Title: Development and evaluation of double locus sequence typing for molecular epidemiological investigations of Clostridium difficile.

Authors: Stojanov M, Magalhaes B, Terletsky V, Basset P, Prod'hom G, Greub G, Senn L, Blanc DS

Journal: European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology

Year: 2016 Feb

Volume: 35

Issue: 2

Pages: 175-81

DOI: 10.1007/s10096-015-2525-4

Development and evaluation of double locus sequence typing for molecular epidemiological investigations of *Clostridium difficile*.

Miloš Stojanov[1], Bàrbara Magalhaes[1], Valery Terletsky[1], Patrick Basset[1], Guy Prod'hom[2], Gilbert Greub[2], Laurence Senn[1], Dominique S. Blanc[1,2].

[1]Service of Hospital Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland

[2]Institute of Microbiology, Lausanne University Hospital, Lausanne, Switzerland

Corresponding author:

Miloš Stojanov

Institute of Microbiology

University Hospital of Lausanne

Rue du Bugnon 46

1011 Lausanne

Switzerland

Email: milos.stojanov@chuv.ch

Abstract

Despite the development of novel typing methods based on whole genome sequencing, most laboratories still rely on classical molecular methods for outbreak investigation or surveillance. Reference methods for *Clostridium difficile* include ribotyping and pulsed-field gel electrophoresis, which are band-comparing methods often difficult to establish and which require reference strain collections. Here we present the double locus sequence typing (DLST) scheme as a tool to analyse *C. difficile* isolates. Using a collection of clinical *C. difficile* isolates recovered during a one-year period, we evaluated the performance of DLST and compared the results to multilocus sequence typing (MLST), a sequence-based method that has been used to study the structure of bacterial populations and highlight major clones. DLST had a higher discriminatory power compared to MLST (Simpson's index of diversity of 0.979 versus 0.965) and successfully identified all isolates of the study (100% typeability). Previous studies showed that discriminatory power of ribotyping was comparable to that of MLST, thus DLST might be more discriminatory than ribotyping. DLST is easy to establish and provides several advantages, including absence of DNA extraction (PCR is performed on colonies), no specific instrumentation, low cost and unambiguous definition of types. Moreover, implementation of DLST typing scheme on an Internet database, such previously done for *Staphylococcus aureus* and *Pseudomonas aeruginosa* (http://www.dlst.org), will allow users to easily obtain the DLST type by submitting directly sequencing files and will avoid problems associated with multiple databases.

Introduction

During the last few decades *Clostridium difficile* has arisen as a major human pathogen mainly associated with nosocomial infections [1-3]. Disruption of the gut microbiota homeostasis due to use of antibiotics allows *C. difficile* to colonize the colon and cause a whole range of intestinal diseases, ranging from mild diarrhea to life-threatening diseases like pseudomembranous colitis [3, 4]. Novel genotypes associated to more severe clinical outcomes and outbreaks were increasingly reported throughout Europe and United States [5, 6].

Molecular typing of clinical isolates, allowing rapid epidemiological tracking of *C. difficile* infections (CDI), could lead to development of more effective infection control measures that might reduce the spread of *C. difficile* between patients, yet reducing the risk of outbreaks. Recent advances in sequencing technology allowed the use of whole genome sequencing as tool for epidemiological tracking of CDI [7, 8]. Nevertheless, this technology remains restricted to few centers for research purposes and classical molecular typing methods remain essential for epidemiological investigations. The actual reference method for *C. difficile* is PCR-ribotyping, which relies on the amplification of variable DNA segments comprised between 16S and 23S rRNA genes of *rrn* operons [9]. After electrophoresis, DNA banding patterns are compared to those of reference strains and a PCR ribotype is assigned. The requirement of a reference strain collection and the lack of standardization make *de novo* implementation of this method challenging and comparison of results between laboratories often difficult. In contrast to band pattern methods, sequence-based methods are portable and definitive, offering good intra- and inter-laboratory reproducibility [10]. Multilocus sequence typing (MLST) uses the nucleotide sequence data of several (generally seven) housekeeping genes. This method is considered the gold standard to understand the global population

structure of a bacterial species [11, 12]. However, it is rather expensive and its discriminatory power is often relatively low to investigate local epidemiology.

We recently developed a typing scheme involving single strand sequencing of small fragments of only two highly variable loci (double locus sequence typing, DLST). This typing scheme allowed us to investigate the epidemiology of two major nosocomial pathogens, *Staphylococcus aureus* and *Pseudomonas aeruginosa* [13, 14]. Using this approach, a definitive type is assigned to strains, based on the sequence of the two alleles, and typing results can be unambiguously compared between laboratories with the help of a web-based database (http://www.dlst.org). In this study, we developed the DLST typing scheme for *C. difficile* in order to investigate the epidemiology of this bacterium. To validate the method, DLST results were compared to MLST using a collection of strains isolated at the University Hospital of Lausanne during a one-year period.

Material and methods

Bacterial isolates

A total of 109 *C. difficile* clinical isolates (toxigenic and non toxigenic) were collected from hospitalized patients at the Lausanne University Hospital during the year 2012 (Supplementary file 1). Stools of symptomatic patients were tested for the presence of *C. difficile* glutamate dehydrogenase (GDH) antigen and the A/B toxins with an immunochromatographic test (C. Diff. Quik Chek Complete®, Alere, Orlando, FL, USA). If positive, stools were cultured and *C. difficile* isolates were identified using standard microbiological methods. In addition to clinical isolates, a collection of 18 strains (Supplementary file 1) with known PCR ribotypes was included in the study (strains were kindly provided by F. Barbut, see Acknowledgments section). Presence of toxins was assessed by a 5-plex PCR assay targeting the toxin genes *tcdA*, *tcdB*, *cdtA* and *cdtB*, in addition to 16S rDNA as previously described [15].

Molecular procedures for DLST

Primers for amplification of loci are shown in Table 1. When required, they were designed using Primer3 software (version 2.3.4, [16]). PCR amplification of the loci was performed with the KAPA 2G Robust HotStart PCR kit (KAPA Biosystems, Cape Town, South Africa). *C. difficile* colonies were used directly as template for PCR, by transferring a small amount of colony biomass in the reaction tubes using sterile toothpicks. PCR amplification was carried out in 30-µl reaction containing 1.25 U of Taq DNA polymerase, 1X Reaction Buffer B, 0.4 µM of each primer, and 0.2 mM of each dNTP. PCR cycling conditions consisted of 3 min of initial denaturation at 95 °C, 30 cycles of 15 sec at 95 °C, 30 sec at 60 °C and 45 sec at 72 °C and a final extension step of 3 min at 72 °C. Sequencing reactions were performed with the Big Dye Terminator kit 3.1 (Applied Biosystems, Carlsbad, CA, USA) and purification of

sequencing products was performed with the BigDye XTerminator kit (Applied Biosystems), according to the manufacturer's instructions. Purified samples were analyzed with the ABI 3130xl sequencer (Applied Biosystems), according to standard protocols. For sequences of unsatisfactory quality, the whole procedure was repeated and if the absence of sequence was again obtained after the second assay, a null allele (0) was assigned to the isolate.

Allele assignment for DLST

Sequences were analyzed using the BioNumerics software version 7.0 (Applied Maths, Sint-Martens-Latem, Belgium). Increasing allele numbers were assigned sequentially to new alleles.

Comparison between DLST and MLST

Chromosomal DNA for MLST analysis was extracted with the GenElute bacterial genomic DNA Kit (Sigma-Aldrich, Buchs, Switzerland), following manufacturer's specifications. MLST was performed according to the typing scheme proposed by Griffiths et al. [17]. Sequencing was performed as for DLST (see above).

Discriminatory power of MLST and DLST was evaluated by calculating the Simpson's index of diversity (ID), which is the probability that two strains sampled randomly in the collection belong to two different types [18]. An ID value of 1 would indicate that the typing method was able to distinguish each isolate and, conversely, an index of 0 would indicate that all isolates belong to an identical type. This coefficient was determined via an online tool (http://biophp.org/stats/discriminatory_power/demo.php).

The degree of congruence between DLST and MLST was calculated using the adjusted Wallace coefficient (AW), which indicates the probability that two strains belonging to one type by one method will also be classified to a same type using the other method [19]. This coefficient was determined via an online tool (http://darwin.phyloviz.net/ComparingPartitions/).

Epidemiological investigation

Probable epidemiological links between CDI cases were suspected when two or more patients were hospitalized in the same ward, under overlapping periods of time. If those patients carried isolates with the same DLST type, these epidemiological links were considered as possible.

Results

Development of DLST scheme

A literature search was conducted to identify highly variable loci present in *C. difficile* genomes [20, 21]. Four loci (TR6, TR10, A6 and C6) were selected and tested *in silico* on eight *C. difficile* genomes available in the NCBI database. Among these, TR6, TR10 and C6 were retained for further analyses (locus A6 was not found in all strains). Specific primers located in conserved parts of C6, TR6 and TR10 loci were used to determine the sequence of the variable region containing the repeat units. As expected, amplicons of different sizes were obtained for each locus, confirming the genomic variability at these sites. Sequencing of amplicons was performed on both ends for a subset of strains and the best-performing sequencing direction (forward or reverse) was consequently chosen (Table 1). A trimming start located in the conserved region was determined for each locus and lengths of alleles were selected according to the variants that had the shortest variable region (Table 2). Allele sequences of C6, TR6 and TR10 loci and MLST types were successfully determined for all strains, with the exception of a null TR10 allele in one isolate (Table 3).

The Simpson's ID was calculated for combinations of two loci and the combination C6+TR6 was found to have the highest discriminatory power, almost the same as the three loci combined together (Table 3). Considering that C6 and TR6 had a typeability of 100%, their combination was the best candidate for the DLST scheme.

The Adjusted Wallace indexes were calculated in order to compare the congruence between DLST and MLST (Table 3). The fact that $AW_{DLST \rightarrow MLST} = 0.877$ and $AW_{MLST \rightarrow DLST} = 0.514$ means that, if two strains are in the same cluster by DLST, they have about 88% chance of having the same MLST type, while conversely, the chance is only about 51%. This reflects the fact that, within our collection of strains, DLST was more discriminatory than MLST and that DLST may subdivide MLST types. Correspondence between DLST and

MLST types, as well as their distribution and presence/absence of toxin A, toxin B, and the binary toxin is shown in Table 4. For isolates with identical ribotypes (Supplementary table 2), identical DLST types were also observed, with two exceptions: a single DLST variant was observed within the 4 isolates of ribotype 027 and within the 3 isolates of ribotype 078-126.

Stability of DLST markers was evaluated by comparing two isolates from the same patient. For 11 patients, a second isolate was available after 11 to 103 days (mean 30 days, median 20 days). For 8 of them, the same DLST type was observed in both isolates, suggesting the stability of DLST markers over this period of time. For the remaining three patients, the second isolate showed a different DLST type and a different sequence type (ST), suggesting the presence of different strains rather than a genetic evolution over time.

Analysis and confirmation of transmissions

The newly developed DLST typing scheme was used to investigate possible *C. difficile* transmissions at the University Hospital of Lausanne during the year 2012. From a total of 98 symptomatic patients diagnosed with toxigenic *C. difficile* stool samples, at least one isolate was successfully recovered in 58 patients (75 isolates in total). Epidemiological maps for these patients were constructed (Supplementary table 2) and 25 possible transmissions between patients could be highlighted. For 23 of them, different DLST types were observed in isolates from linked patients, ruling out transmission event. In only two cases, isolates with the same DLST (and ST) were found for patients with epidemiological links, supporting transmission between patients.

Discussion

Despite the recent development of highly discriminatory typing methods based on whole genome sequencing, classical molecular typing remain the only available methods for most laboratories. Ideally, such typing methods should give fast and unambiguous results. In this study we developed a typing method based on two highly variable loci (DLST) and we used a local epidemiological collection of *C. difficile* isolates to evaluate the method. Our results showed that, using the C6 and TR6 loci, DLST was more discriminatory than MLST, and thus more discriminatory than ribotyping as previous studies showed that MLST had a similar discriminatory power as ribotyping [17]. The good congruence between DLST and MLST ($AW_{DLST \rightarrow MLST} = 0.877$) shows that our method is able to recognize important lineages, such as the ST1 (ribotype 027). The stability of DLST (same DLST in consecutive isolates over several weeks in the same patients) suggests no transmission occurred when two patients carry different DLST types.

Investigation of possible *C. difficile* transmissions between patients in our hospital (CDI patients with overlapping period of hospitalization) with DLST results allowed us to rule out 23 possible transmission events and to confirm only two. Therefore, suspected events of transmission based on epidemiological data can easily be investigated with DLST.

One advantage of DLST is that it assigns a definite characterization of types, allowing ongoing surveillance and thus an early detection of outbreaks or increase frequency of transmission events. However, during the 1-year period of investigation in our hospital, the number of transmissions must have been underestimated. First, for nearly half of CDI patients, no isolate was obtained and could represent the source or the recipient of a transmission event. Second, recent studies showed that a large percentage of new CDI cases resulted of transmission from asymptomatic cases [7, 22, 23]. Third, the persistence of *C. difficile* in the environment might further complicate the establishment of epidemiological

links between patients. Among nine patients belonging to the predominant DLST type in our collection, five were found in the same ward but with no apparent epidemiological link supporting the hypothesis of an environmental reservoir.

Interestingly, we did not observe the international hypervirulent clone rt027 in our collection of clinical isolates, suggesting this clone did not reached our hospital yet. However, 7 patients carried the hypervirulent clone ST11 (rt078), for which four DLST types were found (4-4, 5-4, 17-5 and 17-34) suggesting this clone did not cause outbreak.

Ribotyping is the standard typing method to study the molecular epidemiology of *C. difficile* in Europe [24, 25]. However, the need of a reference strain collection and comparison of banding patterns to discriminate isolates make the setting of this method challenging. We developed a DLST typing method that provides several advantages previously shown [13] including low cost, high portability and definitive typing. Moreover, we were able to avoid the DNA extraction step, by performing the PCR amplification of the two loci directly on *C. difficile* colonies. Standardization of the results can be easily simplified by the implementation of the typing scheme on a centralized Internet database assigning the DLST alleles such it has been done for DLST of *Staphylococcus aureus* and *Pseudomonas aeruginosa* (http://www.dlst.org). Thus, the *C. difficile* DLST typing scheme might represent a valuable alternative for existing molecular typing of this bacterium and should be tested to more diverse strain collections to confirm its promising value.

Acknowledgments

We thank F. Barbut (Hôpital Saint-Antoine, 184 rue du Faubourg Saint-Antoine, 75012 Paris, France) for providing us with *C. difficile* strains with known ribotypes (Supplementary file 1) and C. Choulat for technical assistance.

Competing interests

The authors declare that they have no competing interests.

References

[1] Freeman J, Bauer MP, Baines SD, Corver J, Fawley WN, Goorhuis B, Kuijper EJ, Wilcox MH (2010) The changing epidemiology of *Clostridium difficile* infections. Clin Microbiol Rev 23 (3):529-549

[2] Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC (2015) Burden of *Clostridium difficile* infection in the United States. N Engl J Med 372 (9):825-834

[3] Rupnik M, Wilcox MH, Gerding DN (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol 7 (7):526-536

[4] Peniche AG, Savidge TC, Dann SM (2013) Recent insights into *Clostridium difficile* pathogenesis. Curr Opin Infect Dis 26 (5):447-453

[5] Carroll KC, Bartlett JG (2011) Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. Annu Rev Microbiol 65:501-521

[6] Honda H, Dubberke ER (2014) The changing epidemiology of *Clostridium difficile* infection. Curr Opin Gastroenterol 30 (1):54-62

[7] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS (2013) Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 369 (13):1195-1205

[8] Mac Aogain M, Moloney G, Kilkenny S, Kelleher M, Kelleghan M, Boyle B, Rogers TR (2015) Whole-genome sequencing improves discrimination of relapse from reinfection

and identifies transmission events among patients with recurrent *Clostridium difficile* infections. J Hosp Infect 90 (2):108-116

[9] Bidet P, Barbut F, Lalande V, Burghoffer B, Petit JC (1999) Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. FEMS Microbiol Lett 175 (2):261-266

[10] Aires-de-Sousa M, Boye K, de Lencastre H, Deplano A, Enright MC, Etienne J, Friedrich A, Harmsen D, Holmes A, Huijsdens XW, Kearns AM, Mellmann A, Meugnier H, Rasheed JK, Spalburg E, Strommenger B, Struelens MJ, Tenover FC, Thomas J, Vogel U, Westh H, Xu J, Witte W (2006) High interlaboratory reproducibility of DNA sequence-based typing of bacteria in a multicenter study. J Clin Microbiol 44 (2):619-621

[11] Enright MC, Spratt BG (1999) Multilocus sequence typing. Trends Microbiol 7 (12):482-487

[12] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95 (6):3140-3145

[13] Kuhn G, Francioli P, Blanc DS (2007) Double-locus sequence typing using *clfB* and *spa*, a fast and simple method for epidemiological typing of methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol 45 (1):54-62

[14] Basset P, Blanc DS (2014) Fast and simple epidemiological typing of *Pseudomonas aeruginosa* using the double-locus sequence typing (DLST) method. Eur J Clin Microbiol Infect Dis 33 (6):927-932

[15] Persson S, Torpdahl M, Olsen KE (2008) New multiplex PCR method for the detection of *Clostridium difficile* toxin A (*tcdA*) and toxin B (*tcdB*) and the binary toxin

(*cdtA/cdtB*) genes applied to a Danish strain collection. Clin Microbiol Infect 14 (11):1057-1064

[16] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Res 40 (15):e115

[17] Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJ, Jolley KA, Kirton R, Peto TE, Rees G, Stoesser N, Vaughan A, Walker AS, Young BC, Wilcox M, Dingle KE (2010) Multilocus sequence typing of *Clostridium difficile*. J Clin Microbiol 48 (3):770-778

[18] Hunter PR (1990) Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol 28 (9):1903-1905

[19] Severiano A, Pinto FR, Ramirez M, Carrico JA (2011) Adjusted Wallace coefficient as a measure of congruence between typing methods. J Clin Microbiol 49 (11):3997-4000

[20] Zaiss NH, Rupnik M, Kuijper EJ, Harmanus C, Michielsen D, Janssens K, Nubel U (2009) Typing *Clostridium difficile* strains based on tandem repeat sequences. BMC Microbiol 9:6

[21] Marsh JW, O'Leary MM, Shutt KA, Pasculle AW, Johnson S, Gerding DN, Muto CA, Harrison LH (2006) Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in Hospitals. J Clin Microbiol 44 (7):2558-2566

[22] Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, Griffiths D, Vaughan A, Finney J, Wilcox MH, Crook DW, Peto TE (2012) Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med 9 (2):e1001172

[23] Curry SR, Muto CA, Schlackman JL, Pasculle AW, Shutt KA, Marsh JW, Harrison LH (2013) Use of multilocus variable number of tandem repeats analysis genotyping to

299 determine the role of asymptomatic carriers in *Clostridium difficile* transmission. Clin Infect
300 Dis 57 (8):1094-1102
301 [24] Bidet P, Lalande V, Salauze B, Burghoffer B, Avesani V, Delmee M, Rossier A,
302 Barbut F, Petit JC (2000) Comparison of PCR-ribotyping, arbitrarily primed PCR, and
303 pulsed-field gel electrophoresis for typing *Clostridium difficile*. J Clin Microbiol 38 (7):2484-
304 2487
305 [25] Killgore G, Thompson A, Johnson S, Brazier J, Kuijper E, Pepin J, Frost EH,
306 Savelkoul P, Nicholson B, van den Berg RJ, Kato H, Sambol SP, Zukowski W, Woods C,
307 Limbago B, Gerding DN, McDonald LC (2008) Comparison of seven techniques for typing
308 international epidemic strains of *Clostridium difficile*: restriction endonuclease analysis,
309 pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus
310 variable-number tandem-repeat analysis, amplified fragment length polymorphism, and
311 surface layer protein A gene sequence typing. J Clin Microbiol 46 (2):431-437
312