated suspected nosocomial transmission (SNT) was identified through SNP analysis. The draft WGS identified five clusters, with 16 of 38 strains belonging to these clusters. There were two clusters for ST81 (ST81-SNT-1 and ST81-SNT-2), two for ST183 (ST183-SNT-1 and ST183-SNT-2), and one for ST17 (ST17-SNT-1). ST183-SNT-1 was the largest SNT cluster, encompassing five patients who were associated with Wards A, B, and K. The most frequent room changer was a patient labeled Pt08, who changed rooms seven times in Ward B. Patients Pt36 and Pt10, who were also in Ward B, had multiple admissions and discharges during the study period.

Conclusions Additional culture tests and SNP analysis of *C. difficile* using draft WGS revealed silent transmission within the wards, particularly in cases involving frequent room changes and repeated admissions and discharges. Monitoring *C. difficile* transmission using WGS-based analysis could serve as a valuable marker in infection control management.

Keywords *Clostridioides difficile*, Whole-genome sequencing, SNP analysis, Silent transmission

*Correspondence:

Kotaro Aoki
kotaro.aoki@med.toho-u.ac.jp

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Clostridioides difficile is a Gram-positive, spore-forming, anaerobic bacterium and a pathogen responsible for *C. difficile* infection (CDI) [1]. The occurrence of CDI, caused by toxin A or B-producing *C. difficile*, is concerning not only due to adverse patient outcomes but also due to the increased costs associated with treatment and infection control [2]. In particular, the spore-forming nature of *C. difficile* necessitates different decontamination methods compared to those used for general bacteria and requires more labor-intensive care for CDI patients [2–4]. Tracking the transmission of *C. difficile* between patients and interrupting transmission pathways is crucial for infection prevention in hospitals [5].

C. difficile typing is typically performed using PCR ribotyping (RT) [6, 7]. Strains that belong to RT027 are known for causing CDI with severe symptoms or high mortality rates [8, 9]. With the widespread availability of whole-genome sequencing (WGS) facilitated by massive parallel sequencers, the typing of *C. difficile* has shifted from RT to methods such as multilocus sequencing typing (MLST), core-genome MLST, and single nucleotide polymorphism (SNP)-based typing [10–13]. These sequence-based typing methods have proven to be robust for analysis in fields beyond nosocomial transmission, including food poisoning and One-Health [14, 15]. Notably, SNPs-based typing is considered the best practice for distinguishing bacterial strains.

The enzyme immunoassays (EIA) and nucleic acid amplification testing (NAAT) for detecting *C. difficile* and its Toxin A/B antigens in stool samples are commonly used for diagnosing CDI [1, 16]. Routine *C. difficile* culture tests are often omitted because EIA and NAAT exhibited good concordance with culture test results, take less time, and the culture method requires an additional toxin or toxin-encoding gene detection test, which takes 48 h or more [17]. Consequently, *C. difficile* strain collections with clinical information are valuable for nosocomial transmission analysis using WGS [18].

In this retrospective study, we aimed to detect silent nosocomial transmissions of *C. difficile* during the study period by integrating WGS and SNPs-based analyses with epidemiological data from the hospital wards.

Materials and methods

Patients and strains

This study was conducted as part of the Asian Pacific *C. difficile* Surveillance Study [6, 19]. Thirty-eight patients were included in the study in Fig. 1. Briefly, 1,046 stool samples collected for various medical purposes, including CDI diagnosis (note: there were instances of patient duplication), were submitted to the microbiological laboratory at Toho University Omori Medical Center from April 1st, 2014, to January 20th, 2015. *C. difficile* was

isolated from 199 of these stools using ChromID™ *C. difficile* agar (bioMérieux, France) and incubating at 35 °C for 48 h. The presence of *C. difficile*, including both glutamate dehydrogenase and toxin A or B, was confirmed in 112 strains using an enzyme immune assay with *C. DIFF QUIK COMPLETE*® (Kohjin Bio Co., Ltd., Saitama, Japan) and bacterial cell suspensions. Informed consent was obtained from 38 cases diagnosed with CDI, and first isolate strains from each case during the study period were used for this study (Table S1).

Draft whole-genome sequencing analysis

DNA was extracted using a combination of achromopeptidase treatment and phenol/chloroform, followed by purification with the Wizard® SV Gel and PCR Clean-Up Kit System. DNA libraries were prepared for sequencing on the MiSeq platform (Illumina Inc., CA, USA) using the Nextera XT DNA Library Preparation Kit (Illumina) and the Illumina DNA Prep (M) Tagmentation Library Preparation Kit (Illumina). These libraries were sequenced using the MiSeq Reagent Kit V3-600 cycles, allowing 300 bp paired-end reads (Illumina Inc.). Draft genome contigs were generated by *de novo* assembly using SPAdes version 3.15.5 [20]. Genome annotation and species identification were performed using Fast Average Nucleotide Identity with type strain genomes facilitated by the DNA Data Bank of Japan's Fast Annotation and Submission Tool [21]. Gene identification and alignment analysis for the following genes: *tcdA* encoding toxin A (TcdA); *tcdB* encoding toxin B (TcdB); *cdtA* encoding binary toxin A (CdtA); *cdtB* encoding binary toxin B (CdtB); *tcdC* encoding the negative regulator of the *tcdA* and *tcdB*, were performed using Nucleotide BLAST [22] and Jalview version 2 [23]. *C. difficile* strain 630 (*tcdA+tcdB+cdtA/cdtB-*, accession no. NC_009089) and strain CD196 (*tcdA+tcdB+cdtA/cdtB+*, accession no. NC_013315) were used as reference genome of toxin gene sequences. MLST was performed using *C. difficile* MLST databases in PubMLST.org (<https://pubmlst.org/cdifficile/>). The draft WGS data has been deposited in the DDBJ/ENA/GenBank under BioProject accession number PRJNA1036794, with individual sample accession numbers listed in Table S1.

Core-genome SNPs-based phylogenetic analysis

Phylogenetic analysis based on core-genome SNPs was performed according to our previous report [11]. Briefly, sequencing reads were aligned to the reference genomes of genetically closest strains for each ST using the Burrows-Wheeler Aligner with the 'SW' algorithm [24]. The core-genome sequence were extracted using the Sequence Alignment/Map (SAMtools) software, version 1.1 [25], with the "mpileup" option, and VarScan version 2.3.7, using the "mpileup2cns" option [26]. The

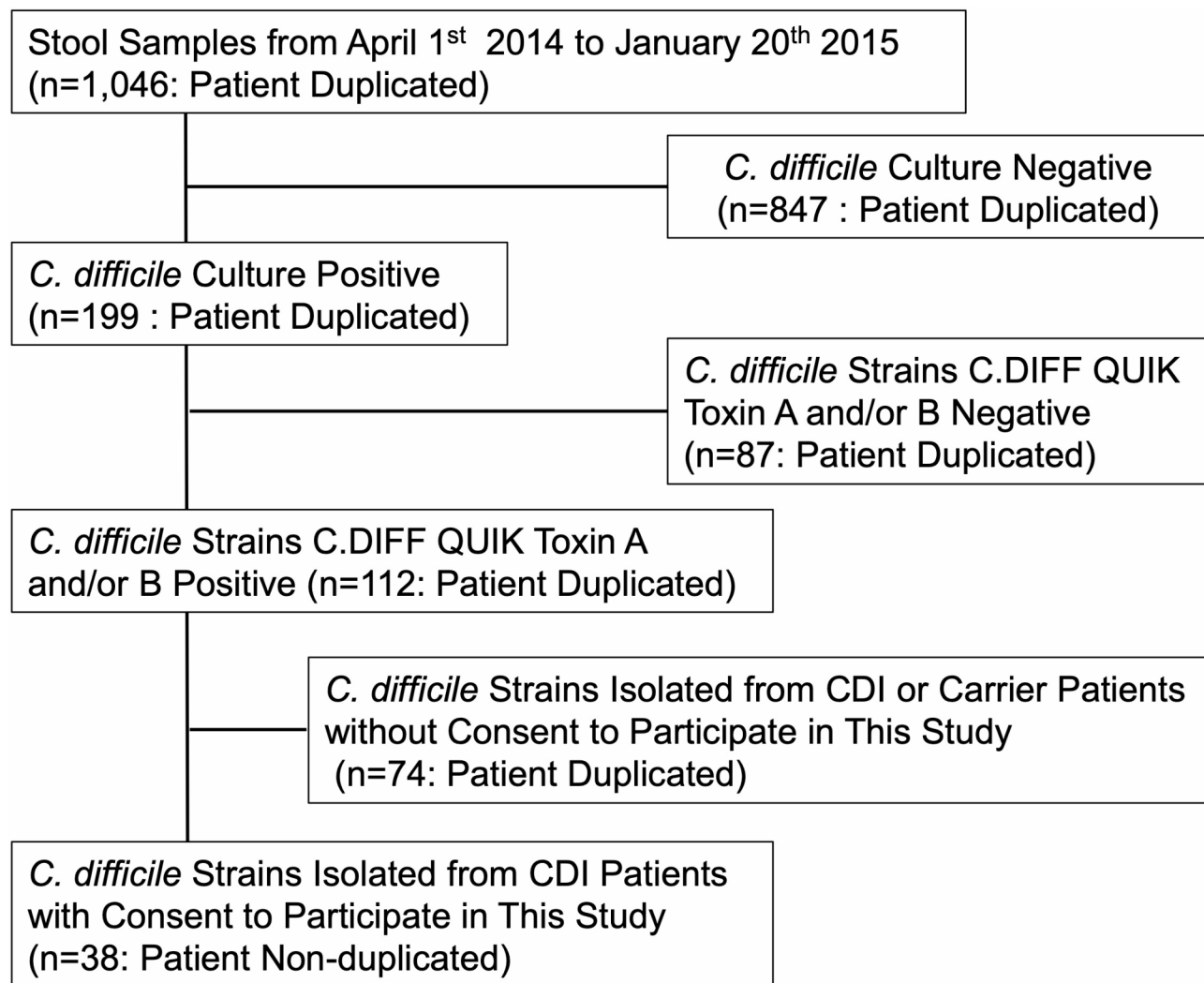


Fig. 1 Sampling workflow of *C. difficile* strains in this study

Table 1 Sequence type of *C. Difficile* and the hospital wards where the patients were admitted

| MLST | Hospital ward | | | | | | | | | | | Total |
|-------|---------------|---|---|---|---|---|---|---|---|---|---|-------|
| | A | B | C | D | E | F | G | H | I | J | K | |
| ST81 | 1 | 2 | 1 | 1 | 2 | | 1 | | 1 | 1 | 1 | 11 |
| ST183 | 1 | 8 | | | | | | | | | | 10 |
| ST17 | | | | 5 | | | | | | 1 | 1 | 7 |
| ST5 | | | 1 | | | | 1 | | | | | 2 |
| ST8 | | | | | | 1 | | 1 | | | | 2 |
| ST2 | | | | 1 | | | | | | | | 1 |
| ST3 | | | | | | | | 1 | | | | 1 |
| ST37 | | | 1 | | | | | | | | | 1 |
| ST42 | | | | | | | | | | 1 | | 1 |
| ST100 | | | | | | | | 1 | | | | 1 |
| ST185 | 1 | | | | | | | | | | | 1 |

homologous recombination regions were estimated using ClonalFrameML and were excluded from the core-genome. A phylogenetic tree was generated using RAxML, based on SNPs within the core-genome, excluding the homologous recombination regions.

Results

Draft whole-genome sequencing and molecular characterizing of *C. difficile*

Results and characteristics of the draft WGS are presented in Table S1. MLST classified 38 strains into 11 STs. The most prevalent were strains belonging to ST81 ($n=11$), followed by ST183 ($n=10$) and ST17 ($n=7$), as shown in Table 1. Strains from other STs were detected in fewer than two instances. ST81 strains were isolated from patients across nine wards, whereas ST183 and ST17 were predominantly found in 3 wards (Table 1). Specifically, of the 10 ST183 strains, eight were isolated from patients in Ward B, which is the Department of

Hematology and Oncology. Similarly, of 7 ST17 strains, five were from Ward D, the Department of Respiratory Medicine. Toxin type varied depending on the ST, except for ST5. Strains of ST183 and ST17 were characterized as A+B+CDT⁻, while ST81 was A-B+CDT⁻, as detailed in Table S1.

Core-genome SNPs-based phylogenetic analysis

The rate of core-genome size relative to reference genome size was 65.9% for ST81, 49.1% for ST183, and 29.9% for ST17. We established a distinct cutoff for Suspected Nosocomial Transmission (SNT) for each ST due to their varying core-genome size: ≤ 3 SNPs for ST81, ≤ 2 SNPs for ST183, and SNPs ≤ 1 for ST17. Based on these criteria, two SNT groups were identified in ST81 (ST81-SNT-1 comprising CD013 and CD159; ST81-SNT-2 comprising CD054, CD082, and CD304), two in ST183 (ST183-SNT-1 comprising CD014, CD015, CD055, CD086, and CD303; ST183-SNT-2 comprising CD164, CD296, and CD302), one in ST17 (ST17-SNT-1 comprising CD085, CD155, and CD160), as shown in Fig. 2. In ST183, the number of SNPs detected across all strains was fewer than in ST81 and ST17 (< 5 SNPs). The retrospective draft WGS identified five clusters, and 16 of 38 strains (42.1%) were genetically related to any of them.

Moving patients between wards in the hospital

ST17 was isolated consistently over a three-month period, whereas ST183 was isolated sporadically over a ten-month period. Our focus was on ST183, for which we visualized the hospitalization periods and room assignments of patients in Ward B (Fig. 3). Patients associated with ST183-SNT-1, which comprised five individuals, had stayed in 6 rooms within Ward B, one room in Ward A, and another in Ward K. Those associated with ST183-SNT-2, consisting of 3 patients, had stayed in 4 different rooms within Ward B. Notably, patients linked to both ST183-SNT-1 and ST183-SNT-2 frequently moved rooms within the ward. Pt08, who belonged to ST183-SNT-1 and moved room 7 times, was the most frequent mover. Pt36, also a part of ST183-SNT-1, was hospitalized and discharged five times. Pt10 moved rooms five times, was then discharged, re-hospitalized, and subsequently moved rooms an additional two times. Pt08, Pt10, and Pt36 had multiple overlaps in room occupancy during different periods. Patients of ST183-SNT-2 were frequently changing rooms, though they did not often share rooms.

Discussions

We identified silent and multiple nosocomial transmission of *C. difficile* through draft WGS of isolates from additional culture tests over a period of 10 months in a single center. These patients frequently moved between

rooms within the ward and experienced repeated discharges and re-hospitalizations. Implementing aggressive additional culture tests for *C. difficile* and conducting retrospective WGS analysis of the isolated strains could be instrumental in detecting and monitoring the effectiveness of routine infection control measures.

In this study, the main genetic lineage identified were ST81, ST183, and ST17, which have been reported as the top five frequently isolated lineages in Japan [11]. Patients isolated with ST183 were suggested to have higher risks for acquiring or transmitting *C. difficile*, potentially due to frequent room changes. This is supported by a previous report indicating that patients staying in the same room as a CDI patient have a higher risk of developing CDI [27]. Patients with CDI received antimicrobial treatment, and half of them continued to carry *C. difficile*. Moreover, their skin and surroundings could be contaminated with *C. difficile* [28].

The SNP accumulation rate of *C. difficile* is reported to be approximately 1.7×10^{-6} mutations per site per year [29]. This translates to about one SNP emerging every two months, or seven SNPs per year, based on the genome size of *C. difficile* type strain ATCC 9689=DSM 1296 (GenBank accession number CP011968.1), which is 4,109,692 bp. The core-genome rate, or the core-genome size relative to the reference genome size, is crucial when interpreting genetic relationships among bacterial strains. In our previous report, the core-genome rate was 58.4% for ST17, 32.0% for ST8, 34.1% for ST2, 72.6% for ST81, and 41.8% for ST183, using the *C. difficile* strain CD630 belonging to ST54 (accession number: NC_009089) as a reference genome. This study found no major differences in ST81 (65.9%) and ST183 (49.1%) but approximately half in ST17 (29.9%). The absence of complete genome sequence data for *C. difficile* ST17 in GenBank, and the closest hit being a ST3 strain genome in MINTyper analysis, suggests a genetic distance between ST17 and ST3 as a possible reason for the small core-genome rate in ST17. Therefore, the cut-off for interpreting core-genome SNP analysis results should be variable and dependent on the core-genome rate. In other words, it is important to note that if the core-genome size is halved, the detectable SNPs may also be halved. We adjusted the criterion for the number of SNPs suspected of transmission according to the core-genome rate by ST to avoid overestimating low SNP numbers.

This retrospective study offers an opportunity to consider the silent transmission of *C. difficile* in a hospital. A previous study conducted in a hospital in China showed a similar genetic relation rate among included *C. difficile* strains, with 43.8% (110/241 strains) being comparable to the 42.1% (16/38 strains) found in this study [30]. Although *C. difficile* ST17 was isolated within a short period, it was not considered a case of nosocomial

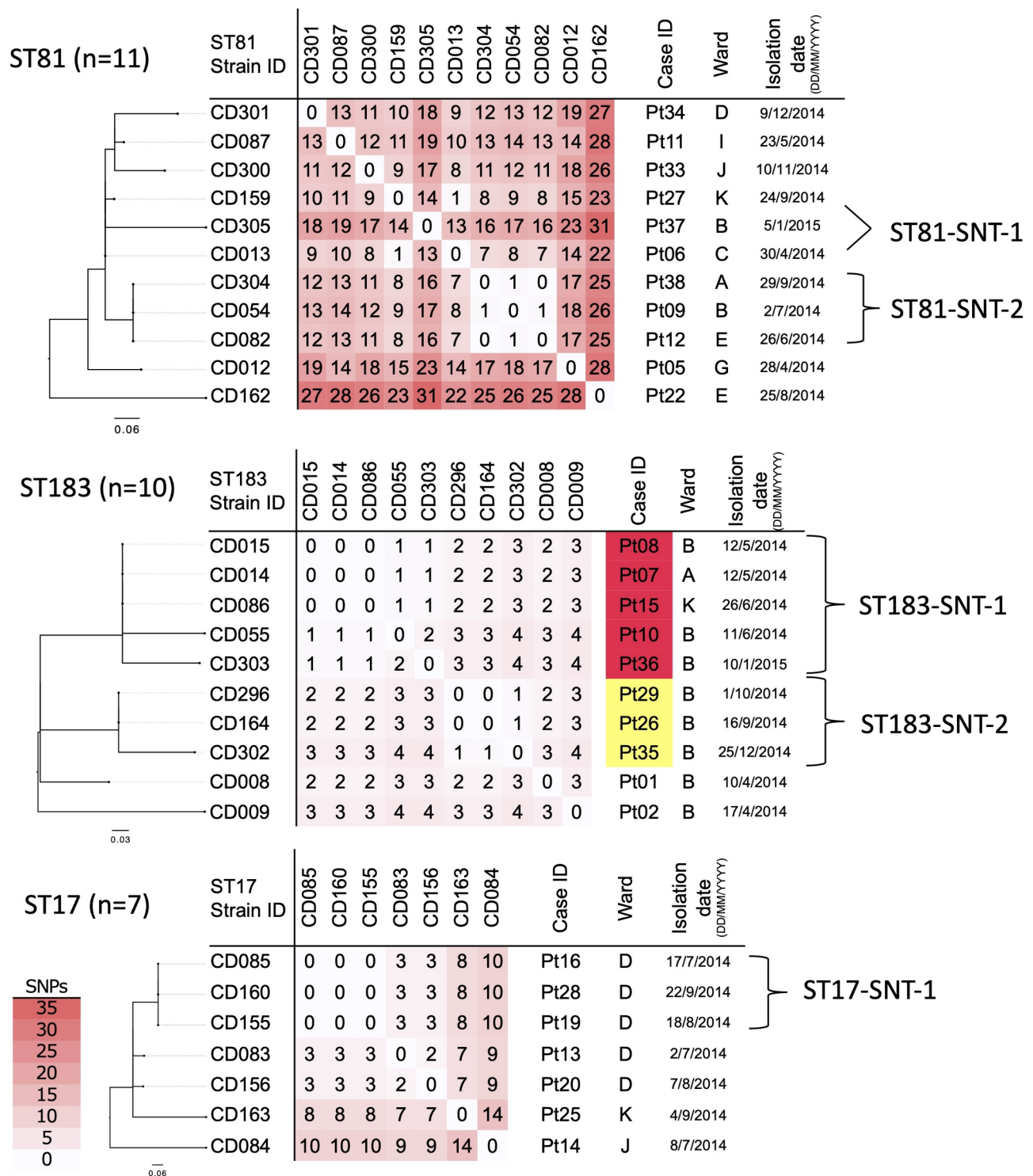


Fig. 2 Phylogenetic analysis of major sequence types (STs) of *C. difficile* based on core-genome single nucleotide polymorphisms (SNPs). The number of SNPs is displayed in a matrix. The core-genome size rate (%), defined as the ratio of the core-genome size to the reference genome, is presented as follows; 65.9% for ST81, calculated as 2,830,796/4,293,712 bp (reference CE91-St50 [ST81, NZ_AP025558.1]); 49.1% for ST183, calculated as 2,008,677/4,089,134 bp (reference S-0253 [ST8, NZ_CP076401.1]); 29.9% for ST17, calculated as 1,230,694/4,109,635 bp (reference FDAARGOS_267 [ST3, NZ_CP020424.2]). Notably, strains of Suspected Nosocomial Transmission (SNT) are highlighted in red (for ST81-SNT-1, ST183-SNT-1) and yellow (for ST183-SNT-2), corresponding to Fig. 3

Competing interests

The authors declare no competing interests.

Author details

¹Infection Control Section, Toho University Omori Medical Center, Tokyo, Japan

²Department of General Medicine and Emergency Care, Toho University School of Medicine, Tokyo, Japan

³Department of Microbiology and Infectious Diseases, Toho University School of Medicine, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan

⁴Department of Microbiology and Infection Control and Prevention, Toho University Graduate School of Medicine, Tokyo, Japan

⁵Department of Laboratory Medicine, Faculty of Medicine, Toho University School of Medicine, Tokyo, Japan

⁶Center for the Planetary Health and Innovation Science (PHIS), The IDEC Institute, Hiroshima University, Higashi-Hiroshima, Japan

Received: 22 April 2024 / Accepted: 29 August 2024

Published online: 17 September 2024

References

- Martin JSH, Monaghan TM, Wilcox MH. *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. *Nat Rev Gastroenterol Hepatol*. 2016;13:206–16.
- Carling PC, Parry MF, Olmstead R. Environmental approaches to controlling *Clostridioides difficile* infection in healthcare settings. *Antimicrob Resist Infect Control*. 2023;12:94.
- Schnitzlein MK, Young VB. Capturing the environment of the *Clostridioides difficile* infection cycle. *Nat Rev Gastroenterol Hepatol*. 2022;19:508–20.
- Nazarko L. *Clostridium difficile*: dealing with a silent menace. *Br J Community Nurs*. 2007;12(290):292–5.
- Janezic S, Rupnik M. Development and implementation of whole genome sequencing-based typing schemes for *Clostridioides difficile*. *Front Public Health*. 2019;7:309.
- Lew T, Putsathit P, Sohn KM, Wu Y, Ouchi K, Ishii Y et al. Antimicrobial susceptibilities of *Clostridium difficile* isolates from 12 Asia-Pacific Countries in 2014 and 2015. *Antimicrob Agents Chemother*. 2020;64.
- Abad-Fau A, Sevilla E, Martín-Burriel I, Moreno B, Bolea R. Update on commonly used molecular typing methods for *Clostridioides difficile*. *Microorganisms*. 2023;11.
- MacCannell DR, Louie TJ, Gregson DB, Laverdiere M, Labbe A-C, Laing F, et al. Molecular analysis of *Clostridium difficile* PCR ribotype 027 isolates from Eastern and Western Canada. *J Clin Microbiol*. 2006;44:2147–52.
- Kuijper EJ, Coignard B, Tüll P. ESCMID Study Group for *Clostridium difficile*, EU Member States, European Centre for Disease Prevention and Control. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clin Microbiol Infect*. 2006;12(Suppl 6):2–18.
- Seth-Smith HMB, Biggel M, Roloff T, Hinic V, Bodmer T, Risch M, et al. Transition from PCR-Ribotyping to whole genome sequencing based typing of *Clostridioides difficile*. *Front Cell Infect Microbiol*. 2021;11:681518.
- Aoki K, Takeda S, Miki T, Ishii Y, Tateda K. Antimicrobial susceptibility and molecular characterization using whole-genome sequencing of *Clostridioides difficile* collected in 82 hospitals in Japan between 2014 and 2016. *Antimicrob Agents Chemother*. 2019;63:e01259–19.
- Baktash A, Corver J, Harmanus C, Smits WK, Fawley W, Wilcox MH, et al. Comparison of whole-genome sequence-based methods and PCR ribotyping for Subtyping of *Clostridioides difficile*. *J Clin Microbiol*. 2022;60:e0173721.
- Janezic S, Rupnik M. Genomic diversity of *Clostridium difficile* strains. *Res Microbiol*. 2015;166:353–60.
- Uelze L, Grützke J, Borowiak M, Hammer JA, Juraschek K, Deneke C, et al. Typing methods based on whole genome sequencing data. *One Health Outlook*. 2020;2:3.
- Franz E, Rotariu O, Lopes BS, MacRae M, Bono JL, Laing C, et al. Phylogeographic Analysis Reveals Multiple International transmission events have driven the global emergence of *Escherichia coli* O157:H7. *Clin Infect Dis*. 2019;69:428–37.
- Kraft CS, Parrott JS, Cornish NE, Rubinstein ML, Weissfeld AS, McNult P et al. A Laboratory Medicine Best practices systematic review and Meta-analysis of nucleic acid amplification tests (NAATs) and algorithms including NAATs for the diagnosis of *Clostridioides (Clostridium) difficile* in adults. *Clin Microbiol Rev*. 2019;32.
- Crobach MJT, Planche T, Eckert C, Barbut F, Terveer EM, Dekkers OM, et al. European Society of Clinical Microbiology and Infectious diseases: update of the diagnostic guidance document for *Clostridium difficile* infection. *Clin Microbiol Infect*. 2016;22(Suppl 4 Supplement 4):S63–81.
- Al-Zahrani IA. *Clostridioides (Clostridium) difficile*: a silent nosocomial pathogen. *Saudi Med J*. 2023;44:825–35.
- Collins DA, Sohn KM, Wu Y, Ouchi K, Ishii Y, Elliott B, et al. *Clostridioides difficile* infection in the Asia-Pacific region. *Emerg Microbes Infect*. 2020;9:42–52.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
- Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y, Arita M. DFAST and DAGA: web-based integrated genome annotation tools and resources. *Biosci Microbiota Food Health*. 2016;35:173–84.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–76.
- Echaz JF, Veras L, Zervos M, Dubberke E, Johnson L. Hospital roommates and development of health care-onset *Clostridium difficile* infection. *Am J Infect Control*. 2014;42:1109–11.
- Sethi AK, Al-Nassir WN, Nerandzic MM, Bobulsky GS, Donskey CJ. Persistence of skin contamination and environmental shedding of *Clostridium difficile* during and after treatment of C-difficile infection. *Infect Control Hosp Epidemiol*. 2009;31:21–7.
- Gibson B, Eyre-Walker A. Investigating evolutionary rate variation in Bacteria. *J Mol Evol*. 2019;87:317–26.
- Wen X, Shen C, Xia J, Zhong L-L, Wu Z, Ahmed, MAE-GE-S, et al. Whole-genome sequencing reveals the high nosocomial transmission and Antimicrobial Resistance of *Clostridioides difficile* in a single Center in China, a four-year retrospective study. *Microbiol Spectr*. 2022;10:e0132221.
- Behroozian AA, Chludzinski JP, Lo ES, Ewing SA, Waslawski S, Newton DW, et al. Detection of mixed populations of *Clostridium difficile* from symptomatic patients using capillary-based polymerase chain reaction ribotyping. *Infect Control Hosp Epidemiol*. 2013;34:961–6.
- Dayananda P, Wilcox MH. A review of mixed strain *Clostridium difficile* colonization and infection. *Front Microbiol*. 2019;10:692.
- Hell M, Permoser M, Chmelizek G, Kern JM, Maass M, Huhulescu S, et al. *Clostridium difficile* infection: monoclonal or polyclonal genesis? *Infection*. 2011;39:461–5.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.