*Journal of Medical Microbiology* (2008), **57**, 771–775 DOI 10.1099/jmm.0.47743-0

# Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (*tcdB*) sequences

Richard A. Stabler, Lisa F. Dawson, Leslie T. H. Phua and Brendan W. Wren

Correspondence
Brendan W Wren
brendan.wren@lshtm.ac.uk

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

The reported incidence and mortality of *Clostridium difficile*-associated disease has increased significantly, which in part is likely to be due to the emergence of a new, highly virulent strain in North America and Europe. This epidemic strain, referred to as BI/NAP1/027, has increased virulence, attributed to overexpression of the two toxin-encoding genes, *tcdA* and *tcdB*, which may be due to truncation of the negative regulator (*tcdC*) by a 1 bp deletion. In a previous study of whole-genome comparisons using microarray analysis of 75 *C. difficile* isolates, it was noted that the 20 027 strains, which formed a hypervirulent clade, possessed a unique hybridization pattern for the 7 toxin B microarray reporters. This unique pattern was conserved in all of these 027 strains. The pattern was different for the 55 non-027 strains tested. These data, along with the knowledge that 027 strains are toxinotype III (i.e. possess a complete *tcdB* gene of comparable size to toxin reference strain VPI 10463), suggest that the sequence of the N-terminal binding domain of toxin B must be divergent from *C. difficile* strain 630 (and the other 55 strains tested). Additionally, these 027 strains had comparable hybridization patterns across the whole microarray, as well as for *tcdB*. Therefore, it was suggested that they share a similar, novel N-terminal binding domain. The aim of this study was to ascertain the sequence variation in *tcdB* from eight characterized BI/NAP1/027 strains. The study confirmed significant sequence variation of *tcdB* from the sequenced strain 630 and slight variation in *tcdB* among the eight 027 strains. These results suggest that toxin B from 027 strains may have a different binding capacity compared with its less-virulent counterparts and may, in addition to the mutated *tcdC* regulator, be responsible for the increased virulence of 027 strains.

Received 2 November 2007
Accepted 21 December 2007

## INTRODUCTION

*Clostridium difficile* is known to produce a number of factors that contribute to its virulence, including two related toxins: toxin A, an enterotoxin, and toxin B, a cytotoxin (Lyerly *et al.*, 1986). The toxin A- and B-encoding genes (*tcdA* and *tcdB*) are part of a 19.6 kb pathogenicity locus consisting of five genes including *tcdC*, which negatively regulates toxin synthesis (Matamouros *et al.*, 2007). However, in recent years, increasing numbers of pathogenic *C. difficile* strains have been reported with truncated versions of toxin A and/or B, and there appears to be considerable variation in the pathogenicity locus (Borriello *et al.*, 1992; Depitre *et al.*, 1993; Pituch *et al.*, 2003; Toyokawa *et al.*, 2003; van den Berg *et al.*, 2004).

Disturbingly, the reported incidence of *C. difficile*-associated disease has increased significantly in the last decade, with a new group of highly virulent strains causing outbreaks of increased severity in North America and Europe (Loo *et al.*, 2005; McDonald *et al.*, 2005; Pépin *et al.*, 2005; Warny *et al.*, 2005). The origin of these strains (called BI/NAP1/027 hypervirulent strains) is uncertain, although it has been proposed that increased use of fluoroquinolone antibiotics may give the strains a selective advantage (Pépin *et al.*, 2005). It has been proposed that the increased virulence could be attributed to inactivation of the negative regulator (*tcdC*), resulting in overexpression of *tcdA* and *tcdB*. However, an 18 bp deletion within *tcdC* has been identified in both hypervirulent and non-hypervirulent isolates (McDonald *et al.*, 2005), and is not responsible for hypertoxigenicity (Matamouros *et al.*, 2007). Interestingly, a 1 bp deletion at the 5′ end of *tcdC*,

Abbreviation: REA, restriction endonuclease analysis.

Primer sequences, and nucleotide and amino acid sequence alignments of regions e–g, are available as supplementary data with the online version of this paper

which results in truncation of the protein, has been described for BI/NAP1/027 isolates, but was also found in other distantly related isolates (Curry *et al.*, 2007). There are conflicting data for several other non-027 strains carrying a similar deletion in the toxin regulatory gene. Thus, other factors are likely to play a role in the increased virulence of 027 strains.

In a previous study, we carried out whole-genome comparison of 75 *C. difficile* strains by microarray analysis (Fig. 1). We found a distinctive lack of hybridization of the 3′ end of *tcdB* in 027 strains (Fig. 2) compared with most other strains, including the sequenced strain 630 (Stabler *et al.*, 2006). Given the importance of toxin B to the pathogenesis of *C. difficile*, the aim of this study was to ascertain whether the hypervirulent 027 *tcdB* genes were distinct and whether there was sequence microheterogeneity within *tcdB* among the 027 strains.

## METHODS

*C. difficile* was cultured using standard methods in an anaerobic cabinet. All *C. difficile* isolates were kindly supplied by Dale Gerding (Edward Hines Jr Veterans Affairs Hospital, Hines, IL, USA). Genomic DNA was prepared using a Wizard genomic DNA purification kit according to the manufacturer's guidelines. PCR amplifications were performed using standard methods with primers described in Supplementary Table S1 (available with the online journal). Sequencing reactions were set up using an ABI BigDye terminator cycle sequencing kit according to the manufacturer's guidelines, and processed using an ABI Prism 3730 DNA Analyzer. Chromas software version 2.31 was used to extract the sequence information, which was aligned using CLUSTAL_W.

## RESULTS AND DISCUSSION

Whole-genome analysis of 75 *C. difficile* strains using a microarray has previously identified four distinct lineages

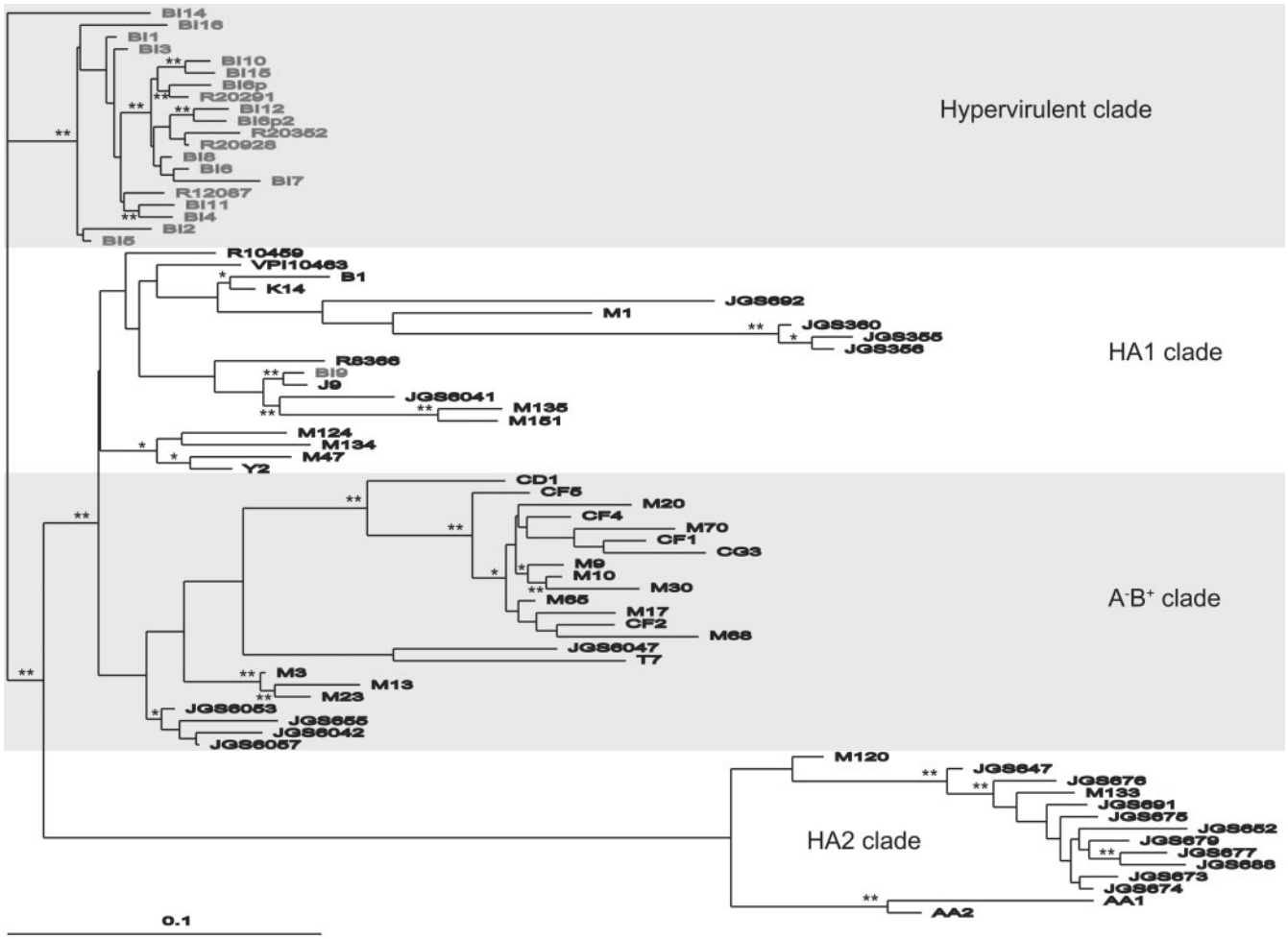


**Fig. 1.** Phylogenetic relationship of strains associated with different clinical outcomes and animal sources represented as four major clades. Strains are designated at the end of the branches and are shaded according to ribotype. Grey, ribotype 027; black, non-027 ribotypes. Branches with ** have a *P* value of 1.0 and represent 100 % of all phylogenies showing a given topology; * indicates a *P* value of 0.98. Adapted from Stabler *et al.* (2006).

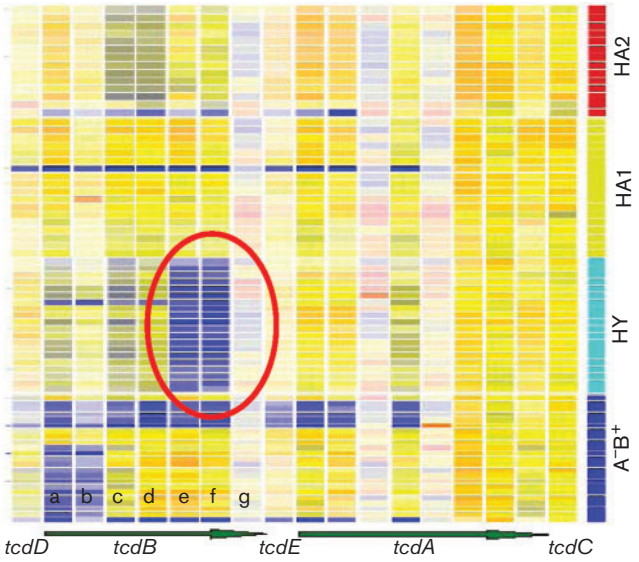


**Fig. 2.** Gene map of the toxin pathogenicity locus (*tcdD, tcdB, tcdE, tcdA* and *tcdC*). Each horizontal line represents a single-strain competitive genomic hybridization, whilst the vertical colour bars represent the presence (yellow lines) or absence/high divergence (blue lines) of each gene from CD0659 (*tcdD*) on the left through to CD0664 (*tcdC*) on the right. Clade blocks: $A^-B^+$ (dark blue) represents strains in the $A^-B^+$ clade; HY (light blue) represents strains in the hypervirulent clade, HA1 (yellow) represents strains in the human and animal 1 clade; and HA2 (red) represents strains in the human and animal 2 clade. *tcdB* was subdivided into seven microarray reporters (regions) labelled a–g. The red ellipse indicates deletion/high divergence in *tcdB* (regions e, f and g) found only in the hypervirulent clade strains. Adapted from Stabler *et al.* (2006).

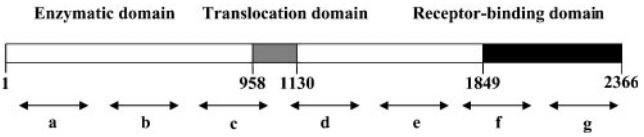


**Fig. 3.** Schematic representation of the location of BμG@S *C. difficile* 630 *tcdB* microarray PCR reporters on the BμG@S microarray (http://bugs.sgul.ac.uk). Structure adapted from Rupnik *et al.* (2005) showing aa 1–2366; aa 958–1130 indicate the hydrophobic region and aa 1849–2366 the repetitive domain. Arrows indicate the approximate location of microarray reporter regions a–g: region a, aa 10–269; region b, aa 415–678; region c, aa 755–1010; region d, aa 1080–1326; region e, aa 1389–1626; region f, aa 1704–1964; and region g, aa 2089–2346.

(Stabler *et al.*, 2006). One clade comprised 20 of the 21 ribotype 027 strains tested and was named the hypervirulent clade (Fig. 1). Seventeen of the hypervirulent isolates represented seventeen subcategories of the BI restriction endonuclease analysis (REA) group (BI-1 to BI-17). Other studies have demonstrated that different BI REA patterns are more than 80 % related by dendrographic analysis (McDonald *et al.*, 2005). REA subtypes BI-1 to BI-5 were found in historic isolates obtained before 2001 (McDonald *et al.*, 2005), which are fluoroquinolone-sensitive. REA subtypes BI-6 to BI-17 were isolated after 2001; these are fluoroquinolone-resistant and are known as the epidemic hypervirulent isolates (McDonald *et al.*, 2005). Microarray analysis of the hypervirulent clade identified a distinct set of genes for these isolates that were classified as 'absent/highly divergent' including a unique deletion/divergence pattern for the 3′ end of *tcdB* (Fig. 2) (Stabler *et al.*, 2006). It has been shown that BI/NAP1/027 strains are toxinotype III (McDonald *et al.*, 2005) with an $A^+B^+$ toxinotype. The pathogenicity locus of toxinotype III strains is almost identical to VPI 10463, the toxinotype 0 reference strain, the only difference being the loss of a few restriction sites.

**Table 1.** Nucleotide sequence identity of hypervirulent isolate (test) *tcdB* genes (regions e–g) compared with *C. difficile* 630 (reference)

BI-9 was an apparent ribotype 027 that clustered in HA1 clade (Fig. 1), BI-1 to BI-3 were representative of historic fluoroquinolone-sensitive strains, BI-10 to BI-14 were representative of epidemic fluoroquinolone-resistant strains and BI-14 was a hypervirulent clade outlier (Fig. 1).

| Reference | Test | Identity (%) |
|---|---|---|
| **Region e** | | |
| CD630 | BI-1 | 89 |
| CD630 | BI-2 | 89 |
| CD630 | BI-3 | 89 |
| CD630 | BI-9 | 100 |
| CD630 | BI-10 | 89 |
| CD630 | BI-11 | 89 |
| CD630 | BI-13 | 88 |
| CD630 | BI-14 | 89 |
| **Region f** | | |
| CD630 | BI-1 | 87 |
| CD630 | BI-2 | 87 |
| CD630 | BI-3 | 88 |
| CD630 | BI-9 | 100 |
| CD630 | BI-10 | 87 |
| CD630 | BI-11 | 87 |
| CD630 | BI-13 | 88 |
| CD630 | BI-14 | 88 |
| **Region g** | | |
| CD630 | BI-1 | 91 |
| CD630 | BI-2 | 92 |
| CD630 | BI-3 | 92 |
| CD630 | BI-9 | 99 |
| CD630 | BI-10 | 92 |
| CD630 | BI-11 | 92 |
| CD630 | BI-13 | 91 |
| CD630 | BI-14 | 95 |

**Table 2.** Nucleotide sequence identity of hypervirulent isolate (test) *tcdB* genes (regions e–g) compared with *C. difficile* BI-1 (reference)

See Table 1 for isolate descriptions.

| Reference | Test | Identity (%) |
|---|---|---|
| **Region e** | | |
| BI-1 | BI-2 | 99 |
| BI-1 | BI-3 | 99 |
| BI-1 | BI-9 | 89 |
| BI-1 | BI-10 | 99 |
| BI-1 | BI-11 | 99 |
| BI-1 | BI-13 | 98 |
| BI-1 | BI-14 | 99 |
| **Region f** | | |
| BI-1 | BI-2 | 96 |
| BI-1 | BI-3 | 99 |
| BI-1 | BI-9 | 87 |
| BI-1 | BI-10 | 99 |
| BI-1 | BI-11 | 99 |
| BI-1 | BI-13 | 99 |
| BI-1 | BI-14 | 99 |
| **Region g** | | |
| BI-1 | BI-2 | 100 |
| BI-1 | BI-3 | 100 |
| BI-1 | BI-9 | 91 |
| BI-1 | BI-10 | 100 |
| BI-1 | BI-11 | 100 |
| BI-1 | BI-13 | 99 |
| BI-1 | BI-14 | 93 |

The loci are the same size in both toxinotypes, indicating that there are no deletions present in toxinotype III strains (Rupnik *et al.*, 1998). However, microarray probes did not bind to the 3′ receptor-binding domain of *tcdB* (Fig. 3), which was most likely due to sequence divergence of the 3′ receptor-binding domain of *tcdB*, as the presence of this region was confirmed by PCR. Therefore, the sequence was sufficiently different from the *C. difficile* 630 sequence to prevent binding to the microarray reporters (Fig. 2). The regions showing sequence divergence were e, f and g (Fig. 3). These three regions were sequenced in this study from a selection of historic and epidemic BI subtypes including BI-9, which was of interest as it did not cluster in the hypervirulent clade (Fig. 1) and did not possess the unique *tcdB* microarray hybridization pattern (Fig. 2).

The sequence data for *tcdB* regions e, f and g identified deviations from the *C. difficile* 630 sequence (Table 1, Supplementary Fig. S1a–c available with the online journal). All BI REA sequences tested except BI-9 showed sequence deviation from *C. difficile* 630 of greater than 5 % in all three regions (Table 1). This variation in sequence would be sufficiently different under the stringent microarray conditions to fail to hybridize to the reported elements on the microarray and therefore the sequence would appear to be absent/highly divergent. These data explain why the microarray data presented previously by Stabler *et al.*

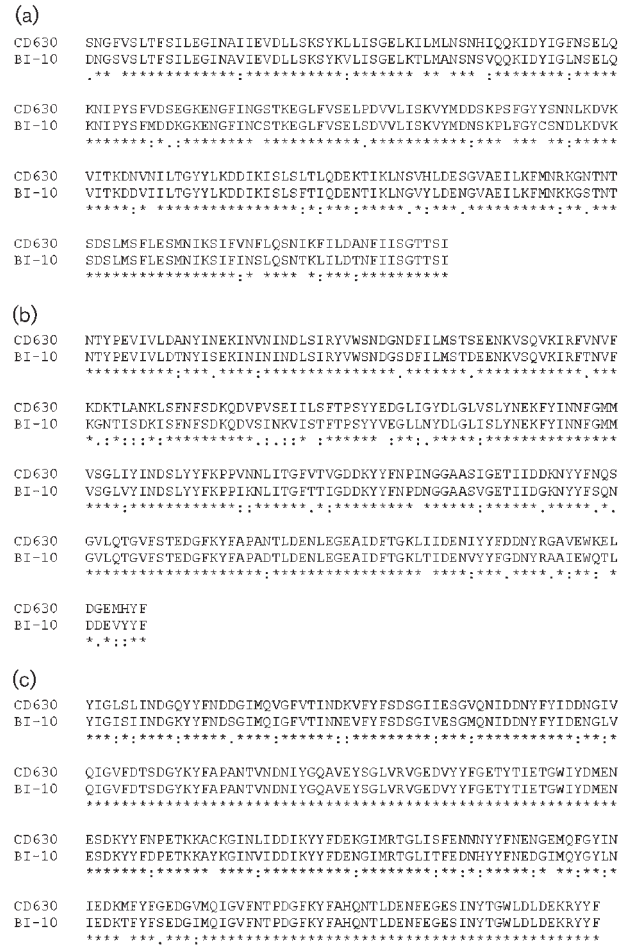


**Fig. 4.** Alignment of the predicted amino acid sequence from *C. difficile* 630 and BI-10 for TcdB regions e (a), f (b) and g (c).

(2006) identified these regions of *tcdB* as absent/highly divergent. Sequence identity between REA type BI-1 and the other BI REA types tested, with the exception of BI-9, was 98–99 % for section e, 96–99 % for section f and 99–100 % for section g (Table 2). The exception was BI-14 for region f, which had only 93 % identity to BI-1 (Table 2), but had a higher identity to 630 (95 %; Table 1). This was not unexpected, as BI-14 was a distinct outlier in the hypervirulent clade (Fig. 1) and appears to be a mid-point between the two *tcdB* genotypes for this region. Sequence analysis of the three regions of *tcdB* for the REA BI-9 isolate demonstrated 100 % identity for regions e and f, and 99.7 % for region g to *C. difficile* 630. This is also consistent with the microarray data and the fact that BI-9 is not genotypically similar to the other BI isolates. The alteration in *tcdB* sequence resulted in a markedly altered TcdB protein sequence but did not include any translation terminators that would lead to a truncated protein (Fig. 4, Supplementary Fig. S2a–c available with the online journal).

In conclusion, seven 027 strains were found to have a different 3′-end *tcdB* sequence extending over 2.9 kb. By

extrapolation of previous microarray data (Stabler *et al.*, 2006), it appears that this is a unique feature of the hypervirulent 027 strains with the exception of BI-9. It is possible that the divergent C-terminal binding domain may affect the binding capability of toxin B, potentially increasing the cell range to which the toxin binds. It is possible that the increased virulence observed in 027 strains may be due an altered binding avidity and/or alteration in host-cell specificity of toxin B, in addition to increased toxin production through deletions in *tcdC*.

## ACKNOWLEDGEMENTS

This research was supported by the Wellcome Trust.

## REFERENCES

**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.

**Curry, S. R., Marsh, J. W., Muto, C. A., O'Leary, M. M., Pasculle, A. W. & Harrison, L. H. (2007).** *tcdC* genotypes associated with severe TcdC truncation in an epidemic clone and other strains of *Clostridium difficile*. *J Clin Microbiol* **45**, 215–221.

**Depitre, C., Delmee, M., Avesani, V., L'Haridon, R., Roels, A., Popoff, M. & Corthier, G. (1993).** Serogroup F strains of *Clostridium difficile* produce toxin B but not toxin A. *J Med Microbiol* **38**, 434–441.

**Loo, V. G., Poirier, L., Miller, M. A., Oughton, M., Libman, M. D., Michaud, S., Bourgault, A. M., Nguyen, T., Frenette, C. & other authors (2005).** A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med* **353**, 2442–2449.

**Lyerly, D. M., Phelps, C. J., Toth, J. & Wilkins, T. D. (1986).** Characterization of toxins A and B of *Clostridium difficile* with monoclonal antibodies. *Infect Immun* **54**, 70–76.

**Matamouros, S., England, P. & Dupuy, B. (2007).** *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol Microbiol* **64**, 1274–1288.

**McDonald, L. C., Killgore, G. E., Thompson, A., Owens, R. C., Jr, Kazakova, S. V., Sambol, S. P., Johnson, S. & Gerding, D. N. (2005).** An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med* **353**, 2433–2441.

**Pépin, J., Saheb, N., Coulombe, M. A., Alary, M. E., Corriveau, M. P., Authier, S., Leblanc, M., Rivard, G., Bettez, M. & other authors (2005).** Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. *Clin Infect Dis* **41**, 1254–1260.

**Pituch, H., van Belkum, A., van den Braak, N., Obuch-Woszczatynski, P., Sawicka-Grzelak, A., Verbrugh, H., Meisel-Mikolajczyk, F. & Luczak, M. (2003).** Clindamycin-resistant, toxin A-negative, toxin B-positive *Clostridium difficile* strains cause antibiotic-associated diarrhea among children hospitalized in a hematology unit. *Clin Microbiol Infect* **9**, 903–904.

**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., Pabst, S., Rupnik, M., von Eichel-Streiber, C., Urlaub, H. & Soling, H. D. (2005).** Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* toxin B (TcdB) by host cells. *Microbiology* **151**, 199–208.

**Stabler, R. A., Gerding, D. N., Songer, J. G., Drudy, D., Brazier, J. S., Trinh, H. T., Witney, A. A., Hinds, J. & Wren, B. W. (2006).** Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *J Bacteriol* **188**, 7297–7305.

**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.

**van den Berg, R. J., Claas, E. C., Oyib, D. H., Klaassen, C. H., Dijkshoorn, L., Brazier, J. S. & Kuijper, E. J. (2004).** Characterization of toxin A-negative, toxin B-positive *Clostridium difficile* isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping. *J Clin Microbiol* **42**, 1035–1041.

**Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E. & McDonald, L. C. (2005).** Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**, 1079–1084.