

Clostridium difficile Genotyping Based on *slpA* Variable Region in S-Layer Gene Sequence: an Alternative to Serotyping

Tuomo Karjalainen,¹ Nicolas Saumier,¹ Marie-Claude Barc,¹ Michel Delmée,² and Anne Collignon^{1*}

Université de Paris-Sud, Faculté de Pharmacie, Département de Microbiologie, 92296 Châtenay-Malabry Cedex, France,¹ and Université Catholique de Louvain, Faculté de Médecine, Unité de Microbiologie, Brussels 1200, Belgium²

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 enables the differentiation of 10 major serogroups, which are represented by capital letters (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.

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) were examined. The strains included the 10 serogroup reference strains and clinical isolates (Table 1). Serogroups had been previously determined by slide agglutination with rabbit antisera (13) or enzyme-linked immunosorbent assay and confirmed by typing by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (14). Toxin A production was determined by the *C. difficile* toxin A test (Oxoid). In vitro cytotoxin (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 (Gentra Systems, Minneapolis, Minn.; www.gentra.com).

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'-ATGAATAAGAAAAAYWTAGCAATGCG-3') and slpAV5 (5'-TCTATTCTATCTTCTCCWGCTAC-3'), where Y = CT, W = AT, and R = AG. For amplification of the entire gene, primers slpAV3 and slpAC5 (5'-AGCKATACCTTTACCWACTTG-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).

* Corresponding author. Mailing address: Université de Paris-Sud, Faculté de Pharmacie, Département de Microbiologie, 5 rue JB Clement, 92296 Châtenay-Malabry Cedex, France. Phone: (33) 1 46 83 55 49. Fax: (33) 1 46 83 55 37. E-mail: anne.collignon@cep.u-psud.fr.

TABLE 1. *C. difficile* isolates studied and their *slpA* RFLP patterns^a

Strain	Serogroup	Origin	Toxins A and B in vitro	Size of <i>slpA</i> variable domain (nt/a.a.)	Size(s) (nt) of fragment(s) obtained with:				GenBank accession no
					<i>Dra</i> I	<i>Hha</i> II	<i>Pvu</i> II	<i>Rsa</i> I	
ATCC 43594 (W1194)	A1	Brussels, Belgium	A ⁺ B ⁺	1,182,394	282, 366, 456	126, 136, 191, 729	391, 791	1,130	AF458877
Kohn	A ^b	Nottingham, United Kingdom	A ⁻ B ⁻	1,065/355	102, 216, 723	248, 346, 471	100, 154, 794	189, 798	AF448119
9354	A ^b	Bondy, France	A ⁺ B ⁺	1,053/351	945	227, 826	274, 779	367, 686	AF448120
TO005	A10	Toronto, Canada	A ⁺ B ⁺	1,185/395	288, 363, 456	111, 126, 923	391, 794	1,133	AF458878
ATCC 43593 (1351)	B	Brussels, Belgium	A ⁻ B ⁻	1,059/353	189, 216, 327, 327	200, 201, 274, 312	1,059	144, 260, 583	AF458879
EX560	B	Mauscron, Belgium	A ⁻ B ⁻	1,059/353	189, 216, 327, 327	200, 201, 274, 312	1,059	144, 260, 583	AF448121
CO109	B	Jouy-en-Josas, France	A ⁻ B ⁻	1,059/353	189, 216, 327, 327	200, 201, 274, 312	1,059	144, 260, 583	AF448122
ATCC 43596 (545)	C	Namur, Belgium	A ⁺ B ⁺	1,077/359	402, 651	126, 175, 776	200, 265, 612	144, 335, 568	AF448123
630	C	London, United Kingdom	A ⁺ B ⁺	1,077/359	402, 651	126, 175, 776	200, 265, 612	144, 335, 568	AF448124
C253	C	Rome, Italy	A ⁺ B ⁺	1,077/359	402, 651	126, 175, 776	200, 265, 612	144, 335, 568	AF458880
ATCC 43597 (3232)	D	Brussels, Belgium	A ⁻ B ⁻	1,167/389	141, 201, 753	171, 230, 681	1,167	155, 370, 642	AF458880
90-111	D	Bondy, France	A ⁻ B ⁻	1,167/389	141, 201, 753	171, 230, 681	1,167	155, 370, 642	AF458881
93-136	D	Bondy, France	A ⁻ B ⁻	1,167/389	141, 201, 753	171, 230, 681	1,167	155, 370, 642	AF458882
ATCC 43598 (1470)	F	Brussels, Belgium	A ⁻ B ⁺	1,062/354	1,038	111, 172, 273, 491	200, 862	113, 159, 173, 190, 332	AF448125
R7404	F	London, United Kingdom	ND	1,062/354	1,038	111, 172, 273, 491	200, 862	113, 159, 173, 190, 332	AF300677
GA195600	F	Kanazawa, Japan	A ⁻ B ⁺	1,062/354	1,038	111, 172, 273, 491	200, 862	113, 159, 173, 190, 332	AF448126
GA195601	F	Kanazawa, Japan	A ⁻ B ⁺	1,062/354	1,038	111, 172, 273, 491	200, 862	113, 159, 173, 190, 332	AF448127
ATCC 43599 (2022)	G	Brussels, Belgium	A ⁺ B ⁺	1,065/355	846, 219	485, 580	1,065	1,065	AF448128
R8366	G	London, United Kingdom	ND	1,065/355	846, 219	485, 580	1,065	1,065	AF300676
96-392	G	Paris, France	A ⁺ B ⁺	1,065/355	846, 219	485, 580	1,065	1,065	AF448129
ATCC 43600 (2149)	H	Brussels, Belgium	A ⁺ B ⁺	1,089/363	265, 824	410, 679	222, 299, 562	407, 628	AF448365
89-638	H	Bondy, France	A ⁺ B ⁺	1,089/363	265, 824	410, 679	222, 299, 562	407, 628	AF448366
90-204	H	Bondy, France	A ⁺ B ⁺	1,089/363	265, 824	410, 679	222, 299, 562	407, 628	AF448367
ATCC 43601 (7322)	I	Brussels, Belgium	A ⁻ B ⁻	1,017/339	993	171, 701	920	136, 222, 538	AF458883
56026	I	Brussels, Belgium	A ⁻ B ⁻	1,017/339	993	171, 701	920	136, 222, 538	AF458884
54823	I	Brussels, Belgium	A ⁻ B ⁻	1,017/339	993	171, 701	920	136, 222, 538	AF458885
ATCC 43602 (4811)	K	Brussels, Belgium	A ⁻ B ⁻	1,065/355	102, 159, 233, 568	286, 699	185, 265, 522	196, 265, 409	AF448368
94-416	K	Paris, France	A ⁺ B ⁺	1,065/355	102, 159, 233, 568	286, 699	185, 265, 522	196, 265, 409	AF448369
48-515	K	Brussels, Belgium	A ⁺ B ⁺	1,065/355	102, 159, 233, 568	286, 699	185, 265, 522	196, 265, 409	AF448370
ATCC 43603 (5036)	X	Brussels, Belgium	A ⁻ B ⁻	1,038/346	507, 531	162, 268, 525	231, 380, 427	101, 153, 778	AF448371
12-934	X	Verriers, Belgium	A ⁻ B ⁻	1,038/346	507, 531	162, 268, 525	231, 380, 427	101, 153, 778	AF448372
36-678	X	Brussels, Belgium	A ⁻ B ⁺	1,038/346	507, 531	162, 268, 525	231, 380, 427	101, 153, 778	AF448373

^a Serogroup reference strains are in bold. Classification is according to Deln e et al. (13, 14); Fragment sizes of <100 bp are not indicated. nt, nucleotides; a.a., amino acids; NID, not determined.^b Serogroup subgroup unknown.

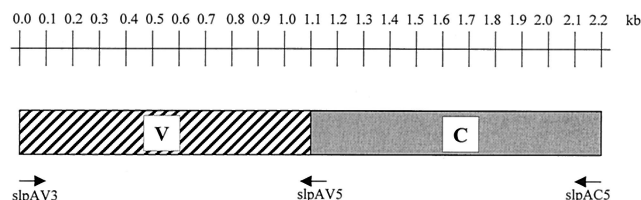


FIG. 1. Structural organization of the *slpA* gene of *C. difficile*. The conserved (C) domain is in gray, and the variable (V) domain is represented by the hatched bar. Positions of the primers used for amplification of the domains are indicated below the gene structures. Gene length is indicated in kilobases above the gene structures.

Initial denaturation was carried out at 95°C for 5 min. Thirty-five cycles of amplification were performed in a Perkin-Elmer GeneAmp PCR system 2400 thermocycler. Each cycle consisted of three steps: denaturation at 95°C (30 s), annealing at 45°C (1 min), and extension at 72°C (1 min to amplify the conserved or variable domain and 2 min to amplify the entire gene). An additional step of extension for 10 min at 72°C was performed at the end of the amplification to complete the extension of the primers. Samples (3 µl) of amplified products were analyzed by electrophoresis in a 1.0% (wt/vol) agarose gel with a 100-bp ladder (Amersham-Pharmacia Biotech) as the molecular size marker.

RFLP analysis. Five microliters of PCR-amplified products, purified with the Wizard PCR Preps DNA purification system (Promega), were digested with the restriction enzymes *DraI*, *HinI*, *PvuII*, and *RsaI* in accordance with the vendor's (Life Technologies) recommendations. The digested amplified products were analyzed by electrophoresis in a 1.0% (wt/vol) agarose gel.

Sequence analysis. PCR products were purified with the QIAquick PCR purification kit (Qiagen, Hilden, Germany). DNA sequencing was carried out with the BigDye terminator DNA sequencing kit (PE Applied Biosystems, Warrington, England). The samples were analyzed with an automated DNA sequencer, the ABI PRISM 310 genetic analyzer (Perkin-Elmer). Initial sequencing was carried out with the same primers as used for PCR. More sequence was acquired by using primers the sequences of which were derived from the internal sequence data obtained for the variable region.

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 possibil-

ity, 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*, *HinI*, *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 *HinI* 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

FIG. 2. Multiple alignments of the predicted amino acid sequences encoded by variants of the variable region of the *slpA* gene from Delmée's serogroup reference strains A1, B, C, D, F, G, H, I, K, and X. Furthermore, three additional *SlpA* sequences from serogroup A (Kohn, 9354, and TO005) are included in the alignment. Additional sequence data were obtained for at least two isolates in each serogroup (Table 1). Analysis of the data showed that the *slpA* gene sequence from strains EX560 and CO109 was 100% identical to that of the serogroup B reference strain; that of strains 630 and C253 was 100% identical to that of the serogroup C reference strain; that of strains 90-111 and 93-136 was 100% identical to that of the serogroup D reference strain; that of strains R7404, GAI95600, and GAI95601 was 100% identical to that of the serogroup F reference strain; that of strains R8366 and 93-392 was 100% identical to that of the serogroup G reference strain; that of strains 89-638 and 90-204 was 100% identical to that of the serogroup H reference strain; that of strains 56026 and 54823 was 100% identical to that of the serogroup I reference strain; that of strains 94-416 and 48-515 was 100% identical to that of the serogroup K reference strain; and that of strains 12-934 and 36-678 was 100% identical to that of the serogroup X reference strain. Identical residues are indicated by asterisks below the alignment; functionally identical residues are indicated by colons or periods.

A1 AAPVFAAASDVISLODG-TNDKYTVSNTKASDLVKDILAAQNLTGAVILNKDTKVTFYD 59
A10 AAPVFAAASDVISLODG-TNDKYTVSNTKASDLVKDILAAQNLTGALCRKTKDTESYFLC 59
Kohn A AAPVFAVDAPVKAPEGTDYVDNYTTVQSKYEKALKKLDGGLVNGE-----IKSIDVYFD 54
9354_A AAPVFAADEQVK-----YQNTYTVVQSKYEKALKDMQKGIITDKK-----IKSIAISYE 48
B AAPVFAATAVPVDKE-----YITVQKDYKETLKKIQAGIIVDKS-----ISNIDIKYD 47
C AAPVFAATTGTQG-----YTVVKNWKKKAVKQLQDGLKDNS-----IGKITVSFN 45
D AAPVFAATTTPSEDNAG-----KYTVVKDKYKIDFDEVKLVKENDAAIDAYLKENNSTR 55
F AAPVFAADST-TPG-----YTVVKNWKKKAVKQLQDGLKKNK-----ISTIKVSFN 44
G AAPVFAADTKVETG-----DQGYTVVQSKYKKAQEQLQKGLDGS-----ITEIKVFE 49
H AAPVFAADTTVKEEG-----YTVVQDKYEKLKELKAKIKDGT-----ITSVGEVFD 46
I AAPVFAANVVT-GSEG-----YTVSQKDYKKAQKQDGIKKNI-----ITSINIYFD 46
K AAPVFAASSFTADYN-----YTVVQGKYQKVITGLQDGLKNGK-----ITNIDVIFD 47
X AAPVFAAD---VKAE-----YITVQKDYKDTLKKIQAGIKDGS-----ITNLVVTFYD 44
: * : :

A1 ANE-KDSSTPTGD-KK-----VYSEQTLTTANGNEDYVKTTLKNLDAGEYAI 104
A10 MQMRKMSSTATLEIKE-----VYSEQILTANGNEDYVKTTLKNLDAGEYAI 106
Kohn A G---SKIAQAITD-----KDELFR-----LVDAKLDGLDGDSDVQ 86
9354_A GKPVTITITVADMDTKG-----KTSTKEELASALLKTTVNDKLDNLDGDDYVD 95
B G-KLVS--EYGISGT-----KVSELTDESVKFYNLVEAQLKNMDDGDGTF 89
C DGUVVEVAPKSN-K-----KADRDAAAEKLYNLVNTQLDKLDGDDYVD 88
D DAAWALAGTNSLNIINITVKDSDDSVVVPKTVVKTNTVVDTDYVENKLDNLDYVD 115
F GNSVGEVTPASSGAK-----KADRDAAAEKLYNLVNTQLDKLDGDDYVD 88
G GTLASTIKVSGEL-----NAADASKLLFTQVNDKLDNLDGDDYVD 89
H GKPITITLAPKADG-----SDKDAIAEQLETLTKNQLKGLDGKGYVD 87
I GTKVST---VNVAG-----DQDARDKIADDFNNTVEGKLDLSLDGDKYVD 87
K GSSIGEVPV-----GSDAAAAATKLKSLVDDKLDNLDGDKGYVD 85
X KDKEVANYNYKTNAT-----TADAKEVAATTLNVLVDSKLDNLDGDDLVS 89
. . . * :

A1 IDLTYNNAKTVEIKVVAASEKTVVSSDAKNSAKDIAEKYVFEDKDLKWLPLKTINASDF 164
A10 IDLTYNNAKTVEIKVVAASEKTVVSSDAKNSAKDIAEKYVFEDKDLN-ALKTINASDF 165
Kohn A FTVHYA--GDAGKYT-LETLDALKTAASASAEAI PPGG-----EGVVGVITNSD 133
9354_A FDIITYV--GDADRLT-AGDLNTFAKGLAESTEKKI PPAK-----GSNYGVAKTN- 141
B FIITYN--TNNKEYS-KAELEDLKTQLENKVVAAPATNG-----GNGAVME--GE 134
C FSVVDYN--LENKIITNQADAEAVTKLNSLNEKTLIDIAATKDT-----FGMVSKTQDS 139
D FDIITKASTVSSKYTAANIYDLAANNGKILIDALKNVTEKQLK-----DGGILGEVAKT 170
F EVTYN--LATQIIT-KAEAEAVLTKLQYNDKVLINSATDTV-----KGMVSDTQ-V 137
G FLITSP--GGQDKIT-TSKLVALKD-LTGASADAIAGT-----SSADGVVNTG 135
H FKITYG--AKAEVPAASLADDIQYADQINASEKILVEVAA-----SEAGIAKFNVS 139
I FEISYN--KGVVEEY-TLELENYKLLDNKVVIPKASGV-----NPGAVKEKSGS 134
K FNVYTT--TKS-IIT-KAELKNYYNQLESSEKDRILIGNEPQDTG-----TKGLIKADT-D 135
X FNIKYD--AAEFKHT-KDEMDALKTRLENKEIVKPASET-----TAGLVMA--DG 134
: :

A1 SKTDSYQQVLYPKGKRLQGFSTYRATNYNEGTAIGNTPVILTLKSTSKSNLKTAVEELO 224
A10 SKTDSYQQVLYPKGKRLQGFSTYRATNYNEGTAIGNTPVILTLKSTSKSNLKTAVEELO 225
Kohn A PSG--TGTNAKTALKYSDIVKTAFTGN--GVVLTQLDSSLSSVKLGEVKASANTYETVT 189
9354_A -SG--TGKLTDDTAIVISTIEG-KVEGN--NLTI SLKDAPS--KVGVI GANN--DTLA 190
B SGKAKSADRSTIGSDVYDFVIVEDSVSGEWTLKAEPKASELAALNAVYKFTQSF--DDG 192
C EGNVAATKALKVKVDVATFG-LKSGGSEDTGYVEMKAG-AVEDKYGKVGDS--TAGIAI 195
D ISGATTPSAPTGDTFASYFTVGTGVTNGKVALEINIAEPASTVLVKTDAELT----- 223
F DSKNVAAN-PLKVSMDYTI-P-SAITGSDSGSYIAKPEKTTSLLYGTGVDGA--TAGKAI 193
G AAGGSTETNSAGTKLAMSADFTAYTDSSETAVKIKITAKMDNDTKFKGAGETT---YST 191
H NNKVIAGDAPLKVVDVATVTTNGSNKK--VLTISAAAGLSGFSYGTLDKDTG--ASSS 194
I ADEAVAANDNDIKGSDLYNTTVEADTTNGGYKLSITAKTISDV-----KYGT-----IG 182
K GTTAVAADAPLKSIDI FTFSYDEVTVGLKAEPTSKVSAGKVQGLKYGNTGATNYTSGAEI 195
X VTNSKKAADKSLYAKDVIKFDVVS DTIG--YKLTATPISDAQLATLKATYKYAN-----NT 187

A1 KLNASYSNTTTTLAGDDRIQTAIEISKEYYNNNDGEKSDHSADV KENVKNVVLVGANALVDG 284
A10 KLNASYSNTTTTLAGDDRIQTAIEISKEYYNNNDGEKQDHSADV KENVKNVVLVGANALVDG 285
Kohn A KVAFDKNTLTKTIVSVGDK-LDLKSLPLVIKEIKNGTLIIDLPENTSSSSDSSN-LKV 247
9354_A DVTFADDAKLT--VSVGDP-KIDLAKSFIF-DTKTGKLG-GIVEKENDATEHAYV--RV 242
B TSTFAGATAFTVNTPTTQVVKSSKSLNLTSLANTTGQVGDVLTENIVPNTK-AVSVKI 251
C NLPSTG-LEYAGKGTIDFNKT-LKVVDVTGGSTPSAVAVSGFVTKDDTDLAKSGTINVRV 253
D -TSPTTQDKMSFANAKITLTGDDRLDFSKPSIVD GALGDFAKAAATTPGKQQTINVRV 282
F TVDTASNEAFAGNGKVIDYKNS-FKATVQGDGT---VKTSGVVLKASDMAATGTIKVRV 249
G GLTFEDGSTEK-IVKLGDSDIIDITKALKTLVPGSKATVKFAEKTSPASVQPVITKLRI 250
H DVDAITLDTNATITEGDTKVLDFDMSFKFNESTKKVGLVLTENTNTPADPGTKTIVVRV 250
I AGKVPDAKVITATGDAVLKGTVDISASYALEASTGNVSGLSLTDNTPSSDS--VNVRI 244
K SVPTTGLTLTADTTATDVNISDVMSAFKFNGT---DTISGFAGSS---ASTLRASIKV 249
X KVEFASATELAATDGSAREVAKGKEYNATGSLVFDG-ATGKTSNINVDPLTNKGDTVVVK 246

A1 LVAAPLAAEKDAPILLTSKDKLDSSVKSEIKRVLDLKTSTEVTKTVYIAGGVNSVAKEV 344
A10 LVAAPLAAEKDAPILLTSKDKLDSSVKSEIKRVLDLKTSTEVTKTVYIAGGVNSVSKVEV 345
Kohn A VNAKEQITIDLDASSYKSAEDINQLKAFDNEVKNLTAIEELQKD-NVKTNKVQLIDGKY 306
9354_A INAKEQITIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 302
B INAKEQITIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 302
C INAKEESIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 308
D INAKEQETVKAT--DYDALKAVSKYKFDSTEIGRVYDEAKDLNDKDKLDGSSYDK-DGTY 312
F TSAKEESIDVDSSSYISAEENLAKKYVENPKVESEAYNAIVALQND-GIESDLVQLVNGKY 308
G INAKEETIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 307
H IKAVEKITIDVSSNSTTKAKDLAKQYVFT-DVSDTDPESLSYMLKNINDGKVAVKNFDDGY 313
I INAKEITIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 291
K INAKEESIDVDSSSHRTAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 308
X INAKEITIDLDASSYKSAEDLAKAYAFDNEVKNLTYTEIEAYOKDSNNKTNKVQIVDGKY 302
* : . : . :

A1 VTELESMLKVERFSGDDRYETSLKIADEIGLDNDKAYVVGGTGLADAMS 394
A10 VTELESMLKVERFSGDDRYETSLKIADEIGLDNDKAYVVGGTGLADAMS 395
Kohn A QVTLWADGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 351
9354_A QVTLWADGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 351
B EVVLYPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 353
C QVIFYPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 359
D KAVFFAEGKRLQGFSSYKGFSTDEAESGLADGNAALKLVIESTDEDDFD 389
F QVIFYPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 354
G QVILYPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 355
H EVTIFPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 363
I QVEIFADGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 339
K QVVLFAQKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 355
X QVALYPEGKRLTTKSAQPSKLAGDN-SPLCLVIKADKLKLDKYVEDLR 346
. . . * : . : . :

TABLE 2. Pairwise comparisons of amino acid sequence homology of the variable region of the SlpA protein between *C. difficile* serogroups

Serogroup	% Identity									
	A	B	C	D	F	G	H	I	K	X
A	100	30	28	14	28	35	25	11	29	31
B	30	100	32	20	33	32	28	36	26	47
C	28	32	100	23	64	36	33	38	37	32
D	14	20	23	100	21	22	23	25	20	14
F	28	33	64	21	100	35	34	40	37	32
G	35	32	36	22	35	100	28	38	31	34
H	25	28	33	23	34	28	100	32	30	26
I	11	36	38	25	40	38	32	100	33	34
K	29	26	37	20	37	31	30	33	100	28
X	31	47	32	14	32	34	26	34	28	100

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,

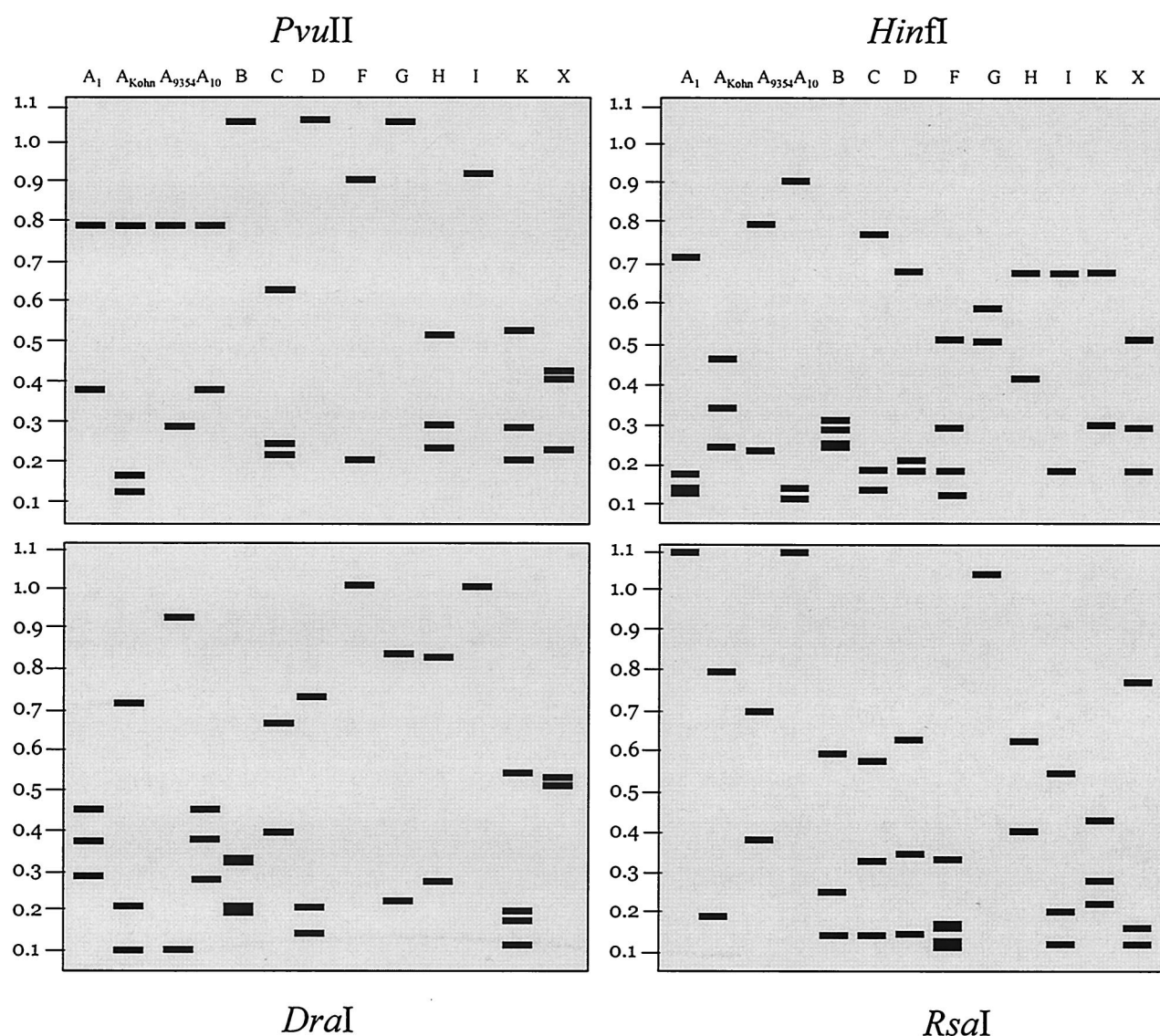


FIG. 3. Computer-generated restriction enzyme digestion profiles of the 1-kb *slpA* variable region from the serogrouped strains indicated. A simulation with the four enzymes indicated is shown. Sizes of the expected restriction products are indicated in kilobases on the left.

7). 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 Delmée'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 most diagnostic laboratories and some strains are untypeable or coagulable 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 highly flagellated 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 pro-

files 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

- Barbut, F., N. Mario, M. Delmee, J. Gozian, and J. C. Petit. 1993. Genomic fingerprinting of *Clostridium difficile* isolates by using a random amplified polymorphic DNA (RAPD) assay. *FEMS Microbiol. Lett.* **114**:161-166.
- Barbut, F., N. Mario, M. C. Meyohas, D. Binet, J. Frottier, and J. C. Petit. 1994. Investigation of a nosocomial outbreak of *Clostridium difficile*-associated diarrhoea among AIDS patients by random amplified polymorphic DNA (RAPD) assay. *J. Hosp. Infect.* **26**:181-189.
- Beall, B., G. Gherardi, M. Lovgren, R. R. Facklam, B. A. Forwick, and G. J. Tyrrell. 2000. *emm* and *sof* gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. *Microbiology* **146**:1195-1209.
- Bidet, P., F. Barbut, V. Lalande, B. Burghoffer, and J. C. Petit. 1999. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. *FEMS Microbiol. Lett.* **175**:261-266.
- Blair, D. F., and S. K. Dutcher. 1992. Flagella in prokaryotes and lower eukaryotes. *Curr. Opin. Genet. Dev.* **2**:756-767.
- Borriello, S. P., H. A. Davies, S. Kamiya, P. J. Reed, and S. Seddon. 1990. Virulence factors of *Clostridium difficile*. *Rev. Infect. Dis.* **12**(Suppl. 2):S185-S191.
- Brazier, J. S. 2001. Typing of *C. difficile*. *Clin. Microbiol. Infect.* **7**:428-431.
- Calabi, E., S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, A. Dell, G. Dougan, and N. Fairweather. 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol. Microbiol.* **40**:1187-1199.
- Casademont, I., D. Chevrier, and J. L. Guesdon. 1998. Cloning of a *sapB* homologue (*sapB2*) encoding a putative 112-kDa *Campylobacter fetus* S-layer protein and its use for identification and molecular genotyping. *FEMS Immunol. Med. Microbiol.* **21**:269-281.
- Cerquetti, M., A. Molinari, A. Sebastianelli, M. Diociaiuti, R. Petruzzelli, C. Capo, and P. Mastrantonio. 2000. Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microb. Pathog.* **28**:363-372.
- Davies, H. A., and S. P. Borriello. 1990. Detection of capsule in strains of *Clostridium difficile* of varying virulence and toxigenicity. *Microb. Pathog.* **9**:141-146.
- Delmée, M., V. Avesani, N. Delferrière, and G. Burtonboy. 1990. Characterization of flagella of *Clostridium difficile* and their role in serogrouping reactions. *J. Clin. Microbiol.* **28**:2210-2214.
- Delmée, M., M. Homel, and G. Wauters. 1985. Serogrouping of *Clostridium difficile* strains by slide agglutination. *J. Clin. Microbiol.* **21**:323-327.
- Delmée, M., Y. Laroche, V. Avesani, and G. Cornelis. 1986. Comparison of serogrouping and polyacrylamide gel electrophoresis for typing of *Clostridium difficile*. *J. Clin. Microbiol.* **24**:991-994.
- Hennequin, C., A. Collignon, and T. Karjalainen. 2001. Analysis of expression of GroEL (Hsp60) of *Clostridium difficile* in response to stress. *Microb. Pathog.* **31**:255-260.
- Hennequin, C., F. Porcheray, A.-J. Waligora, A. Collignon, P. Bourlioux, and T. Karjalainen. 2001. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology* **147**:87-96.
- Karjalainen, T., A.-J. Waligora-Dupriet, M. Cerquetti, P. Spigagli, P. Mauri, and P. Mastrantonio. 2001. Molecular and genomic analysis of two genes encoding surface-anchored proteins from *Clostridium difficile*. *Infect. Immun.* **69**:3442-3446.
- Kuijper, E. J., J. H. Oudbier, W. N. Stuijbergen, A. Jansz, and H. C. Zanen.

1987. Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. *J. Clin. Microbiol.* **25**:751–753.
19. McCoubrey, J., and I. R. Poxton. 2001. Variation in the surface layer proteins of *Clostridium difficile*. *FEMS Immunol. Med. Microbiol.* **31**:131–135.
20. Meinersmann, R. J., L. O. Helsel, P. I. Fields, and K. L. Hiett. 1997. Discrimination of *Campylobacter jejuni* isolates by *fla* gene sequencing. *J. Clin. Microbiol.* **35**:2810–2814.
21. Pantosti, A., M. Cerquetti, and P. M. Gianfrilli. 1988. Electrophoretic characterization of *Clostridium difficile* strains isolated from antibiotic-associated colitis and other conditions. *J. Clin. Microbiol.* **26**:540–543.
22. Poilane, I., T. Karjalainen, M. C. Barc, P. Bourlioux, and A. Collignon. 1998. Protease activity of *Clostridium difficile* strains. *Can. J. Microbiol.* **44**:157–161.
23. S  ra, M., and U. B. Sleytr. 2000. S-layer proteins. *J. Bacteriol.* **182**:859–868.
24. Seddon, S. V., and S. P. Borriello. 1992. Proteolytic activity of *Clostridium difficile*. *J. Med. Microbiol.* **36**:307–311.
25. Stubbs, S. L., J. S. Brazier, G. L. O'Neill, and B. I. Duerden. 1999. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. *J. Clin. Microbiol.* **37**:461–463.
26. Tasteyre, A., M.-C. Barc, A. Collignon, and T. Karjalainen. 2001. Role of flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun.* **69**:7937–7940.
27. Tasteyre, A., M.-C. Barc, T. Karjalainen, P. Dodson, S. Hyde, P. Bourlioux, and P. Borriello. 2000. A *Clostridium difficile* gene encoding flagellin. *Microbiology* **146**:957–966.
28. Tasteyre, A., T. Karjalainen, V. Avesani, M. Delm  e, A. Collignon, P. Bourlioux, and M.-C. Barc. 2000. Phenotypic and genotypic diversity of the flagellin gene (*flaC*) among *Clostridium difficile* isolates from different serogroups. *J. Clin. Microbiol.* **38**:3179–3186.
29. Tasteyre, A., T. Karjalainen, V. Avesani, M. Delm  e, A. Collignon, P. Bourlioux, and M.-C. Barc. 2001. Molecular characterization of *ftiD* gene encoding flagellar cap and its expression among *Clostridium difficile* isolates from different serogroups. *J. Clin. Microbiol.* **39**:1178–1183.
30. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
31. Toma, S., G. Lesiak, M. Magus, H.-L. Lo, and M. Delm  e. 1988. Serotyping of *Clostridium difficile*. *J. Clin. Microbiol.* **26**:426–428.
32. Ventura, M., M. L. Callegari, and L. Morelli, L. 2000. S-layer gene as a molecular marker for identification of *Lactobacillus helveticus*. *FEMS Microbiol. Lett.* **189**:275–279.
33. Waligora, A.-J., C. Hennequin, P. Mullany, A. Collignon, P. Bourlioux, and T. Karjalainen. 2001. Characterization of a cell surface protein of *C. difficile* with adhesive properties. *Infect. Immun.* **69**:2144–2153.