"Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping."

Karjalainen, Tuomo ; Saumier, Nicolas ; Barc, Marie-Claude ; Delmée, Michel ; Collignon, Anne

ABSTRACT

Recent investigations of Clostridium difficile cell wall components have revealed the presence of an S-layer encoded by the slpA gene. The aim of this study was to determine whether slpA genotyping can be used as an alternative to serotyping. The variable regions of slpA were amplified by PCR from serogroup reference strains and various clinical isolates chosen randomly. Amplified products were analyzed after restriction enzyme digestion and DNA sequencing. The sequences of the variable region of the SlpA protein were found to be strictly identical within a given serogroup but divergent between serogroups. These preliminary results suggest that PCR-restriction fragment length polymorphism, in conjunction with DNA sequencing of the slpA variable region, could constitute an alternative typing method for determining C. difficile serotypes.

CITE THIS VERSION


Le dépôt institutionnel DIAL est destiné au dépôt et à la diffusion de documents scientifiques émanant des membres de l'UCLouvain. Toute utilisation de ce document à des fins lucratives ou commerciales est strictement interdite. L'utilisateur s'engage à respecter les droits d'auteur liés à ce document, principalement le droit à l'intégrité de l'œuvre et le droit à la paternité. La politique complète de copyright est disponible sur la page Copyright policy.

DIAL is an institutional repository for the deposit and dissemination of scientific documents from UCLouvain members. Usage of this document for profit or commercial purposes is strictly prohibited. User agrees to respect copyright about this document, mainly text integrity and source mention. Full content of copyright policy is available at Copyright policy.

Available at: http://hdl.handle.net/2078.1/27760  [Downloaded 2022/05/11 at 00:58:47 ]
**Clostridium difficile** Genotyping Based on slpA Variable Region in S-Layer Gene Sequence: an Alternative to Serotyping

Tuomo Karjalainen, Nicolas Saumier, Marie-Claude Barc, Michel Delmée, and Anne Collignon

Université de Paris-Sud, Faculté de Pharmacie, Département de Microbiologie, 92296 Châtenay-Malabry Cedex, France, and Université Catholique de Louvain, Faculté de Médecine, Unité de Microbiologie, Brussels 1200, Belgium

Received 10 January 2002/Returned for modification 16 February 2002/Accepted 13 April 2002

Recent investigations of *Clostridium difficile* cell wall components have revealed the presence of an S-layer encoded by the *slpA* gene. The aim of this study was to determine whether *slpA* genotyping can be used as an alternative to serotyping. The variable regions of *slpA* were amplified by PCR from serogroup reference strains and various clinical isolates chosen randomly. Amplified products were analyzed after restriction enzyme digestion and DNA sequencing. The sequences of the variable region of the *slpA* protein were found to be strictly identical within a given serogroup but divergent between serogroups. These preliminary results suggest that PCR-restriction fragment length polymorphism, in conjunction with DNA sequencing of the *slpA* variable region, could constitute an alternative typing method for determining *C. difficile* serotypes.

*Clostridium difficile* is the etiological agent of antibiotic-associated pseudomembranous colitis and is considered the major cause of nosocomial diarrhea. Its pathogenicity is mediated by two toxins, A and B, both of which damage the human colonic mucosa and are potent tissue-damaging enzymes (6). Confirmed and putative accessory virulence factors that could play a role in intestinal colonization have been identified, including a capsule (11), proteolytic enzymes (22, 24), and adhesins involved in mucus and cell association (15–17, 26–29, 33).

Many bacteria express a surface-exposed proteinaceous layer, termed the S-layer, that forms a regular two-dimensional array visible by electron microscopy. Previous studies have shown that a high degree of variability exists in the molecular masses of the two proteins composing the *C. difficile* S-layer (8, 10, 19). Each strain carries a higher-molecular-mass protein of 48 to 56 kDa, encoded by the C-terminal, conserved part of the *slpA* gene, and a lower-molecular-mass protein of 36 to 45 kDa, coded for by the variable N-terminal part of the gene (10, 17, 19). The lower-molecular-mass S-layer protein, referred to as the P36 protein, appears to be located on the exterior surface of the bacteria and has adhesive properties (10, 17). Interestingly, the gene encoding the S-layer precursor is present in a genetic cluster locus carrying 17 open reading frames (ORFs), 11 of which carry a similar two-domain architecture, likely to encode surface-anchored proteins (17).

Most epidemiological studies of *C. difficile* have been performed by using several typing systems. Serogrouping by slide agglutination or enzyme-linked immunosorbent assay with rabbit antisera (12) or toxin A test (Oxoid). In vitro cytotoxicity (toxin B) determination was performed by using tissue culture cells as previously described (28).

Bacteria were grown under anaerobic conditions in Tryptone-glucose-yeast infusion broth (Difco Laboratories, Detroit, Mich.) for 48 h.

Genomic DNA isolation. DNA was extracted from 10 ml of an overnight anaerobic culture of *C. difficile* in accordance with the protocol provided in the Puregene DNA gram-positive bacterium and yeast DNA extraction kit (Centra Systems, Minneapolis, Minn.; www.gentra.com).

**MATERIALS AND METHODS**

**Bacterial strains, media, and growth conditions.** Thirty-two *C. difficile* isolates belonging to 10 different serogroups (A, B, C, D, F, G, H, I, K, and X) (14, 31). In serogroup A, another 20 subgroups (subgroups A1 to A20) can be distinguished by polyacrylamide gel electrophoresis (14); these subgroups possess serogroup-specific somatic antigens but have a flagellar antigen in common that is responsible for cross-agglutination on slides (12). Recently, new molecular techniques have been developed for *C. difficile* typing (7, 18).

The aim of this work was to study the genotypic variability of the *slpA* gene encoding the outwardly exposed domain of the major *C. difficile* surface protein. Amplicons obtained by PCR from serogroup reference strains and various clinical isolates were analyzed by restriction fragment length polymorphism (RFLP) analysis and nucleotide sequencing.

PCR amplification of *slpA*. For amplification of the variable domain of the *slpA* gene (Fig. 1) from various *C. difficile* isolates, the primers used were SlpAV3 (5′-ATGAAATAAGAAAAAYWTAGACATRG-C3′) and SlpAV5 (5′-CTCTTTCTATCTTTCTCCWGCTAC-3′), where Y = CT, W = AT, and R = AG. For amplification of the entire gene, primers slpAV3 and slpAC5 (5′-AGCKATACTTTACCCWACHTTG-3′), where K = TG, were used. DNA amplification by PCR was performed in a reaction volume of 50 μl consisting of 1 μl of purified genomic DNA (1 μg/μl), 1 μl each of the 5′ and 3′ primers at 20 pmol/μl, 25 μl of water, and 25 μl of Ready-Mix Taq PCR Reaction Mix with MgCl₂ (Sigma).
<table>
<thead>
<tr>
<th>Strain</th>
<th>Serogroup</th>
<th>Origin</th>
<th>Toxins A and B</th>
<th>Size(s) (nt) of fragment(s) obtained with:</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>101</td>
<td>HpaII</td>
<td>103, 176, 556, 687, 794</td>
<td>AF448119</td>
</tr>
</tbody>
</table>

**TABLE 1.** Different boxes inside and their respective patterns.
Nucleotide sequence accession numbers.

Computer analyses. Nucleotide and protein sequence alignments were performed with the DNA CLUSTAL W program (30).

Nucleotide sequence accession numbers. The GenBank nucleotide sequence accession numbers of the slpA variable region from the C. difficile isolates studied are given in Table 1.

RESULTS

PCR amplification and sequence analysis of the slpA variable region from serogroup reference strains. It has been previously demonstrated that the C. difficile slpA gene carries a conserved 3' half and a variable 5' half in the strains studied (Fig. 1) (17). The gene encodes a larger precursor protein that is probably cleaved into distinct S-layer proteins. The protein corresponding to the N-terminal half (referred to as P36) is exposed to the external milieu and thus could play a role in determining serogroup specificity. To investigate this possibility, the variable region of the slpA gene from 10 serogroup reference strains was amplified by PCR with the primers indicated in Fig. 1 (Table 1). A PCR product of approximately 1 kb was obtained from all of the strains (Table 1).

The PCR product was digested with the restriction enzymes DraI, HinII, PvuII, and RsaI. Each serogroup revealed a different RFLP profile after electrophoresis in an agarose gel of the restriction products (not shown), suggesting that the slpA variable-region sequence is different in each serogroup. For this reason, the slpA gene variable region from the reference strains was sequenced. The length of the variable region varied from 1,017 to 1,185 nucleotides (339 to 395 amino acids) (Table 1). As shown in Fig. 2, alignment of the amino acid sequences of the variable region revealed that each sequence was unique, with interserogroup sequence homology ranging from 11% (between serogroups A and I) to 64% (between serogroups C and F) (Table 2).

The presence of nucleotide sequence differences in the slpA genes of C. difficile reference strains allowed the establishment of a precise computer-generated RFLP profile for each serogroup. As shown in Fig. 3 and Table 1, with the exception of serogroup A, unique restriction patterns were produced for each serogroup by digestion with the restriction enzymes HinII and RsaI. PvuII digestion could not distinguish serogroups B, D, and G, and DraI digestion could not differentiate serogroups F and I.

Sequence analysis of the slpA gene variable region from C. difficile clinical isolates. The unique slpA sequence and RFLP patterns of each serogroup suggested that nucleotide sequence analysis and/or PCR-RFLP of the variable region of this gene could be employed to differentiate between serogroups. To test this hypothesis, we sequenced the variable region of the slpA genes from 22 clinical isolates whose serogroups were known (at least two sequences per serogroup) and which were chosen fortuitously and represented diverse geographic locations (Table 1). Analysis of the sequences obtained revealed that, with the exception of serogroup A, the nucleotide and deduced protein sequences were 100% identical within a given serogroup (Fig. 2). In contrast, the four strains belonging to serogroup A each had a unique slpA variable region, although the SlpA N-terminal regions in A1 and A10 were 93% identical.

DISCUSSION

Different methods have been developed to type C. difficile strains. A correlation has been observed between the electrophoretic patterns of proteins extracted from clinical isolates.
and their virulence, strains isolated during hospital outbreaks and symptomatic disease having a different profile from that observed in isolates from healthy carriers or infants (21). Delmée et al. (14) compared serogrouping of *C. difficile* by slide agglutination with rabbit antisera and polyacrylamide gel electrophoresis of whole-cell proteins, permitting correlation between the two typing systems and establishment of a single classification. This serogrouping is also carried out by the enzyme-linked immunosorbent assay technique.

Recently, new molecular techniques have been developed to type *C. difficile* strains based on DNA polymorphism, such as arbitrary primed PCR, a genotypic method permitting detection of polymorphisms within the target genome without prior knowledge of the target nucleotide sequence. This method has been used as an efficient discriminative method for investigation of nosocomial outbreaks of *C. difficile*-associated diarrhea but has certain drawbacks, e.g., lack of reproducibility (1, 2, 5,

<table>
<thead>
<tr>
<th>Serogroup</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>28</td>
</tr>
<tr>
<td>D</td>
<td>14</td>
</tr>
<tr>
<td>E</td>
<td>28</td>
</tr>
<tr>
<td>F</td>
<td>28</td>
</tr>
<tr>
<td>G</td>
<td>35</td>
</tr>
<tr>
<td>H</td>
<td>25</td>
</tr>
<tr>
<td>I</td>
<td>11</td>
</tr>
<tr>
<td>K</td>
<td>29</td>
</tr>
<tr>
<td>X</td>
<td>31</td>
</tr>
</tbody>
</table>

and their virulence, strains isolated during hospital outbreaks and symptomatic disease having a different profile from that observed in isolates from healthy carriers or infants (21). Delmée et al. (14) compared serogrouping of *C. difficile* by slide agglutination with rabbit antisera and polyacrylamide gel electrophoresis of whole-cell proteins, permitting correlation between the two typing systems and establishment of a single classification. This serogrouping is also carried out by the enzyme-linked immunosorbent assay technique.

Recently, new molecular techniques have been developed to type *C. difficile* strains based on DNA polymorphism, such as arbitrary primed PCR, a genotypic method permitting detection of polymorphisms within the target genome without prior knowledge of the target nucleotide sequence. This method has been used as an efficient discriminative method for investigation of nosocomial outbreaks of *C. difficile*-associated diarrhea but has certain drawbacks, e.g., lack of reproducibility (1, 2, 5,
Pulsed-field gel electrophoresis has been applied to *C. difficile*, but this method is costly and technically complex and some strains are untypeable (7). PCR ribotyping is being used routinely by the United Kingdom Anaerobe Reference Unit in Cardiff, Wales, to type *C. difficile* strains (7). This method relies on DNA pattern profiles, obtained by PCR amplification of a specific chromosomal region such as the rRNA gene (4) or the 16S-23S rRNA gene intergenic spacer region (25). Each of the strains belonging to one of Delmé’s serogroups gives a different banding pattern. PCR ribotyping appears to be more discriminatory than the arbitrary primed PCR or pulsed-field gel electrophoresis method in epidemiological studies of *C. difficile* outbreaks (7). However, most of the aforementioned typing techniques are discriminating but labor intensive. Serotyping with specific antisera gives adequate levels of discrimination for epidemiological purposes and is not technically difficult. However, the reagents required for serotyping are not readily available to diagnostic laboratories and some strains are untypeable or coagglutinable with this technique.

The epitopes for the serotypes are probably parts of bacterial surface proteins. In group A streptococci, *sof* (serum opacity factor gene) or *emm* (M protein gene) sequence-based analysis has been used more successfully than serological analysis for strain subtyping (3). These proteins are surface associated. Until now, no investigations have explored the usefulness of S-layer gene sequencing for subtyping of strains. A study of *Campylobacter fetus* (9) and *Lactobacillus helveticus* (32) demonstrated that amplification of the S-layer gene by PCR can be used for identification of strains. Furthermore, for *C. fetus*, each strain exhibited a different Southern blotting pattern when hybridized with the PCR product. This suggested that genotyping based on *slpA* gene structure could be useful for typing of strains. Since the P36 S-layer protein of *C. difficile* is located on the surface of the bacteria and because S-layer proteins are the most abundant bacterial proteins (23), it would be logical to conclude that they play a major role in determining serogroup specificity. This was confirmed by PCR-RFLP analysis and nucleotide sequencing: DNA and deduced amino acid sequences of the *slpA* variable region were 100% identical within a given serogroup, whereas interserogroup identity was, in general, fairly low. However, further sequencing of more strains is necessary to confirm these data. The exception is serogroup A, which is known to carry 20 subgroups. It is evident that these subserogroups may not be completely specified by the S-layer, since the *slpA* sequences were quite similar in serogroups A1 and A10. The subserogroup specificity of highlyflagellated serogroup A could be attributable to other surface proteins, most likely the flagella.

These data indicate that the *slpA* gene constitutes a reliable target for differentiation of *C. difficile* isolates and could be used as an alternative method to serotyping, at least in an outbreak setting until more diagnostic data become available. The gene could be easily amplified from various strains by PCR by using primers described here and then sequenced. Alternatively, the amplified DNA could be digested with restriction enzymes and profiles could be compared. However, sequence-based methods have advantages over the more commonly used RFLP methods, which are difficult to standardize. Furthermore, the relatively small size of certain restriction fragments obtained after digestion may render interpretation of the profiles difficult in clinical laboratories. Sequence data give more reliable results, thus eliminating interlaboratory variation due to, for example, gel mobility differences (20). The technology for generating DNA sequence data has become readily available for on-site or commercial service companies. Finally, this methodology should have a relatively low cost; the overall cost for generating one sequence can be estimated to be less than $10 U.S. However, because of country-to-country variations in labor and reagent costs, this estimate may not be valid for some countries. Sequencing can be expected to produce the same results in different laboratories, even with the use of different methods. This, combined with the ease of generation of template DNA for sequencing by PCR, makes sequence-based molecular typing a promising alternative to the more traditional methods.

ACKNOWLEDGMENT

We thank H. Kato, Kanazawa University, Kanazawa City, Japan, for the gift of *C. difficile* strains GAI95600 and GAI95601.

REFERENCES
