Detection of *Clostridium difficile* toxin genes by PCR: sequence variation may cause false-negative results

The genes *tcdA* and *tcdB* have been detected by PCR to investigate the epidemiological characteristics of *Clostridium difficile*. PCR-based detection has also been applied to diagnose *C. difficile* infection. However, published primers have been partially evaluated. In this study, 50 primer sets were scanned in 180 *tcdA/B* sequences and 75 no match events were identified. These results indicate that new *C. difficile* detection targets are necessary.

*Clostridium difficile* is a common nosocomial pathogen that causes hospital-acquired diarrhoea and in severe cases it can cause lethal pseudomembranous colitis. *C. difficile* infection (CDI) diagnosis usually depends on the detection of two major virulence factors, namely, toxins A and B, or the coding genes *tcdA* and *tcdB* that are located in the pathogenicity locus (PaLoc). Toxins A and B or toxin B only are produced by toxigenic strains because the 3′-end of the *tcdA* gene is truncated (Rupnik, 2008). Toxins can be detected by various methods. For instance, PCR-based toxin gene detection is preferred rather than immunoassays detecting toxins because the former exhibits a higher sensitivity than the latter. Furthermore, PCR-based methods provide not only identification tools but also typing approaches, such as toxinoyping. Using PCR amplification of part of the *tcdA* and *tcdB* genes followed by digestion with restriction enzymes (PCR-RFLP), a toxinoyping method was established that can classify the strains into different groups (Rupnik, 2008, 2010). Some sequence changes, such as deletion in *tcdA*, can be also detected by the primers (Kato *et al.*, 1998). Based on the produced toxins, two major phenotypes are classified: toxin A positive, toxin B positive (*A*+*B*+) and toxin A negative, toxin B positive (*A*−*B*+). These phenotypes are partially, but not entirely, correlated with deletions in *tcdA*.

These methods combined with other typing methods, such as PCR ribotyping and multi-locus sequence typing, are commonly used to describe the epidemiological characteristics of CDI. However, the genomic diversity of *C. difficile* is significantly high, and nucleotide mutations have occurred in *tcdA/B* genes; as such, false-negative results may be obtained in the detection of *tcdA* or *tcdB* (He *et al.*, 2010; Cheng *et al.*, 2011; Du *et al.*, 2014). Using seven available PaLoc sequences, we found three base mismatches in the primer set NK104/NK105, which is widely used to detect *tcdB*. Two other sets of *tcdA* primers contain more than three mismatches per set, which may cause false-negative results in *tcdA* detection (Cheng *et al.*, 2011). Therefore, these primers should be evaluated in a wider scope. We sequenced 70 strains, including 66 from five provinces in China (11 in Shandong, 7 in Henan, 12 in Guangdong, 21 in Beijing, 15 in Shanghai), and one each from USA, UK, France and Japan. We also obtained sequences of 25 strains from GenBank: one in Switzerland, one in Ireland, one in Belgium, three in UK, two in France (one *tcdA* sequence is absent), seven in Canada (three *tcdA* sequences are absent), and ten in USA (six *tcdA* sequences are absent). We obtained 85 *tcdA* sequences and 95 *tcdB* sequences; we then divided these sequences into 16 types denoted A01–A16 and B01–B16, respectively (Du *et al.*, 2014). In the current research, 24 pairs of published PCR primers for *tcdA* and 26 for *tcdB* were characterized. We detected single nucleotide polymorphisms (SNPs) in *tcdA* and *tcdB* sequence types distinguished previously in the regions corresponding to these PCR primers; we counted the number of mismatches and then evaluated the primers by bioinformatics analysis. SNP analysis results revealed that mutations in *tcdA* and *tcdB* are widely distributed in the whole sequences of these two genes, not in particular regions (Fig. 1). The total SNP rate in *tcdA* is lower than that in *tcdB*, and the numbers of mismatches with the *tcdA* primers are less than those with the *tcdB* primers (Fig. 1a and Table S1, available in the online Supplementary Material). For *tcdA*, 21 out of 24 primer sets could bind most of the *tcdA* sequence types (14/16), particularly A01–A14, and one or two base mismatches could be detected. The deletions in the 3′-end caused 22 no match events in 19 sets with two other sequence types, namely, A15 and A16. Two primer sets, A1F/A1769R and CD-tcdA-F/CD-tcdA-R, completely matched all of the *tcdA* sequence types, which are the most efficient choices for *tcdA* detection. To distinguish A+ and A− strains, we found NK11/NK9 completely matched all of the A+ strains but could not map A− strains, which is also efficient. The no match events caused by deletions could not cause false-negative results and biases such as those caused by base mismatches in primers; therefore, these events are not important in *tcdA* detection.

Sequence variation and base mismatches in the primers for *tcdB* detection were higher than those for *tcdA* detection (Fig. 1b and Table S2). Only eight out of 26 primer sets could bind all 16 *tcdB* sequence types. All of the sets exhibited mismatches with all of the types, and 53 no match events from 14 sets were identified. The mismatches mainly occurred in B09–B16, particularly in B16 of strain 8864, corresponding to a high variation in the whole sequence. Furthermore, a real-time PCR primer set of tcdB-1F/tcdB-1R and a multiplex PCR primer set of CD-tcdB-F1/F2 and CD-tcdB-R1/R2 formed by two forward primers and two reverse primers, contained fewer mismatches than the other sets. In particular, both primer sets displayed one or two mismatches in the four sequence types. These two sets are the most efficient tools to detect *tcdB*. Therefore, no primer could completely cover all of the *tcdB* sequence types.
Base mismatches and complete no match events were widely distributed in the primers used to detect tcdA and tcdB, particularly tcdB. Although many PCR primers have been designed, only a few of these primers can be used to map all of the toxin gene sequence types. The conditions of primer matching in tcdA are more efficient than those in tcdB, this observation is consistent with the different divergences of these two genes. In particular, tcdA is more conserved than tcdB.
specifically one sequence from VPI10463
include incorrect designation of primers
2014). Other causes of mismatches may
we could choose other conserved regions
performances to diagnose many variant
sequences revealed that sufficient reference
many primers exhibited the best matches
respectively). Our results showed that
been obtained from strains 1470 and 5340
accession nos Z23277 and AF217292,
respectively); two tcdB sequences have also
GeneOhm Cdiff assay (BD Diagnostics), have shown good
many variant C.
difficile strains (Eastwood et al., 2009; Zidaricˇ et al., 2011); maybe conserved
regions have been chosen as targets in
The mortality caused by CDI has
increased, and epidemiological
characteristics have changed remarkably
since new emerging PCR ribotypes became prevalent, such as RT017; most strains of
this ribotype are ST37 (Freeman et al., 2010). RT017 is prevalent in Asia and
China; the hypervirulent strain RT027 (ST1) was identified for the first time in
mainland China recently (Collins et al., 2013; Wang et al., 2014). Considering the high rate of mismatches in primers of tcdB, we could choose other conserved regions as targets or degenerate primers. Hence, an accurate and sensitive detection method is the key to early diagnosis, timely treatment and epidemiological survey. Furthermore, whole genome sequencing and analysis can be potentially applied in typing and epidemiological survey (Baker & Ellington, 2014).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (grant nos. 81101218, 81201322 and 81301402).

Bo Cao,†† Yuanyuan Zhang,†† Jinfeng Bian,†† Ying Cheng,†† Jing Wang,† Hongbing Jia,‡ Wen Zhang,†,‡ Jinxing Lu,†,‡ Chen Chen†,‡ and Pengcheng Du†,‡
†State Key Laboratory for Infectious Disease Prevention and Control, and National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, PR China
‡Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310003, PR China
‡Key Laboratory of Surveillance and Early-warning on Infectious Disease, China-Japan Friendship Hospital, Beijing 100029, PR China

Correspondence: Pengcheng Du (dupengcheng@icdc.cn)
Chen Chen (chenchen@icdc.cn)
†These authors contributed equally to this work.

Abbreviations: CDI, C. difficile infection; PaLoc, pathogenicity locus; SNP, single nucleotide polymorphism.

One supplementary table is available with the online Supplementary Material.