Genetic diversity and epidemiology of accessory gene regulator loci in Clostridioides difficile

Yuta Okada¹, Shu Okugawa¹*, Mahoko Ikeda¹,², Tatsuya Kobayashi¹, Ryoichi Saito³, Yoshimi Higurashi² and Kyoji Moriya¹,²

Abstract

Quorum sensing is known to regulate bacterial virulence, and the accessory gene regulator (agr) loci is one of the genetic loci responsible for its regulation. Recent reports examining Clostridioides difficile show that two agr loci, agr1 and agr2, regulate toxin production, but the diversity of agr loci and their epidemiology is unknown. In our study, in silico analysis was performed to research genetic diversity of agr, and C. difficile isolates from clinical samples underwent multilocus sequence typing (MLST) and PCR analysis of agr loci. To reveal the distribution of agr among different strains, phylogenetic analysis was also performed. In our in silico analysis, two different subtypes, named agr2M and agr2R, were found in agr2, which were previously reported. PCR analysis of 133 C. difficile isolates showed that 131 strains had agr1, 61 strains had agr2R, and 26 strains had agr2M; agr2R was mainly found in clade 1 or clade 2 organisms, whereas agr2M was only found in clade 4. With rare exception, agr1-negative sequence types (STs) belonged to clade C-I and C-III, and one clade 4 strain had agr2R. Our study revealed subtypes of agr2 not previously recognized, and the distribution of several agr loci in C. difficile. These findings provide a foundation for further functional and clinical research of the agr loci.

DATA SUMMARY

The genomic sequences of ‘agrD’ and ‘agrB’ on the whole-genome sequence of C. difficile strain 630 (GenBank accession number: AM180355.1) and the sequences of CDR20291_3187, CDR20291_3187A, CDR20291_3188, CDR20291_3189 on the whole-genome sequence of R20291 (GenBank accession number: FN545816.1) were used as templates to search for similar genetic sequences on the C. difficile whole-genome sequence registered in the nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) database.

INTRODUCTION

Clostridioides difficile is an obligate anaerobic bacterium with the ability to form spores. First described in 1978 as a clinical entity, C. difficile infection is often characterized by gastrointestinal symptoms like diarrhea. Moreover, potentially lethal manifestations of this disease, such as toxic megacolon or sepsis, are sometimes observed [1, 2]. More than 40 years have passed since C. difficile infection was first reported, and now toxin A and B produced by C. difficile are known to be the culprit of this usually nosocomial disease [2]. However, the mechanism regulating the production and extracellular release of these toxins is not fully understood, and knowledge leading to development of novel drugs is of importance.

Bacterial quorum sensing is a topic of rapidly expanding interest. Quorum sensing refers to dynamic control of bacterial density, metabolism and various physiological activities mediated by signalling molecules that the bacteria themselves produce [3]. As the signalling molecule accumulates in the surrounding environment, bacteria sense the increase of cell density and thus may change their collective behaviour accordingly. In some bacterial species, a relationship between quorum sensing and specific virulence factors has been observed [4–6].

The agr system is one such quorum-sensing system in bacteria that has been described in the literature. A few reports show that dysfunction of the agr system leads to decreased production and activity of C. difficile toxin production [7–9]. The agr...
system and its orthologues are known to be present across firmicutes including *Staphylococcus aureus* [10]. In *S. aureus*, links between many virulence factors and the *agr* system are reported, and such associations are also investigated in other organisms possessing genes akin to staphylococcal *agr* [11, 12].

In contrast, in *C. difficile*, the presence of strains only having *agrB* and *agrD* (*agr1* locus) and strains that additionally have a complete set of four *agr* genes, *agrC, agrA, agrB* and *agrD* (traditionally called the *agr2* locus), was previously reported in clinical samples. It was suggested that most strains fall into the latter group [13]. However, the genetic diversity of the *agr* loci remain to be fully understood. Recent phylogenetic and evolutionary studies of *C. difficile* has been performed based on MLST, and groups of strains that are phylogenetically close to each other called ‘clades’ are now widely accepted, but the linkage of STs and clades with *agr* subtypes is not known to date.

The function of *agr* genes in *C. difficile* is also not well understood. Some previous studies showed that *C. difficile* *agr* genes positively enhance toxigenicity through a quorum-signalling substance supposed to have a thiolactone structure that the auto-inducer-peptides (AIPs) produced by *S. aureus* *agr* system also has [7–9]. Insertional deletion of *agrA* by the

---

**Table 1. Primers used for screening of *agr1, agr2R* and *agr2M***

| No. | Primer name | Sequence (5’→3’) |
|-----|-------------|------------------|
| 1   | agr1_BD_F   | GGCTGTGATAAATGCAAGGACAGGTACG |
| 2   | agr1_BD_R   | GCTTTCATAGTAAATACACCCATGC |
| 3   | agr2R_AC_F  | GACCTACTGAGAAGCCTTAGC |
| 4   | agr2R_AC_R  | GAGTAAAGGGCTTGAACCTGC |
| 5   | agr2M_AC_F  | GTGAATTTCAAGTTTTCAGATGC |
| 6   | agr2M_AC_R  | AGCTAAACCTTTCCCCATC |

---

**Fig. 1.** Genomic organization of (a) *agr1*, (b) *agr2R* and (c) *agr2M* loci. The targeted regions of the primer pairs in Table 1 are also shown, with the expected length of amplicons shown.
ClosTron system in the agr2 locus of strain R20291 led to underexpression of flagellar biosynthesis genes in C. difficile [7].

The aim of this study was to investigate the epidemiology of agr genes in the context of a MLST-based phylogeny, laying the groundwork for future microbiological and clinical research of C. difficile agr.

METHODS

In silico analysis of C. difficile agr loci on GenBank

Using genomic data of the agr1 locus of C. difficile strain CD630 (GenBank accession number: AM180355.1) and agr2R locus of strain R20291 (GenBank accession number: FN545816.1) as template sequences [9], BLAST search for agr homologues was performed. In addition to previously known agr loci, the newly identified agr2M locus has the same four-component (agrA, agrB, agrC, agrD) structure as agr2R but with ~80% nucleotide identity to agr2R. These three agr loci were targeted in the subsequent analysis by PCR.

C. difficile isolates

A total of 133 non-duplicate C. difficile clinical isolates were obtained from the University of Tokyo Hospital between 2013 to 2014 [14]. Laboratory reference strains ATCC BAA-1382 and NCTC 13366 (hereafter referred to as CD630 and R20291, respectively) were also subjected to identical isolation methods and subsequent analysis.

Genetic analysis of agr loci in C. difficile strains

Both clinical and reference lab strains of C. difficile in our study were anaerobically cultured by our previous reported method [14]. C. difficile colonies on culture agar were inoculated into sterile water and then boiled at 95°C for 10 min to make DNA templates for subsequent PCR using the Emerald Amp PCR Master Mix kit (Takara Bio, Shiga, Japan). A 100bp DNA ladder (Toyobo, Osaka, Japan) was used to evaluate product size during gel electrophoresis of PCR amplicons. MLST and toxigenicity analysis were performed as described in previous reports [14–16]. As part of MLST analysis, PCR amplicons underwent DNA sequencing, and STs and clades were determined based on DNA sequencing data using the PubMLST sequence query page (https://pubmlst.org/cdifficile/). PubMLST was also used to determine STs and clades of some strains found in our in silico analysis. PCR of agr loci was performed on laboratory strains and clinical isolates targeting genetic regions overlapping with agrB1 and agrD1 of CD630 (AM180355.1) using primer pairs agr1_BD_F/ agr1_BD_R, agr2R_C and agr2R_A of R20291 (FN545816.1) using pairs agr2R_AC_F/agr2R_AC_R, and agr2M_C and agr2M_A of M68 (FN5458375.1) using pairs agr2M_AC_F/ agr2M_AC_R (Table 1, Fig. 1). Amplicons recovered from PCR underwent DNA sequencing by the Sanger method at Eurofins Genomics (Ebersberg, Germany) to confirm that the amplicons match the targeted genetic region. For samples negative for agr1-screening, additional PCR examination was performed to amplify the genomic region entirely, including agr1. Concatenated nucleotide sequences of housekeeper genes used to determine MLST was obtained from PubMLST website, and these datasets underwent phylogenetic analysis using PhyML via the Bio.Phylo library integrated with Biopython [17–19]. A phylogenetic tree was drawn using iTOL v5 (https://itol.embl.de/) [20], and clades of STs not specified on PubMLST were determined based on the tree.

RESULTS

In silico analysis of agr subtypes among C. difficile

Using the agr1 sequence of CD630 (GenBank accession number: AM180355.1) and the agr2R sequence of R20291 (GenBank accession number: AM545816.1) as templates, BLAST search was performed to find agr homologues using the genomic data of C. difficile strains registered at GenBank. Regarding agr1, almost all strains had genetic regions with over 96% nucleotide-based identity to agr1 of CD630. On the other hand, BLAST search using the nucleotide sequence of the R20291 agr2R locus showed that, besides whole-genome sequences with genetic regions of over 95% identity to agr2R, there were a group of strains that had genetic regions with around 80% identity compared to the agr2R of strain R20291 (Table 2). One of these strains is M68 (GenBank accession number: FN668375.1), which is a toxin A-negative, toxin B-positive strain belonging to ST37/ clade 4. To distinguish this from agr2R, the agr2R locus in strains of M68 on GenBank was named agr2M. The genetic organization of the agr loci is shown in Fig. 1. Fig. 2 shows that the amino acid sequences predicted from ORF analysis

| Strain          | Accession No. | ST  | Clade |
|-----------------|---------------|-----|-------|
| DSM 29629       | CP016104.1    | ST39| Clade 4|
| DSM 29627       | CP016102.1    | ST37| Clade 4|
| CD161           | CP029154.1    | ST37| Clade 4|
| CDT4            | CP029152.1    | ST37| Clade 4|
| M68             | FN668375.1    | ST37| Clade 4|
| CFS             | FN665652.1    | ST76| Clade 4|
| DSM28669        | CP012323.1    | ST109| Clade 4|
| BJ08            | CP003939.1    | "| "|
| DSM29637        | CP016106.1    | ST83| Clade 1|
| CBA7204         | CP029566.1    | ST203| Clade 1|

*WGS of BJ08 did not have adk (one of the seven housekeeper genes used in MLST), but the sequence of the other six housekeeper genes used to determine sequence type were identical to that of ST37.
Fig. 2. Multiple sequence alignment of Agr2R and Agr2M by Clustal Omega. Comparisons between amino acid sequences of AgrB, AgrD, AgrC and AgrA are shown. Symbols are as follows: an asterisk indicates identical, a colon means strongly similar, and a comma indicates weakly similar.
of \textit{agr}R and \textit{agr}M shares many amino acids, but \textit{Agr}D and \textit{Agr}A show 10–20\% difference in length between the two loci.

**Epidemiology of \textit{agr} genes in clinical isolates of \textit{C. difficile}**

PCR was performed on a total of 133 \textit{C. difficile} isolates and two laboratory strains of CD630 and R20291 to confirm the presence of \textit{agr}1, \textit{agr}2R and \textit{agr}2M. The position of the sequences corresponding to the primers is shown in Fig. 1. The whole list of STs/clades, \textit{agr} patterns and toxigenicity revealed in our study is shown in Table 3; the relationships between clades and \textit{agr} patterns in clinical isolates of \textit{C. difficile} was elucidated. Previous studies on \textit{C. difficile} \textit{agr} mainly focused on regulatory effects of \textit{agr} genes on virulence factors, such as toxin production in laboratory strains [7–9], so the finding that not only toxigenic strains but also nontoxigenic strains possess these genes is of note and warrants further analysis of unknown functions for \textit{agr} loci in clinical isolates of \textit{C. difficile}. Potential new functions of \textit{agr}M are also of interest, and we are the first to report this locus.

As of laboratory reference strains, PCR analysis confirmed that CD630 and R20291 both have a genetic region corresponding to \textit{agr}1, but CD630 does not have \textit{agr}R or \textit{agr}M. In contrast, R20291 has the \textit{agr}2R region. Of the total 133 clinical isolates, 131 isolates were positive for \textit{agr}1, 61 isolates were positive for \textit{agr}2R, and 26 isolates were positive for \textit{agr}2M. The tree shows no branching point dividing STs with \textit{agr}1+\textit{agr}2R and \textit{agr}1+\textit{agr}2M, respectively.

As of laboratory reference strains, PCR analysis confirmed that CD630 and R20291 both have a genetic region corresponding to \textit{agr}1, but CD630 does not have \textit{agr}R or \textit{agr}M. In contrast, R20291 has the \textit{agr}2R region. Of the total 133 clinical isolates, 131 isolates were positive for \textit{agr}1, 61 isolates were positive for \textit{agr}2R, and 26 isolates were positive for \textit{agr}2M. The whole list of STs/clades, \textit{agr} patterns and toxigenicity revealed in our study is shown in Table 3; the relationships between clades and \textit{agr} patterns in clinical isolates of \textit{C. difficile} was elucidated. Previous studies on \textit{C. difficile} \textit{agr} mainly focused on regulatory effects of \textit{agr} genes on virulence factors, such as toxin production in laboratory strains [7–9], so the finding that not only toxigenic strains but also nontoxigenic strains possess these genes is of note and warrants further analysis of unknown functions for the \textit{agr} loci of \textit{C. difficile}. Potential new functions of \textit{agr}M are also of interest, and we are the first to report this locus.

Our study not only adds to previous reports showing the universality of \textit{agr}1 and the distribution of \textit{agr}2R among clinical and laboratory \textit{C. difficile} isolates [13, 27] but also revealed the distribution of \textit{agr}2M. The discrepancy of \textit{agr} patterns within a single ST observed in some STs may provide clues to help uncover the evolutionary pathway of \textit{C. difficile}. A previous study by Darkoh \textit{et al.} [9] showed that examined laboratory strains of \textit{C. difficile} could live without the \textit{agr}1 locus. The presence of \textit{agr}1-negative isolates in our study also suggests that loss of \textit{agr}1 in \textit{C. difficile} is not lethal, and the same thing can be said about \textit{agr}2R and \textit{agr}2M. The subject of our phylogenetic analysis does not represent the whole population of \textit{C. difficile}, but still, several

**DISCUSSION**

This is the first study focusing on a phylogenetic analysis of the \textit{agr} locus in clinical isolates of \textit{C. difficile}. Based on our finding that two relatively different subtypes of \textit{agr} exists in \textit{agr}2, the epidemiology of \textit{agr} in clinical isolates of \textit{C. difficile} was elucidated. Previous studies on \textit{C. difficile} \textit{agr} mainly focused on regulatory effects of \textit{agr} genes on virulence factors, such as toxin production in laboratory strains [7–9], so the finding that not only toxigenic strains but also nontoxigenic strains possess these genes is of note and warrants further analysis of unknown functions for the \textit{agr} loci of \textit{C. difficile}. Potential new functions of \textit{agr}M are also of interest, and we are the first to report this locus.

The phylogenetic tree also shows that the closest ST to ST402 in the phylogenetic tree is ST204, which was previously mentioned to be clade C-III [26]; other STs of clade C-III are also relatively close to ST402.

![Fig. 3. Visualization of PCR bands from \textit{agr}1, \textit{agr}2R and \textit{agr}2M screening. 1: strain CD630 (laboratory strain); 2: strain R20291 (laboratory strain); 3: ST54/clade 1 (clinical isolate); 4: ST17/clade 1 (clinical isolate); and 5: ST81/clade 4 (clinical isolate).](image-url)
notable findings about \textit{agr} were obtained. First, a deviation of \textit{agr} patterns in some local parts of the phylogenetic tree/clades was observed. For example, almost all \textit{agr2R}-positive strains were in clade 1 or 2, and all STs with \textit{agr2M} were in clade 4. Consistent with these results, a previous phylogenetic study of \textit{C. difficile} suggests that \textit{agr2M} is mainly found in clade 4 and rarely in clade 1 [28]. Interestingly, the subclade-like group in clade 4, which the \textit{agr2R}-positive ST198 belongs to, includes two \textit{agr2R}-negative STs without \textit{agr2M}, whereas another subclade in clade 4, including ST37 and ST81, does not have any STs with \textit{agr2R}.

The fact that single \textit{agr1}-negative isolates were found in both clade C-I and clade C-III suggests that loss of \textit{agr1} may have occurred at some point during their evolution. The validity of bacterial identification of clade C-I strains was questioned in a previous report based on the fact that the average nucleotide identity score between strains of clade C-I and clade 1–5 were low [22]. Not only genetic but also functional characterization of \textit{C. difficile} clades is expected to provide further insight into bacteriological evolution and taxonomy, and analysis of \textit{agr} function may play a key role in moving the field forward.

| Sequence types | Clade | No. of samples | \textit{agr1} | \textit{agr2R} | \textit{agr2M} | toxigenicity |
|----------------|-------|----------------|------------|------------|------------|-------------|
| ST17           | 1     | 12             | +          | −          | +          | +           |
| ST109          | 4     | 11             | +          | −          | +          | −           |
| ST81           | 4     | 10             | +          | −          | +          | +           |
| ST2            | 1     | 9              | +          | −          | −          | +           |
| ST15           | 1     | 9              | +          | −          | −          |           |
| ST54           | 1     | 9              | +          | −          | −          | +           |
| ST8            | 1     | 8              | +          | −          | −          | +           |
| ST3            | 1     | 5              | +          | −          | +          | +           |
| ST37           | 4     | 4              | +          | −          | +          | +           |
| ST100          | 1     | 4              | +          | −          | −          | −           |
| ST26           | 1     | 3              | +          | −          | −          | −           |
| ST35           | 1     | 3              | +          | −          | −          | +           |
| ST33           | 1     | 3              | +          | −          | −          |           |
| ST11           | 5     | 2              | +          | −          | +          | +           |
| ST14           | 1     | 2              | +          | −          | −          | +           |
| ST48           | 1     | 2              | +          | −          | +          | +           |
| ST55           | 1     | 2              | +          | −          | −          | +           |
| ST401          | 4     | 2              | +          | −          | −          | −           |
| ST407          | C-I   | 2              | +          | −          | −          | −           |
| ST5            | 3     | 1              | +          | −          | −          | +           |
| ST28           | 1     | 1              | +          | −          | −          | −           |
| ST41           | 2     | 1              | +          | −          | −          | +           |
| ST42           | 1     | 1              | +          | −          | −          | +           |
| ST49           | 1     | 1              | +          | −          | −          | +           |
| ST58           | 1     | 1              | +          | −          | −          | +           |
| ST63           | 1     | 1              | +          | −          | −          | +           |
| ST66           | 1     | 1              | +          | −          | −          | +           |
| ST123          | 2     | 1              | +          | −          | −          | +           |
| ST129          | 1     | 1              | +          | −          | −          | +           |
| ST153          | 1     | 1              | +          | −          | −          | +           |
| ST159          | 4     | 1              | +          | −          | −          | +           |
| ST183          | 1     | 1              | +          | −          | −          | +           |
| ST198          | 4     | 1              | +          | −          | −          | +           |
| ST243          | 4     | 1              | +          | −          | −          | −           |

Table 3. Continued

| Sequence types | Clade | No. of samples | \textit{agr1} | \textit{agr2R} | \textit{agr2M} | toxigenicity |
|----------------|-------|----------------|------------|------------|------------|-------------|
| ST247          | 1     | 1              | +          | −          | +          | +           |
| ST278          | 1     | 1              | +          | −          | −          | +           |
| ST297          | C-I   | 1              | −          | −          | −          | −           |
| ST301          | 1     | 1              | +          | −          | −          | +           |
| ST303          | C-I   | 1              | +          | −          | −          | −           |
| ST304          | 1     | 1              | +          | −          | −          | +           |
| ST400          | 1     | 1              | +          | −          | +          | −           |
| ST402          | C-III | 1              | −          | −          | −          | −           |
| ST403          | 4     | 1              | +          | −          | −          | −           |
| ST404          | 1     | 1              | +          | −          | −          | +           |
| ST405          | 4     | 1              | +          | −          | −          | +           |
| ST406          | 1     | 1              | +          | −          | −          | +           |
| ST408          | 1     | 1              | +          | −          | −          | +           |

Table 4. Relationship between \textit{agr} patterns and clades

| Clade | 1 | 2 | 3 | 4 | 5 | C-I | C-III | total |
|-------|---|---|---|---|---|-----|-------|-------|
| \textit{agr1} | 32 | 0 | 2 | 5 | 1 | 3  | 0     | 44    |
| \textit{agr1}+\textit{agr2R} | 57 | 3 | 0 | 1 | 1 | 0  | 0     | 61    |
| \textit{agr1}+\textit{agr2M} | 0  | 0 | 0 | 26| 0 | 0  | 0     | 26    |
| Negative | 0 | 0 | 0 | 0 | 0 | 1  | 1     | 2     |
| Total   | 89| 3 | 2 | 32| 2 | 4  | 1     | 133   |

Continued
In summary, our study sheds light on the previously unknown genetic diversity and molecular epidemiology of the agr loci in C. difficile. Together with our MLST-based phylogenic analysis, our findings may offer novel insights into the agr loci, leading to further functional, evolutionary and clinical research of this gene.

Fig. 4. Maximum-likelihood phylogenetic tree (model TN93, 1000 replicates) based on analysis by PhyML. STs found in our clinical samples; ST1 (corresponding to R20291 strain) and STs used in previous studies as references for clade C-I samples; ST1 (corresponding to R20291 strain) and STs used in replicates) based on analysis by PhyML. STs found in our clinical research of this gene.

Funding information
This work was supported by JSPS KAKENHI Grant Number 19K08949. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgements
We would like to thank Editage (www.editage.jp) for English language editing.

Author contributions
Conceptualization: Y.O. and S.O. Methodology: Y.O. and S.O. Software: Y.O. Validation: Y.O. Formal analysis: Y.O. and S.O. Investigation: Y.O. and S.O. Resources: Y.O. and S.O. Data curation: Y.O. and S. O. Writing – original draft: Y.O. Writing – review and editing: Y.O., S.O., M.I., T.K., R.S., Y.H., K.M. Visualization: Y.O. Supervision: K.M. Project Administration: S.O. Funding: S.O.

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

Ethical statement
This study was approved by the Institutional Ethics Committee (Number Ethical statement
The authors declare that there are no conflicts of interest.

References
1. Kuehne SA, Cartman ST, Hepa JT, Kelly ML, Cockayne A et al. The role of toxin A and toxin B in Clostridium difficile infection. Nature 2010;467:711–713.
2. Bartlett JG, Chang TW, Gurwith M, Gorbach SL, Onderdonk AB. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. N Engl J Med 1978;298:531–534.
3. Whiteley M, Diggle SP, Greenberg EP. Progress in and promise of bacterial quorum sensing research. Nature 2017;551:313–320.
4. Ohtani K, Hayashi H, Shimizu T. The luxS gene is involved in cell-cell signalling for toxin production in Clostridium perfringens. Mol Microbiol 2002;44:171–179.
5. Sperandio V, Li CC, Kaper JB. Quorum-Sensing Escherichia coli regulator A: a regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic E. coli. Infect Immun 2002;70:3085–3093.
6. Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL et al. Quorum-sensing regulators control virulence gene expression in Vibrio cholerae. Proc Natl Acad Sci U S A 2002;99:3129–3134.
7. Martin MJ, Clare S, Goulding D, Faulds-Pain A, Barquist L et al. The agr locus regulates virulence and colonization genes in Clostridium difficile 027. J Bacteriol 2013;195:3672–3681.
8. Darkoh C, DuPont HL, Norris SJ, Kaplan HB. Toxin synthesis by Clostridium difficile is regulated through quorum signaling. mBio 2015;6:e02569.
9. Darkoh C, Odo C, DuPont HL. Accessory gene Regulator-1 locus is essential for virulence and pathogenesis of Clostridium difficile. mBio 2016;7:e01237-16.
10. Wuster A, Babu MM. Conservation and evolutionary dynamics of the agr cell-to-cell communication system across Firmicutes. J Bacteriol 2008;190:743–746.
11. Qin X, Singh KV, Weinstock GM, Murray BE. Effects of Entero- coccus faecalis fsr genes on production of gelatinase and a serine protease and virulence. Infect Immun 2000;68:2579–2586.
12. Autret N, Raynaud C, Dubail I, Berche P, Charbit A. Identification of the agr locus of Listeria monocytogenes: role in bacterial virulence. Infect Immun 2003;71:4463–4471.
13. Marsden GL, Davis IJ, Wright VJ, Sebaihia M, Kuiper EJ et al. Array comparative hybridisation reveals a high degree of similarity between UK and European clinical isolates of hypervirulent Clostridium difficile. BMC Genomics 2010;11:389.
14. Okada Y, Yagihara Y, Wakabayashi Y, Igawa G, Saito R et al. Epide- miology and virulence-associated genes of Clostredioides difficile isolates and factors associated with toxin EIA results at a university hospital in Japan. Access Microbiol 2020;2.
15. Kuwata Y, Tanimoto S, Sawabe E, Shima M, Takahashi Y et al. Molecular epidemiology and antimicrobial susceptibility of Clostridium difficile isolated from a university teaching hospital in Japan. Eur J Clin Microbiol Infect Dis 2015;34:763–772.

16. Persson S, Torpdahl M, Olsen KE. New multiplex PCR method for the detection of Clostridium difficile toxin A (tcdA) and toxin B (tcdB) and the binary toxin (cdtA/cdtB) genes applied to a Danish strain collection. Clin Microbiol Infect 2008;14:1057–1064.

17. Talevich E, Invergo BM, Cock PJ, Chapman BA. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in BioPython. BMC Bioinformatics 2012;13:209.

18. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ et al. BioPython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422–1423.

19. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;59:307–321.

20. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B et al. Toward automatic reconstruction of a highly resolved tree of life. Science 2006;311:1283–1287.

21. 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.

22. Chen R, Feng Y, Wang X, Yang J, Zhang X et al. Whole genome sequences of three clade 3 Clostridium difficile strains carrying binary toxin genes in China. Sci Rep 2017;7:43555.

23. 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.

24. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of Clostridium difficile. Clin Microbiol Rev 2015;28:721–741.

25. Ramirez-Vargas G, López-Ureña D, Badilla A, Orozco-Aguilar J, Murillo T et al. Novel clade C-I Clostridium difficile strains escape diagnostic tests, differ in pathogenicity potential and carry toxins on extrachromosomal elements. Sci Rep 2018;8:13951.

26. 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.

27. Stabler RA, He M, Dawson L, Martin M, Valiente E 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:R102.

28. Riedel T, Wetzel D, Hofmann JD, Pioron SPEO, Dannheim H et al. High metabolic versatility of different toxigenic and non-toxigenic Clostridioides difficile isolates. Int J Med Microbiol 2017;307:311–320.

---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.