Defining criteria to interpret multilocus variable-number tandem repeat analysis to aid *Clostridium difficile* outbreak investigation

George Broukhanski,1,2 Andrew Simor1 and Dylan R. Pillai1,2

1Department of Laboratory Medicine & Pathobiology, University of Toronto, Toronto, Ontario, Canada
2Ontario Agency for Health Protection and Promotion, Toronto, Ontario, Canada

PFGE is currently the North American standard for surveillance for *Clostridium difficile* but lacks discriminatory power to aid outbreak investigation. A further limitation to PFGE is the high baseline rate of the epidemic North American pulsotype (NAP) 1 strain in hospitals. Multilocus variable-number tandem repeat analysis (MLVA) appears to have superior discriminatory power but criteria to define clonality have not been set. We conducted surveillance for toxin-positive *C. difficile* infection (CDI) at a single academic health sciences centre between September 2009 and April 2010. Seventy-four patient specimens resulting in 86 discrete CDI episodes were subjected to PFGE and MLVA. Results were analysed using Bionumerics software to generate phylogenetic trees and coupled to patient demographic data. Amongst the NAP1 strains, two distinct clusters were identified by MLVA using 90 % similarity as a cut-off by Manhattan distance-based clustering, four clusters using 95 % and seven clusters using 97 %. Population analysis conducted on multiple colonies (*n* = 25) demonstrated that 1–3 % difference in MLVA types was typical for a single individual. Typing was also conducted in the context of institutional outbreaks (*n* = 42, three outbreaks) in order to determine clusters within the NAP1 strain. By combining longitudinal surveillance with epidemiological information, single specimen population analysis and typing in the context of institutional outbreaks, we conclude that the use of the Manhattan distance-based clustering with a cut-off of 95–97 % is capable of distinguishing outbreak clones from sporadic isolates.

**INTRODUCTION**

*Clostridium difficile* infection (CDI) is a major hospital-acquired infection in patients treated with antibiotics. Currently outbreak investigation of CDI in North America is performed using PFGE, and despite wide use there are factors which limit its utility due to a lack of discriminatory power in outbreak investigation, especially in geographical regions where baseline rates of the North American pulsotype (NAP) 1 prevalence are high (Pillai et al., 2010; Mulvey et al., 2010). For this reason, alternative typing methods might be better suited for this purpose. It was previously shown that multilocus variable-number tandem repeat analysis (MLVA) is a highly useful technique for investigation of outbreaks of *C. difficile* in hospitals (Marsh et al., 2006; Fawley et al., 2008; Goorhuis et al., 2009), especially in the case of clonal lineages, but criteria to distinguish outbreak from sporadic isolates are not yet defined. We assessed the utility of MLVA for hospital outbreak typing and propose criteria to define clusters during outbreaks. In this study, we use a combination of longitudinal surveillance of CDI, nosocomial outbreak typing and population analysis of stool specimens in order to validate cut-off values used to determine true clusters in the context of institutional outbreaks.

**METHODS**

Stool specimens (*n* = 86) were obtained from 74 consecutive patients with CDI at an academic health sciences centre over a period of 8 months (September 2009–April 2010). Specimens positive by Tox A/B II enzyme immunoassay (TechLab) were processed using ethanol for spore selection and plated on *C. difficile* moxalactam–norfloxacin (CDMN) agar to grow under anaerobic conditions for 48 h (Aspinall & Hutchinson, 1992). Grown colonies were tested with UV light, Gram staining and L-proline aminopeptidase test (PRO disk; Hardy Diagnostics) to confirm their identity. For PFGE typing, Peptone Yeast Glucose Broth (Anaerobe Systems) was inoculated with a single colony. PFGE was performed following standard procedures (Killgore et al., 2008). For MLVA, one colony from a CDMN plate was resuspended in 100 µl InstaGene (Bio-Rad), heated at 100 °C for
10 min, vortexed for 10 s and centrifuged at 8000 g for 2 min. Two microlitres of the supernatant was used for amplification. A previously published MLVA protocol (van den Berg et al., 2007) was modified by eliminating F3Cd and H9Cd loci from analysis and modifying primer sequences to further accommodate sequence variation in C. difficile strains (Broukhanski et al., 2011). To aid identification of presumptive NAP1, NAP7 and NAP8 strains, the deletion within the final extension step of 30 min at 60 °C was used. Amplicons were diluted 1:20 with distilled water and 1 µl was transferred to 11 µl formamide mixed with 0.1 µl LIZ500 DNA ladder (Life Technologies) per tube. The plate was heated for 3 min at 95 °C and snap-cooled to 4 °C. The Applied Biosystems 3130xl Genetic Analyzer (Life Technologies) was used to separate the amplicons and GENESCAN software (Life Technologies) was used to identify their length. Repeat numbers per locus were entered into BioNumerics software v. 6 (Applied Maths) to perform cluster analysis and coupled to patient demographic data, such as hospital unit location, date of admission and date of positive result by enzyme immunoassay. The Manhattan distance measure algorithm (a numerical coefficient which sums differences in repeat units and size of tcdC deletion) and Ward clustering were used to generate the dendrograms (Ichikawa et al., 2010). Size of deletion in the tcdC gene (18 bp in NAP1 and 38 bp in NAP7/8 strains) was treated as a number (90 %, 39) the diversity of strains is minimal (2–3 % difference using the Manhattan distance measure) and is limited to one locus (most often A6Cd or C6Cd). Only in one specimen did all 25 colonies have an identical MLVA profile. Few specimens (<10 %, n=4) had strains with a difference in more than one locus, which, most likely, represent polyclonal infections (Fig. 3). Clustering analysis of MLVA data of all isolates collected over the surveillance

RESULTS

For the hospital-based surveillance arm of this study, demographic breakdown revealed that the mean age of patients was 72 and 62 % were females. Sixty-four patients (86.5 %) had single episodes with 10 (14.5 %) patients experiencing recurrent CDI. PFGE analysis demonstrated that 32 CDI cases were due to the NAP1 pulsotype (37 %), 3 (3.5 %) due to NAP7, 2 (2.3 %) due to NAP4, 1 (1.2 %) due to NAP2, with the remainder (56 %) due to arbitrarily assigned pulsotypes. Typing with MLVA demonstrated variability among NAP1 strains (Fig. 1). Two clusters of NAP1 could be identified when 90 % similarity was used as a cut-off by Manhattan distance-based clustering, four clusters when 95 % similarity was used as a cut-off and seven clusters when 97 % was used as a cut-off. MLVA of specimens from three other hospital outbreaks (Fig. 2) was used to evaluate the degree of variation of MLVA profiles in isolated C. difficile strains. These results showed that the degree of heterogeneity of MLVA types within an outbreak varies by hospital. Nevertheless, observed clusters in these three outbreaks had MLVA profiles with less than 5 % difference. Typing multiple colonies per specimen (population analysis) with MLVA demonstrated that usually (90 %, n=39) the diversity of strains is minimal (2–3 % difference using the Manhattan distance measure) and is limited to one locus (most often A6Cd or C6Cd). Only in one specimen did all 25 colonies have an identical MLVA profile. Few specimens (<10 %, n=4) had strains with a difference in more than one locus, which, most likely, represent polyclonal infections (Fig. 3). Clustering analysis of MLVA data of all isolates collected over the surveillance

![Fig. 1. Comparison of PFGE (right) and MLVA (left) for NAP1 clinical isolates. The numerical values represent repeat units at five separate loci based on the MLVA sequencing method, except for the final value, which represents the 18 bp deletion in the tcdC gene found in NAP isolates. The vertical lines represent three cut-off values at 15 %, 9 % and 3 % difference based on the Manhattan distance algorithm.](image-url)
period shows two well-defined groups of NAP1 strains (Fig. 4). In an attempt to identify whether any demographic parameters can be linked to these groups, we paired MLVA results with patient demographic data. The best match was observed when MLVA data were coupled with the date of admission. Two MLVA clusters using a 5% cut-off based on the Manhattan distance-based clustering correlated with admission in January and March (Fig. 5).

**DISCUSSION**

The optimal method of typing *C. difficile* strains in the context of institutional outbreaks has yet to be defined. PFGE and ribotyping are the two methods used most commonly in North America and Europe, respectively. Other typing methods such as restriction endonuclease analysis, toxino typing and multilocus sequence typing are used most often for research purposes. MLVA was identified as the most discriminatory typing method (Kuijper *et al.*, 2009), and when seven typing techniques were compared, it was shown that MLVA and restriction endonuclease analysis are the only methods suitable for outbreaks investigation (Killgore *et al.*, 2008). However, MLVA is thought to be too discriminatory for this purpose and its ability to identify multiple MLVA types within a single specimen could obscure epidemiological links if only one isolate per specimen is typed (Tanner *et al.*, 2010). Therefore, the challenge is to define cut-offs that empower MLVA to define clusters within the context of hospital outbreaks. We show here that changing the cut-off values from 90% to 97% based on the Manhattan distance-based clustering results in the identification of two or seven clusters, respectively (Fig. 1). Hence, the cut-off value chosen can dramatically alter the interpretation of clonality.

It is known that the *C. difficile* genome is highly plastic (Sebaihia *et al.*, 2006; Scaria *et al.*, 2010) and therefore overly discriminatory cut-off values may exaggerate differences between clinical isolates. To help identify what is the most appropriate cut-off value, we analysed MLVA profiles among specimens collected during institutional outbreaks where presumably a single clone was spreading.
Results of this study, performed on specimens from three representative hospitals, demonstrate that three possible MLVA patterns can be observed (Fig. 2). In those sets of specimens, when a single ‘clone’ was present during the outbreak based on epidemiological information, natural variation was observed in the range of 3–5% within the outbreak cluster. Results of population analysis of multiple isolates from a single specimen demonstrated that the majority of specimens (90%) have variation in one or two loci (especially A6Cd and C6Cd). This variation was approximately 1–3% difference based on the Manhattan distance-based clustering, suggesting that the cut-off value should not be less than 3%. Combination of these data (MLVA of hospital isolates and population analysis of multiple isolates per specimen) indicates that a cut-off value greater than 3% but less than 5% based on the Manhattan distance-based clustering identifies true outbreak clones. Using this cut-off value range, MLVA typing was used during longitudinal hospital surveillance and, when coupled to epidemiological information, was able to define clusters that correlated with the date of admission thus validating the interpretive criteria proposed here (Fig. 5). These results will enable infection control investigators to distinguish outbreak clones from sporadic isolates during a hospital outbreak.

Conclusions
MLVA is a highly discriminatory typing method capable of defining clusters of NAP1 isolates when the correct cut-off
values are used. Population analysis suggests that a cut-off value of <3% based on the Manhattan distance-based clustering is overly discriminatory due to the natural variation of *C. difficile* genotypes in a single specimen, but a cut-off of 3–5% should capture true clusters based on the five loci we have used. Hospital surveillance using clinical and epidemiological data reinforces that a cut-off of 3–5% is able to identify true clusters of NAP1. Further multi-centre studies relying on both MLVA data and clinical epidemiological information are required to validate the 3–5% cut-off value proposed here.

**ACKNOWLEDGEMENTS**

We thank the medical laboratory technicians in the Molecular Surveillance department at the Public Health Laboratory in Toronto for their expert technical assistance. We thank the participating institutions for their active role in *C. difficile* surveillance. We also thank Kate Henbest for technical assistance.

**REFERENCES**


