Multilocus Variable-Number Tandem-Repeat Analysis for Investigation of *Clostridium difficile* Transmission in Hospitals

Jane W. Marsh,^{1*} Mary M. O'Leary,¹ Kathleen A. Shutt,¹ A. William Pasculle,² Stuart Johnson,^{4,5}
Dale N. Gerding,^{4,5} Carlene A. Muto,^{1,3} and Lee H. Harrison¹

Infectious Diseases Epidemiology Research Unit, University of Pittsburgh School of Medicine and Graduate School of Public Health, Pittsburgh, Pennsylvania¹; Division of Microbiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania²; Division of Hospital Epidemiology and Infection Control, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania³; Hines VA Hospital, Hines, Illinois⁴; and Loyola University Chicago Stritch School of Medicine, Maywood, Illinois⁵

Received 11 November 2005/Returned for modification 25 January 2006/Accepted 8 April 2006

Clostridium difficile is a major cause of antibiotic-associated gastrointestinal illness. Recently, an increased incidence of hospital-acquired infections with severe outcomes has been reported in North America and Europe. Current molecular-typing methods for detection of outbreaks and nosocomial transmission are labor-intensive, subjective, or insufficiently discriminatory to differentiate between closely related strains. This report describes the development of multilocus variable-number tandem-repeat (VNTR) analysis (MLVA) for molecular subtyping of *C. difficile*. Seven VNTR loci were identified from the *C. difficile* 630 genome by screening an isolate collection of various restriction endonuclease analysis (REA) types. The stability of the loci for short-term epidemiologic investigations was determined by performing MLVA on consecutive isolates of the same REA type from individual patients collected over as many as 90 days. Validation of MLVA for molecular genotyping was performed by direct comparison with REA results obtained from Hines Veterans Affairs Hospital on a combined collection of 40 *C. difficile* isolates from two different sources. The ability of MLVA to detect outbreaks was demonstrated on a collection of tertiary-care hospital isolates from a defined *C. difficile* outbreak in 2001. MLVA successfully clustered *C. difficile* isolates of the same REA type and discriminated isolates of unique REA type. Thus, MLVA is an objective, portable genotyping method that permits reliable detection of *C. difficile* outbreaks and can aid epidemiologic investigations of nosocomial transmission.

Clostridium difficile is a major cause of antibiotic-associated gastrointestinal disease in the United States and worldwide. Increased incidence of infections leading to serious outcomes, including pseudomembranous colitis, colectomy, and death, has recently been reported (5, 7, 22, 25). A recent report estimated the cost of *C. difficile*-associated disease to the health care system in the United States to be \$1.1 billion per year (17).

Nosocomial transmission of *C. difficile* is common. Epidemiologic investigations of hospital-acquired disease require rapid, reliable, and discriminatory genotyping methods to track transmission and identify the emergence of new pathogenic variants. Current *C. difficile*-genotyping methods include PCR ribotyping, arbitrarily primed PCR (AP-PCR), amplified fragment length polymorphism (AFLP), pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST), and restriction endonuclease analysis (REA) (2, 4, 13, 15, 18, 28). PCR ribotyping and AP-PCR do not provide sufficient discriminatory power to distinguish among the circulating genotypes and can therefore obscure epidemiologic investigations (3). In addition, AP-PCR can suffer from lack of reproducibility (6). PFGE is not a consistently reliable subtyping tool for *C. difficile* due to DNA degradation of some strains, resulting in frequent genotyping failures, and studies comparing AFLP with PFGE

have demonstrated that AFLP is a more reliable *C. difficile*-genotyping tool than PFGE (15). The application of REA for studying the epidemiology of *C. difficile*, first described by Kuijper et al., is considered a standard in *C. difficile* genotyping and has been instrumental in describing outbreak strains, including the current REA BI clone that has been found in multiple hospitals in North America and parts of Europe (16, 20, 21, 31). However, the method is not readily transportable between laboratories, since interpretation of REA banding patterns is subjective and confirmation of REA matches requires analysis of isolates on the same gel.

A genotyping method that is amenable to *C. difficile* surveillance and the generation of a *C. difficile* genotypic database is necessary to efficiently track transmission of the organism both locally and globally. Multilocus variable-number tandem-repeat (VNTR) analysis (MLVA) is a useful genotyping tool that has demonstrated application to epidemiologic studies and outbreak detection for a variety of bacterial organisms (14, 23). This report describes the development and application of MLVA for the surveillance and detection of nosocomial *C. difficile* outbreaks.

MATERIALS AND METHODS

Bacterial strains. *C. difficile* strains were streaked from frozen glycerol or meat broth onto sheep blood agar plates that had been preincubated anaerobically for 48 h (prereduced). Anaerobic conditions were generated using Oxoid Anaerogen packets in sealed culture jars (Oxoid Ltd., Hampshire, United Kingdom). *C. difficile* cultures were grown at 37°C for 48 h anaerobically, subcultured from a single colony onto fresh prereduced sheep blood agar plates, and grown for 48 h at 37°C. Prereduced tryptic soy broth was inoculated with each bacterial culture

* Corresponding author. Mailing address: University of Pittsburgh School of Medicine, 865 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261. Phone: (412) 648-3102. Fax: (412) 648-8455. E-mail: marshj@dom.pitt.edu.

FIG. 1. Venn diagram depicting the collection of 86 *C. difficile* isolates comprising three study groups. Study group A, 20 serial patient isolates from UPMC; study group B, 40 isolates, 20 from HVA and 20 from UPMC; study group C, 56 UPMC 2001 outbreak isolates. HVA isolates, white; UPMC isolates, gray.

and incubated anaerobically for 18 h at 37°C for subsequent genomic-DNA isolation for either REA or MLVA.

A total of 86 *C. difficile* clinical isolates were studied, 66 from the University of Pittsburgh Medical Center (UPMC) and 20 toxigenic reference isolates from Hines Veterans Affairs Hospital (HVA). These 20 isolates were selected from an extensive collection of more than 6,000 isolates dating from the early 1980s to the present that have been categorized by REA typing into >100 REA groups and >400 specific REA types (4, 9). These strains represent the most common toxigenic REA groups, including toxin variant groups CF, AA, BK, and BI and the VPI 10463 reference strain. The UPMC isolates were chosen randomly from 135 toxigenic *C. difficile* isolates collected sequentially over a 7-month period in 2001 and were typed by REA at the University of Pittsburgh (Pitt REA types) as part of an infection control investigation of a large *C. difficile* outbreak (22).

To validate MLVA, the study isolates were classified into three study groups (A, B, and C) that were not mutually exclusive (Fig. 1 and Table 1). Study group A consisted of 10 pairs of serial isolates collected on different dates from the same patient at UPMC. Each pair was indistinguishable by REA typing performed at either the University of Pittsburgh or HVA. These isolates were used to determine the stability of the VNTR loci over time and to determine the ability of MLVA to identify highly related isolates. Study group B consisted of 40 *C. difficile* isolates—20 HVA reference isolates and 20 UPMC isolates representing the most prevalent Pitt REA types from the 2001 outbreak that were REA typed at HVA as part of another investigation (21). These isolates were used to correlate MLVA types with the well-established HVA REA-typing scheme and to determine the discriminatory power of MLVA relative to REA. Study group C consisted of 56 of the 66 UPMC isolates that were collected as part of an infection control investigation of the 2001 outbreak (22). This study group was used to determine the utility of MLVA for epidemiologic investigation of a large *C. difficile* outbreak in a hospital setting. Twenty-seven of the isolates fell into more than one study group—7 isolates were common to study groups A and C only, 17 isolates were common to study groups B and C only, and 3 isolates were common to all three study groups (Fig. 1 and Table 1).

REA typing. Restriction endonuclease analysis was performed at the University of Pittsburgh and at HVA as previously described (4). Briefly, 10 µg of purified genomic DNA was subjected to restriction digestion with 20 U of HindIII and electrophoresed on a 0.7% agarose gel for 12 h. The gels were stained with ethidium bromide and photographed under UV light. The study group B strains were typed according to the HVA REA-typing scheme, designating REA group and type alphanumerically (e.g., J9 or BI9). Plasmid analysis was performed by restriction digestion of isolated plasmid DNA with HindIII, followed by electrophoresis with the corresponding restricted genomic DNA. Plasmid and genomic-DNA restriction patterns were compared to identify bands of plasmid origin. Unique plasmids were designated by the letter “p” followed by a numeric designation (e.g., J9p4; Table 1) to distinguish the different plasmid patterns. REA groups and types were assigned by visual inspection of banding patterns and comparison with known *C. difficile* reference strains. Isolates as-

signed the same REA type had indistinguishable banding patterns compared to the reference strain. These results were confirmed by running the reference and test strains on the same gel. REA types were assigned at the University of Pittsburgh to the serial patient isolates comprising study group A using similar typing methods; however, Pitt REA types were given sequential numeric values (e.g., 1, 2, or 3) based on the unique *C. difficile* REA patterns observed and were therefore unique to the University of Pittsburgh (see Table 3).

VNTR identification. Tandem Repeats Finder software was used to detect repeat elements in the *C. difficile* 630 chromosome (1). These sequence data were produced by the *C. difficile* Sequencing Group at the Sanger Institute and can be obtained from <http://ftp.sanger.ac.uk/pub/pathogens/cd/>. Tandem repeats with a period size of ≤20 bp (to facilitate automated fragment detection), a copy number of ≥2, and a percent consensus match of ≥85% were initially selected for screening by PCR from a collection of isolates with different REA types. The low percent consensus match of 85% was required to provide sufficient tandem repeats for screening due to the low G+C content of the organism. Primers that flanked the tandem-repeat region were designed using DNASTAR Primer Select software (DNASTAR, Madison, WI). The VNTR screen was initially performed by PCR amplification of each candidate tandem-repeat locus on genomic DNA from seven isolates comprising four frequently identified Pitt REA types. The variability of each tandem-repeat locus was assessed by gel electrophoresis on a 1% agarose gel and sequence analysis to determine the size of the resulting PCR products and the tandem-repeat copy number, respectively. The 5′ and 3′ flanking regions of each tandem repeat were confirmed by comparison with the available *C. difficile* genomic sequence, and the tandem-repeat sequences were counted manually to determine the repeat copy number at each locus for each *C. difficile* isolate. Tandem-repeat loci that generated more than one allele (copy number) were selected for further evaluation of the defined *C. difficile* study groups, A, B, and C.

MLVA. PCR amplification and sequence analysis of the seven selected *C. difficile* repeat (CDR) loci was performed on the 86 *C. difficile* isolates from UPMC and HVA with the following primer sets: CDR4F (5′-ATTAATCATATCCTACAGAACACGA-3′) and CDR4R (5′-TAAACAAATGATATAAACTGAAAG-3′); CDR5F (5′-AATTTTAAGTTAACGTTTTTCTACAT-3′) and CDR5R (5′-AGCCATTTTATCAATCCTTTCTAT-3′); CDR9F (5′-TCTGGGATGTAACTAGCGACTTGT-3′) and CDR9R (5′-TCTTAGGGGAATTTATTGGAGGAA-3′); CDR48F (5′-AGGAGCTTTATATGGACATTCAGGTAG-3′) and CDR48R (5′-AATCTCTTTCAAACCTTCAATCTCAAT-3′); CDR49F (5′-ACATATTTAGGCATTTAGTC-3′) and CDR49R (5′-GAGTATTATTTATCATTTGTGGGTATTA-3′); CDR59F (5′-GTAGAAGGGGCAATTTATATGA-3′) and CDR59R (5′-CCTTCTGGCTTCTCTGTGA-ATA-3′); and CDR60F (5′-GGTGACATGCTGCTCCTG-3′) and CDR60R (5′-AACGCATTAAATTTAC-TCCTCATAC-3′). MLVA on the 20 HVA isolates was performed in a blinded fashion. Genomic *C. difficile* DNA was isolated from 10-ml tryptic soy broth cultures grown for 18 h anaerobically using the QIAGEN DNeasy MiniKit according to the manufacturer's instructions for gram-positive organisms (QIAGEN, Valencia, CA). DNA amplification was performed on 1 µl of purified genomic DNA. PCRs were performed in 50-µl volumes containing 1× AmpliTaq Gold PCR buffer (15 mM Tris-HCl, pH 8.0, 50 mM KCl), 2.5 mM MgCl₂, 0.2 mM each deoxynucleoside triphosphate, 0.2 µM of each primer, and 1.5 units AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA). The PCR cycle conditions were as follows: 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 48°C for 1 min, and 72°C for 1 min 30 s, and a final extension at 72°C for 5 min. The PCR products were prepared for sequencing by precipitation with 60 µl of 20% polyethylene glycol 8000 (Sigma-Aldrich, St. Louis, MO), 2.5 M NaCl solution for 1 h at room temperature, followed by centrifugation at 2,750 × g for 1 h at 4°C. The polyethylene glycol 8000-precipitated PCR product was washed twice with 70% ethanol, followed by centrifugation at 2,750 × g for 10 min at 4°C. The cleaned PCR product was resuspended in 30 µl distilled H₂O, and 1 µl of the product was used for sequencing with the ABI Big Dye Terminator Kit v.3.1 (Applied Biosystems) with the same primers employed in the PCR. Automated sequence detection was performed on an ABI 3730 sequence detection system. Electropherograms from forward and reverse primer pairs for each VNTR locus were edited, and the number of tandem repeats at each locus was manually determined using DNASTAR SeqMan software (DNASTAR, Madison, WI).

Data analysis. Copy numbers at each of the seven CDR loci were concatenated to generate an MLVA type for each isolate (Table 1). The genetic relationships of the isolates was determined by clustering them according to MLVA type using the summed absolute distance as the coefficient for calculating the minimum-spanning tree available in the Bionumerics 4.01 software program (Applied Maths, Austin, TX). The summed absolute distance between two

TABLE 1. *C. difficile* MLVA types, VNTR copy numbers, REA types, and study groups

MLVA type	No. of isolates	No. of VNTR for each CDR locus							HVA REA type	Study group(s) ^a
		CDR4	CDR5	CDR9	CDR48	CDR49	CDR59	CDR60		
1	1	29	16	6	6	14	6	14	J7	B
2	1	38	3	6	10	18	4	6	BI 8	B
3	1	25	7	8	8	21	7	10	B1	B
4	1	37	6	6	5	18	7	3	Y2	B
5	1	13	3	10	6	14	6	8	Z1	B
6	1	35	16	6	6	13	6	19	J1	B
7	1	33	9	8	4	14	6	2	A1	B
8	1	25	3	14	9	11	4	9	BI 6	B
9	1	33	3	14	9	11	4	10	BI 1	B
10	1	28	3	10	5	8	6	25	Z3	B
11	1	31	8	10	6	21	7	9	K1	B
12	1	16	16	6	5	18	6	12	J9	B
13	1	3	0	16	8	7	7	6	CF 4	B
14	1	24	6	12	5	31	7	9	K14	B
15	1	3	0	27	7	7	7	6	CF 2	B
16	1	3	0	33	8	17	7	7	CF 1p	B
17	1	33	6	10	4	11	6	3	Y1	B
18	1	24	3	7	4	22	4	9	W1	B
19	1	Null	Null	10	13	24	5	2	AA2	B
20	1	41	6	8	5	11	7	2	G1	B
21	1	17	5	7	3	21	9	15		C
22	1	32	16	6	6	14	6	16		C
23	1	6	16	6	6	15	6	15		C
24	2	23	16	6	6	15	6	15	J28/J29	BC(2)
25	1	Null	Null	13	9	15	6	2	BK6	BC
26	1	32	6	2	7	10	6	15		C
27	1	7	16	6	6	14	6	14	J30	BC
28	1	31	16	6	6	14	6	16		C
29	1	38	16	6	6	15	6	15	J9p4	BC
30	1	18	3	14	9	11	4	9	BI 9	BC
31	1	18	16	6	6	18	6	14	J9p0	BC
33	2	3	0	27	7	7	7	7		A ¹ , AC ¹
34	1	41	16	6	6	14	6	15		C
35	4	19	3	15	9	11	4	9	BI 10	A ² , AC ² , BC(2)
36	3	20	3	13	9	11	4	9	BI 9	BC(2), C
39	1	17	8	21	8	16	7	6	BR 2	BC
40	1	6	3	8	6	8	7	13	CL	BC
41	1	20	8	9	6	19	7	10	K	BC
42	1	31	16	6	6	16	6	17	J9p6	BC
43	2	37	3	13	9	15	4	7	BI 11p1	A ¹⁰ , ABC ¹⁰
45	1	12	3	15	9	11	4	10	BI 6	BC
47	1	13	3	16	9	11	4	9		C
48	1	13	3	15	9	9	4	9		C
52	1	19	3	12	9	12	4	9		C
53	1	34	6	2	7	10	6	8	BM	BC
54	1	34	16	6	7	16	6	16	J31	ABC ⁹
56	1	35	16	6	7	16	6	15		A ⁹
57	1	23	8	9	13	15	6	2		AC ⁶
59	1	25	6	6	5	12	7	4		C
61	1	24	6	11	5	17	7	3		C
64	4	20	3	14	9	11	4	9		A ³ , AC ³ , C(2)
65	2	14	3	15	9	11	4	9		C(2)
66	2	18	3	17	9	11	4	9		C(2)
70	1	24	8	9	13	15	6	2		A ⁶
72	2	18	3	8	9	15	4	9		A ⁴ , AC ⁴
73	1	22	3	9	3	27	4	7		C
74	1	12	3	15	9	11	4	9		C
75	1	18	3	13	9	12	4	9		C
76	1	19	3	14	9	12	4	9		C
83	2	Null	Null	12	8	15	6	2		A ⁵ , AC ⁵
84	1	33	3	9	3	25	4	6		C
89	1	33	3	9	3	17	4	7		C
100	1	14	3	16	9	11	4	9		C
101	1	13	3	15	9	11	4	9		C
102	1	18	6	10	3	1	9	27		C
108	1	40	6	9	5	19	7	10		C
112	1	15	3	17	9	11	4	5	BI 6	ABC ⁷
113	1	15	3	17	10	11	4	5		A ⁷
115	1	15	3	16	8	12	4	9		C
119	1	25	6	3	4	26	8	12		AC ⁸
121	1	25	6	3	4	25	8	12		A ⁸

^a Superscripts 1 to 10 indicate patient numbers from study group A; (2) indicates two isolates.

TABLE 2. Characteristics of *C. difficile* VNTR loci

VNTR locus	Location ^a	Size (bp)	Copy no. range ^b	No. of alleles ^b	Repeat sequence
CDR4	755721	6	3–41	28	TTGCTC
CDR5	692929	8	0–17	9	TATATTG/AG
CDR9	664660	8	3–34	17	TAAAAGAG
CDR48	167124	7	3–13	9	ATAGATT
CDR49	3688632	7	1–31	21	AT/AC/TTTCT
CDR59	771338	11	5–10	6	TAAG/ATATA/GGAT/C
CDR60	677132	17	2–27	18	G/ATAAA/GTAGGATG/ATAAAA

^a Genome location based on *C. difficile* 630 sequence (<http://www.sanger.ac.uk>).

^b Based on analysis of 86 isolates in this study.

MLVA-typed isolates is the calculated summed tandem-repeat difference (STRD) at all seven VNTR loci. The minimum-spanning-tree algorithm associates MLVA types with the smallest STRD. Priority rules within the Bionumerics software were set to link MLVA types that had the most single-locus variants (SLVs) first. These rules are based on the eBURST algorithm for analysis of MLST data from clonal bacterial populations in which descendants of a founding genotype develop as a result of genetic change at a single locus (8). Color coding of strains with STRDs of ≤ 10 was performed to facilitate visual identification of the most genetically related isolates and to discriminate less related, sporadic *C. difficile* isolates.

RESULTS

Identification and characterization of VNTR loci and MLVA types. A total of 383 repeated sequence elements were identified from analysis of the 4.3-Mb *C. difficile* 630 chromosome. Sixty-one repeats met the selection criteria of repeat size, copy number, and consensus match. There were 30 repeats that met the selection criteria for which oligonucleotide primers could not be designed due to the low G+C contents (29%) of the nucleotide sequences flanking these regions. Primers that flanked each of the remaining 31 different repeats were designed for PCR analysis of each tandem repeat from a collection of clinical *C. difficile* isolates of various REA types. From this screen, seven CDR loci were obtained that generated reliable PCR products and that were variable across the isolates examined (Table 2). The loci are clustered within 0.6 to 0.8 Mb on the chromosome, except for one VNTR, CDR49, which is located at 3.7 Mb. Two of the VNTR loci, CDR4 and CDR60, reside within predicted protein-coding regions, CD0632 and CD0568, respectively. However, a homology search based on the nucleotide sequences of these predicted open reading frames did not identify any known bacterial protein-coding regions. The remaining tandem repeats reside within predicted noncoding regions.

A total of 71 unique MLVA types were identified among the 86 *C. difficile* isolates examined (Table 1). The CDR4 locus was the most diverse VNTR, with 28 alleles, and the CDR59 locus showed the least amount of diversity, with 6 alleles (Table 2). For 4 of the 86 isolates studied, CDR4 and CDR5 had null alleles that did not generate an amplification product. Interestingly, the CDR4 and CDR5 null alleles always occurred together with two repeats at CDR 60 (Table 1, MLVA types 19, 25, and 83) (2). In addition, five isolates had a CDR5 allele of 0 (Table 1, MLVA types 13, 15, 16, and 33) (2). In each case, the CDR5 locus amplified and the nucleotide sequence corresponded to the 5' and 3' flanking regions, but the CDR5 repeat

TABLE 3. Study group A characteristics

Patient no.	REA type ^a	MLVA type	Culture date ^b	Interval (days)	No. of different loci	STRD
1	13	33	5/18/01	19	0	0
1	13	33	6/6/01			
2	4	35	6/14/01	14	0	0
2	4	35	6/28/01			
3	4	64	8/31/01	10	0	0
3	4	64	9/9/01			
4	2	72	7/10/01	58	0	0
4	2	72	9/6/01			
5	28	83	7/20/01	2	0	0
5	28	83	7/22/01			
6	30	57	6/13/01	19	1	1
6	30	70	7/2/01			
7	2	112	9/27/01	2	1	1
7	2	113	9/29/01			
8	4	119	10/29/01	4	1	1
8	4	121	11/2/01			
9	7	54	6/11/01	2	2	2
9	7	56	6/13/01			
10	BI 11p1	43	6/1/01	92	0	0
10	BI 11p1	43	9/1/01			

^a Numeric values are Pitt REA types; alphanumeric values are HVA REA types.

^b Month/day/year.

was absent. Interestingly, three out of five of these isolates were REA typed and found to belong to the CF REA group. The other two isolates (MLVA type 33) were highly related to the CF2 REA type by MLVA (an SLV with an STRD of 1). These data suggest that sufficient diversity of alleles at the CDR loci exists to permit the discrimination of unrelated isolates while allowing identification of genetically related isolates.

Stability of VNTR loci (study group A isolates). Serial isolates from individual patients that were cultured on different dates were subjected to MLVA to determine if the seven VNTR loci were sufficiently stable to identify genetically related isolates collected over time. There were 10 pairs of study group A isolates with intervals between isolate collection dates for individual patients ranging from 2 to 92 days (Table 3). The isolates within each pair were indistinguishable by either Pitt or HVA REA typing (Table 3). For six of these pairs, each of the isolates within each pair had identical MLVA types (Tables 1 and 3, patient no. 1 to 5 and 10). Three pairs of isolates generated MLVA types that were SLVs of one another, with each pair varying by one tandem repeat at different loci (Tables 1 and 3, patient no. 6 to 8). One pair of isolates was a double-locus variant (DLV) with an STRD of 2 (Tables 1 and 3, patient no. 9), which was the maximum STRD observed by MLVA for all of the study group A isolates. Taken together, these data indicate that the seven CDR loci are relatively stable over time and tend to mutate one tandem repeat at a time. Furthermore, these data establish that MLVA types with an STRD of ≤ 2 are indicative of a high degree of genetic relatedness among *C. difficile* isolates. Based on this observation, MLVA typing correlated with the REA-typing data for all 10 pairs of serial patient isolates, demonstrating that MLVA correctly identified genetically related *C. difficile* isolates.

FIG. 2. Minimum-spanning-tree analysis of MLVA data from 40 REA-typed *C. difficile* strains (study group B). The circles represent unique MLVA types (numeric value). HVA REA types are designated alphanumerically. MLVA types representing one or two isolates are white or gray, respectively. The numbers between the circles represent the STRDs between MLVA types. Solid lines represent single- or double-locus variants, dashed lines represent triple-locus variants, and dotted lines represent ≥ 4 -locus variants. Green and pink shading identify isolates belonging to the J or BI REA group with STRDs of ≤ 10 , respectively.

Of note was the finding that isolates belonging to the same REA group had the same copy number at several of the seven VNTR loci examined. For instance, isolates belonging to the J REA group were invariant at CDR5, CDR9, and CDR59 (Table 1). Thus, the CDR loci required for discrimination among J-group strains by MLVA are CDR4, CDR48, CDR49, and CDR60. Isolates belonging to the BI REA group

FIG. 3. Minimum-spanning-tree analysis of MLVA data from 56 *C. difficile* patient isolates (study group C). The circles represent unique MLVA types (numeric value). HVA REA types are designated alphanumerically. MLVA types with one, two, or three isolates are white, gray, or blue circles, respectively. The numbers between the circles represent the STRDs between MLVA types. Solid lines represent single- or double-locus variants, dashed lines represent triple-locus variants, and dotted lines represent ≥ 4 -locus variants. The shaded areas (A, B, C, D, E, and F) denote genetically related clusters defined by an STRD of ≤ 10 .

were invariant at CDR5 and CDR59, and a majority of the isolates had the same copy number at CDR48, CDR49, and CDR60. Thus, the VNTR loci required for discrimination of isolates belonging to the BI REA group are CDR4 and CDR9. Similarly, the three isolates belonging to the CF REA group were invariant at CDR4, CDR5, and CDR59. The major discriminatory locus for the CF REA group was CDR9. Based on a relatively limited number of isolates, these invariant loci are unique for each REA group and establish an objective signature genotype.

In summary, MLVA generated objective genotypes with easily recognizable VNTR patterns. When minimum-spanning-tree analysis was applied to these data, the 40 *C. difficile* study group B isolates clustered according to the REA-typing scheme for the majority of the strains. MLVA identified 37 different *C. difficile* genotypes, while REA identified 36 unique types. MLVA was less discriminatory than REA within the J REA group and more discriminatory than REA among the BI REA group. In addition, while MLVA was able to discriminate between unique REA types, it also revealed genetic relation-

ships among the different REA types that were previously unrecognized.

Molecular epidemiologic assessment by MLVA of 56 *C. difficile* isolates collected from a tertiary-care hospital during an outbreak (study group C isolates). MLVA was performed on the 56 UPMC study group C isolates to determine the utility of MLVA for detection of genetically related *C. difficile* strains and assessment of nosocomial transmission. Minimum-spanning-tree analysis of the MLVA data revealed that these isolates consisted of several related complexes (Fig. 3A to F). The largest complex consisted of 26 isolates corresponding to the BI REA group, which is associated with multiple outbreaks in North America and Europe (Fig. 3A) (12, 20–22, 25, 31). A majority of these isolates generated MLVA types that were SLVs or DLVs of each other with an STRD of ≤ 2 , indicating that they were highly related to one another. Furthermore, the fact that these isolates were collected from the same hospital within a 7-month period suggests that this degree of similarity among MLVA types is highly indicative of nosocomial transmission. There was one isolate of REA type BI11 that had an

STRD of 23 and was a triple-locus variant (TLV) of MLVA type 36 (BI9) (Fig. 3, MLVA type 43). This isolate may be genetically related to isolates of MLVA type 36, but no conclusions regarding nosocomial transmission of the isolate can be made.

The second-largest complex identified by MLVA consisted of nine isolates corresponding to the J REA group. These isolates had not previously been recognized as genetically related by Pitt REA typing (Fig. 3B) (22). Two of these isolates were SLVs with an STRD of 1 and were therefore considered to be highly related and representative of nosocomial transmission (Fig. 3B, MLVA types 22 and 28). Four isolates that were DLVs with an STRD of ≤ 4 were considered to be genetically related and possibly nosocomially acquired (Fig. 3B, MLVA types 29, 34, 42, and 54). There were two isolates that were TLVs of each other, with an STRD of 3 (Fig. 3E, MLVA types 23 and 27). One of the isolates was REA type J30 and a DLV of REA type J9p0 (MLVA type 31), with an STRD of 15. The MLVA data suggest that this pair of isolates is genetically related to the other nine isolates corresponding to the J REA group.

Three additional pairs among the study group C isolates were found to be genetically related by MLVA (Fig. 3C, D, and F). Isolate pair C (Fig. 3C) has MLVA types that are DLVs with an STRD of 2 and are therefore considered to be highly related and highly suspect for nosocomial transmission. Both isolate pairs D and F (Fig. 3D and F) are DLVs, with an STRD of 9. While these isolates are genetically related, conclusions regarding nosocomial transmission cannot be made.

In summary, the MLVA results indicate that the UPMC 2001 isolate collection consists primarily of two major complexes of highly related isolates, suggestive of several concomitant outbreaks. The BI outbreak appeared to be more clonal than the J outbreak based on the fact that isolates within the BI cluster had more SLVs and DLVs with an STRD of ≤ 2 than the J outbreak. In addition, MLVA differentiated among BI6 and BI9 isolates and was therefore more discriminatory than REA for isolates belonging to the BI REA group. In addition, MLVA identified several additional pairs of genetically related isolates. These data demonstrate that MLVA is a useful genotyping tool for the investigation of nosocomial transmission in a hospital setting.

DISCUSSION

The tandem-repeat search of the *C. difficile* 630 chromosome identified numerous repeat elements. However, most of these loci did not meet the criteria of small period size, high copy number, and relatively high repeat consensus. In addition, the remarkably low chromosomal G+C content precluded PCR primer design and amplification of many loci. Therefore, of the 61 repeat elements that met the original criteria, only 30 tandem repeats were screened for VNTR.

Of these loci, only seven were found to be reliably amplified and variable among a collection of isolates of various REA types. The other loci were either not variable as determined by sequence analysis or the PCR generated either nonspecific products or products that were too large to be useful for subsequent automation.

Tandem-repeat loci in many bacteria often reside within

genes and can be associated with antigenic variation (11, 19, 27). The *C. difficile* VNTR loci described in this report do not reside within genes of known function. Thus, their functions in the *C. difficile* chromosome are uncertain. Six of the seven VNTR loci are clustered between 0.15 and 0.75 MB, while CDR49 is located at 3.70 MB on the 4.3-MB *C. difficile* 630 chromosome. Thus, the MLVA loci are clustered within less than 25% of the *C. difficile* chromosome. The significance of this clustering of repeat elements is not known. VNTR locations for MLVA of other pathogenic organisms are ideally distributed throughout the chromosome (14, 23). The rationale for this design is that genotype-defining changes in DNA sequence are more likely to be identified with greater coverage of the chromosome. The seven VNTR loci described in this study have demonstrated utility for the epidemiologic investigation of *C. difficile* outbreaks and nosocomial transmission at UPMC. In addition, these loci can be utilized to reveal genetic relationships among established REA types. However, further studies are required to determine the utility of these clustered repeat elements for global *C. difficile* epidemiology.

The relative stability of the seven loci for short-term epidemiologic investigations is supported by the MLVA data obtained from serial patient isolates (study group A). These data demonstrate that all seven VNTR loci are stable over several months and that MLVA correctly identifies genetically related isolates. While copy number differences were observed within several pairs of consecutive isolates, tandem-repeat loci that varied differed only by one repeat, suggesting that mutations that occur in the CDR loci often occur one tandem repeat at a time. This finding is consistent with our previous observations of mutational events at tandem-repeat loci for *Escherichia coli* O157:H7 (24).

Validation of MLVA as a useful *C. difficile*-genotyping tool is provided by direct comparison of MLVA with REA typing. MLVA clustered a collection of *C. difficile* strains according to the HVA-defined REA groups and types. There was one example where MLVA could not discriminate between two REA types (J28 and J29) and two examples where MLVA discriminated between isolates within the BI REA group (BI6 and BI9). These minor variations emphasize the clonal nature of *C. difficile* outbreaks. Both methods successfully identified outbreaks, yet MLVA provided a measure of genetic relatedness that was not easily revealed by REA typing. The minimum-spanning-tree analysis of the study group B MLVA data reveals the genetic relationships among the HVA REA groups. HVA REA groups CF, AA, and BK clustered on the same branch of the minimum-spanning tree (Fig. 2), indicating a genetic relatedness among these isolates that was not apparent by HVA REA typing. The unifying feature of these isolates is that they are all phenotypically A negative and B positive with regard to production of the large clostridial toxins (9, 10). Previous studies using AFLP have demonstrated genetic relatedness among a large collection of TcdA-negative strains from seven countries (29). In this study, MLVA identified genetic relationships among established REA groups that were previously unrecognized. MLVA data, in conjunction with additional genotypic and phenotypic information, will be useful for understanding the molecular evolution of *C. difficile* and disease pathogenesis. A recent study of *Bordetella pertussis* combining MLVA and multiple-antigen sequence typing to

characterize recent whooping cough epidemics in Holland identified a postvaccination clonal expansion (26). Thus, MLVA is expected to be a valuable typing tool for the tracking and characterization of emerging virulent *C. difficile* strains, such as those recently reported in the United States, Canada, the United Kingdom, The Netherlands, Belgium, and other European countries (12, 22, 25, 30).

The utility of MLVA for detection of outbreaks was demonstrated by analysis of a group of 56 isolates collected during the 2001 *C. difficile* outbreak at UPMC. These isolates represent a fraction of the 2001 outbreak collection. Thus, while the data set is not complete, the MLVA results clearly demonstrate the genetic relationships among the isolates and confirm the presence of two major outbreaks and several smaller clusters at UPMC. The largest outbreak (Fig. 3A), was initially identified by Pitt REA typing and later confirmed by HVA as part of the global BI epidemic strain. However, the smaller outbreak (Fig. 3B) was not initially detected by REA typing performed at the University of Pittsburgh. The discovery that these isolates were part of another outbreak was made after MLVA demonstrated their genetic relatedness and HVA REA typing confirmed that they belonged to the REA J group. This oversight illustrates the subjective nature of REA typing. In contrast, MLVA provides objective numeric data, which eliminates the variability of band-based *C. difficile*-genotyping methods and can provide uniformity of typing between laboratories. In addition, MLVA is sufficiently discriminatory to distinguish isolates that are REA unique.

In this study, MLVA types with an STRD of ≤ 10 were color coded to illustrate genetic relatedness. An STRD of ≤ 10 was selected based on the minimum-spanning-tree analysis of study group B strains. The 20 HVA isolates were selected for validation of MLVA and are genetically diverse. The 20 UPMC isolates were representative of Pitt REA types that were prevalent during the 2001 outbreak. Thus, the combined MLVA data set from these isolates is not a continuum and cannot be considered complete. MLVA types may exist that would further minimize the genetic distances (STRDs) between *C. difficile* MLVA types on the minimum-spanning tree. Further studies of consecutive isolates collected over time are required to define STRDs that are epidemiologically relevant.

This study takes into account both the STRD and the number of locus differences to assess the genetic relatedness of *C. difficile* isolates by MLVA. Minimum-spanning trees were calculated so that isolates with the most SLVs were linked first. This priority rule was adapted from the eBURST algorithm for MLST, which describes the evolution of clonal complexes of closely related bacteria and assumes that SLVs will diversify over time to yield DLVs and TLVs (8). Further studies are required to determine the mutation rate of each CDR locus, as this variable should be considered when calculating genetic distances by MLVA. In addition, epidemiologic studies of well-characterized *C. difficile* populations will help determine the genetic and evolutionary relationships among clonal populations and establish uniform genotyping databases for improved detection and surveillance of *C. difficile*-associated disease.

This study demonstrates that MLVA is useful for investigations of nosocomial *C. difficile* infections. The method, including DNA extraction, PCR amplification, and sequence analysis of the seven CDR loci, can be performed on 12 samples in

approximately 36 h and at a cost of approximately \$40.00 per sample, and it has replaced REA as the primary *C. difficile*-genotyping method at UPMC. The resulting MLVA type provides objective genotypic information that has the potential to facilitate the comparison of *C. difficile* strains between institutions and laboratories. Future epidemiologic investigations of prevalent MLVA genotypes may elucidate evolutionary trends in *C. difficile* disease. MLVA types associated with particular virulence factors, including toxins and drug resistance determinants, can be easily tracked, assuming that clinical laboratories are culturing for *C. difficile*, enabling appropriate infection control and medical interventions that may lead to improved outcomes and reduced health care costs. In summary, this study demonstrates that MLVA is a valuable genotyping method that can be applied to outbreak detection and epidemiologic investigations of *C. difficile*-associated disease.

ACKNOWLEDGMENTS

This work was supported in part by a career development award to L. H. Harrison from the National Institute of Allergy and Infectious Diseases (K24 AI52788) and a U.S. Department of Veterans Affairs Research Service grant to D. N. Gerding.

We gratefully acknowledge the expert technical assistance of Sujata Patel.

REFERENCES

- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573–580.
- Bidet, P., F. Barbut, V. Lalande, B. Burghoffer, and J. C. Petit. 1999. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. *FEMS Microbiol. Lett.* **175**:261–266.
- Bidet, P., V. Lalande, B. Salauze, B. Burghoffer, V. Avesani, M. Delmee, A. Rossier, F. Barbut, and J. C. Petit. 2000. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *J. Clin. Microbiol.* **38**:2484–2487.
- Clabots, C. R., S. Johnson, K. M. Bettin, P. A. Mathie, M. E. Mulligan, D. R. Schaberg, L. R. Peterson, and D. N. Gerding. 1993. Development of a rapid and efficient restriction endonuclease analysis typing system for *Clostridium difficile* and correlation with other typing systems. *J. Clin. Microbiol.* **31**:1870–1875.
- Dallal, R. M., B. G. Harbrecht, A. J. Boujoukas, C. A. Sirio, L. M. Farkas, K. K. Lee, and R. L. Simmons. 2002. Fulminant *Clostridium difficile*: an underappreciated and increasing cause of death and complications. *Ann. Surg.* **235**:363–372.
- Ellsworth, D. L., K. D. Rittenhouse, and R. L. Honeycutt. 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Bio-Techniques* **14**:214–217.
- Fawley, W. N., P. Parnell, P. Verity, J. Freeman, and M. H. Wilcox. 2005. Molecular epidemiology of endemic *Clostridium difficile* infection and the significance of subtypes of the United Kingdom epidemic strain (PCR ribotype 1). *J. Clin. Microbiol.* **43**:2685–2696.
- Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**:1518–1530.
- Geric, B., S. Johnson, D. N. Gerding, M. Grabnar, and M. Rupnik. 2003. Frequency of binary toxin genes among *Clostridium difficile* strains that do not produce large clostridial toxins. *J. Clin. Microbiol.* **41**:5227–5232.
- Johnson, S., S. P. Sambol, J. S. Brazier, M. Delmee, V. Avesani, M. M. Merrigan, and D. N. Gerding. 2003. International typing study of toxin A-negative, toxin B-positive *Clostridium difficile* variants. *J. Clin. Microbiol.* **41**:1543–1547.
- Jordan, P. W., L. A. Snyder, and N. J. Saunders. 2005. Strain-specific differences in *Neisseria gonorrhoeae* associated with the phase variable gene repertoire. *BMC Microbiol.* **5**:21.
- Katikireddi, V. 2005. UK launches inquiry into *Clostridium difficile* outbreak. *CMAJ* **173**:138.
- Kato, H., N. Kato, K. Watanabe, K. Ueno, H. Ushijima, S. Hashira, and T. Abe. 1994. Application of typing by pulsed-field gel electrophoresis to the study of *Clostridium difficile* in a neonatal intensive care unit. *J. Clin. Microbiol.* **32**:2067–2070.
- Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones. 2000. Multiple-locus vari-

- able-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182**:2928–2936.
15. Klaassen, C. H., H. A. van Haren, and A. M. Horrevorts. 2002. Molecular fingerprinting of *Clostridium difficile* isolates: pulsed-field gel electrophoresis versus amplified fragment length polymorphism. J. Clin. Microbiol. **40**:101–104.
 16. Kuijper, E. J., J. H. Oudbier, W. N. Stuijbergen, A. Jansz, and H. C. Zanen. 1987. Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. J. Clin. Microbiol. **25**:751–753.
 17. Kyne, L., M. B. Hamel, R. Polavaram, and C. P. Kelly. 2002. Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. Clin. Infect. Dis. **34**:346–353.
 18. Lemee, L., A. Dhalluin, M. Pestel-Caron, J. F. Lemeland, and J. L. Pons. 2004. Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. J. Clin. Microbiol. **42**:2609–2617.
 19. Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling, and D. L. Kasper. 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. Proc. Natl. Acad. Sci. USA **93**:4131–4136.
 20. McDonald, L. C., G. E. Killgore, A. Thompson, R. C. Owens, Jr., S. V. Kazakova, S. P. Sambol, S. Johnson, and D. N. Gerding. 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N. Engl. J. Med. **353**:2433–2441.
 21. McEllistrem, M. C., R. J. Carman, D. N. Gerding, C. W. Genheimer, and L. Zheng. 2005. A hospital outbreak of *Clostridium difficile* disease associated with isolates carrying binary toxin genes. Clin. Infect. Dis. **40**:265–272.
 22. Muto, C. A., M. Pokrywka, K. Shutt, A. B. Mendelsohn, K. Nouri, K. Posey, T. Roberts, K. Croyle, S. Krystofiak, S. Patel-Brown, A. W. Pasculle, D. L. Paterson, M. Saul, and L. H. Harrison. 2005. A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. Infect. Control Hosp. Epidemiol. **26**:273–280.
 23. Noller, A. C., M. C. McEllistrem, A. G. Pacheco, D. J. Boxrud, and L. H. Harrison. 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. J. Clin. Microbiol. **41**:5389–5397.
 24. Noller, A. C., M. C. McEllistrem, K. A. Shutt, and L. H. Harrison. 2006. Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Microbiol. **44**:374–377.
 25. Pepin, J., L. Valiquette, M. E. Alary, P. Villemure, A. Pelletier, K. Forget, K. Pepin, and D. Chouinard. 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. CMAJ **171**:466–472.
 26. Schouls, L. M., H. G. van der Heide, L. Vauterin, P. Vauterin, and F. R. Mooi. 2004. Multiple-locus variable-number tandem repeat analysis of Dutch *Bordetella pertussis* strains reveals rapid genetic changes with clonal expansion during the late 1990s. J. Bacteriol. **186**:5496–5505.
 27. Skuce, R. A., T. P. McCorry, J. F. McCarroll, S. M. Roring, A. N. Scott, D. Brittain, S. L. Hughes, R. G. Hewinson, and S. D. Neill. 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology **148**:519–528.
 28. Tang, Y. J., S. T. Houston, P. H. Gumerlock, M. E. Mulligan, D. N. Gerding, S. Johnson, F. R. Fekety, and J. Silva, Jr. 1995. Comparison of arbitrarily primed PCR with restriction endonuclease and immunoblot analyses for typing *Clostridium difficile* isolates. J. Clin. Microbiol. **33**:3169–3173.
 29. van den Berg, R. J., E. C. Claas, D. H. Oyib, C. H. Klaassen, L. Dijkshoorn, J. S. Brazier, and E. J. Kuijper. 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.
 30. Van Steenberghe, J., S. Debast, E. van Kregten, R. van den Berg, D. Notermans, and E. J. Kuijper. 2005. Isolation of *Clostridium difficile* ribotype 027, toxinotype III in the Netherlands after increase in *C. difficile* associated diarrhoea. Eurosurveillance Wkly. **10**:E050714.1. [Online.] <http://www.eurosurveillance.org/ew/2005/050714.asp#1>.
 31. Warny, M., J. Pepin, A. Fang, G. Killgore, A. Thompson, J. Brazier, E. Frost, and L. C. McDonald. 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.