Genetic markers for *Clostridium difficile* lineages linked to hypervirulence

Cornelis W. Knetsch,1 Marjolein P. M. Hensgens,1 Céline Harmanus,1 Madelon W. van der Bijl,2 Paul H. M. Savelkoul,2 Ed J. Kuijper,1 Jeroen Corver1 and Hans C. van Leeuwen1

1Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands
2Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlands

Rapid identification of hypervirulent *Clostridium difficile* strains is essential for preventing their spread. Recent completion of several full-length *C. difficile* genomes provided an excellent opportunity to identify potentially unique genes that characterize hypervirulent strains. Based on sequence comparisons between *C. difficile* strains we describe two gene insertions into the genome of hypervirulent PCR ribotypes 078 and 027. Analysis of these regions, of 1.7 and 4.2 kb, respectively, revealed that they contain several interesting ORFs. The 078 region is inserted intergenically and introduces an enzyme that is involved in the biosynthesis of several antibiotics. The 027 insert disrupts the thymidylate synthetase (*thyX*) gene and replaces it with an equivalent, catalytically more efficient, *thyA* gene. Both gene insertions were used to develop ribotype-specific PCRs, which were validated by screening a large strain collection consisting of 68 different PCR ribotypes supplemented with diverse 078 and 027 strains derived from different geographical locations and individual outbreaks. The genetic markers were stably present in the hypervirulent PCR ribotypes 078 and 027, but were also found in several other PCR ribotypes. Comparative analysis of amplified fragment length polymorphisms, PCR ribotype banding patterns and toxin profiles showed that all PCR ribotypes sharing the same insert from phylogenetically coherent clusters. The identified loci are unique to these clusters, to which the hypervirulent ribotypes 078 and 027 belong. This provides valuable information on strains belonging to two distinct lineages within *C. difficile* that are highly related to hypervirulent strains.

INTRODUCTION

*Clostridium difficile* is a spore-producing anaerobic bacillus that resides in the gut, and has rapidly emerged as a leading cause of antibiotic-associated diarrhoeal disease in humans. Antibiotic treatments that suppress growth of the intestinal microbial flora, but spare *C. difficile*, allow this pathogen to colonize the gut. Upon colonization, *C. difficile* produces toxins that cause intestinal damage and inflammation, and consequently clinical disease (Kelly et al., 1994). Persistence of *C. difficile* in hospitals is enabled by humans shedding highly resistant and infectious spores (Rupnik et al., 2009). In the past decade, the incidence, complications and mortality of *C. difficile*-associated infections (CDIs) have increased dramatically due to the emergence of the hypervirulent *C. difficile* strains PCR ribotypes 078 (NAP7/8) and 027 (NAP01) (Goorhuis et al., 2008; McDonald et al., 2005; Pépin et al., 2004, 2005). We define hypervirulent strains as those that are associated with more severe clinical disease. In addition to the production of toxin A (TcdA) and toxin B (TcdB), the presence of binary toxin (CdtA and CdtB), altered antimicrobial resistance patterns and increased sporulation capacity can contribute to the virulence of certain *C. difficile* strains (Akerlund et al., 2008; Lyras et al., 2009; Merrigan et al., 2010; Schwan et al., 2009; Voth & Ballard, 2005).

The virulence of *C. difficile* strains can be of importance when controlling infections. Although common infection control measures, such as good hygiene, cleaning of the patient environment and restricting antimicrobial therapy, are important in all cases of CDI, there are additional measures necessary to prevent the spread of certain PCR ribotypes; for example, restricting the use of fluoroquinolones to control CDI caused by the hypervirulent PCR ribotype 027 (Kallen et al., 2009; Warny et al., 2005). Therefore, it is important to identify which *C. difficile* strain is causing the infection at an early stage.

Abbreviations: AFLP, amplified fragment length polymorphism; CDI, *C. difficile*-associated infection; DHFR, dihydrofolate reductase; MLST, multilocus sequence typing.
C. difficile is a heterogeneous species, which can be typed based on various methods, including PCR ribotyping, amplified fragment length polymorphism (AFLP), multi-locus sequence typing (MLST), multiple-locus variable number tandem repeat analysis (MLVA), toxotyping, serotyping, repeat analysis and DNA microarrays (Bakker et al., 2010; Brazier, 2001; Griffiths et al., 2010; Killgore et al., 2008; Rupnik et al., 2001; Stabler et al., 2006). Currently, the most frequently used typing method in Europe is PCR ribotyping, which exploits the variability of the intergenic spacer region between the 16S and 23S rDNA. This variability, in combination with multiple copies of the rDNA, results in different amplions after PCR amplification, which are referred to as PCR ribotypes (Bidet et al., 1999; Rupnik et al., 2009). Recently, PCR ribotyping was adapted to enable direct typing of C. difficile from stool samples (Janezic et al., 2011). Besides identification of the hypervirulent PCR ribotypes 078 and 027, this method is currently capable of identifying over 300 different PCR ribotypes. PCR ribotyping depends on comparison of agarose/acylamide gel patterns that requires standardization, a computer database containing each ribotype-specific pattern and computer-assisted analysis for correct interpretation. Occasionally, expertise is needed to interpret the generated PCR ribotype banding patterns. These requirements often limit interlaboratory reproducibility. Therefore, it would be very useful to identify genetic markers that can be used to develop a simple, robust and quick diagnostic method by which hypervirulent C. difficile strains are typed.

Comparative genome analysis has shown that C. difficile genomes are highly variable, which contributes to the distinct phenotypic differences (He et al., 2010; Sebaihia et al., 2006). The epidemic C. difficile 027 genome has 50 regions of genetic difference with a total of 234 additional genes compared with PCR ribotype 012 (Stabler et al., 2009). These variable regions can provide genetic markers for the identification of specific strains. A suitable marker should be stably integrated into the genome. However, many of the above-mentioned regions are not stably integrated, as they encode conjugative transposons or antibiotic resistance, or are sequences of phage origin that are involved in recombination, reshuffling or horizontal gene transfer (Sebaihia et al., 2006). These events make identification of stable strain-specific genetic markers challenging.

In this study we identified genetic markers that are stably present in the hypervirulent PCR ribotypes 078 and 027 and in closely related PCR ribotypes. Strains containing these markers represent specific lineages of C. difficile that are linked to hypervirulence.

METHODS

**In silico comparison of full-length genomes.** Whole genomes of PCR ribotypes 078 and 027 were compared with the PCR ribotype 012 genome using the National Center for Biotechnology Information (NCBI) BLAST application. The genomes of interest (PCR ribotypes 078 and 027) were translated by BLASTX and compared with a C. difficile protein database. This protein database, containing gene-identifier information for each ribotype together with corresponding protein sequences, was created by using the Choose Search Set functionality of BLASTX. By selecting C. difficile NAP08 (taxid: 525259) and C. difficile R20291 (taxid: 645463) the BLAST search was restricted to the sequences in the database from the organism that we selected. To discard genes that were abundantly present in the genome, we entered the following query into the Entrez query syntax: (NOT ribosomal[text] NOT tRNA[text] NOT tetracycline[definition] NOT transposon[text] NOT phage[text] NOT recombinase[text] NOT translocase[notext] NOT conjugation[notext]). A cut-off value of 1e-06 (E-value) was used to indicate a positive hit. Whenever a hit in protein sequence was found, the corresponding gene identifier was maintained. Next, we translated the genome of C. difficile 630 (PCR ribotype 012) with BLASTX using C. difficile NAP08 (taxid: 525259) or C. difficile R20291 (taxid: 645463) annotation. This was done to circumvent the problem of differences in gene annotation between C. difficile strains. This resulted in a second set of gene identifiers with annotation similar to that used for generating the first set. The two sets were compared and differences were selected as potential markers for the specific strains. An additional manual screen was performed on the markers found in order to discard genes that were not excluded by the initial query but still were likely to transfer to other genomes, for example any antibiotic-resistance gene. Then, we analysed the genomic region of the potential markers for the presence of genes associated with mobile elements. The presence of such genes in the vicinity of markers could indicate that they are not stably present and therefore potential markers associated with such genes were discarded. The remaining markers were screened for their uniqueness by performing a BLASTN and a BLASTP homology search of the selected region against all other sequenced C. difficile genomes (630 a.m.;180355.1; CD196-FN538970.1; R20291-FN543816.1; 200785-FN665654.1; B11-FN668941.1; M120-FN665653.1; B19-FN668944.1; M68-FN668975.1; CFS5-FN665652.1). Markers with no significant homology other than to the selected strain were considered the most promising. Finally, genome views were produced using the xBASE genome database (Chaudhuri et al., 2008).

**PCR amplification.** Purification of genomic DNA was performed by using the QIAamp DNA Blood Mini kit (Qiagen) according to the manufacturer’s protocol. Part of the markers was amplified by PCR using 20 pmol forward and reverse primers (Table 1), 1 U GoTaq DNA polymerase (Promega) in 25 μl 1 × Green GoTaq Flexi Buffer PCR buffer (Promega), 0.2 mM dNTPs and 3 mM MgCl₂. The amplification was performed at 94 °C for 2 min, 30 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, followed by a final extension at 72 °C for 10 min. The PCR products were visualized by electrophoresis on an agarose gel (1.5 %) containing ethidium bromide.

**Validation of PCR ribotype 078- and 027-specific markers.** The presence of the specific markers was validated by screening 36 clinical isolates of ribotype 078 and 24 of ribotype 027 originating from various geographical locations and times. In addition, a library of 68 unique representative PCR ribotypes (Leeds–Leiden collection), including the 20 most frequently found PCR ribotypes in The Netherlands, was screened for the presence of the genetic markers. This library was validated (i.e. confirmation of the PCR ribotype) by several specialized laboratories in Europe and Canada. If other PCR ribotypes were found to be positive for the insert, additional strains from these PCR ribotypes were screened. PCR products obtained from several PCR ribotypes were sequenced to confirm amplification of the intended regions.
Identification of inserts specific for *C. difficile* PCR ribotypes 078 and 027

Whole genome sequencing of various PCR ribotypes revealed 4.1–4.3 Mb chromosomes with a mosaic of mobile elements (He *et al.*, 2010). Matching up the genome of reference strain 630 (PCR ribotype 012) with PCR ribotype 078 and 027 strains (see Methods) revealed several potentially distinctive markers, including (parts of) transposable elements (Sebaihia *et al.*, 2006), sequences of phage origin (defined by *blast* hits in clostridium phages) and CRISPRs (clustered regularly interspaced short palindromic repeats)-associated proteins (annotated in 078 strain *C. difficile* QCD-23m63). The above-mentioned markers and several other inserts encoding genes involved in recombination were discarded. Of the remaining markers, we selected the regions with no significant homology other than to the strain of interest (PCR ribotype 078 and 027 strains) as the most promising markers. As a result, we identified two potentially unique inserts, one present in PCR ribotype 078 and the other present in PCR ribotype 027. Fig. 1(a) shows the result of a *blast* nucleotide database search of the selected 078 and 027 regions, including flanking regions. The flanking regions show good alignment with all full-length sequenced *C. difficile* strains, whereas the unique inserts are only found in the corresponding ribotype, 078 (M120 strain) and 027 (R20291;B11;CD196;2007855 strains), respectively.

Characterization of PCR ribotype 078- and 027-specific inserts

The region selected for PCR ribotype 078 is 1675 bp long and contains three ORFs (see Fig. 1b, Table 2). The insertion of the 078 locus is precisely in the intergenic region between *guaA* (GMP synthase) and a putative nucleotidase (CD0198 and CD0199, respectively; annotation according to reference strain 630). Homology searches in the NCBI protein database revealed that the first predicted ORF of the 078 insert, ZP_05399691, has high sequence identity to proteins encoded by a Gram-negative sulphate-reducing bacterium, *Desulfovibrio fructosovorans* (E-value 3e-37, and data not shown). Profile analysis of protein sequences (Söding, 2005) showed that ZP_05399691 belongs to a family of antibiotic biosynthesis monooxygenases (E-value 0.0019; PFAM03992) involved in the biosynthesis of several antibiotics (Sciara *et al.*, 2003). Profile analysis of the predicted ORF ZP_05399692 did not produce significant

---

### Table 1. Primers used for the PCRs developed and for AFLP

<table>
<thead>
<tr>
<th>Primer</th>
<th>Product size (nt)</th>
<th>Target(s)</th>
<th>Sequence (5’→3’)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>078_insert</td>
<td>707</td>
<td>ZP_05399691, ZP_05399692, ZP_05399693</td>
<td>Fw, AAAAGAACAGCTAGGGAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rv, TGGGACTGGAATTAGG</td>
</tr>
<tr>
<td>027_insert</td>
<td>691</td>
<td>CDR20291_0045</td>
<td>Fw, TGGCTCGAGAGTAGTTCTGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rv, TGGGCACTGTACGATAAG</td>
</tr>
<tr>
<td><strong>AFLP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pst-0</td>
<td>–</td>
<td>PstI adaptor</td>
<td>GACTGCGTACATGCAG</td>
</tr>
<tr>
<td>Mse-C</td>
<td>–</td>
<td>MseI adaptor</td>
<td>GATGAGTCCTGAGTAAC</td>
</tr>
</tbody>
</table>

*Fw, forward; Rv, reverse.*
hits (data not shown), whereas BLASTP analysis of the ORF ZP_05399693 resulted in a GCN5-acetyltransferase as the closest homologue. It should be noted that an isolated transposase-like protein B is situated 1.4 kb downstream of the 078 insert, and is not part of the insert itself.

The PCR ribotype 027 insert is 4156 bp long and contains four ORFs (see Fig. 1b, Table 2). Unlike the 078 insert, the 027 insert disrupts the thyX gene (flavin-dependent thymidylate synthetase), in which it has inserted (see Fig. 1b). The first two ORFs (CDR20291_0043 and CDR20291_0044) of

**Fig. 1.** In silico identification of unique inserts for *C. difficile* PCR ribotypes 078 and 027. (a) Result of BLASTN homology search of region 249 237–250 912 of *C. difficile* ribotype 078 strain M120 (GenBank/EMBL/DDBJ accession no. FN665653.1) and region 80 832–84 988 of *C. difficile* ribotype 027 strain R20291 (accession no. FN545816.1) including the 300–500 nt flanking region. The resulting graphic summary is shown. (b) Pairwise comparisons between PCR ribotype 012 (630) and 078 (top) and between 012 and 027 (bottom). Shown are diagrammatic representations of the genes surrounding the inserts using the web-based database xBASE (colour based on percentage DNA G+C content). The blue dashed box indicates the inserted region and arrows indicate the PCR primers used for amplification; bar, 1 kb size marker.

C. W. Knetsch and others

---

**Microbiology**

157
the 027 insert encode the enzymes thymidylate synthetase A (ThyA) and dihydrofolate reductase (DHFR), both of which are involved in the biosynthesis of the DNA base thymine (dTMP). The third ORF (CDR20291_0045) has homology to the Sir2 family of protein deacetylases (E-value 2.2e−26). The last ORF encodes the thiamine biosynthesis protein ThiC (CDR20291_0046).

The insertion of the 027 region into the gene encoding ThyX has resulted in a non-functional thyX gene, which is subjected to accumulation of mutations. The relative age of this non-functional gene is reflected in its number of such mutations. Compared with PCR ribotype 012, containing a functional thyX, 38 mutations have occurred in this gene (540 nt). With an estimated synonymous mutation rate of 0.0002 per genome/per generation in C. difficile (note: including functional ORFs) (He et al., 2010), it is clear that the 027 insert occurred early in the Clostridium lineage and has been present ever since.

Validation of PCR ribotype 078- and 027-specific inserts

To confirm the uniqueness of the 078/027-specific inserts, we developed a PCR with a primer pair amplifying part of the insert (see Methods). Sixty-eight representative strains from a validated PCR ribotype library (collaboration of University of Leeds and University of Leiden) were analysed for the presence of the 078- and 027-specific inserts. Surprisingly, although 63 tested negative for the presence of the 078-specific region, four other PCR ribotypes showed a PCR amplification product (see Fig. 2a). Positive for the 078 insert were PCR ribotypes 033, 045, 066, 078 and 126. To confirm these results we used the database of the reference laboratory of the Leiden University Medical Center for C. difficile to select additional isolates for these respective ribotypes obtained from various geographical locations and outbreaks. Seven PCR ribotype 033, 11 PCR ribotype 045, one PCR ribotype 066, 32 PCR ribotype 078 and nine PCR ribotype 126 strains were tested. In addition, we included a strain belonging to PCR ribotype 127 in our analysis, as this ribotype has a banding pattern similar to that of PCR ribotype 078 (see Fig. 4, below). All these isolates, including several animal strains, tested positive for the presence of the 078 marker (see Fig. 2c, and data not shown).

We also tested our PCR ribotype library for the presence of ribotypes containing the 027 region. As shown in Fig. 2(b), besides PCR ribotype 027, PCR ribotypes 016, 036, 075, 122 and 153 contained the 027-specific region, as evidenced by the presence of a PCR amplicon. The presence of the 027 region in these PCR ribotypes was confirmed by testing additional isolates. Four PCR ribotype 016, three PCR ribotype 036, four PCR ribotype 075, four PCR ribotype 153, one PCR ribotype 122 and 24 PCR ribotype 027 strains were tested. In addition, we supplemented our collection with strains belonging to PCR ribotypes 019, 111 and 156, as two recent studies have indicated that these types are closely related to PCR ribotype 027 (Griffiths et al., 2010; Zaiss et al., 2009). Furthermore, we included three strains belonging to PCR ribotypes 176, 208 and 273, which have PCR ribotype patterns similar to that of PCR ribotype 027 (see Fig. 4 below). All these strains tested positive in our PCR for the presence of the 027 marker (see Fig. 2d, and data not shown). In summary, after screening a large collection of different PCR ribotypes, we found that the 078 and 027 inserts are also present in several other PCR ribotypes.

Comparative analyses of strains containing the genetic marker

The presence of the 078 and 027 markers in several PCR ribotypes suggests that a common ancestor had these respective inserts. Potential relationships among these strains were examined through cluster analysis (see Methods) of AFLP patterns and PCR ribotype banding patterns. The AFLP patterns generated are presented in a dendrogram (Fig. 3). Strains positive for the 078 and 027 markers had similar profiles and evidently clustered together into distinct groups with a relative similarity of approximately 85% (Fig. 3, blue for the 027 group, green for the 078 group), corresponding to the recommended genetic marker.
cut-off value used for the identification of identical strains (Killgore et al., 2008). As for the AFLP patterns, we presented the generated ribotype banding patterns in a dendrogram (Fig. 4). Strains positive for the 078 marker clustered together in a distinct group (relative similarity >85%). Strains positive for the 027 marker also clustered together in a single group, with only PCR ribotype 156 showing a divergent pattern. Thus, our markers for 078 and 027 recognize PCR ribotypes that are clustered in distinct and genetically related groups, supporting the use of these inserts to identify these ‘078-like’ and ‘027-like’ types.

To further study potential relationships among the 078-like and 027-like strains, we analysed the toxin profiles (Table 3). All strains in the 078-like group, with the exception of one strain belonging to PCR ribotype 033, were positive for toxin A (tcdA), toxin B (tcdB) and binary toxin (cdtA). The
PCR ribotype 033 strain was negative for \textit{tcdB}. Others have shown that PCR ribotype 033 can also be positive for \textit{tcdB} (Pituch et al., 2006). In addition, all strains in the 078-like group possessed the 39 bp \textit{tcdC} deletion, which is a distinct marker for PCR ribotype 078 (Goorhuis et al., 2008), supporting the notion that these strains are closely related. All strains in the 027-like group were positive for \textit{tcdA}, \textit{tcdB} and \textit{cdtA}, with the exception of a PCR ribotype 208 strain, which was negative for \textit{cdtA}. Furthermore, we observed heterogeneity for the presence of the 18 bp \textit{tcdC} deletion within the 027-like group, as reported by another group (Janvilisri et al., 2009).

Next, we analysed the prevalence of PCR ribotypes that were positive for the 078/027 insert. Hypervirulent strains 078 and 027 comprise 10 and 11 \% of the total number of samples (\(n=4383\)) submitted to the Dutch National Reference Laboratory, respectively (Fig. 4). All other types tested were less frequently found (<1.3 \%). Although 078-like and 027-like strains are only found sporadically, we analysed whether these strains were as hypervirulent as their counterparts, based on the number of deaths after 90 days. Mortality in patients with CDI due to PCR ribotype 078 and 078-like was 26 \% (28/104) and 27 \% (7/26), respectively, after 90 days, higher than in patients who suffered from CDI due to other ribotypes (20 \%, 166/812). This difference, however, was not statistically significant (\(P=0.13\) and \(P=0.46\), respectively). Mortality in patients with CDI due to type 027 and 027-like was 34 \% (24/70) and 29 \% (4/14), respectively, after 90 days. Mortality in patients with CDI due to PCR ribotype 027 was significantly higher than that in patients with CDI due to other types (\(P=0.01\)). Patients with CDI due to ribotype 027-like also showed higher mortality than patients with CDI due to other types, although the difference was not significant (\(P=0.50\)).

\textbf{DISCUSSION}

Gene transfer between different species allows bacteria to acquire new traits, potentially conferring evolutionary adaptations to the selective conditions encountered. We have exploited these genes, aiming to develop specific PCRs for identification of the two \textit{C. difficile} hypervirulent PCR ribotypes 078 and 027. We have identified genetic markers for these hypervirulent PCR ribotypes and closely related PCR ribotypes. The 078 and 027 insertions are not part of a module involved in conjugation or excision (Sebaihia et al., 2006) and are probably stably integrated, as is exemplified by their presence in 078/027 strains collected from different outbreaks throughout the world (14 European countries and Australia) over a period of 4 years (between 2006 and 2010).
Comparative analyses show that the PCR ribotypes with the 078 insert (033, 045, 066, 078, 126 and 127) form a phylogenetically coherent group, which is reflected by similarities in AFLP patterns, similarities in ribotype patterns, presence of the binary toxin gene cdta and the presence of the characteristic tcdC deletion (39 bp). Similarly, the PCR ribotypes with the 027 insert (016, 019, 027, 036, 075, 111, 122, 153, 156, 176, 208 and 273) are also closely related to each other, which is confirmed by clustering of the AFLP patterns into a distinct group, similarities in ribotype patterns and the presence of cdta. PCR ribotypes 078 and 027 are the most prevalent types in their cluster, while the 078-like and 027-like strains are found in low numbers among humans (Fig. 4).

Clearly, the presence of a unique insert within seemingly diverse PCR ribotypes suggests that they share a common ancestor. Complementary to our comparative analysis, sequence analysis of the 16S rRNA genes of ribotypes 033, 045, 066 and 078 has shown that they all contain a unique base change compared with other strains, confirming that they represent a well-defined subpopulation within the C. difficile species (Rupnik et al., 2001). In the past few years it has become clear that C. difficile has also emerged as an animal pathogen. The PCR ribotype most prominently present in farm animals is 078, accounting for more than 90 % of porcine and bovine isolates (Keel et al., 2007; Weese et al., 2010). Overlap between 078 isolates from animals and humans led to the hypothesis of a zoonotic transmission route (Bakker et al., 2010), but this awaits confirmation. Other ribotypes found in livestock are 002, 033, 045, 066, 077 and 126 (Avbersek et al., 2009; Keel et al., 2007; Pirs et al., 2008). Except for 002 and 077, all these ribotypes tested positive in our 078 PCR, representing more than 95 % of the species found in livestock. This indicates that the 078 marker can be used for typing 078 and 078-like strains.

Our comparative analysis shows that the PCR ribotypes with the 027 marker are closely related to each other. Clustering of the PCR ribotype group 027 has been partially confirmed by two other research groups (see Table 3). Griffiths et al. (2010) identified a phylogenetically coherent group of C. difficile strains using MLST. This group, which was marked as hypervirulent, consists of PCR ribotypes 027, 019, 036 and 153. In addition, Zaiss et al.
(2009) performed typing of C. difficile based on tandem repeat sequences. In their study, PCR ribotype 027 clustered together with ribotypes 019, 111 and 156 into a distinct group. These data, which correlate with our own, confirm that the 027 marker identified in this study can be used to recognize PCR ribotype 027 and 027-like strains. Finally, the 027 genetic marker is noted in a microarray study by Marsden et al. (2010), where it was reported that this insert of four genes is present in PCR ribotypes 001, 078 and 027. Our results and in silico genome comparisons (see Figs 1a and 2) clearly show that the insert is present in PCR ribotype 027 but not in PCR ribotypes 001, 078 or 027.

Rapid and reliable identification of C. difficile types is necessary for controlling CDI. A novel typing method based on the described markers has some limitations when compared with other typing methods such as PCR ribotyping. The presence of the markers in strains other than the hypervirulent PCR ribotypes 078 and 027 reduces its discriminatory power. Nevertheless, the markers can be useful in a rapid pre-screen for detecting potentially hypervirulent strains, especially when considering that the hypervirulent strains are the most prevalent strains of their lineages and that the other strains are much less frequently found. The result of the newly proposed pre-screen is easy to interpret, i.e. marker present or not present, and therefore no expertise or software and databases are required to perform it. This makes it suitable for use in clinical laboratories worldwide. In addition, use of the markers can provide the user with additional information on relatedness between strains, as it is capable of identifying hypervirulent PCR ribotypes 078 and 027 and closely related ribotypes with possibly similar virulence potential. This additional information may be of interest when taking into account how new PCR ribotypes can arise. New PCR ribotypes are assigned when a ribotype pattern is found which differs by at least one band from already defined types. Recently, it has been shown that mechanisms such as homologous recombination (Dingle et al., 2011) or slipped-strand mispairing during DNA replication can be responsible for variation in the intergenic spacer region (Indra et al., 2010). Such events make it plausible that newly identified PCR ribotypes will arise from hypervirulent C. difficile strains, and these may share virulence potential and clinical manifestations.

Analysis of the genes that are located on the 078 and 027 inserts could elucidate the potential impact of the insert on the phenotype. One of the ORFs present on the 078 insert belongs to a family of genes that are involved in biosynthesis of several antibiotics, which might confer a competitive advantage upon PCR ribotype 078.

Insertion of the 027 marker in thyX causes disruption of this gene (see Fig. 1b). At first this seems puzzling, as thyX is an essential gene required for biosynthesis of dTMP. Its important role in microbial DNA replication and presence in several pathogenic bacteria makes it an attractive potential target for antibiotic drug development (Esra Önen et al., 2008; Koehn et al., 2009). Remarkably, the

<table>
<thead>
<tr>
<th>PCR ribotype</th>
<th>AFPL clustering (Fig. 3)</th>
<th>Ribotype clustering (Fig. 4)</th>
<th>TcdA/TcdB/CdtA</th>
<th>Grouped with 078/027 (Griffiths et al., 2010) (MLST)</th>
<th>(Zaiss et al., 2009) (tandem repeat sequences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribotype 078</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>033</td>
<td>Yes</td>
<td>Yes</td>
<td>+/−/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>045</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>066</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>078</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>126</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>127</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>Ribotype 027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>016</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>No</td>
</tr>
<tr>
<td>019</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>027</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>036</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>No</td>
</tr>
<tr>
<td>075</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>111</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>122</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>153</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>156</td>
<td>Yes</td>
<td>No</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Yes</td>
</tr>
<tr>
<td>176</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>208</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/−</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>273</td>
<td>Yes</td>
<td>Yes</td>
<td>+/+/+</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
</tbody>
</table>
027 insert replaces ThyX by a different, although functionally equivalent, enzyme, thymidylate synthetase A (ThyA), together with an additional enzyme, dihydrofolate reductase (DHFR) (see Table 2). The two thymidylate synthetases use different mechanisms in the biosynthesis of dTMP (Koehn et al., 2009). In contrast to ThyX, which produces tetrahydrofolate (THF), ThyA forms dihydrofolate (DFH) and requires DHFR to produce THF (Koehn & Kohen, 2010). It has been shown that the ThyA enzyme is catalytically more efficient than the ThyX enzyme (Escartin et al., 2008), and consequently microbes that contain the highly active ThyA could have a growth advantage. Growth curve analysis with isogenic strains containing either gene (thyX or thyA together with DHFR) should clarify whether there is a difference in growth rate.

This study provides a comprehensive overview of the genetic lineages of C. difficile that are related to hypervirulent PCR ribotypes 078 and 027. PCR analysis of the genetic markers provides a rapid and easy way to type the hypervirulent PCR ribotypes 078 and 027 and closely related PCR ribotypes, irrespective of host and geographical location.

ACKNOWLEDGEMENTS
This work was supported by The Netherlands Organization for Scientific Research (NWO) (ZonMW grant 50-50800-98-079) and HYPERDIFF-The Physiological Basis of Hypervirulence in Clostridium difficile: a Prerequisite for Effective Infection Control (Health-F3-2008-223585).

REFERENCES


