Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*

Ludovic Lemée,¹ Ingrid Bourgeois,¹ Elodie Ruffin,¹ Anne Collignon,² Jean-François Lemeland¹ and Jean-Louis Pons¹

Correspondence

Ludovic Lemée

ludovic.lemee@chu-rouen.fr

¹Groupe de Recherche sur les Antimicrobiens et les Micro-organismes (GRAM EA 2656, IFR 23), Université de Rouen, Faculté de Médecine-Pharmacie, 22 Boulevard Gambetta, F-76183 Rouen Cedex, France

²Département de Microbiologie-Immunologie, Faculté de Pharmacie Paris XI, Châtenay-Malabry, France

A multilocus sequence analysis of ten virulence-associated genes was performed to study the genetic relationships between 29 *Clostridium difficile* isolates of various origins, hosts and clinical presentations, and selected from the main lineages previously defined by multilocus sequence typing (MLST) of housekeeping genes. Colonization-factor-encoding genes (*cwp66*, *cwp84*, *fbp68*, *fliC*, *fliD*, *groEL* and *slpA*), toxin A and B genes (*tcdA* and *tcdB*), and the toxin A and B positive regulator gene (*tcdD*) were investigated. Binary toxin genes (*cdtA* and *cdtB*) were also detected, and internal fragments were sequenced for positive isolates. Virulence-associated genes exhibited a moderate polymorphism, comparable to the polymorphism of housekeeping genes, whereas *cwp66* and *slpA* genes appeared highly polymorphic. Isolates recovered from human pseudomembranous colitis cases did not define a specific lineage. The presence of binary toxin genes, detected in five of the 29 isolates (17 %), was also not linked to clinical presentation. Conversely, toxigenic A–B+ isolates defined a very homogeneous lineage, which is distantly related to other isolates. By clustering analysis, animal isolates were intermixed with human isolates. Multilocus sequence analysis of virulence-associated genes is consistent with a clonal population structure for *C. difficile* and with the lack of host specificity. The data suggest a co-evolution of several of the virulence-associated genes studied (including toxins A and B and the binary toxin genes) with housekeeping genes, reflecting the genetic background of *C. difficile*, whereas flagellin, *cwp66* and *slpA* genes may undergo recombination events and/or environmental selective pressure.

Received 26 April 2005

Revised 27 June 2005

Accepted 4 July 2005

INTRODUCTION

Since its recognition as the main cause of pseudo-membranous colitis (PMC) in 1978, *Clostridium difficile* has emerged as an important enteropathogen (Bartlett *et al.*,

1978; Johnson & Gerding, 1998; Larson *et al.*, 1978). *C. difficile* is currently responsible for virtually all cases of PMC and for 10–25 % of antibiotic-associated diarrhoea (AAD) (Bartlett, 1994; Kelly *et al.*, 1994). It is also the leading cause of nosocomial diarrhoea, and many hospital outbreaks have been described throughout the world (Barbut *et al.*, 1994; Cartmill *et al.*, 1994; McEllistrem *et al.*, 2005). The two main virulence factors are exotoxins, toxins A (enterotoxin) and B (cytotoxin), both of which damage the human colonic mucosa and are potent tissue-damaging enzymes (Bongaerts & Lysterly, 1994; Borriello, 1998). Until recently, it was presumed that a strain must produce both toxins to be considered fully pathogenic. However, it is still unclear (i) why strains that do not produce toxin A (A–B+ variant strains) are yet responsible for diarrhoea and even PMC (Alfa *et al.*, 2000; Toyokawa *et al.*, 2003), and (ii) why some patients infected by toxinogenic strains will present a mild diarrhoea, whereas others will present a fulminant PMC.

Abbreviations: AAD, antibiotic-associated diarrhoea; MLST, multilocus sequence typing; PaLoc, pathogenicity locus; PMC, pseudo-membranous colitis; ST, sequence type; UPGMA, unweighted pair-group method with arithmetic means.

Nucleotide sequences of the internal fragment genes analysed in this work will be deposited in GenBank under accession numbers DQ102375–DQ102379 (for *cdtA*), DQ117049–DQ117053 (for *cdtB*), DQ117054–DQ117074 (for *cwp66*), DQ117075–DQ117103 (for *cwp84*), DQ117104–DQ117132 (for *fbp68*), DQ117133–DQ117161 (for *fliC*), DQ117162–DQ117189 (for *fliD*), DQ117190–DQ117218 (for *groEL*), DQ117219–DQ117240 (for *slpA*), DQ117241–DQ117265 (for *tcdA*), DQ117266–DQ117288 (for *tcdB*) and DQ117289–DQ117311 (for *tcdD*).

Differences in the levels of production of toxins A and B cannot alone account for the wide spectrum of clinical presentations. Additional virulence factors (i.e. colonization factors and binary toxin) and/or host susceptibility could explain differences in clinical presentations. Adherence determinants are probably largely implicated in the colonization of the gastrointestinal tract by *C. difficile*, and numerous studies have characterized factors involved in the adherence and colonization processes, such as S-layer proteins (Calabi *et al.*, 2001, 2002; Cerquetti *et al.*, 2000), the Cwp66 adhesin (Waligora *et al.*, 2001), the Fbp68 fibronectin-binding protein (Hennequin *et al.*, 2003), flagella (Tasteyre *et al.*, 2001), the GroEL heat-shock protein (Hennequin *et al.*, 2001) and hydrolytic enzymes (Poilane *et al.*, 1998; Seddon *et al.*, 1990). In addition, some strains of *C. difficile* also produce an actin-specific ADP-ribosyltransferase (binary toxin), which may act as an additional virulence factor and which may function in synergy with the large clostridial toxins A and B (Barbut *et al.*, 2005).

In a previous work, we delineated phylogenetic relationships within *C. difficile* strains by developing a multilocus sequence typing (MLST) approach (Lemée *et al.*, 2004a). We found that isolates recovered from PMC and from AAD did not cluster in distinct lineages, and thus that no hypervirulent lineage could be characterized within the population of toxigenic human isolates (Lemée *et al.*, 2004a). Since the previous study was based on the polymorphism of housekeeping genes, which reflects the genetic background of the strains, the aim of the present work was to analyse the polymorphism of virulence-associated genes encoding colonization factors, toxins A and B, the sigma factor TcdD, and binary toxin, and to examine the comparative evolution of these virulence-associated and housekeeping genes within a collection of 29 *C. difficile* isolates selected as representative of the main clusters previously defined by MLST.

METHODS

Bacterial isolates. Twenty-nine *C. difficile* isolates, selected as representative of the main clusters previously defined by MLST (Lemée *et al.*, 2004a), and recovered from various hosts and geographic sources, were studied. Twenty-four isolates were recovered from human stools: ten from patients with AAD, eight from patients with PMC and four from patients with asymptomatic carriage (the data for two human A–B+ isolates were unknown). Five isolates were recovered from animal hosts suspected to have clostridial intestinal infection. Isolates were identified as *C. difficile* by Gram stain, colony morphology and fluorescence, API 20A (BioMérieux) biochemical profiles, and for some isolates by sequencing the first 500 bp of 16S rDNA and of an internal fragment of the *tpi* gene (Dhalluin *et al.*, 2003) to confirm their species identification. The lack of a direct epidemiological link between these 29 strains had previously been confirmed by PCR ribotyping (Lemée *et al.*, 2004a), following a procedure described elsewhere (Bidet *et al.*, 2000). Toxigenic A/B types were determined by multiplex PCR targeting the toxin A and B genes (Lemée *et al.*, 2004b). Among the 29 *C. difficile* isolates, 17 harboured the *tcdA* gene (encoding toxin A) and

the *tcdB* gene (encoding toxin B), eight harboured a deleted variant of the *tcdA* gene and the *tcdB* gene (A–B+ variants), and four lacked the *tcdA* and the *tcdB* genes.

Multilocus sequence analysis of virulence-associated genes.

Ten virulence-associated genes were analysed by PCR-sequencing of internal fragments; these included seven colonization factor genes and three genes of the pathogenicity locus (PaLoc). The seven genes encoding colonization factors were as follows (Table 1): *cwp84* (Cwp84 protease), *cwp66* (Cwp66 adhesin), *fbp68* (fibronectin-binding protein), *fliC* (flagellin protein), *fliD* (flagellar cap protein), *groEL* (heat-shock protein) and *slpA* (S-layer protein); these genes have previously been characterized in *C. difficile* (Calabi *et al.*, 2001, 2002; Cerquetti *et al.*, 2000; Hennequin *et al.*, 2001, 2003; Poilane *et al.*, 1998; Seddon *et al.*, 1990; Tasteyre *et al.*, 2001; Waligora *et al.*, 2001). Only one copy of each of the seven genes was found on the *C. difficile* 630 genome (<http://www.sanger.ac.uk/>). The three PaLoc genes analysed were *tcdA* (encoding toxin A), *tcdB* (toxin B) and *tcdD* (encoding a sigma factor that upregulates the expression of toxins A and B) (Mani & Dupuy, 2001; Moncrief *et al.*, 1997). In addition, binary toxin genes were screened by PCR amplification, using primers targeting the binary toxin genes *cdtA* (forward 5'-TGAACCTGGAAA-AGGTGATG-3' and reverse 5'-AGGATTATTTACTGGACCATTTG-3') and *cdtB* (forward 5'-CTTATTGCAAGTAAATACTGAG-3' and reverse 5'-ACCGGATCTCTTGCTTCAGTC-3') (Stubbs *et al.*, 2000). When isolates were positive for the presence of *cdtA* and/or *cdtB* genes, sequencing of the amplified fragment was performed using the same PCR primers. The genomic locations of all the genes studied here and of the seven housekeeping genes analysed in a previous MLST study (Lemée *et al.*, 2004a) are shown in Fig. 1.

To prepare a DNA sample for PCR amplification, a bacterial colony was taken from blood agar culture and resuspended in 1 ml distilled water in a microcentrifuge tube. The sample was then boiled for 20 min prior to being centrifuged for 2 min to settle bacterial debris, and 10 µl of supernatant, containing the genomic DNA, was used for subsequent PCR amplification.

Internal fragments of the selected genes were amplified using the primers listed in Table 1. PCRs were performed on a GeneAmp System 2400 thermal cycler (Applied Biosystems) in a final volume of 50 µl containing 0.5 µM of each primer, 200 µM of each deoxynucleoside triphosphate, and 1.25 U of *Taq* DNA polymerase (Applied Biosystems) in a 1 × amplification buffer (10 mM Tris/HCl, pH 8.3, 50 mM KCl, 2.5 mM MgCl₂). The PCR mixtures were heated for 3 min at 95 °C, then a touch-down procedure followed, consisting of 30 s at 95 °C, annealing for 30 s at temperatures decreasing from 65 °C to 55 °C during the first 11 cycles (with 1 °C decremental steps in cycles 1 to 11), and ending with an extension step at 72 °C for 30 s. A total of 40 cycles were performed. PCR products were then purified with a Qiaquick Gel Extraction kit (Qiagen) and sequenced (200–500 ng DNA) with PCR forward or reverse primers by using an ABI-PRISM Big Dye terminator sequencing kit on an ABI-PRISM 310 Genetic Analyser (Applied Biosystems). Different sequences of a given locus were given allele numbers, and each unique combination of alleles (multilocus allelic profile) was assigned a sequence type (ST). Single point polymorphisms were assessed by resequencing DNA from two separate PCR experiments.

Computer analysis of sequence data. Clustering of the 29 isolates from the matrix of pairwise similarities between the allelic profiles was performed using the START program (<http://www.mlst.net>) by the unweighted pair-group method with arithmetic means (UPGMA). Nucleotide sequences were aligned by using BioEdit Sequence Alignment Editor (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Mean numbers of nucleotide differences between alleles and ratios of nonsynonymous to synonymous substitutions (*dN/dS*) were calculated to test the degree of selection operating on

Table 1. Genetic polymorphism of the ten virulence-associated genes analysed by MLST

Fragment positions are based on the genome of *C. difficile* strain 630. *dN/dS*, ratio of nonsynonymous to synonymous substitutions. I, inosine.

Gene	PCR and sequencing primers (5'→3')	Position of the fragment analysed within each gene	No. of alleles	No. of polymorphic sites	Percentage of polymorphic sites	Mean genetic distance between alleles	Mean G+C content (%)	<i>dN/dS</i>
<i>cwp66</i>	Forward: GTTACTCAIATTGGTGGIITAGG Reverse: TTIGITCTTCCCCATCTAGAIAC	1234–1745	11	293	57.2	0.326	28.86	0.509
<i>cwp84</i>	Forward: GGGGTAAACATGAGAAAG Reverse: GGAAGCTCCATTTACTACT	556–1127	17	30	5.4	0.022	31.91	0.038
<i>fbp68</i>	Forward: TTAGCATAAGTAGAGTTCGTCAAG Reverse: GTCTATCAAGCTTAACAGATATACTC	509–929	15	34	8.0	0.022	27.43	0.162
<i>fliC</i>	Forward: ATGAGAGTTAATACAAATGTAAGTGC Reverse: CTATCCTAATAATTGTAAGTCC	361–794	10	67	15.1	0.101	31.65	0.408
<i>fliD</i>	Forward: ATGTCAAGTATAAGTCCAGTAAG Reverse: TTCAACAAAASCTCTTAAYTCATC	54–490	12	57	13.0	0.057	30.46	0.197
<i>groEL</i>	Forward: ATACTGAATTAGATGCTGTTGAAG Reverse: TCTGAAGTAGTTGCTCTACTTG	584–1054	9	13	2.8	0.010	30.21	0.024
<i>slpA</i>	Forward: ATGAATAAGAAAAIITAGCAATIGC Reverse: TTTTAAATCATCTACATAATCTTTTATITC	115–533	16	387	98.7	0.658	30.81	1.929
<i>tcdA</i>	Forward: AGATTCCATATTTACATGACAATAT Reverse: GTATCAGGCATAAAGTAATATACTTT	7658–7990	8	19	5.7	0.056	29.43	0.052
<i>tcdB</i>	Forward: GTGTAACCTACTTTCATAACACCAG Reverse: GGTGGAGCTGCTTCAATTGGAGAG	5622–5972	5	10	2.9	0.009	29.32	0.857
<i>tcdD</i>	Forward: TAACTCAGTAGATGATTGCAAG Reverse: AGCCTTATTAACAGCTTGCTCTAG	92–476	7	12	3.1	0.046	20.36	0.052

Fig. 1. Genomic locations of the loci analysed in the *C. difficile* 630 genome. The ten loci analysed in this study are shown in bold; the other genes were analysed previously by MLST (Lemée *et al.*, 2004a). The *cdtA* and *cdtB* genes are also shown, but they were found in a complete form in only five of the 29 strains studied.

a locus, using the START program. Gene trees were constructed using the neighbour-joining method and bootstrapping algorithms contained in the MEGA software version 2.1 (Kumar *et al.*, 2001).

The index of association (I_A) between alleles (Smith *et al.*, 1993) was used to test for linkage disequilibrium between alleles of the ten loci analysed (colonization factors and PaLoc genes). The observed variance in the distribution of allelic mismatches in all pairwise comparisons of the allelic profiles was compared to that expected in a freely recombining population (linkage equilibrium). The significance of the difference in the observed and expected variances was evaluated by computing the maximum variance in the distribution of allelic mismatches obtained using 100 randomizations of the dataset (<http://www.mlst.net>).

Finally, the present data, obtained from the multilocus study of virulence-associated genes, were compared to the data obtained from previous MLST analysis of *C. difficile* housekeeping genes (Lemée *et al.*, 2004a).

Nucleotide sequence accession numbers. Nucleotide sequences of the internal fragment genes analysed in this work will be deposited in the GenBank database under accession numbers DQ102375–DQ102379 (for *cdtA*), DQ117049–DQ117053 (for *cdtB*), DQ117054–DQ117074 (for *cwp66*), DQ117075–DQ117103 (for *cwp84*), DQ117104–DQ117132 (for *fbp68*), DQ117133–DQ117161 (for *fliC*), DQ117162–DQ117189 (for *fliD*), DQ117190–DQ117218 (for *groEL*), DQ117219–DQ117240 (for *slpA*), DQ117241–DQ117265 (for *tcdA*), DQ117266–DQ117288 (for *tcdB*) and DQ117289–DQ117311 (for *tcdD*).

RESULTS AND DISCUSSION

Allelic variation in the virulence-associated genes of *C. difficile*

Data reporting the allelic variations in the ten virulence-associated genes are summarized in Table 1. The number of individual alleles for each gene ranged from five for *tcdB* to 17 for *cwp84* [some amplification attempts remained unsuccessful for *cwp66* (eight isolates), *fliD* (one isolate), *slpA* (seven isolates) and *tcdB* (two isolates), despite the use of various primers spanning the whole gene, and were interpreted as null alleles]. The number of polymorphic sites on a given locus varied from 10 (2.9 %) for *tcdB* to 387 (98.7 %) for *slpA*, and the mean genetic distance between alleles of a given locus varied from 0.009 for *tcdB* to 0.658 for *slpA*. Most polymorphisms resulted in synonymous substitutions, except for *cwp66*, *tcdB* and *slpA*, for which the ratios of nonsynonymous to synonymous substitutions ($dN/dS=0.509$, 0.857 and 1.929 , respectively) indicate a potential role of environmental selective pressure in their evolution.

Overall, six of the ten virulence-associated genes (*cwp84*, *fbp68*, *groEL*, *tcdA*, *tcdB* and *tcdD*) exhibit a moderate polymorphism. Concerning the PaLoc (which includes *tcdA*, *tcdB* and *tcdD*), only twenty different toxinotypes have previously been described in *C. difficile* (Rupnik *et al.*, 2003). Our data confirm this moderate polymorphism, and give new information about the genetic variability of the *tcdD* regulator gene, which has never been explored. Concerning colonization factors, the *cwp84*, *fbp68* and *groEL* genes have been reported as highly conserved, whereas *fliC*, *fliD*, and especially *cwp66* and *slpA*, revealed a higher polymorphism (Hennequin *et al.*, 2001, 2003; Savariau-Lacomme *et al.*, 2003). *Cwp66* and *slpA* harbour a two-domain structure, with the cell-surface-exposed domain exhibiting high genetic variability (Calabi *et al.*, 2001; Savariau-Lacomme *et al.*, 2003). We encountered difficulties in the amplification of precisely this variable domain in several strains; thus, further studies are needed to design extragenic primers, which would allow a better amplification of the variable domains of the *cwp66* and *slpA* genes.

Multilocus sequence data analysis of virulence-associated genes

Multilocus allelic profile analysis generated 22 different STs among the 29 isolates. The majority of these (21/22 STs) were represented by single isolates. This trend reflects both the great diversity of the isolates, selected from various origins, and the high polymorphism of several of the virulence-associated genes studied in this work. The results of clustering of the allelic profiles by UPGMA are shown in Fig. 2.

Among human isolates, PMC isolates belonged to eight different STs, which did not define PMC-specific clusters in the population studied. In addition, no correlation could be established between virulence-associated gene patterns

and clinical presentations. Although toxins A and B are recognized as the main virulence factors of *C. difficile* (Borriello, 1998), additional virulence factors and/or the host response (Kyne *et al.*, 2000, 2001) could explain the various clinical presentations. Determinants involved in differences in intrinsic virulence between toxigenic strains could be linked to polymorphism in toxin A and B genes (Rupnik *et al.*, 1998), to the level of toxin A and B production (Borriello *et al.*, 1987; Spigaglia & Mastrantonio, 2002), to various patterns of genes encoding colonization factors (Eveillard *et al.*, 1993; Hennequin *et al.*, 2001, 2003; Poilane *et al.*, 1998; Tasteyre *et al.*, 2001; Waligora *et al.*, 2001), or to additional production of a binary toxin (Barbut *et al.*, 2005; McEllistrem *et al.*, 2005). However, despite previous analyses of patterns of colonization factors (Savariau-Lacomme *et al.*, 2003; Tasteyre *et al.*, 2000), toxins A and B (Rupnik *et al.*, 1998), regulators of toxins A and B (Spigaglia & Mastrantonio, 2002), and even binary toxin (Barbut *et al.*, 2005), no correlation has been established to date between genotype and clinical presentation. Since MLST of housekeeping genes also does not define hypervirulent lineages (Lemée *et al.*, 2004a), the genetic determinants of virulence are probably linked not only to the genes encoding virulence factors, but also to genes involved in the regulation of their expression.

Toxigenic A–B+ variant isolates clustered together in ST1 (although recovered from Japan, the USA and from different French hospitals), and defined a specific lineage which diverged by more than 0.9 linkage distance from the other isolates. Toxin A variant (A–B+) strains have been reported from a number of countries (Alfa *et al.*, 2000; Borriello *et al.*, 1992; Brazier *et al.*, 1999; Kato *et al.*, 1998; Lysterly *et al.*, 1998). The majority of them are similar to the reference strain of serogroup F (Delmee *et al.*, 1985) and share the same genetic changes in the toxin A gene (a 1.7 kb deletion in the repetitive 3'-end domain and a nonsense mutation at position 47) (Rupnik *et al.*, 1998). They also share the same DNA profile by PCR ribotyping (Stubbs *et al.*, 1999) and PFGE (Pituch *et al.*, 2001), although a few different genotypes have recently been described (Rupnik *et al.*, 2003). These results, together with previous MLST data (Lemée *et al.*, 2004a), support the hypothesis of a low genetic diversity of A–B+ variant strains and of the international spread of this phylogenetic lineage. However, the origin and evolutionary history of this lineage within the *C. difficile* species need further investigation.

Animal isolates (ST3, ST4, ST14, ST16 and ST21) were intermixed with human isolates, and thus did not constitute a distinct subpopulation. It has been speculated that domestic animals could constitute a reservoir of *C. difficile* isolates and a potential source for human acquisition (O'Neill *et al.*, 1993). Since the present multilocus analysis of virulence-associated genes, MLST analysis of housekeeping genes (Lemée *et al.*, 2004a) and PCR ribotyping (Arroyo *et al.*, 2005) have not characterized any host specificity, we can also presume that animal isolates could constitute a source for human community infections.

Fig. 2. Dendrogram showing clustering analysis (UPGMA) of the 29 *C. difficile* isolates after data compilation from the ten virulence-associated genes. *AAD, antibiotic-associated diarrhoea; D, diarrhoea; PMC, pseudomembranous colitis; AC, asymptomatic carriage; U, unknown. †H, human; A, animal.

Composite sequence-based analysis of virulence-associated genes

Since a single nucleotide substitution in a given locus is sufficient to generate a new allele, some of the loci studied generated numerous different alleles. However, there may be a bias in the estimation of the genetic distance between isolates in this approach, which considers alleles differing by only one point mutation or by multiple polymorphic sites in the same manner. Therefore, clustering based on a 3402 bp composite sequence obtained from the different loci analysed should be more appropriate to characterize the relationships between isolates (Fig. 3). Only twenty-three of the 29 isolates were analysed, because of the lack of toxin gene sequence data for non-toxigenic (A-B-) isolates and for 1599 and FM18 isolates. The identity between the 23 composite sequences was found to be between 93.3 and 100 %. A dendrogram created from the matrix of pairwise sequence divergences of composite sequences is shown in

Fig. 3(A). Four divergent lineages were characterized: cluster 2 contains all A-B+ variants (including one PMC isolate), cluster 4 contains two PMC isolates (FM16 and 630) and one animal (dog) isolate (654778), cluster 3 corresponds to two human AAD isolates (4984 and CD3), and cluster 1 contains all the remaining isolates, with intermixing of human and animal isolates and of PMC and AAD isolates. Only cluster 4 was not retrieved from composite sequence analysis of housekeeping genes (Fig. 3B). Overall, clustering generated from composite sequence analysis was congruent with clustering of multilocus allelic profiles.

Comparative evolution of virulence-associated genes and housekeeping genes

A quantitative analysis of the linkage between alleles from the different virulence-associated loci was performed by calculating the index of association (I_A) (Smith *et al.*, 1993). When all the 29 isolates were included in the analysis,

Fig. 3. Dendrograms showing genetic relationships of *C. difficile* isolates based on the composite sequence of eight virulence-associated genes (*cwp84*, *fbp68*, *fliC*, *fliD*, *groEL*, *tcdA*, *tcdB* and *tcdD*) (A) or six housekeeping genes (B) (Lemée *et al.*, 2004a). In (A), 23 of the 29 isolates are shown (isolates 1446, 669984, FM5 and 650299 lack the pathogenicity locus and *tcdA*, *tcdB*, *tcdD* and *tcdB* sequences are lacking for the 1599 and FM18 isolates). In (B), the 29 *C. difficile* isolates are shown.

significant linkage disequilibrium was detected (I_A , 5.17). At the level of STs (one isolate from each ST, to avoid bias due to a possible epidemic population structure), significant linkage disequilibrium was also detected (I_A , 2.01). Previous observations have shown that isolates from diverse geographic origins can be included in the same toxinotype (Rupnik *et al.*, 1998), PCR ribotype (Stubbs *et al.*, 1999) or ST (Lemée *et al.*, 2004a), and that PCR ribotyping and toxinotyping display a strong correlation (Rupnik *et al.*, 2001). All these data are consistent with a strong clonal population structure for *C. difficile*.

When examining monolocus dendrograms of housekeeping genes and virulence-associated genes (some of which are

shown in Fig. 4), topologies derived from all the housekeeping genes and from *fbp68*, *groEL*, *tcdA*, *tcdB* and *tcdD* loci were found to be congruent, suggesting a probable co-evolution. Conversely, two discrepancies appeared from the analysis of these monolocus dendrograms. First, *fliC* (Fig. 4) and *fliD* loci exhibited global clustering comparable to housekeeping genes, except for three isolates (630, FM16 and 654778), which were here closely related to the A–B+ isolates. Second, *cwp66* and *slpA* (Fig. 4) loci displayed a much greater polymorphism than the other virulence-associated loci and generated a dendrogram with a topology distinct from the other loci: some isolates were closely related in *cwp66* and *slpA* trees and remote from each other in the other trees (1599 and 630, FM5 and 6761), and

Fig. 4. Dendrograms showing genetic relationships of the *C. difficile* isolates based on allele sequences of individual housekeeping loci (*aroE*, shikimate dehydrogenase, *dutA*, dUTP pyrophosphatase) and virulence-associated loci. Twenty-five isolates are shown for the *tcdA* locus (four isolates are A–B–). Twenty-two isolates are shown for the *slpA* locus (lack of amplification from the other isolates).

Fig. 5. Dendrograms showing genetic relationships between *C. difficile* binary toxin *cdtA* and *cdtB* alleles characterized in the present work (in bold type) and extracted from GenBank.

conversely, some isolates closely related in other trees were remote in *cwp66* and *slpA* trees (1599 and FM18, 1446 and 669984). This suggests that the evolution of these clusters of genes (*fliC/fliD* and *cwp66/slpA*) includes recombinational events and/or strong environmental selective pressure (*slpA* exhibits a high *dN/dS* ratio), and that the variability of these genes may allow them to overcome the host immune response. In addition, the present data for the *slpA* gene suggest that serogroups, which correlate with the polymorphism of the variable domain of *slpA* (Karjalainen *et al.*, 2002), depend also on recombinational events and selective pressure.

Detection of the binary toxin genes in the 29 *C. difficile* isolates

Binary toxin genes were detected in five (17%) of the 29 isolates. Four were A+B+ toxigenic isolates (three AAD and one PMC human isolates) and one was a non-toxigenic isolate (isolate 650299, recovered from a cow). Both *cdtA* and *cdtB* genes were detected in positive isolates. The genetic relationships between the five alleles identified by sequencing of internal fragments and several homologous sequences deposited in GenBank were examined (Fig. 5). Alleles corresponding to the present isolates clustered in two groups (1599/FM18 and CD3/4985/650299), which were also well characterized by housekeeping genes and by virulence-associated genes (except *cwp66* and *slpA*). This also suggests a vertical co-evolution of binary toxin genes with housekeeping genes and several virulence-associated genes, including the PaLoc genes.

Multilocus analysis of virulence-associated genes versus MLST of housekeeping genes

The present multilocus analysis of virulence-associated genes was found to be more discriminant than a previous MLST analysis of housekeeping genes (Lemée *et al.*, 2004a). MLST is considered to be a molecular tool for epidemiological purposes, whose main advantage is to provide unambiguous sequence data. However, its discriminatory power is limited by the choice of housekeeping genes analysed. Therefore, multilocus analysis of virulence-associated genes has been proposed as an alternative sequence-typing method (Zhang *et al.*, 2004), reflecting

here short-term epidemiology, whereas MLST reflects long-term epidemiology. The selection of two or three highly variable virulence-associated genes, such as *slpA*, *cwp66*, and/or *fliC*, would be well suited to short-term epidemiological sequence-based analysis of *C. difficile*.

Finally, the aim of this multilocus analysis was to study the evolution of virulence-associated genes encoding colonization factors, toxins A and B, the sigma factor TcdD, and binary toxin, and to compare it with the evolution of housekeeping genes. The analysis confirms the clonal population structure of *C. difficile*, and reveals a co-evolution of several of the virulence-associated genes studied (including toxins A and B and binary toxin genes) with housekeeping genes, whereas flagellin, *cwp66* and *slpA* genes may undergo recombination events and/or environmental selective pressure.

ACKNOWLEDGEMENTS

The authors thank Karine Maillard (Caen, France), Frederic Barbut (Paris, France), Francine Mory (Nancy, France) and Haru Kato (Tokyo, Japan) for supplying strains.

REFERENCES

- Alfa, M. J., Kabani, A., Lyerly, D. & 7 other authors (2000). Characterization of a toxin A-negative, toxin B-positive strain of *Clostridium difficile* responsible for a nosocomial outbreak of *Clostridium difficile*-associated diarrhea. *J Clin Microbiol* **38**, 2706–2714.
- Arroyo, L. G., Kruth, S. A., Willey, B. M., Staempfli, H. R., Low, D. E. & Weese, J. S. (2005). PCR ribotyping of *Clostridium difficile* isolates originating from human and animal sources. *J Med Microbiol* **54**, 163–166.
- Barbut, F., Mario, N., Meyohas, M. C., Binet, D., Frottier, J. & Petit, J. C. (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.
- Barbut, F., Decre, D., Lalande, V. & 9 other authors (2005). Clinical features of *Clostridium difficile*-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. *J Med Microbiol* **54**, 181–185.

- Bartlett, J. G. (1994). *Clostridium difficile*: history of its role as an enteric pathogen and the current state of knowledge about the organism. *Clin Infect Dis* 18, S265–S272.
- Bartlett, J. G., Moon, N., Chang, T. W., Taylor, N. & Onderdonk, A. B. (1978). Role of *Clostridium difficile* in antibiotic-associated pseudo-membranous colitis. *Gastroenterology* 75, 778–782.
- Bidet, P., Lalande, V., Salauze, B., Burghoffer, B., Avesani, V., Delmee, M., Rossier, A., Barbut, F. & Petit, J. C. (2000). Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *J Clin Microbiol* 38, 2484–2487.
- Bongaerts, G. P. & Lyerly, D. M. (1994). Role of toxins A and B in the pathogenesis of *Clostridium difficile* disease. *Microb Pathog* 17, 1–12.
- Borriello, S. P. (1998). Pathogenesis of *Clostridium difficile* infection. *J Antimicrob Chemother* 41, 13–19.
- Borriello, S. P., Ketley, J. M., Mitchell, T. J., Barclay, F. E., Welch, A. R., Price, A. B. & Stephen, J. (1987). *Clostridium difficile* – a spectrum of virulence and analysis of putative virulence determinants in the hamster model of antibiotic-associated colitis. *J Med Microbiol* 24, 53–64.
- Borriello, S. P., Wren, B. W., Hyde, S., Seddon, S. V., Sibbons, P., Krishna, M. M., Tabaqchali, S., Manek, S. & Price, A. B. (1992). Molecular, immunological, and biological characterization of a toxin A-negative, toxin B-positive strain of *Clostridium difficile*. *Infect Immun* 60, 4192–4199.
- Brazier, J. S., Stubbs, S. L. & Duerden, B. I. (1999). Prevalence of toxin A negative/B positive *Clostridium difficile* strains. *J Hosp Infect* 42, 248–249.
- Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G. & Fairweather, N. (2001). Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol Microbiol* 40, 1187–1199.
- Calabi, E., Calabi, F., Phillips, A. D. & Fairweather, N. F. (2002). Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect Immun* 70, 5770–5778.
- Cartmill, T. D., Panigrahi, H., Worsley, M. A., McCann, D. C., Nice, C. N. & Keith, E. (1994). Management and control of a large outbreak of diarrhoea due to *Clostridium difficile*. *J Hosp Infect* 27, 1–15.
- Cerquetti, M., Molinari, A., Sebastianelli, A., Diociaiuti, M., Petruzzelli, R., Capo, C. & Mastrantonio, P. (2000). Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microb Pathog* 28, 363–372.
- Delmee, M., Homel, M. & Wauters, G. (1985). Serogrouping of *Clostridium difficile* strains by slide agglutination. *J Clin Microbiol* 21, 323–327.
- Dhalluin, A., Lemée, L., Pestel-Caron, M., Mory, F., Leluan, G., Lemeland, J. F. & Pons, J. L. (2003). Genotypic differentiation of twelve *Clostridium* species by polymorphism analysis of the triosephosphate isomerase (*tpi*) gene. *Syst Appl Microbiol* 26, 90–96.
- Eveillard, M., Fourel, V., Barc, M. C., Kerneis, S., Coconnier, M. H., Karjalainen, T., Bourlioux, P. & Servin, A. L. (1993). Identification and characterization of adhesive factors of *Clostridium difficile* involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Mol Microbiol* 7, 371–381.
- Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collignon, A., Barc, M., Bourlioux, P. & Karjalainen, T. (2001). GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology* 147, 87–96.
- Hennequin, C., Janoir, C., Barc, M. C., Collignon, A. & Karjalainen, T. (2003). Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* 149, 2779–2787.
- Johnson, S. & Gerding, D. N. (1998). *Clostridium difficile*-associated diarrhea. *Clin Infect Dis* 26, 1027–1034.
- Karjalainen, T., Saumier, N., Barc, M. C., Delmee, M. & Collignon, A. (2002). *Clostridium difficile* genotyping based on *slpA* variable region in S-layer gene sequence: an alternative to serotyping. *J Clin Microbiol* 40, 2452–2458.
- Kato, H., Kato, N., Watanabe, K. & 7 other authors (1998). Identification of toxin A-negative, toxin B-positive *Clostridium difficile* by PCR. *J Clin Microbiol* 36, 2178–2182.
- Kelly, C. P., Pothoulakis, C. & LaMont, J. T. (1994). *Clostridium difficile* colitis. *N Engl J Med* 330, 257–262.
- Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245.
- Kyne, L., Warny, M., Qamar, A. & Kelly, C. P. (2000). Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. *N Engl J Med* 342, 390–397.
- Kyne, L., Warny, M., Qamar, A. & Kelly, C. P. (2001). Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *Lancet* 357, 189–193.
- Larson, H. E., Price, A. B., Honour, P. & Borriello, S. P. (1978). *Clostridium difficile* and the aetiology of pseudomembranous colitis. *Lancet* 1, 1063–1066.
- Lemée, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J. F. & Pons, J. L. (2004a). Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *J Clin Microbiol* 42, 2609–2617.
- Lemée, L., Dhalluin, A., Testelin, S., Matrat, M. A., Maillard, K., Lemeland, J. F. & Pons, J. L. (2004b). Multiplex PCR targeting *tpi* (triose phosphate isomerase), *tcdA* (toxin A), and *tcdB* (toxin B) genes for toxigenic culture of *Clostridium difficile*. *J Clin Microbiol* 42, 5710–5714.
- Lyerly, D. M., Neville, L. M., Evans, D. T. & 7 other authors (1998). Multicenter evaluation of the *Clostridium difficile* TOX A/B TEST. *J Clin Microbiol* 36, 184–190.
- Mani, N. & Dupuy, B. (2001). Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc Natl Acad Sci U S A* 98, 5844–5849.
- McElliott, M. C., Carman, R. J., Gerding, D. N., Genheimer, C. W. & Zheng, L. (2005). A hospital outbreak of *Clostridium difficile* disease associated with isolates carrying binary toxin genes. *Clin Infect Dis* 40, 265–272.
- Moncrief, J. S., Barroso, L. A. & Wilkins, T. D. (1997). Positive regulation of *Clostridium difficile* toxins. *Infect Immun* 65, 1105–1108.
- O'Neill, G., Adams, J. E., Bowman, R. A. & Riley, T. V. (1993). A molecular characterization of *Clostridium difficile* isolates from humans, animals and their environments. *Epidemiol Infect* 111, 257–264.
- Pituch, H., van den Braak, N., van Leeuwen, W., van Belkum, A., Martirosian, G., Obuch-Woszczatynski, P., Luczak, M. & Meisel-Mikolajczyk, F. (2001). Clonal dissemination of a toxin-A-negative/toxin-B-positive *Clostridium difficile* strain from patients with antibiotic-associated diarrhea in Poland. *Clin Microbiol Infect* 7, 442–446.
- Poilane, I., Karjalainen, T., Barc, M. C., Bourlioux, P. & Collignon, A. (1998). Protease activity of *Clostridium difficile* strains. *Can J Microbiol* 44, 157–161.
- Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C. & Delmee, M. (1998). A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *J Clin Microbiol* 36, 2240–2247.
- Rupnik, M., Brazier, J. S., Duerden, B. I., Grabnar, M. & Stubbs, S. L. (2001). Comparison of toxinotyping and PCR ribotyping of

Clostridium difficile strains and description of novel toxinotypes. *Microbiology* **147**, 439–447.

Rupnik, M., Kato, N., Grabnar, M. & Kato, H. (2003). New types of toxin A-negative, toxin B-positive strains among *Clostridium difficile* isolates from Asia. *J Clin Microbiol* **41**, 1118–1125.

Savariau-Lacomme, M. P., Lebarbier, C., Karjalainen, T., Collignon, A. & Janoir, C. (2003). Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of *Clostridium difficile*. *J Bacteriol* **185**, 4461–4470.

Seddon, S. V., Hemingway, I. & Borriello, S. P. (1990). Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model. *J Med Microbiol* **31**, 169–174.

Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.

Spigaglia, P. & Mastrantonio, P. (2002). Molecular analysis of the pathogenicity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among *Clostridium difficile* clinical isolates. *J Clin Microbiol* **40**, 3470–3475.

Stubbs, S. L., Brazier, J. S., O'Neill, G. L. & Duerden, B. I. (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.

Stubbs, S., Rupnik, M., Gibert, M., Brazier, J., Duerden, B. & Popoff, M. (2000). Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol Lett* **186**, 307–312.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmee, M., Collignon, A., Bourlioux, P. & Barc, M. C. (2000). Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different serogroups. *J Clin Microbiol* **38**, 3179–3186.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H. & Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun* **69**, 7937–7940.

Toyokawa, M., Ueda, A., Tsukamoto, H., Nishi, I., Horikawa, M., Sunada, A. & Asari, S. (2003). Pseudomembranous colitis caused by toxin A-negative/toxin B-positive variant strain of *Clostridium difficile*. *J Infect Chemother* **9**, 351–354.

Waligora, A. J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. & Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect Immun* **69**, 2144–2153.

Zhang, W., Jayarao, B. M. & Knabel, S. J. (2004). Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Appl Environ Microbiol* **70**, 913–920.