JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 2003, p. 5389–5397 Vol. 41, No. 12
0095-1137/03/$08.00+0 DOI: 10.1128/JCM.41.12.5389–5397.2003

Multilocus Variable-Number Tandem Repeat Analysis Distinguishes Outbreak and Sporadic *Escherichia coli* O157:H7 Isolates

Anna C. Noller,[1,2] M. Catherine McEllistrem,[1] Antonio G. F. Pacheco,[3,4] David J. Boxrud,[5] and Lee H. Harrison[1]*

Infectious Diseases Epidemiology Research Unit, University of Pittsburgh Graduate School of Public Health and School of Medicine,[1] and Department of Infectious Diseases and Microbiology[2] and Department of Epidemiology,[3] University of Pittsburgh Graduate School of Public Health and School of Medicine, Pittsburgh, Pennsylvania; Departamento de Epidemiologia e Métodos Quantitativos em Saúde, Escola Nacional de Saúde Pública, FIOCRUZ, Rio de Janeiro, Brazil[4]; and Microbiology Laboratory, Minnesota Department of Health, Minneapolis, Minnesota[5]

Received 16 April 2003/Returned for modification 7 July 2003/Accepted 11 September 2003

***Escherichia coli* O157:H7 is a major cause of food-borne illness in the United States. Outbreak detection involves traditional epidemiological methods and routine molecular subtyping by pulsed-field gel electrophoresis (PFGE). PFGE is labor-intensive, and the results are difficult to analyze and not easily transferable between laboratories. Multilocus variable-number tandem repeat (VNTR) analysis (MLVA) is a fast, portable method that analyzes multiple VNTR loci, which are areas of the bacterial genome that evolve quickly. Eighty isolates, including 21 isolates from five epidemiologically well-characterized outbreaks from Pennsylvania and Minnesota, were analyzed by PFGE and MLVA. Strains in PFGE clusters were defined as strains that differed by less than or equal to one band by using *Xba*I and the confirmatory enzyme *Spe*I. MLVA was performed by comparing the number of tandem repeats at seven loci. From 6 to 30 alleles were found at the seven loci, resulting in 64 MLVA types among the 80 isolates. MLVA correctly identified the isolates from all five outbreaks if only a single-locus variant was allowed. MLVA differentiated strains with unique PFGE types. Additionally, MLVA discriminated strains within PFGE-defined clusters that were not known to be part of an outbreak. In addition to being a simple and validated method for *E. coli* O157:H7 outbreak detection, MLVA appears to have a sensitivity equal to that of PFGE and a specificity superior to that of PFGE.**

Escherichia coli O157:H7 has emerged as an important food-borne pathogen infecting thousands of people per year (17). Most *E. coli* O157:H7 infections are caused by exposure to bovine feces-contaminated food or water. The clinical syndromes caused by this organism include bloody diarrhea and hemolytic-uremic syndrome (4). There have been numerous large food-borne outbreaks of *E. coli* O157:H7-related bloody diarrhea and hemolytic-uremic syndrome (1, 5, 6, 21).

The public health impact of *E. coli* O157:H7 has created a need for improved preventative food-handling techniques and enhanced surveillance for outbreaks. In addition to traditional epidemiological investigations, pulsed-field gel electrophoresis (PFGE) is used to discriminate between outbreak and sporadic strains (2). Although PFGE has been successful, several factors have led researchers to search for alternative methods. The PFGE method, while simple and inexpensive, takes several days to complete, produces results that are suboptimal for interlaboratory comparisons, and can be subjective because it is based on banding patterns (19).

Sequenced-based methods, such as multilocus sequence typing (MLST), are becoming powerful subtyping tools in molecular epidemiology. These methods have the advantage of being easily standardized and automated. MLST, while successful for the differentiation of other organisms (9, 16, 18, 25), has been unable to discriminate among *E. coli* O157:H7 isolates (19). In one study, no variation was detected in seven housekeeping genes and little variation was noted in two surface protein genes (19).

Given the poor discriminatory power of MLST for *E. coli* O157:H7, we decided to target short tandem repeats (TRs), which are areas of the bacterial genome that evolve rapidly. Targeting of these elements, which often vary in number among different strains of the same species (the definition of a variable-number TR [VNTR]), has successfully been used to discriminate between strains of prokaryotes (24). Multiple-locus VNTR analysis (MLVA) involves determination of the number of repeats at multiple loci, thereby providing a powerful tool for assessing the genetic relationships between bacterial strains of the same species. In a study of the highly clonal organism *Bacillus anthracis*, 426 isolates that were previously homogeneous by other molecular subtyping methods, including PFGE, were separated into 89 distinct genotypes by MLVA (14). MLVA has several advantages over PFGE because, like MLST, the output is highly objective, making the data amenable to automated computer analysis for the rapid detection of outbreaks and easy to compare across laboratories.

The two completely sequenced *E. coli* O157:H7 genomes have allowed us to identify many TRs (11, 20). We initially focused on short TRs that varied in the number of times that they were repeated between the two reference genomes. We were then able to compare MLVA and PFGE for their abilities to detect outbreaks. The high discriminatory power of PFGE demands that a competing technique be equal to PFGE, if not superior to PFGE, in its ability to differentiate between iso-

* Corresponding author. Mailing address: Infectious Diseases Epidemiology Research Unit, University of Pittsburgh Graduate School of Public Health and School of Medicine, 521 Parran Hall, 130 DeSoto St., Pittsburgh, PA 15261. Phone: (412) 624-3137. Fax: (412) 624-2256. E-mail: lharriso@edc.pitt.edu.

TABLE 1. Years and states of isolation and outbreak numbers for isolates included in this study[a]

Group[b]	Yr	Location	Outbreak no.[c]	Strain designation
1	1995	Minnesota (daycare)	4	E96001161
				E96001162
				E96001177
				I96001815
	1996	Minnesota (daycare)	1	I96003168
				I97000024
				I97000025
				I97000027
	1996	Minnesota (daycare)	2	I97001003
				I97001017
				I97001180
				I97001040
	1996	Minnesota (daycare)	3	I97000770
				I97001176
	2001	Pennsylvania (restaurant)	5	PHIDL 51
				PHIDL 52
				PHIDL 53
				PHIDL 54
				PHIDL 57
				PHIDL 59
				PHIDL 60
2	1995	Minnesota		E96000049
	1996	Minnesota		E97001162
				E97001249
	1999	Pennsylvania		PHIDL 5
				PHIDL 7
				PHIDL 11
				PHIDL 14
				PHIDL 15
				PHIDL 18
	2000	Pennsylvania		PHIDL 19
				PHIDL 21
				PHIDL 25
				PHIDL 26
				PHIDL 29
				PHIDL 33
				PHIDL 34
				PHIDL 36
				PHIDL 37
				PHIDL 38
				PHIDL 41
				PHIDL 42
				PHIDL 43
				PHIDL 44
				PHIDL 45
				PHIDL 42
				PHIDL 43
				PHIDL 44
				PHIDL 45
	2001	Pennsylvania		PHIDL 46
				PHIDL 47
				PHIDL 48
				PHIDL 49
				PHIDL 55
				PHIDL 56
				PHIDL 58
				PHIDL 61
				PHIDL 62
3	1996	Minnesota		E97001568
	1999	Pennsylvania		PHIDL 2
				PHIDL 3
				PHIDL 4
				PHIDL 8
				PHIDL 9
				PHIDL 10
				PHIDL 12
				PHIDL 13
				PHIDL 16
				PHIDL 17
	2000	Pennsylvania		PHIDL 20
				PHIDL 22
				PHIDL 23
				PHIDL 24
				PHIDL 27
				PHIDL 28
				PHIDL 30
				PHIDL 31
				PHIDL 35
				PHIDL 39
				PHIDL 40
	2001	Pennsylvania		PHIDL 50
				PHIDL 63
Reference	1982	Michigan		EDL933
	1993	CDC[d]		G5244
	1996	Japan		Sakai RIMD 0509952

[a] A total of 80 isolates were tested.
[b] See text for definitions of groups 1 to 3.
[c] Information is provided only for outbreak isolates.
[d] CDC, Centers for Disease Control and Prevention.

lates. In this study, we sought to develop an MLVA assay that is useful for detecting outbreaks while being at least as discriminatory as PFGE and easier to perform than PFGE.

MATERIALS AND METHODS

***E. coli* O157:H7 strains.** All *E. coli* O157:H7 strains ($n = 58$) collected by the Allegheny County Health Department from 1999 to 2001 were provided to the Public Health Infectious Disease Laboratory (PHIDL) at the University of Pittsburgh (Table 1). These strains were not associated with known outbreaks, with the exception of seven isolates from a single restaurant-associated outbreak in August and September 2001. Two strains (PHIDL isolates 27 and 28) collected from the Allegheny County Health Department were Shiga toxin-positive *E. coli* O157:NM (NM, nonmotile). A sample of isolates from the Minnesota Department of Health was also included; these were isolates from four outbreaks ($n = 14$) and sporadic isolates ($n = 4$) collected from 1995 and 1996 (Table 1).

American Type Culture Collection strain EDL933 and strain RIMD 0509952 from Sakai, Japan, were used as reference strains for MLVA (11, 20), while G5244 from the Centers for Disease Control and Prevention was used as the reference strain for PFGE.

Each isolate was classified into one of three groups. Group 1 isolates were from known outbreaks and were associated with a specific PFGE cluster. Strains that had two or more band differences by PFGE with *Xba*I and *Spe*I and that were not known to be associated with an outbreak were classified as group 2. Finally, strains that were different by one band or less by PFGE and that were not associated with a known outbreak were classified as group 3.

PFGE. PFGE analysis was performed according to the Centers for Disease Control and Prevention PulseNet protocol, with minor variations, as described previously (19). The bacterial DNA was restricted with *Xba*I or the confirmatory enzyme *Spe*I (New England Biolabs, Beverly, Mass.). The switch times for *Xba*I and *Spe*I were 3 to 40 and 3 to 20 s, respectively, and the PFGEs with both enzymes ran for 21 h. Dendrograms were created with the Molecular Analyst program (Bio-Rad, Hercules, Calif.) by using the Dice coefficient and a position tolerance of 1.3%. Isolates were classified as belonging to the same PFGE cluster if they had one band difference or less with both *Xba*I and *Spe*I.

Potential VNTRs. More than 100 potential TRs were found in the two fully sequenced *E. coli* O157:H7 genomes, EDL933 (GenBank accession no. AE005174) and Sakai (GenBank accession no. BA000007), with Tandem Repeats Finder software (3). After identification of all TRs that were common to both strains, we chose six TRs that were different in number between the two strains. Among the TRs that were not variable between the two reference genomes, some were found to be variable among the study isolates. For example, because of success with several 6-bp TRs that were variable between the reference genomes, we tested some that were not variable and found them to be variable among the study isolates, such as TR5.

DNA isolation and PCR amplification and sequencing. DNA was isolated by the Prepman Ultra Protocol (Applied Biosystems, Foster City, Calif.). All Allegheny County and Minnesota isolates were analyzed at seven loci. Primers were based on the sequences of the Sakai and EDL933 genomes (11, 19) and were designed by using sequences that were found on the Primer Finder website.

Primers were designed (IDT Inc., Coralville, Iowa) for the amplification and sequencing of the targeted repeat region (Table 2) to verify that the differences seen were due to the variability in the TR region rather than another genetic event (proof-of-concept primers). Each 30-μl PCR mixture contained 3 μl of 10× PCR buffer, 1.5 mM $MgCl_2$, 0.33 μM each primer, 25 μM each deoxyribonucleotide, 1.5 U of the recombinant *Taq* DNA polymerase (Invitrogen, Carlsbad, Calif.), and 1 μl of DNA template. All steps in the PCR thermocycling program except the annealing temperatures were identical for the seven reactions. The annealing temperatures are presented in Table 2. The samples were placed on a GeneAmp PCR System 9700 (Applied Biosystems) and the temperature was raised to 94°C for 4 min, followed by 35 cycles of 94°C for 45 s, 50 to 57°C for 45 s, and 72°C for 1 min. The final hold was for 5 min at 72°C. The PCR products were purified with Exo-Sap It (U.S. Biochemical Corporation, Cleveland, Ohio).

The forward and reverse strands of the PCR products were sequenced with an ABI PRISM 3700 genetic analyzer (Applied Biosystems) and the Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) by the protocol described previously (19). Contigs were created using the base calling and fragment assembling software programs, Phred and Phrap (7, 8). Once the sequences were aligned, the numbers of repeats were counted by using the ClustalX (13) or Chromas (Technelysium Pty. Ltd.) program.

Data analysis. The unweighted pair group method with arithmetic means (UPGMA) was used to generate the PFGE and MLVA dendrograms.

The sensitivity and specificity of MLVA for the detection of outbreaks were calculated by using the pairwise distances between isolates after being analyzed by UPGMA (also known as cophenetic distances) to determine which cutoff point would yield the highest values for both isolates in each pair. Sensitivity, a measure of the ability to detect outbreaks, was defined as the ability of the MLVA-derived dendrogram to classify a pair of group 1 isolates as belonging to an outbreak. Specificity, a measure of the power to discriminate unrelated isolates, was defined as the ability of an MLVA-derived dendrogram to classify a pair of group 2 isolates as not belonging to an outbreak.

We observed that the single-locus variants (SLVs) that occurred during outbreaks differed by only a single TR. To test the hypothesis that only a single TR difference would likely occur during an outbreak, we determined the likelihood that such a difference would occur between group 2 isolates. This was achieved by constructing an empirical distribution of the distances in that group after logarithmic transformation was performed to account for normality and to allow negative values. Using another approach, we compared the mean distance

TABLE 2. VNTR locus-specific primers and characteristics[a]

TR name	Forward primer sequence	Reverse primer sequence	Annealing temp (°C)	Tandem repeat sequence	No. of repeats		No. of alleles	Diversity[b]	Found in *E. coli* K-12	Inside ORF[c]
					Minimum	Maximum				
TR1	ACTGCATGATAAGCCTCAGG	GGAAGGAAGCTGATAGGT	57	AAATAG	4	20	12	0.89	No	No
TR2	CGCAGTTGATACCTACGG	TGATTAAGCGTGTACTGA	53	TGGCTC	7	58	30	0.96	No	Yes
TR3	TCTTGTCAATATAGATTGG	GCCACACTGCGAGTATAGAG	50	TATCTT	3	10	8	0.71	No	Yes
TR4	GGTGATGGCTTGATATTGA	GGACAACTTGTAGTACAAG	53	TGCAAA	2	9	7	0.58	Yes	No
TR5	GTTGATTATCATGGTATGTC	TAGCAGATGTTCGTTCCT	51	AAGGTG	6	21	13	0.86	No	Yes
TR6	GATGGTTCGACTAACCGTTAT	TGCTGAAACTGACGACCAGT	53	TTAAATAATCTACAGAAG	7	12	6	0.72	Yes	Yes
TR7	CGCAGTGATCATTATTAGC		50	GACCAC	4	9	6	0.68	Yes	Yes

[a] Primers used for the initial amplification and sequencing of the selected tandem repeats for all isolates and characteristics of each TR locus.
[b] Diversity is based on Nei's marker diversity, which is $1 - \Sigma(\text{allele frequency})^2$, and on 63 or 64 unique genotypes.
[c] Most of the open reading frames (ORFs) were hypothetical based on the sequence of either Sakai or EDL933 in the National Center for Biotechnology Information database (11, 19).

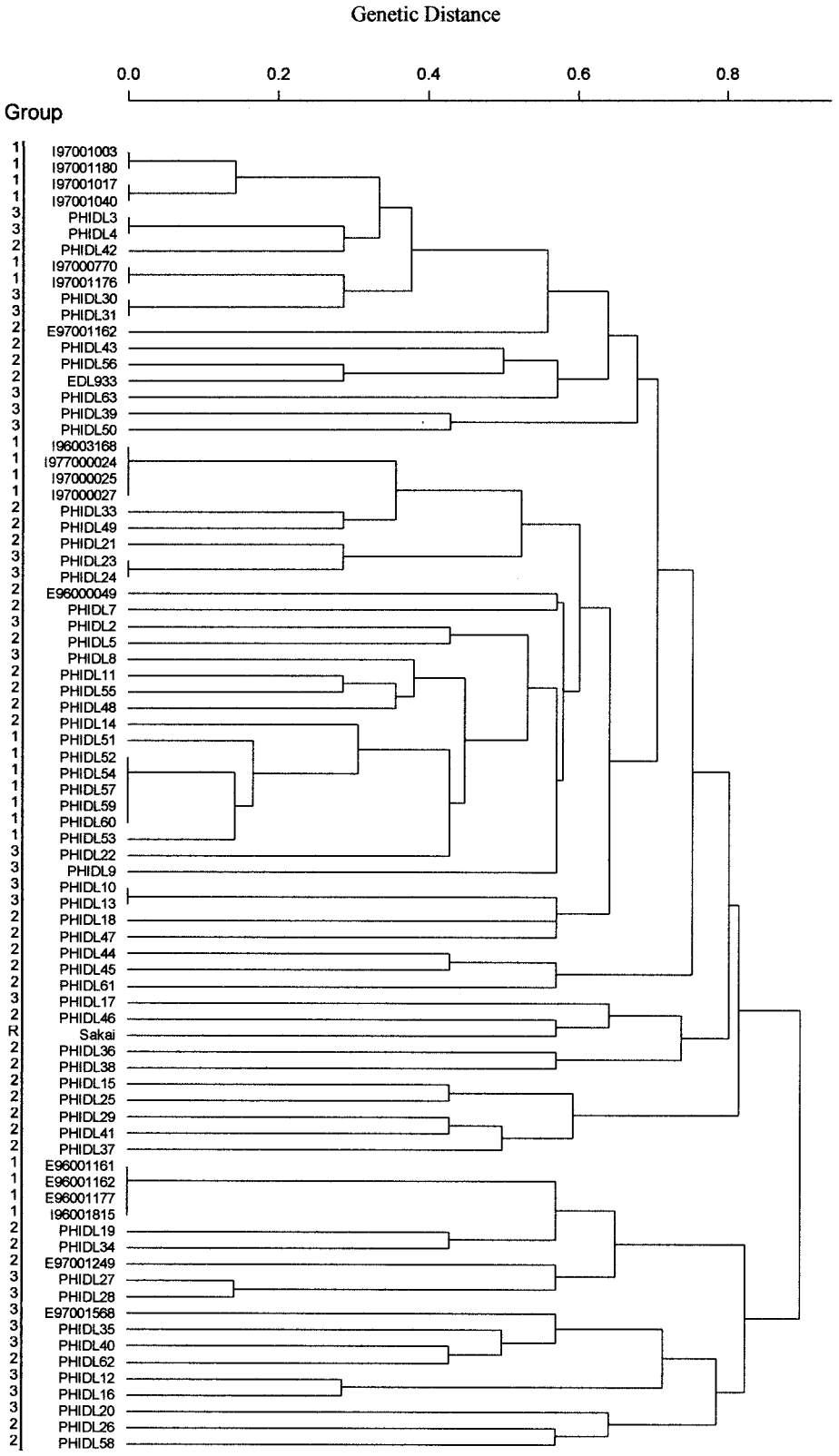
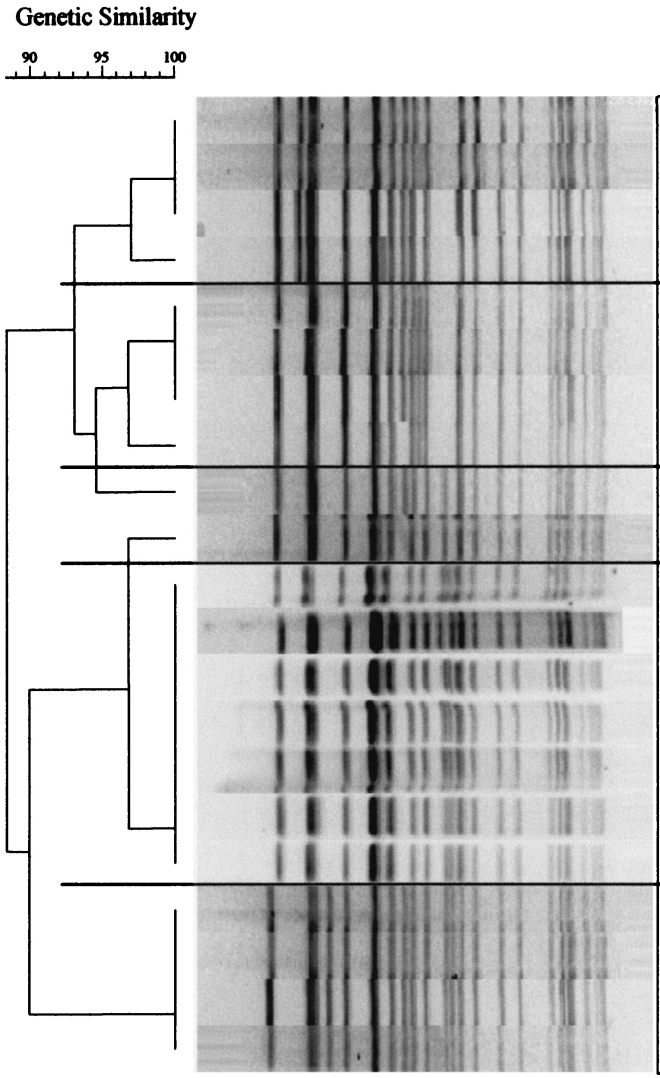


FIG. 1. MLVA dendrogram based on the allelic profiles of the 80 *E. coli* O157:H7 isolates. See Table 1 for isolate details.



MLVA Results

Isolate	Outbreak Number	TR1	TR2	TR3	TR4	TR5	TR6	TR7
E96001161	4	9	40	7	4	7	11	7
E96001177	4	9	40	7	4	7	11	7
E96001162	4	9	40	7	4	7	11	7
I96001815	4	9	40	7	4	7	11	7
I97001003	2	11	30	7	5	9	11	7
I97001180	2	11	30	7	5	9	11	7
I97001040	2	11	31	7	5	9	11	7
I97001017	2	11	31	7	5	9	11	7
I97000770	3	10	24	7	5	9	11	7
I97001176	3	10	24	7	5	9	11	7
PHIDL60	5	15	25	7	5	11	10	6
PHIDL51	5	15	25	7	5	11	9	6
PHIDL52	5	15	25	7	5	11	10	6
PHIDL53	5	16	25	7	5	11	10	6
PHIDL54	5	15	25	7	5	11	10	6
PHIDL57	5	15	25	7	5	11	10	6
PHIDL59	5	15	25	7	5	11	10	6
I97000024	1	13	19	7	5	11	10	5
I97000025	1	13	19	7	5	11	10	5
I96003168	1	13	19	7	5	11	10	5
I97000027	1	13	19	7	5	11	10	5

FIG. 2. Types of all group 1 isolates from five outbreaks obtained by PFGE with *Xba*I and the corresponding MLVA types. The numbers under each TR locus reflect the number of times that the TR was found in that isolate. The horizontal lines through the dendrogram and chart are used to visually demarcate the outbreak isolates.

among pairs of group 1 and group 2 isolates by a Student's *t* test. All analyses were done with the statistical package R (12).

RESULTS

PCR amplification and sequence analysis of potential VNTRs. Initially, 11 loci of a subset of 16 PHIDL isolates were sequenced to determine if the TR locus had sufficient variability (data not shown). If variation at a particular locus existed in this small subset, the loci of the remaining isolates were amplified. We found that seven loci had multiple alleles with substantial variability (Table 2 and Fig. 1). The seven primer sets amplified all isolates at all loci with two exceptions: isolate E96001161 with the TR2 primers and isolate E97001249 with the TR5 primers. These data were counted as missing for the MLVA analysis. We sequenced the seven loci of all of our isolates to confirm that the size variations seen in the PCR products were due to the number of TRs. In all cases, the size variation that we observed was due to the number of TRs. Rarely, there was sequence variation within the repeat.

Locus characteristics. From 6 to 30 alleles were found for the seven loci, with VNTRs repeating as few as 2 times at one locus and as many as 58 times at another (Table 2). The diversity for each locus was calculated on the basis of either 63 or 64 unique genotypes; the former number was used for TR2 and TR5 because of unsuccessful PCR amplification.

MLVA for outbreak detection. Group 1 included organisms from five separate outbreaks, each of which was associated with a specific PFGE cluster (Fig. 2). All isolates from outbreaks 1, 3, and 4 had identical MLVA types. Among the isolates from the remaining two outbreaks, SLVs were a result of single TR differences in all instances. Among the isolates from outbreak 2, two isolates had 30 rather than 31 repeats at locus TR2. Among the isolates from outbreak 5, one isolate had 16 rather than 15 repeats at locus TR4 and another isolate

had 9 rather than 10 repeats at locus TR6. When SLVs were considered part of the same outbreak cluster, the sensitivity of MLVA for the identification of outbreak strains as such by using all seven loci was 100% (21 of 21 strains).

The outbreak 2 isolates differed from the outbreak 3 isolates at two loci. The outbreaks involved person-to-person transmission, were separated in time by 2 weeks in September 1996, and occurred in cities about approximately 75 mi apart. There was no known epidemiological connection between the two outbreaks.

The probability that a pair of isolates not belonging to an outbreak had at most a one TR difference was estimated to be 6.93×10^{-6} when all seven loci were taken into account. The differences in the average distances in groups 1 (0.4) and 2 (14.4) were also highly significant ($P < 0.0001$) These data suggest that intralocus differences that occur during outbreaks occur one TR at a time, whereas unrelated isolates are much more likely to differ by more than one TR.

MLVA for discriminating outbreak isolates from sporadic cases. Each group 2 isolate had a unique MLVA type (Fig. 3). Additionally, these isolates differed from other isolates included in this study by at least two VNTR loci, for a specificity of 100% (35 of 35 isolates). The discriminatory power was less with all possible combinations of six loci. For example, when TR1or TR2 was excluded, PHIDL 14, a group 2 isolate, was included in outbreak 5 if a single-locus difference was allowed. In addition, PHIDL 3 and PHIDL 4, both group 3 isolates, differed from outbreak 2 and 3 isolates by one locus. When TR7 was excluded, PHIDL 30 and PHIDL 31, both of which are group 3 isolates, differed from outbreak 3 isolates by only a single locus. Similar results were encountered with the exclusion of each of the remaining loci.

MLVA for discriminating strains related by PFGE. After restriction with *Xba*I, the 24 group 3 isolates were found to group together in seven PFGE-based clusters, even though they were not part of any identified outbreaks. Some of the isolates were further subgrouped after restriction with *Spe*I by the PulseNet protocol (22; data not shown). Since these strains had not been identified as part of an outbreak, they could not be included in the calculation of sensitivity and specificity. However, the PFGE and MLVA results were compared to provide insights into the relative discriminatory powers of these two methods by using the limited epidemiological information that was available for these isolates (Fig. 4).

The *Xba*I-based cluster containing PHIDL isolates 2, 9, 22, 30, and 31 was subdivided by *Spe*I into two clusters, with one cluster consisting of the two isolates recovered in 1999 and the second cluster consisting of the three isolates (PHIDL isolates 22, 30, and 31) recovered in 2000. MLVA provided further discrimination among some of the isolates from 2000. PHIDL isolates 30 and 31 had identical MLVA types, and these two organisms were also isolated 1 day apart. In contrast, PHIDL 22 differed at three loci from PHIDL isolates 30 and 31, and the time of its recovery was separated by 3 months from those for the other two isolates.

In addition to PHIDL isolates 30 and 31, other isolates that were clustered by PFGE were also highly related by MLVA. For example, PHIDL isolates 3 and 4 were identical by MLVA and were isolated in Allegheny County 1 day apart. Taken together with the analysis of the group 1 isolates, the data suggest that these isolates were part of an unrecognized outbreak.

On the other hand, MLVA also differentiated some group 3 strains. For example, PHIDL isolates 39 and 50 were different at three MLVA loci and were isolated 11 months apart. PHIDL isolates 8 and 23 were also clustered by PFGE (indistinguishable by PFGE with *Xba*I and one band difference by PFGE with *Spe*I), were detected more than a year apart, and differed at five MLVA loci. These data suggest that MLVA is able to distinguish among unrelated strains that may be falsely clustered together by PFGE.

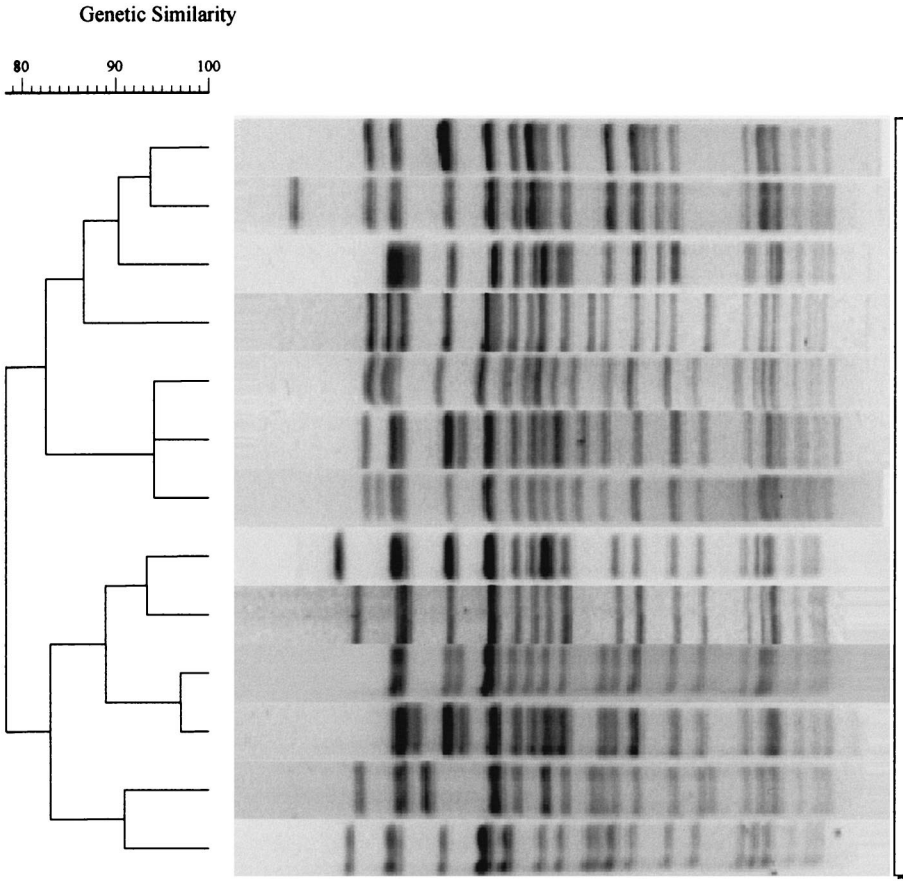
The preliminary data suggest that even with a second enzyme, PFGE is unable to differentiate strains as well as MLVA is. These data suggest that group 3 isolates consist of both previously unrecognized *E. coli* O157:H7 outbreak isolates and unrelated isolates that PFGE erroneously clustered together. If these data are confirmed in future studies, they will indicate that MLVA is more specific than PFGE for the detection of outbreaks caused by this organism.

DISCUSSION

The MLVA assay that we developed was highly sensitive for the identification of *E. coli* O157:H7 outbreaks, while at the same time it was able to accurately discriminate among sporadic isolates. The use of cutoffs of a difference of one locus or less with two TR differences allowed us to correctly classify all group 1 and group 2 isolates. The data for our group 1 isolates, consisting of well-characterized outbreak isolates, suggest that isolates that differ at no more than a single locus are highly related and should be considered part of the same outbreak. It was striking that the SLVs that we identified among group 1 isolates all differed by a single repeat, suggesting that SLVs that occur during outbreaks are likely to differ by a small number of repeats (C. Keys, Z. Jay, A. Fleishman, J. Fox, G. Evans, and P. Keim, Abstr. 103rd Gen. Meet. Am. Soc. Microbiol., poster C-399, 2003). Whether all intraoutbreak SLVs differ by a single repeat remains to be seen. However, the data for our group 3 isolates suggest that the difference may not always be a single TR because PHIDL isolates 27 and 28, which were likely from a point source, differed by two repeats at locus TR2.

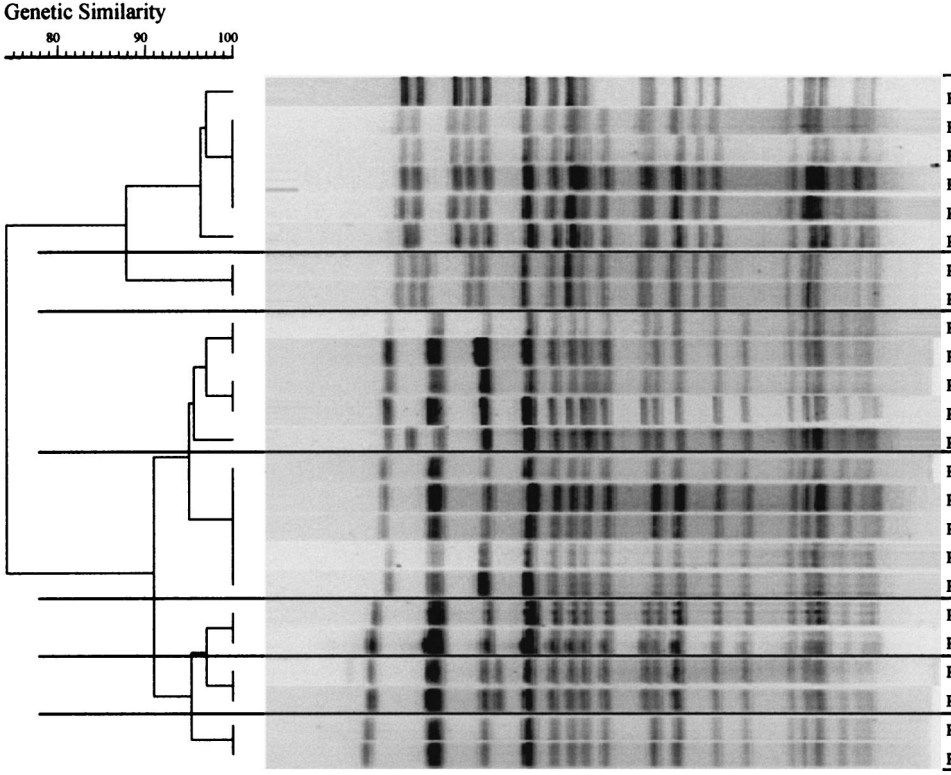
Importantly, MLVA was able to distinguish among some group 3 isolates that appeared to be highly related by PFGE. On the basis of a comparison of the results of these two assays and the available epidemiological information, it appears that this group included both sporadic and outbreak-related strains. Thus, MLVA was more discriminatory than PFGE with the group of isolates that we studied.

The major implication of this finding is that if MLVA is used as part of routine public health surveillance, it may result in fewer false-positive signals suggestive of an outbreak. This finding, in addition to the fact that MLVA has many other advantages over PFGE, suggests that MLVA is superior to PFGE. We are automating this process by analyzing fluorescently tagged PCR amplicons of the seven TR loci on a 3700 DNA analyzer, as described by Keim et al. (14). This will eliminate the sequencing step that was described in this experiment and further reduce user intervention, thereby increasing the efficiency of this protocol.



Isolate	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7
PHIDL15	17	39	4	5	15	8	7
PHIDL41	15	15	4	4	13	8	6
PHIDL37	12	58	4	5	9	8	6
E97001249	8	20	3	4	No	8	7
PHIDL26	11	29	5	4	12	9	6
PHIDL34	11	15	5	4	7	10	7
PHIDL19	12	18	5	4	7	9	7
PHIDL55	13	28	7	5	10	10	6
E96000049	16	17	7	5	10	10	7
PHIDL43	15	28	9	5	8	11	6
PHIDL47	20	35	7	9	14	10	6
PHIDL46	14	26	8	4	10	10	5
PHIDL60	15	25	7	5	11	10	6

FIG. 3. Types of a sample of group 2 isolates obtained by PFGE with *Xba*I and the corresponding MLVA types. The numbers under each TR locus reflect the number of times that the TR was found in that isolate. No, no product from the PCR.



Isolate	Onset Date	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7
E97001568	12/4/1996[1]	10	No	5	3	21	7	8
PHIDL12	8/11/1999	11	14	7	3	21	8	7
PHIDL35	7/21/2000	10	14	5	4	14	8	9
PHIDL16	8/31/1999	11	14	6	3	9	8	7
PHIDL20	3/6/2000[1]	8	23	5	4	18	8	8
PHIDL40	8/7/2000	10	9	5	3	9	8	9
PHIDL27	5/21/2000	9	33	6	4	14	8	7
PHIDL28	5/27/2000	9	35	6	4	14	8	7
PHIDL8	3/13/1999	15	16	8	5	10	10	6
PHIDL23	3/21/2000	17	22	7	5	9	10	5
PHIDL24	3/25/2000	17	22	7	5	9	10	5
PHIDL63	10/5/2001	14	19	8	5	10	11	6
PHIDL17	10/18/1999	15	31	8	5	6	10	4
PHIDL22	3/1/2000	11	11	7	5	11	11	6
PHIDL30	6/27/2000	10	38	7	5	9	11	6
PHIDL31	6/26/2000	10	38	7	5	9	11	6
PHIDL9	5/5/1999	13	18	7	5	8	9	6
PHIDL2	6/21/1999	14	16	7	4	13	10	6
PHIDL39	8/8/2000	15	24	7	7	10	11	6
PHIDL50	6/6/2001	12	29	7	4	10	11	6
PHIDL10	8/6/1999	16	34	7	8	12	10	6
PHIDL13	8/24/1999	16	34	7	8	12	10	6
PHIDL3	7/18/1999	13	10	7	5	9	11	7
PHIDL4	7/17/1999	13	10	7	5	9	11	7

FIG. 4. Types of a sample of group 3 isolates obtained by PFGE with *Xba*I and the corresponding MLVA types. The numbers under each TR locus reflect the number of times that the TR was found in that isolate. The horizontal lines through the dendrogram and the chart are used to visually demarcate the isolates grouped by PFGE. [1], the dates represent the culture date and not the date of the onset of symptoms; No, no product from the PCR.

VNTRs are rapidly evolving genomic elements that have successfully been used for the molecular typing of other pathogens, such as *B. anthracis*, *Yersinia pestis*, and *Mycobacterium tuberculosis* (10, 14, 15). One potential concern is that VNTRs evolve so rapidly that multiple MLVA types would emerge during an outbreak initially caused by a single clone. In fact, we observed SLVs among the isolates from two of the five outbreaks that we studied. This is similar to findings obtained by PFGE, by which differences of up to several bands can be observed by PFGE during outbreaks (23). Whether MLVA frequently exhibits a degree of diversity that diminishes its utility for outbreak detection will need to be studied with additional isolates.

We primarily chose relatively short TRs for two reasons. First, shorter repeats may be associated with an increased potential of DNA polymerase slippage, resulting in either the loss or the gain of a TR (24). Second, shorter repeat sizes may facilitate automation by reducing the potential overlap of different loci during the run on the DNA sequencer. Of the seven VNTR loci that we analyzed, a minimum of 6 alleles and a maximum of 30 alleles were found at one locus, which gives MLVA tremendous discriminatory abilities that are superior to those of PFGE on the basis of the results for our isolates.

In conclusion, our data suggest that MLVA should be considered an alternative method for the subtyping of *E. coli* O157:H7 isolates because it is technically superior to PFGE. The MLVA protocol is amenable to the handling of large sample sets and can easily be standardized for comparisons of results among different laboratories.

ACKNOWLEDGMENTS

We thank the Allegheny County Health Department for its support, especially Bruce Dixon, Joan McMahon, and Mary Blazina for assistance with obtaining the isolates and Sharon Silvestri for providing epidemiological data. We also thank Jane Marsh for thoughtful review of the manuscript.

Support for this study was provided in part by the Technical Support Working Group through the Department of Defense (grant N41745-03-C-4005) and a Research Career Award (K24 AI52788 to L.H.H.) from the National Institute of Allergy and Infectious Diseases.

Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not reflect the views of the Public Health Service or the University of Pittsburgh.

REFERENCES

1. **Bell, B. P., M. Goldoft, P. M. Griffin, M. A. Davis, D. C. Gordon, P. I. Tarr, C. A. Bartleson, J. H. Lewis, T. J. Barrett, J. G. Wells, et al.** 1994. A multistate outbreak of *Escherichia col*i O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. JAMA **272**:1349–1353.
2. **Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm.** 1997. Surveillance by molecular subtype for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.
3. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.
4. **Besser, R. E., P. M. Griffin, and L. Slutsker.** 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease. Annu. Rev. Med. **50:**355–367.
5. **Breuer, T., D. H. Benkel, R. L. Shapiro, W. N. Hall, M. M. Winnett, M. J. Linn, J. Neimann, T. J. Barrett, S. Dietrich, F. P. Downes, D. M. Toney, J. L. Pearson, H. Rolka, L. Slutsker, and P. M. Griffin.** 2001. A multistate outbreak of *Escherichia coli* O157:H7 infections linked to alfalfa sprouts grown from contaminated seeds. Emerg. Infect. Dis. **7:**977–982.
6. **Centers for Disease Control and Prevention.** 2002. Multistate outbreak of *Escherichia coli* O157:H7 infections associated with eating ground beef—United States, June-July 2002. Morb. Mortal. Wkly. Rep. **51:**637–639.
7. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.
8. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.
9. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics **154:**1439–1450.
10. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144**(Pt 5)**:**1189–1196.
11. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
12. **Ihaka, R., and R. Gentleman.** 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. **5:**299–314.
13. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson.** 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:**403–405.
14. **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 variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.* J. Bacteriol. **182:**2928–2936.
15. **Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim.** 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pesti*s genome. J. Clin. Microbiol. **39:**3179–3185.
16. **Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulakvelidze.** 2002. Multilocus sequence typing for characterization of clinical and environmental S*almonella* strains. J. Clin. Microbiol. **40**:1626–1635.
17. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5:**607–625.
18. **Nicolas, P., G. Raphenon, M. Guibourdenche, L. Decousset, R. Stor, and A. B. Gaye.** 2000. The 1998 Senegal epidemic of meningitis was due to the clonal expansion of A:4:P1.9, clone III-1, sequence type 5 *Neisseria meningitidis* strains. J. Clin. Microbiol. **38:**198–200.
19. **Noller, A. C., M. C. McEllistrem, O. C. Stine, J. G. Morris, Jr., D. J. Boxrud, B. Dixon, and L. H. Harrison.** 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J. Clin. Microbiol. **41:**675–679.
20. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
21. **Samadpour, M., J. Stewart, K. Steingart, C. Addy, J. Louderback, M. McGinn, J. Ellington, and T. Newman.** 2002. Laboratory investigation of an *E. coli* O157:H7 outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington. J. Environ. Health **64:**16–20, 25, 26.
22. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.
23. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233–2239.
24. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62:**275–293.
25. **Zhou, J., M. C. Enright, and B. G. Spratt.** 2000. Identification of the major Spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing. J. Clin. Microbiol. **38**:977–986.