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

Yuta Okada[1], Shu Okugawa[1,*], Mahoko Ikeda[1,2], Tatsuya Kobayashi[1], Ryoichi Saito[3], Yoshimi Higurashi[2] and Kyoji Moriya[1,2]

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 *agr2R* and *agr2M*, 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.

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

No.	Primer name	Sequence (5'-->3')
1	agr1_BD_F	GGCTGATGAATAATCCAAGGACAGGTACTA
2	agr1_BD_R	GCTTTCATAGTTAATATAACCACCATGC
3	agr2R_AC_F	GACCTACTGCAGAACCTTCAGC
4	agr2R_AC_R	GAGTTAAAGGCTTGAAACTTGC
5	agr2M_AC_F	GTGAATTTGGATTTTTCAGATGCC
6	agr2M_AC_R	AGCTAAACCTTCCCCCATC

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
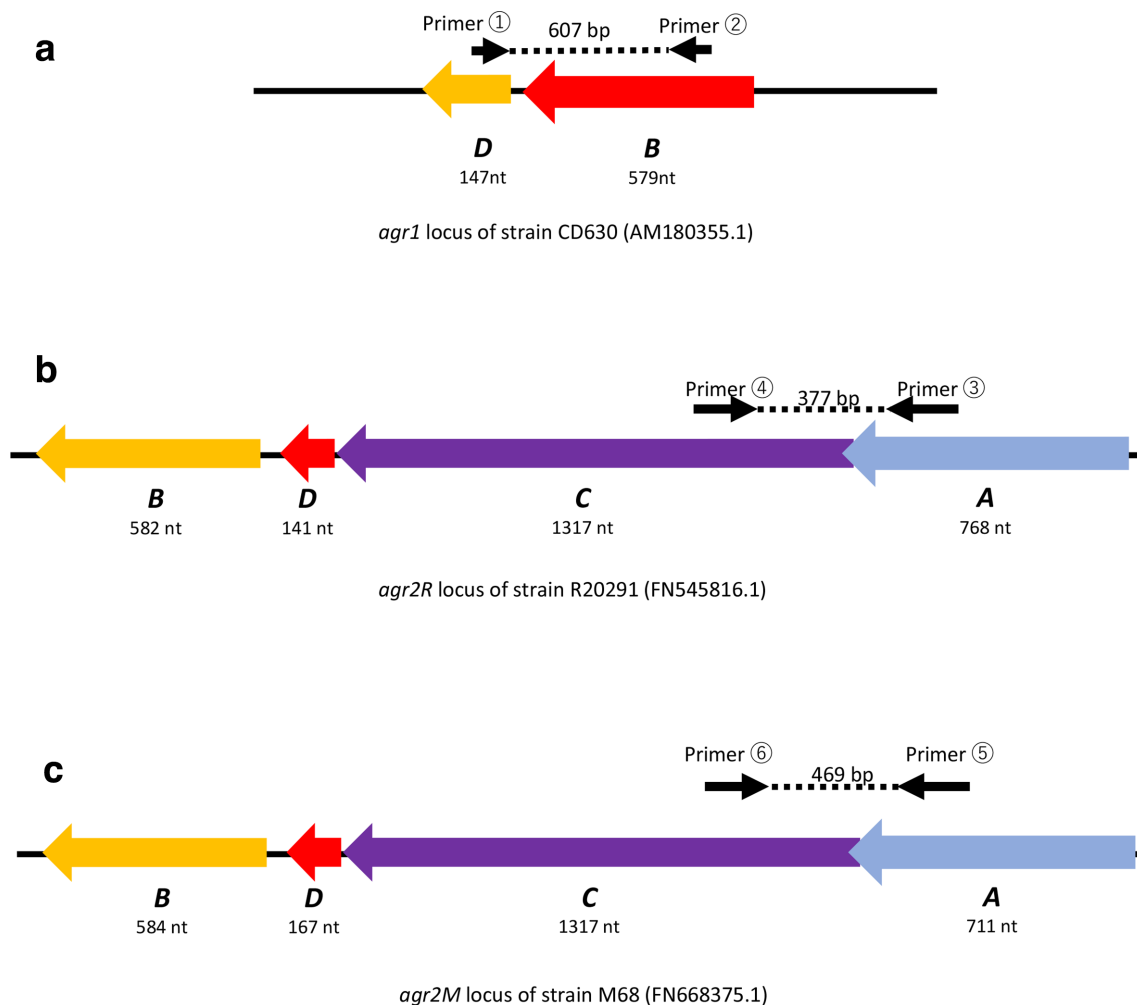
Fig. 1. Genomic organization of (a) *agr1*, (b) *agr2R* and (c) *agr2M* loci. The targeted regions of the primer pairs in Table 1 are also shown, with the expected length of amplicons shown.

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 100 bp 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(FN668375.1) 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

Table 2. List of whole-genome sequences (WGS) of *C. difficile* that likely possess *agr2M*. BLAST search showed that these WGSs have a region with 80.33–80.37 % nucleotide identity compared with *agr2R* region of R20291 (FN545816.1) and >95% identity compared with *agr2M* region of M68 (FN668375.1). Determination of ST/clade was performed using the PubMLST sequence query page

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
CF5	FN665652.1	ST86	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.

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.

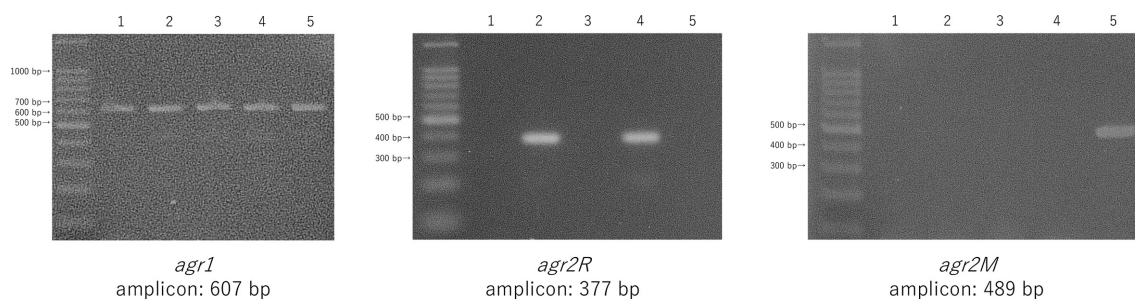## RESULTS

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 *agr2* 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 *agr2* 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

```
Agr2R_A    MNRNFIYLLKIYKKGCLRIVISIGICDDELHYRIKIKDILSEILSSYPINYNIYEFSSGE      60
Agr2M_A    ------------------LINIGICDDELHYRLKIKDILKKVLSSYTMDYNIHEFSSGK      41
                             :*.**********.******..::****.::***.*****.

Agr2R_A    ELLNNYPKDLDILIMDIQMKTINGMDTARKIREFDHKLEIIFVTSFVEFMQEGYEVKAYR     120
Agr2M_A    ELLCNYPKNLDILIIDIQMEILNGIDTSRKIREFDENLEIIFMTSFPEFMQEGYEVKAYR     101
           *** ****:*****::****:.*:**:**:*******.:***:***.*:************

Agr2R_A    YILKPINKEKISKSVLPCINEMMKKRNNYLTINVKNYVDRIKIDSITYIETDRPNILIYT     180
Agr2M_A    YILKPINEKKITKNILPCIDEIMKKRNNYLTIKVKNYVDRIKIDSIIYIETDRPNIVIYT     161
           *******:*:*:**.:****:*:**********:************:**********:***

Agr2R_A    HDDMYITKMSISKIEKILNEYGFFRCHNSYIVNLKLVESMSGSTVIVDGKSIPISKYRVK     240
Agr2M_A    HDNTYTTKMSISKIEKILSEFGFFRCHNSYIINLKLVQSMNGNSVVINGKSIPVSKYRVK     221
           **:.* ************.*.:**********:*****.**.*.:*::.*****:******

Agr2R_A    GLKLAITNILGDIVC     255
Agr2M_A    GLKLAITNILGDIIC     236
           *************:*


Agr2R_B    MFKRLSYKFANILVNNEIVESEDFEIYRYGFETLIYFIVNISVALFIGIIFDRFIHTVIF      60
Agr2M_B    MFKSLSYKFANILVRNEVIEDEDFEIYRYGFETLVYFIINISVALLGIALNKFIQTIIF      60
           *** *********.**::*:.***********:***:*******:** ::.**:*:**

Agr2R_B    LSCYCTLRQFTGGYHARNYKECTLTFAVIYLITIFSANNIDINKYKYLLVLLMIISILTI     120
Agr2M_B    LVCYCTLRQFTGGYHARNYTECTITFALIYSIILVTKNIDIYKFKYILLLLLLLSTVII     120
           * ***************** ***:***:** :::**** *:**:*:**::* : *

Agr2R_B    YKLAPLEHRNKPLSESEKKHYRKTVQKILFVIICLIILCKILNIFQQYVIYAIISIYWIA     180
Agr2M_B    HKVAPLEHRNKPLSIYEKKNYRNIIKKITLSIIIVVTISLIFNIMSEYIIYSSLAVFLIT     180
           :*:*********** .***:**: :**:  : **  :: .*:**:.:*:**:  :::. *:

Agr2R_B    ILIYIGMKVNNDQ--     193
Agr2M_B    ILLLVQIIINFFKK*     194
           **: : : :*  :


Agr2R_C    MLTNLSIGFWNIINLVSGTIDWIIFFLVIHLVGKRKISNIRYIFTTIIIFIFMAYINVVK      60
Agr2M_C    MLETLNINLWDIINIITGCIDWIIFFLVMSILGKKNISFRKYTIYMTLVLVSMSLVNFMS      60
           ** .*.*.:*:***::* *********:. ::**:.**:*:  :::: *: :*:*.

Agr2R_C    VFPNIKILLCMVIGISFYITSYEDKIYKSIIVSLLFWLGLMIAEGSAVGLVVFINKLNNI     120
Agr2M_C    IFPNFKIIICIVIGSLFYIASYKESIYKCILITLLFWLGLMMGEGLAIGIVVVINQLNNI     120
           :***:**:.*:*:***.:***:**:*.***: *::***********.** *:*:**.:**.:**.**:****

Agr2R_C    NITQHENLFRIQAIIMSKVFLFIGFILFKYFKLSLDFKPKDMILIGIPILSNIVSLLLIF     180
Agr2M_C    DIILNGNLFRLESIVISKIFLFIIFMSIKYFKLSIEFKPKDMVLIGLPILSNIVSLLLIF     180
           :*  : ****::*:*::**:***: *:  *****::******:***:************

Agr2R_C    EYNLRVNTITTENLFIFIFAILLILSSSIILLIVIGKIVQDDKLNLEYELINERIKTNYR     240
Agr2M_C    GYNLKNNFVSDDNIFVIVLSTLLIVSSSIIILIVIGKIVQDDKVKLEYELINERIKIDYK     240
            ***: * :: :*:*:*::::: ***.***:*************::*********** :*:

Agr2R_C    NYENINEVHDRLKYVYHDLKNHMICIKNYDTKEEIISYIDNLEFQISEFEHFKNTGNKTL     300
Agr2M_C    SYESINEIHDKLKYVYHDLKNHMICIKNYDTKEEIISYVNNLENKINDFEYFRNTGNKTL     300
           .**.***:*:**:***************************:.:*:**::*.:.**:*.*:*:*:****

Agr2R_C    DLILAEKISICKKYDIQIEDNINISKLNFIENNDICSIFANSLDNAIEACMDINNELEKR     360
Agr2M_C    DIILGEKIYLCKKYNIEFEDSINISKLQFIQDVDICAIFANALDNAIEACMKMNNNTEKR     360
           *:**.*** :****:*::**.******:**:: ***:****.*********:.**. ***

Agr2R_C    IEVKATYINKFAIIKFINTKVNDIKLIDKRIQTSKDDDKIHGIGLASIKYIVNKYGGETI     420
Agr2M_C    IEVKATYVNGFAIIKFINTKVNDIKFIDERIKTSKDNNRVHGIGLASIKYMVSKYDGELI     420
           *******:* ***************:**:**:*****::::*********:*.**.** *

Agr2R_C    VNYSDNEFILKIMIPIKS 438
Agr2M_C    VNYSDNEFILKIMIPIKA 438
           *****************:


Agr2R_D    MKKIALNLLKNISALSFGIAVLSANSASSWVAHQAKEPQALQKLKK---------      46
Agr2M_D    MKKIALSLLKSMSTLSLGISRLSANTASSWIAHQAEEPQALKKLKKNNTSTKFSL      55
           ******.***:.*:**:**:. ****.****:****.*****::*****.****
```

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.

Fig. 3. Visualization of PCR bands from *agr1*, *agr2R* and *agr2M* screening. 1: strain CD630 (laboratory strain); 2: strain R20291 (laboratory strain); 3: ST54/lade 1 (clinical isolate); 4: ST17/clade 1 (clinical isolate); and 5: ST81/clade 4 (clinical isolate).

of *agr2R* and *agr2M* shares many amino acids, but *AgrD* and *AgrA* show 10–20% difference in length between the two loci.

Epidemiology of *agr* genes in clinical isolates of *C. difficile*

PCR was performed on a total of 133 *C. difficile* isolates and two laboratory strains of CD630 and R20291 to confirm the presence of *agr1*, *agr2R* and *agr2M*. The position of the sequences corresponding to the primers is shown in Fig. 1. Fig. 3 shows examples of gel electrophoresis visualizing bands corresponding to 607, 377 and 469 bp amplicons yielded by PCR targeting *agr1*, *agr2R* and *agr2M*, respectively.

As of laboratory reference strains, PCR analysis confirmed that CD630 and R20291 both have a genetic region corresponding to *agr1*, but CD630 does not have *agr2R* or *agr2M*. In contrast, R20291 has the *agr2R* region. Of the total 133 clinical isolates, 131 isolates were positive for *agr1*, 61 isolates were positive for *agr2R*, and 26 isolates were positive for *agr2M*. The whole list of STs/clades, *agr* patterns and toxigenicity revealed in our study is shown in Table 3; the relationships between clades and *agr* patterns are summarized in Table 4.

Some STs showed peculiar patterns of *agr*, such as *agr1*-negative ST297 and ST402, as well as *agr2R*-positive ST198 belonging to clade 4. As most STs contained only a single pattern of *agr* loci, ST3 and ST11 were characteristic in that both *agr2R*-positive and *agr2R*-negative strains belonged to them.

Phylogenetic analysis based on STs from our clinical isolates and STs representing known clades

STs found in this study underwent MLST-based phylogenetic analysis together with ST1 (clade 2) and STs previously reported to be of clade C-I, C-II and C-III [21–26]. Phylogenetic analysis was performed using the maximum-likelihood method with the concatenated sequences of seven housekeeper genes, applying the MLST method. The resulting phylogenetic tree is shown in Fig. 4.

In clade 1, no branching point with a bootstrap value over 70 was shown to be a border between STs with *agr1* and STs

with *agr1*+*agr2R*, suggesting that in terms of bootstrap value, these two groups are not clearly demarcated from each other. This can also be said of clade 4, as the tree shows no branching point dividing STs with *agr1*+*agr2M*, *agr1*+*agr2R* and *agr1* patterns.

The phylogenetic tree also shows that ST297 (previously reported to be a ST of clade C-I [22]) belongs to a group of neighbours in clade C-I. As mentioned above, the tree shows that ST303 and ST407 were confirmed to have *agr1* and also belong to clade C-I. 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.

DISCUSSION

This is the first study focusing on a phylogenetic analysis of the *agr* locus in clinical isolates of *C. difficile*. Based on our finding that two relatively different subtypes of *agr* loci exists in *agr2*, the epidemiology of *agr* in clinical isolates of *C. difficile* was elucidated. Previous studies on *C. difficile* *agr* mainly focused on regulatory effects of *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 *agr* loci of *C. difficile*. Potential new functions of *agr2M* 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 *agr1* and the distribution of *agr2R* among clinical and laboratory *C. difficile* isolates [13, 27] but also revealed the distribution of *agr2M*. The discrepancy of *agr* patterns within a single ST observed in some STs may provide clues to help uncover the evolutionary pathway of *C. difficile*.

A previous study by Darkoh *et al.* [9] showed that examined laboratory strains of *C. difficile* could live without the *agr1* locus. The presence of *agr1*-negative isolates in our study also suggests that loss of *agr1* in *C. difficile* is not lethal, and the same thing can be said about *agr2R* and *agr2M*.

The subject of our phylogenetic analysis does not represent the whole population of *C. difficile*, but still, several

Table 3. Sequence types, MLST-based clade category, toxigenicity (*tcdA* or *tcdB-positivity*), and *agr* status of isolated *C. difficile*

Table 3. Continued

Sequence types	Clade	No. of samples	*agr1*	*agr2R*	*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	+	+ (n=4) − (n=1)	−	+ (n=4) − (n=1)
ST37	4	4	+	−	−	+
ST100	1	4	+	+	−	−
ST26	1	3	+	−	−	−
ST35	1	3	+	+	−	+
ST53	1	3	+	+	−	+
ST11	5	2	+	+ (n=1) − (n=1)	−	+
ST14	1	2	+	−	−	+
ST48	1	2	+	+	−	+ (n=1) − (n=1)
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	+	+	−	+
ST201	3	1	+	−	−	+
ST205	1	1	+	+	−	−
ST223	2	1	+	+	−	+
ST243	4	1	+	−	−	−
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	+	−	−	+

notable findings about *agr* were obtained. First, a deviation of *agr* patterns in some local parts of the phylogenetic tree/ clades was observed. For example, almost all *agr2R*-positive strains were in clade 1 or 2, and all STs with *agr2M* were in clade 4. Consistent with these results, a previous phylo-genetic study of *C. difficile* suggests that *agr2M* is mainly found in clade 4 and rarely in clade 1 [28]. Interestingly, the subclade-like group in clade 4, which the *agr2R*-positive ST198 belongs to, includes two *agr2R*-negative STs without *agr2M*, whereas another subclade in clade 4, including ST37 and ST81, does not have any STs with *agr2R*.

The fact that single *agr1*-negative isolates were found in both clade C-I and clade C-III suggests that loss of *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 *C. difficile* clades is expected to provide further insight into bacteriological evolution and taxonomy, and analysis of *agr* function may play a key role in moving the field forward.

Table 4. Relationship between *agr* patterns and clades

Clade	1	2	3	4	5	C-I	C-III	total
agr1	32	0	2	5	1	3	0	44
agr1 +agr2R	57	3	0	1	1	0	0	61
agr1 +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
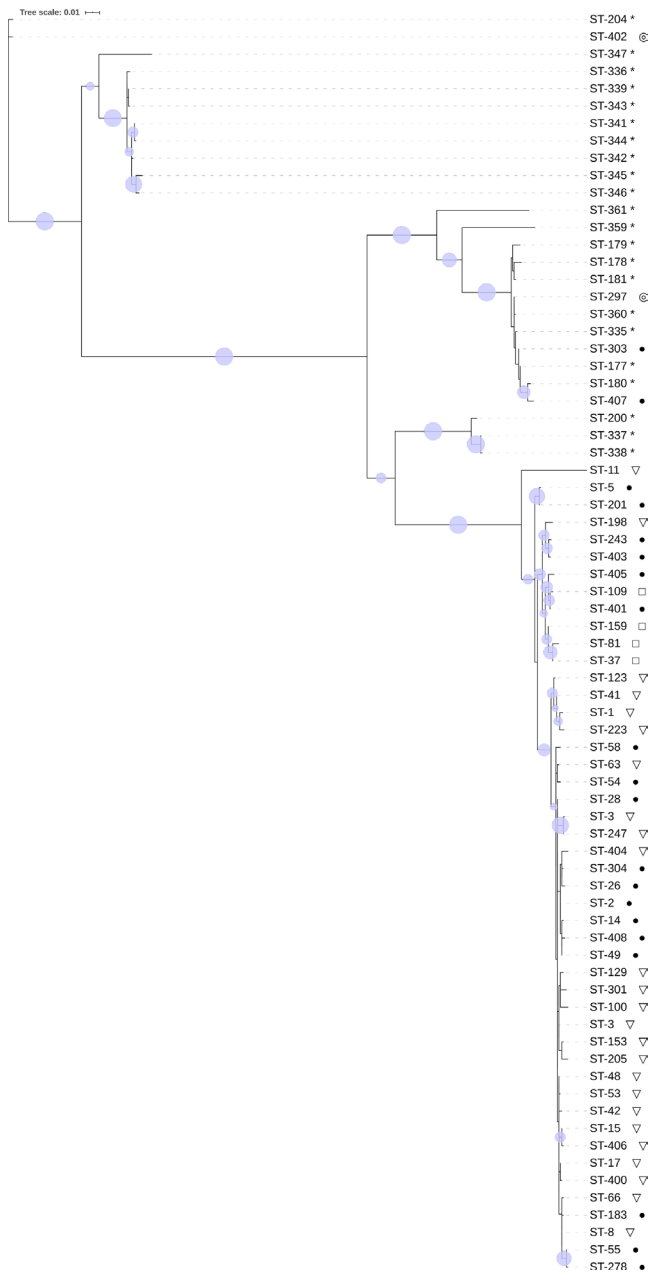
Fig. 4. Maximum-likelihood phylogenetic tree (model TN93, 1000 replicates) based on analysis by PhyML. STs found in our clinical samples; ST1 (corresponding to R20291 strain) and STs used in previous studies as references for clade C-I, C-II and C-III underwent this analysis. Circles are drawn on branches with bootstrap value over 50. ●: *agr1*; ▽: *agr1 +22R*; □: *agr1 +22M* ; ◎: no *agr* locus; *: reference strains.

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.

References

1. **Kuehne SA**, **Cartman ST**, **Heap 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 Enterococcus 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**, **Kuijper 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*. Epidemiology and virulence-associated genes of Clostridioides 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 KEP. 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 PJA, Chapman BA. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics* 2012;13:209.

18. Cock PJA, 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. Ramírez-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, Plorin 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.