Figure 1. Isolation frequency of EAEC, EPEC, ETEC, EIEC, and STEC isolated from patients in Thailand in year 2006-2007.

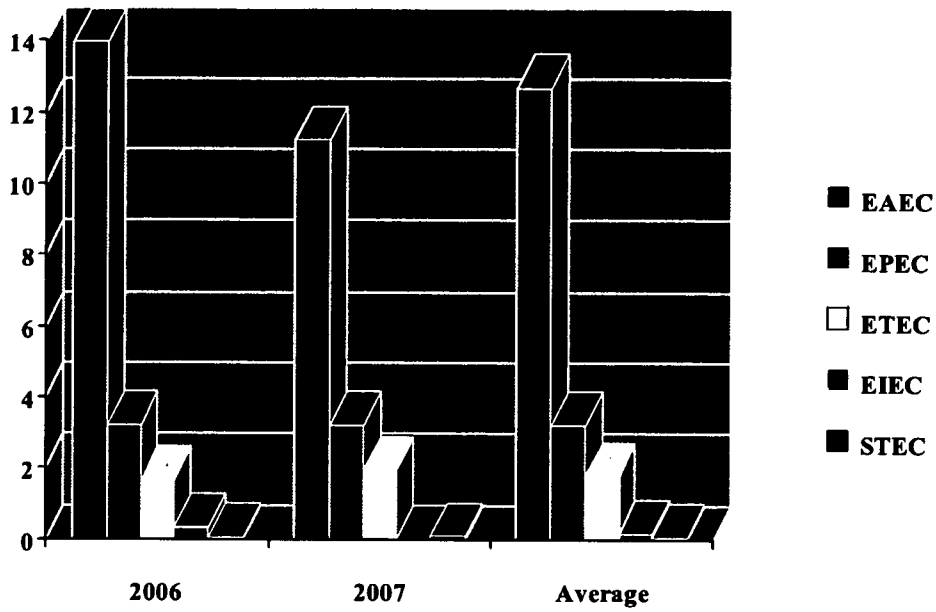
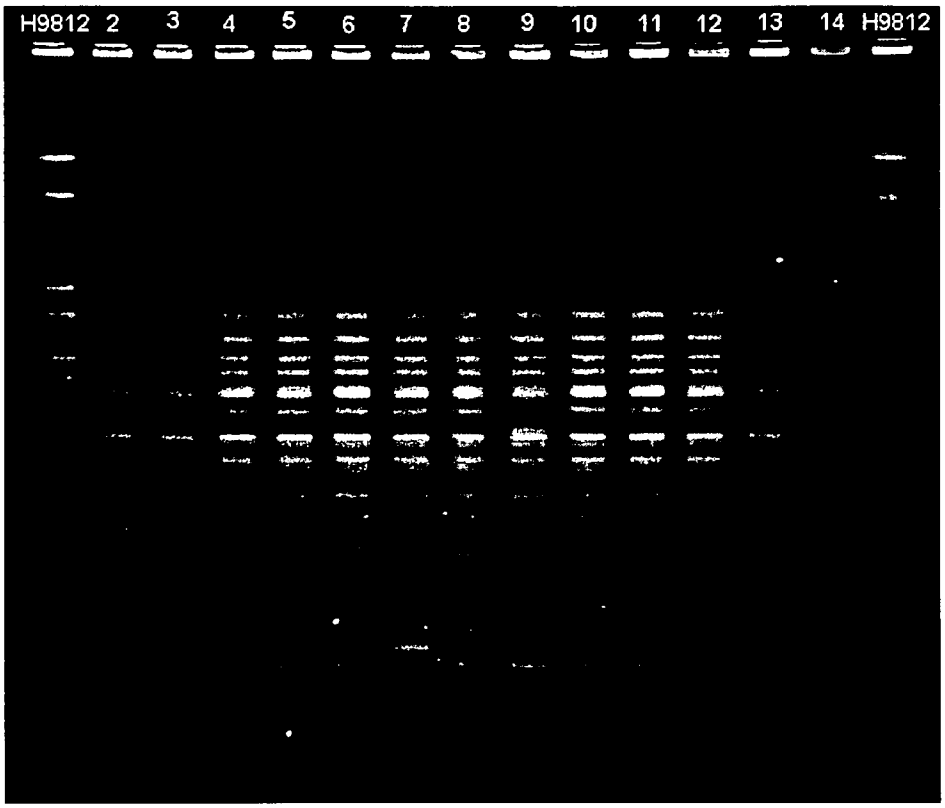


Figure 2. ***xba*I-digested PFGE Pattern of EAEC strains isolated from patients and food handlers in an outbreak of food-borne illness.**



Lane 1, 15 *Salmonella* ser. Braenderup H9812, *DNA marker*
Lane 2-5, EAEC strains collected from food handlers
Lane 6-9 EAEC strains collected from patients
Lane 10-14 EAEC strains collected from food handlers

The Utility of Multilocus Variable-Number Tandem Repeat Analysis as a Molecular Tool for Phylogenetic Study of *Shigella sonnei*

Chien-Shun Chiou[1,2*], Haruo Watanabe[3], You-Wun Wang[1], Wan-Ling Wang[1], Jun Terajima[3], Kwai-Lin Thong[4], Phung Dac Cam[5], Sheng Kai Tung[1]

[1]The Central Region Laboratory, Center of Research and Diagnostics, Centers for Disease Control, Taichung City 40855, Taiwan

[2]Institute of Medicine, Chung Shan Medical University, Taichung City 40201, Taiwan

[3]National Institute for Infectious Diseases, Tokyo, Japan

[4]University of Malaya, 50603 Kuala Lumpur, Malaysia

[5]Division of Enteric Infections, National Institute of Hygiene and Epidemiology, Vietnam

* Corresponding author

Chien-Shun Chiou: nipmcsc@cdc.gov.tw

Haruo Watanabe: haruwata@nih.go.jp

You-Wun Wang: pleione@cdc.gov.tw

Wan-Ling Wang: r93b47307@cdc.gov.tw

Sheng Kai Tung: kbj@cdc.gov.tw

Jun Terajima: terajima@nih.go.jp

Phung Dac Cam: cam@fpt.vn

Kwai-Lin Thong: thongkl@um.edu.my

Running title: MLVA for phylogenetic analysis of *S. sonnei*

ABSTRACT

A multilocus variable-number tandem repeat (VNTR) analysis (MLVA) method comprising 26 VNTR loci has been shown to be a powerful tool for disease outbreak investigation and disease surveillance. The MLVA could also be applied for phylogenetic study of *Shigella sonnei*. A total of 916 isolates were used to evaluate the usefulness of MLVA for phylogenetic study. The isolates were also characterized with pulsed-field gel electrophoresis (PFGE) and inter-IS1 spacer typing (IST) to compare genetic relationships among the isolates constructed with the MLVA data. Of the 916 total isolates, 703 IST1 isolates, which were descended very recently from a common ancestor collected over a period of 3.25 years, were used to evaluate variation rate of the VNTR loci and evolutionary speed of the IST1 clone. The MLVA data for the 703 IST1 isolates revealed two hypervariable, two moderately variable, and 22 low variable loci. The variation rate was strongly correlated with the number of repeat units within a locus. The evolutionary speeds were 0.39 loci, 0.049 loci and 0.0055 loci per year, respectively, for the IST1 clone defined by 26-, 24- (excluding two hypervariable loci) and 22- locus (excluding two hypervariable and two moderately variable loci) profiles. In parallel with the IST and PFGE clonal complexes, the panel of 916 isolates was grouped into 3 distinct MLVA clonal complexes. The results suggest that strains sharing 19 or more identical loci can be grouped into a clonal group. By excluding certain highly variable loci from the panel of 26 loci, MLVA can serve as a useful tool for the phylogenetic analysis of *S. sonnei* over various evolutionary timescales, and thus can help in the tracking of the global transmission of *S. sonnei* clones.

INTRODUCTION

Shigella sonnei is one of the causative agents of shigellosis. Unlike the other three *Shigella* species, *S. dysenterae*, *S. flexneri*, *S. boydii*, which are prevalent in developing countries, *S. sonnei* is predominant in industrialized countries (6) and is one of the major causes for travel-associated diarrheal disease (2). Transmission of *S. sonnei* strains between countries and continents occurs frequently via international traveling (7, 15).

Analyses of *Shigella* isolates by various genotyping methods have provided information on the genetic relatedness of isolates for the purposes of epidemiological investigation and evolutionary study. Among these genotyping methods, pulsed-field gel electrophoresis (PFGE) has been proven to be a powerful tool for discriminating *Shigella* strains and become standardized as a method for constructing an international molecular subtyping network for foodborne disease surveillance (14). However, PFGE is, at times, too discriminatory for investigating clonal relationship among *Shigella* strains evolved over years or decades. In order to study an *S. sonnei* epidemic, we previously developed an inter-IS1 spacer typing (IST) method for the subtyping of *S. sonnei* strains (1). IST is less discriminative than PFGE for *S. sonnei*, but is more suitable than PFGE for investigating the phylogenetic relationship among *S. sonnei* strains in circulation over a time span of a small number of years (17). Clustering analysis of PFGE and IST data is usually presented by a dendrogram, which is not appropriate for defining a clone or clonal complex because of lack of an objective cutoff value of pattern similarity.

Multilocus sequence typing (MLST) is a sequence-based typing method widely adopted for the evolutionary study of the various strains of a number of bacterial pathogens, as listed on the website, http://www.mlst.net/. MLST data can be presented by an eBURST algorithm that is applicable to explore the patterns of evolutionary descent of certain bacterial pathogens (3). However, MLST has not been successfully applied to *S. sonnei*. To date, there is still no an appropriate molecular tool for phylogenetic study of *S. sonnei*. In a previous study, we developed a multilocus variable-number tandem repeat (VNTR) analysis (MLVA) method for fine typing of *S. sonnei* isolates and that provided highly useful information for disease outbreak investigation and disease surveillance. Since the MLVA comprises as many as 26 VNTR loci, it may also be a useful tool for the evolutionary study of *S. sonnei*. In this study, we evaluated the variability of the 26 VNTR loci and the evolutionary speed of

an *S. sonnei* clone with a large number of isolates. The MLVA data for phylogenetic analysis was accessed.

MATERIALS AND METHODS

Bacterial strains: *Shigella* isolates were collected from hospitals or isolated in the laboratories of Centers for Disease Control, Taiwan, from 1996 to 2002. A total of 916 *S. sonnei* isolates were available for characterization. *Salmonella enterica* subsp. *enterica* serotype Braenderup H9812, kindly provided by Dr. B. Swaminathan of the Centers for Disease Control and Prevention, Atlanta, Ga, USA, was used as the PFGE size maker strain.

IST typing: IST typing was performed as previously described (1). The gel images were analyzed using BioNumerics version 4.5 (Applied Maths, Kortrijik, Belgium). Only those DNA fragments with sizes ranging within the size markers (0.1 – 3 kb) were counted for comparison.

PFGE: The PulseNet PFGE protocol for *S. sonnei* and other enterobacteria was used for PFGE analysis (13), except that 5 U of NotI was used to replace XbaI for the restriction digestion. PFGE images were analyzed using the fingerprint analysis software, BioNumerics version 4.5 (Applied Maths). A PFGE pattern with one or more DNA bands different from the others was taken to be a unique PFGE pattern.

Data analysis: Dendrograms constructed with IST and NotI-digested PFGE data were generated by the unweighted pair group method with arithmetic mean (UPGMA), using the Dice-predicted similarity value of two patterns. Phylogenetic trees constructed with MLVA profiles were generated by minimum spanning tree (MST) algorithm. Both of these two clustering programs were provided by BioNumerics version 4.5 (Applied Maths). The discriminatory power of the genotyping methods was presented by discriminatory index (5) and a 95% confidence interval (CI) (4). The average evolutionary distance was calculated as the accumulated number of locus differences of isolates to the ancestor genotype divided by the number of the isolates. The evolutionary speed was calculated as the accumulated number of locus differences of the total isolates to the ancestor divided by the number of total isolates and by the period of time (3.25 years) of the isolates collected.

RESULTS

IST genotypes for the 916 isolates. A total of 28 IST genotypes were identified in the 916 *S. sonnei* isolates, with a discriminatory index of 0.392 (CI, 0.351 to 0.433). Clustering analysis of the 28 IST patterns by UPGMA algorithm revealed 3 distinct clusters, designated as IST1, IST3 and IST6 clonal complexes (Fig. 1), consisting of 12, 12 and 4 IST types, respectively. The most distant genotypes within the IST6 clonal complex shared only 53% pattern similarity; the IST genotypes within the IST1 and IST3 clonal complexes shared more than 75% pattern similarity. IST1 was the most prevalent genotype within the IST1 clonal complex; it was identified in 89% (710) of the 796 isolates belonging to the IST1 clonal complex collected from 1996 to 2004 (Table 1). The IST1 clone was the main contributor to the shigellosis epidemic during the period of 2000-2003. The first identified IST1 isolate was obtained from a traveler returning from India in March 2000. Before the emergence of the IST1 clone, IST11 was the major circulating clone. The IST2 and IST23 strains which emerged in 2000 were derived from the IST1 clone; they were identified in patients from IST1 outbreaks and shared indistinguishable PFGE patterns with the majority of isolates collected from the outbreaks. The same scenario was also the case for IST30, which was derived from IST11. IST3 of the IST3 clonal complex, was another major clone attributed to the epidemic in 2000-2003. IS3, IST4, IST14, IST25, and IST26 were found to be closely related since certain strains from each of the 5 IST genotypes shared an indistinguishable PFGE pattern. IS4 was the earliest identified genotype among the 5 IST types. It is believed to have existed in the area over many years and evolved into new genotypes such as IST3 and IST25. Both IST6 and IST7 of the IST6 clonal complex intermittently emerged in 7 years. In general, most IST clones emerged and then disappeared soon thereafter; they usually caused an outbreak or just a few cases at time and then became extinct. Interestingly, all the imported isolates, obtained from travelers from China, Cambodia, India, Indonesia, Thailand, and Vietnam, had IST genotypes belonging to the IST1 clonal complex.

PFGE genotypes for the 916 isolates. PFGE with NotI identified 144 PFGE genotypes for the 916 isolate, with a discriminatory index of 0.917 (CI, 0.904 to 0.930). A dendrogram constructed with the 144 PFGE patterns presented 3 distinct clusters, designated as the J16N09.0018, J16N09.0019 and J16N09.0118 clonal complexes (Fig. 2). The 3 PFGE clonal complexes were equivalent to the clonal complexes, IST3, IST1 and IST6. The J16N09.0018 clonal complex consisted of all

the isolates belonging to the IST3 clonal complex. J16N09.0019 and J16N09.0118 consisted of all the isolates belonging to the IST1 and IST6 clonal complexes, respectively.

PFGE and MLVA genotypes for the 710 IST1 isolates. The 710 IST1 isolates were selected for comparing the discriminatory power of PFGE and MLVA. In total, 92 PFGE and 82 MLVA-26 genotypes, defined by 26 VNTR locus profiles, were identified for the 710 IST1 isolates, with a discriminatory index of 0.868 (CI, 0.848 to 0.888) for PFGE and of 0.896 (CI, 0.880 to 0.911) for MLVA. Although a greater number of PFGE types were found in the isolates, PFGE was not more discriminative than MLVA. The first two most prevalent PFGE types (J16N09.0019 and J16N09.0023) occupied 41% (293) of the 710 isolates, indicating that the genotypes were quite stable in the level of chromosomal structure. The two genotypes differed at a 220 kb DNA fragment, which was found to be an invasive plasmid. This invasive plasmid is unstable and easy to lose during cell division. *S. sonnei* with the invasive plasmid displayed a phase I phenotype, but, without the plasmid, displayed a phase II phenotype (unpublished data). The 293 isolates with the two PFGE genotypes were further discriminated into 43 MLVA-26 genotypes.

Phylogenetic relationship among the 710 IST1 isolates defined with MLVA-26 profiles. MST clustering analysis of the 82 MLVA-26 genotypes showed a relatively divergent phylogenetic tree. Of the 82 MLVA-26 types, 76 representing (for) 703 isolates were grouped in a large, tight cluster within which each genotype had a minimal distance of 1-2 loci with the others (Fig. 3A). The genotypes within the cluster were considered to have descended very recently from a common ancestor or founder; and were designated to be of the SS26.66 clone complex. The MLVA-26 genotype, SS26.66, was the founder of the clone, but SS26.1 was the most prevalent genotype. SS26.66 had more single locus variants (SLVs) than SS26.1. Most of the isolates collected in 2000 belonged to the SS26.1 genotype. The distance for isolates to the founder increased over time. The rest of the 6 genotypes making up the 7 isolates had a distance of 4 or more loci to the SS26.66 clone. Six of the 7 isolates were imported. The Indian strain had a distance of 4 loci to the clone.

Variability of VNTR loci in the 703 IST1 isolates. The 703 IST1 isolates belonging to a recently expanded clone provided a panel of excellent samples for evaluating variation in the 26 VNTR loci over a period of 3.25 years. The MLVA data indicated that variation had occurred on 10 loci. SS3 and SS6 were hypervariable loci,

with variation rates of 75.82% and 35.99%, respectively (Table 2). SS11and SS9 were moderately variable loci, with variation rates of 7.97%, and 6.12%, respectively. The rest of the 22 loci were of low variability with variation rates of less than 0.85%. Regression analysis indicated that the variation rate was best fit by a second-degree polynomial relationship with the number of repeat units ($r^2 = 0.869$). This result suggested that loci having more repeat units would be more variable. Variation rate was also strongly correlated to the total number of alleles ($r^2 = 0.830$) and to the allele diversity ($r^2 = 0.401$) observed in a collection of 203 diverse non-IST1 isolates. The results indicated that the number of alleles and level of allele diversity for a given locus were predictors for locus variability. SS1 had a high number of alleles and a high level of allele diversity observed in the non-IST1 isolates, but it had a low variation rate observed in the 703 IST1 isolates. Since SS1 in SS26.66 contained only 2 repeat units, the low variation rate in the IST1 isolates fitted the fact that the variation rate was strongly correlated to the repeat number within the locus.

The ancestral genotype, SS26.66, contained 20 repeat units in each of the SS3 and SS6 loci. Of the 703 isolates belonging to the SS26.66 clone, 533 had variation on the SS3 locus and 253 on the SS6 locus. Of the SS3 variants, 76% were the result of insertion; in contrast, 68% of SS6 variants were the result of deletion (Fig. 4). The changes ranged from 1 to 11 repeat units. The variations were largely attributed to insertion or deletion of a small number of repeat units. Single-repeat insertion and deletion contributed 69% and 64% of the variation to the SS3 and SS6 variants, respectively. Changes in 3 or fewer repeats contributed to 92% and 94% of the variations to the SS3 and SS6 variants, respectively.

Phylogenetic relationship among the 710 IST1 isolates constructed with MLVA-24 profiles. When the two hypervariable loci (SS3 and SS6) were excluded, a total of 23 MLVA-24 genotypes, defined by the 24-locus profiles, were identified for the 710 isolates. MST clustering analysis generated a simple phylogenetic tree for the 23 MLVA-24 genotypes (Fig. 3B). SS24.1 was the predominant genotype; it represented 84% of the isolates. All the isolates collected in 2000 belonged to this genotype. SS24.1 had 12 SLVs, among which SS24.2 and SS24.5 were significant. SS24.2 comprised 7% of the isolates and had 1 SLV; isolates of this genotype and its SLV were collected in 2001 and 2002. SS24.5 made up 4% of the isolates and had 3 SLVs; isolates of this genotype and its SLVs were collected in 2002 and 2003. The imported strains had a distance of at least 3 loci to the founder genotype, SS24.1.

Phylogenetic relationship among the 710 IST1 isolates constructed with MLVA-22 profiles. When the 4 highly variable loci (SS3, SS6, SS9 and SS11) were excluded, only 13 MLVA-22 genotypes, defined by the 22-locus profiles, were identified for the 710 isolates. MST clustering analysis generated a quite simple phylogenetic tree for the 13 MLVA-22 genotypes (Fig. 3C). SS22.1 was the predominant genotype; it accounted for 97% of the isolates. All the isolates collected in 2000 belong to the SS22.1 genotype. SS22.1 had 8 SLVs; all of the SLVs emerged in the second year and thereafter. The Indian strain was a double locus variant of SS22.1; other imported strains had a distance of 3 or more loci to SS22.1.

Evolutionary speed of *S. sonnei* clones. The 703 IST1 isolates of the SS26.66 clone were a panel of good samples for evaluating the evolutionary speed of *S. sonnei* clones. For the genotypes defined by the 26-locus profiles, the maximal variation distance from SS26.66 was 4 loci (Table 3). The average evolutionary distance to the ancestral genotype, SS26.66, was 0.23 loci for the isolates collected in 2000. The average distance increased over time from 1.23 loci, then 1.72 loci, to 2.2 loci for those collected in 2001, 2002 and 2003, respectively. Calculated from the total isolates collected in a period of 3.25 years (from October 2000 to 2003), the average evolutionary speed for the SS26.66 clone was 0.39 loci per year. For the genotypes defined by 24-locus profiles by excluding the hypervariable loci, SS3 and SS6 from the panel of 26 loci, the maximal variation distance from SS24.1 was 2 loci. The average evolutionary distance of MLVA-24 genotypes to SS24.1 increased from zero locus for the isolates collected in 2000 to 0.70 loci for those collected in 2003. The average evolutionary speed for the SS24.1 clone was 0.049 loci per year. For the genotypes defined by the 22-locus profiles minus the 4 highly variable loci, SS3, SS6, SS9 and SS11, the average evolutionary distance of MLVA-22 genotypes to SS22.1 also increased over time, from zero locus for the isolates collected in 2001 to 0.1 for those collected in 2003. The average evolutionary speed for the SS22.1 clone was 0.0055 loci per year.

Phylogenetic relationship among 916 isolates. A panel of 916 isolates, including the 710 IST1 and 206 non-IST1 isolates, were used to evaluate the phylogeny of the *S. sonnei* isolates using the MLVA data. A total of 151 MLVA-26, 64 MLVA-24 and 43 MLVA-22 genotypes were identified for the 916 isolates, collected in Taiwan in the period 1996-2004. MST clustering analysis of the 151 MLVA-26 genotypes generated a complex phylogenetic tree (Fig. 5A). The MST tree revealed

that the genetic relatedness among the isolates defined by the IST and PFGE data were quite concordant with that defined by the MLVA data. Isolates within an IST or a PFGE clonal complex shared a greater number of identical alleles. In parallel with the clusters defined by IST and PFGE data, the 916 isolates were designated into 3 (SS26.66, SS26.3 and SS26.6) clonal complexes, with a maximal distance of 7 loci between genotypes within a clonal complex. All of the 24 imported isolates belonged to the SS26.66 clonal complex.

The phylogenetic tree constructed with the 64 MLVA-24 genotypes was simper (Fig. 5B). In parallel with the clusters defined by the IST and PFGE data, the isolates were designated into 3 (SS24.1, SS24.3 and SS24.6) clonal complexes, with a maximal distance of 5 loci between the genotypes within a clonal complex. Similarly, the 43 MLVA-22 genotypes were assigned into 3 (SS22.1, SS22.3 and SS22.6) clonal complexes, with a maximal distance of 3 loci between genotypes within a clonal complex (Fig. 5C). The results suggested a stringency of group definition such that genotypes sharing 19 or more identical loci could be included into a clonal group or clonal complex as genotypes defined with 26-, 24- and 22-locus profiles.

DISCUSSION

MLVA has been demonstrated to be a powerful tool for fine typing of *S. sonnei* for both disease outbreak investigation and disease surveillance (8); however, the use of MLVA as a tool for phylogenetic analysis has been arguing. VNTR locus is considered too variable to serve as a molecular marker for study of the evolution of a bacterial species. In addition, in most cases, only a limited number of VNTRs are currently identified for a bacterial species or a serotype (9-12), which also constrains the value of MLVA as an investigative tool. The criterion for defining a clonal group or a clonal complex is another concern, if MLVA data are used for phylogenetic analysis. In this study, we evaluated the variation rate of 26 VNTR loci with data from a large panel of isolates descending very recently from a common ancestor, estimated the evolutionary speed of an *S. sonnei* clone, and suggested the requisite stringency for defining an *S. sonnei* clonal group or clonal complex.

From the results of MST clustering analysis of the 26-locus profiles (MLVA-26) for the 710 IST1 isolates, we considered 703 of the isolates to have descended very recently from a common MLVA strain (SS26.66) which emerged in 2000 (Fig. 3A). The SS26.66 clone first caused a shigellosis outbreak at a high school in eastern Taiwan in October 2000, after which the strain spread to other areas of the country and evolved into a large number of PFGE genotypes through sustained transmission among hosts over 3.25 years (17). Since these isolates belonged to a line of common evolutionary descent, they were excellent samples to be used to evaluate the variation rate of the 26 VNTR loci and to estimate the evolutionary speed of the *S. sonnei* clone. The MLVA data of the 703 isolates revealed that 10 of the 26 VNTR had varied during a period of 3.25 years and the loci displayed different degrees of variation. The results revealed 2 hypervariable, 2 moderately variable and 22 low variable loci. By excluding the highly variable loci from the panel of 26 VNTR loci, MLVA is applicable to investigating the genetic relatedness among strains over various evolutionary timescales.

The genotypes for the 703 IST1 isolates defined by the 26-locus profiles were quite divergent (Fig. 3A). The high level of diversification among the isolates was largely attributed to the two hypervariable loci, SS3 and SS6. As the calculated evolutionary speed was 0.39 loci per year for the clone, this measurement by analyzing 26 loci could only be used to study the genetic relatedness among strains over a very short timescale. Therefore, MLVA with highly variable loci is not an

appropriate tool for evolutionary study. When SS3 and SS6 were excluded, the genotypes for the 703 IST1 isolates defined by the 24-locus profiles were much less diversified (Fig. 3B). As the calculated evolutionary speed was 0.049 loci per year, the measurement by analyzing the 24 loci was useful for exploring genetic relatedness among strains over an evolutionary timescale of decades. By characterization of the 24 loci, the MLVA could be used for tracking transmission of *S. sonnei* clones between countries and geographical regions. Likewise, by excluding the hypervariable loci (SS3 and SS6) and the moderately variable loci (SS9 and SS11) from the panel of 26 loci, the measurement with the 22 loci detected only a few SLVs among the 703 isolates (Fig. 3C). The calculated evolutionary speed for the clone was 0.0055 loci per year, suggesting measurement with the 22 loci useful for exploring genetic relatedness among strains over an evolutionary timescale of several hundred years. Therefore, by excluding the rapidly variable loci from the panel of 26 loci, MLVA is applicable to the evolutionary study of clones with a longer-term evolutionary timescale.

Regression analysis indicates variation rate is strongly correlated with repeat number in locus, suggesting that locus variation rate has a strain or clone effect. SS1 had a low observed variation rate in the SS26.66 clone, which fit with the fact that SS1 harbored a low number of repeats in the SS26.66 strain. The SS1 variation rate would be expected to differ greatly when a strain harbors a high number of repeats in SS1. Therefore, the value of the variation rate observed in a collection of clone isolates may not be representative of the degree of locus variability in the *S. sonnei* population. Since the variation rate observed in the 703 IST1 isolates is correlated to the number of alleles and the allele diversity calculated from 203 diverse non-IST1 isolates, the latter two variables are predictors of locus variability. Therefore, the number of alleles and/or allele diversity obtained from a large panel of diverse isolates can be used to assess the degree of locus variability. The data from the SS26.66 clone indicated 4 loci (SS3, SS6, SS9 and SS11) are highly variable, which is consistent with the high number of alleles and the high level of allele diversity observed in the non-IST1 isolates. The data also suggested that SS1, SS10 and SS23 were also highly variable (Table 2).

The MLVA data obtained from the 703 IST isolates suggest that SS3, SS6, SS9 and SS11 are highly variable. By excluding the highly variable loci from the panel of 26 loci, MLVA data can be used for the phylogenetic study of *S. sonnei* strains over timescales of a few years, decades or hundreds of years. However, some of the 22 loci

may be still too variable to be used for the study of longer evolutionary timescales. SS1, SS10 and SS23 are three other highly variable loci since they were found in a high number of alleles and with a high level of allele diversity in the 206 non-IST isolates in this study, and in 536 isolates collected in central Taiwan in a previous study (8). Excluding the 7 highly variable loci from the panel of 26 loci, MLVA with the 19 low variable loci would be an appropriate molecular tool for evolutionary study of *S. sonnei* strains over longer-term timescales.

This study showed that MLVA can serve as a valuable tool for the phylogenetic analysis of *S. sonnei* strains. One of the concerns about the use of MLVA method for evolutionary studies is the required stringency of the group (or clone) definition. In this study, the panel of 916 isolates characterized by IST and PFGE were grouped into 3 distinct clusters. The clustering results with PFGE and IST data helped to define an MLVA group. Isolates within each of the 3 IST clusters or the 3 PFGE clusters shared a greater number of identical VNTR loci. In concordance with the IST and PFGE results, we considered a panel of isolates within an IST or a PFGE clonal complex as belonging to an MLVA clonal group (Fig. 5). Therefore, strains within a MLVA clonal complex designated by 26, 24 and 22 loci had a maximal distance of 7, 5 and 3 loci, respectively. As a result, we suggest that strains within a group or clonal complex defined by the 26, 24 or 22 loci shall share 19 or more identical loci. When SS1 and SS10 were also excluded from the panel of loci, genotypes within the MLVA clonal complex defined with the 20-locus profiles has a maximal distance of 2 loci (data not shown), so a clonal group definition for the measurement is a sharing of 18 or more identical loci.

The 916 *S. sonnei* isolates collected in Taiwan over 9 years (from 1996 to 2004) fell into 3 clonal complexes only, suggesting that the *S. sonnei* strains were fairly restrictedly clonal. *S. sonnei* is one of the primary causal agents of travel-associated diarrhea (2), and transmission of clones can quite frequently take place between countries and geographical regions between where people regularly travel. As shown in Fig. 5, all of the imported isolates belonged to a common clonal complex. The imported isolates were derived from China, India and several Southern Asian countries. The data indicated that strains or clones within this clonal complex have been circulating within Asian countries. It seems likely that clones of the clonal complex are also at present circulating in other continents as a result of frequent international traveling. Further studies on the strains collected in various geographical

regions should be conducted to investigate this hypothesis.

From 1996 to 1999, *S. sonnei* infections were reported infrequently in Taiwan but in 2000-2002, *S. sonnei* caused an epidemic in the central and eastern part of the country. The IST1 clone was the major contributor to the epidemic. Since the first emergent IST1 strain to be identified came from India and shared an indistinguishable PFGE pattern with the majority of isolates of the IST1 clone, we concluded in our previously report that the Indian strain was the original source of the epidemic (17). However, the MLVA data show that the Indian strain differed at 4 loci with the epidemic (SS26.66) clone. Since a strain is very unlikely to have varied at 4 loci over such short period of time, the Indian strain can be ruled out as the source of the epidemic. This observation also indicated certain PFGE patterns to be quite stable. In a previous study, we reported a PFGE strain (J16N09.0015) that was identified in isolates collected from 9 epidemiological events in 5 Asian countries over a period of 7 years (8). In the present study, we also have reported two prevalent PFGE types (J16N09.0019 and J16N09.0023) that remained stable in 41% (293) of 710 IST1 isolates. Obviously, it is problematic to use PFGE data alone for tracking the source of an outbreak or epidemic without sufficient epidemiological evidence.

PFGE is a highly discriminatory tool for the subtyping of *S. sonnei* strains. Variation in the PFGE pattern accumulates rapidly (17), indicating that PFGE is not an appropriate tool for the phylogentic analysis of *S. sonnei* strains over a long evolutionary timescale. IST is less discriminatory than PFGE; IST data are more suitable for designating an *S. sonnei* clone or clonal complex (1). PFGE and IST data are typically represented by a dendrogram that provides a convenient way to illustrate the genetic relatedness among isolates. However, the topology of the dendrogram can be arbitrary, and it provides no objective cutoff value of the pattern similarity for assigning a clone or clonal complex. In contrast, MLVA data can be presented by a MST tree that provides information which is essential for discerning the patterns of evolutionary descent of isolates within a clonal complex (3). Nevertheless, the IST and PFGE data did help in assessing the necessary stringency for defining a MLVA clonal group.

The mutation rate and the factors that affect the mutation rate of the VNTR loci of *Escherichia coli* O157:H7 has been investigated. In the study, Vogler et al (16) indicate mutation rate is strongly correlated with the copy number of repeat units within a given locus and with the number of alleles and allele diversity observed in a

large collection of diverse *E. coli* O157/O55 isolates. Furthermore, mutation of the locus is largely attributed to single repeat insertion and deletion events. The variation rate of VNTR loci observed in a large number of isolates of a clone is in fact not equivalent with but rather proportional to the mutation rate. In our study, we found that the variation rate observed in isolates of an IST1 clone was strongly correlated to the copy number of repeat units in the founder strain and to the number of alleles and the allele diversity calculated from a panel of diverse non-IST1 isolates. Our data also show that variation in the two hypervariable loci, SS3 and SS6, was largely attributed to single repeat insertion and deletion events (Fig. 4).

In conclusion, MLVA is a useful tool for the phylogenetic analysis of *S. sonnei*. By excluding certain highly variable loci from the panel of 26 loci, MLVA data can be applied to infer levels of genetic relatedness between strains over various evolutionary timescales, from a period of years to a few hundred years. Variation rate is correlated with copy number of repeat units and also with number of alleles and allele diversity. Therefore, the rapidly variable loci have a higher number of alleles and allele diversity in the *S. sonnei* population. Strains sharing 19 or more identical loci can be grouped into a clonal group. The MLVA-24 and MLVA-22 data are appropriate for constructing patterns of evolutionary descent for strains over decades to hundreds of years; therefore, by excluding the highly variable loci from the panel of 26 loci, MLVA can serve as a useful tool to investigate the global transmission of *S. sonnei* clones.

ACKNOWLEDGMENTS

This work was supported by a grant (DOH 97-DC-) of the Centers for Disease Control, Department of Health, Taiwan, and by a grant (H17-Shinkou-Ippan-019) of Ministry of Health, Labor and Welfare, Japan.

REFERENCES

1. **Chiou, C. S., H. L. Wei, Y. W. Wang, J. C. Liao, and C. C. Li.** 2006. Usefulness of Inter-IS1 Spacer Polymorphisms for Subtyping of *Shigella sonnei* Isolates. J Clin Microbiol **44:**3928-33.
2. **Ekdahl, K., and Y. Andersson.** 2005. The epidemiology of travel-associated shigellosis--regional risks, seasonality and serogroups. J Infect **51:**222-9.
3. **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:**1518-30.
4. **Grundmann, H., S. Hori, and G. Tanner.** 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol **39:**4190-2.
5. **Hunter, P. R.** 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol **28:**1903-5.
6. **Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine.** 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ 77:651-66.
7. **Lee, H. C., K. L. Chen, C. L. Tsai, C. H. Chen, T. N. Yeh, C. R. Yang, Y. L. Wang, H. Y. Chiu, C. L. Lee, H. P. Su, and T. H. Lin.** 2004. Imported infection of *Shigella sonnei* molecular epidemiological investigation of cases of the Bali tours. Epidemiology Bulletin **20:**23-42.
8. **Liang, S. Y., H. Watanabe, J. Terajima, C. C. Li, J. C. Liao, S. K. Tung, and C. S. Chiou.** 2007. Multilocus Variable-Number Tandem Repeat Analysis for Molecular Typing of *Shigella sonnei*. J Clin Microbiol.
9. **Liao, J. C., C. C. Li, and C. S. Chiou.** 2006. Use of a multilocus variable-number tandem repeat analysis method for molecular subtyping and phylogenetic analysis of *Neisseria meningitidis* isolates. BMC Microbiol **6:**44.
10. **Marsh, J. W., M. M. O'Leary, K. A. Shutt, A. W. Pasculle, S. Johnson, D. N. Gerding, C. A. Muto, and L. H. Harrison.** 2006. Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in Hospitals. J Clin Microbiol **44:**2558-66.
11. **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:**5389-97.
12. **Overduin, P., L. Schouls, P. Roholl, A. van der Zanden, N. Mahmmod, A.**

Herrewegh, and D. van Soolingen. 2004. Use of multilocus variable-number tandem-repeat analysis for typing *Mycobacterium avium* subsp. *paratuberculosis*. J Clin Microbiol **42:**5022-8.

13. **Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J. Barrett.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog Dis **3:**59-67.
14. **Swaminathan, B., P. Gerner-Smidt, L. K. Ng, S. Lukinmaa, K. M. Kam, S. Rolando, E. P. Gutierrez, and N. Binsztein.** 2006. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis **3:**36-50.
15. **Terajima, J., N. Tosaka, K. Ueno, K. Nakashima, P. Kitsutani, M. K. Gaynor, S. Y. Park, and H. Watanabe.** 2006. *Shigella sonnei* outbreak among Japanese travelers returning from Hawaii. Jpn J Infect Dis **59:**282-3.
16. **Vogler, A. J., C. Keys, Y. Nemoto, R. E. Colman, Z. Jay, and P. Keim.** 2006. Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. J Bacteriol **188:**4253-63.
17. **Wei, H. L., Y. W. Wang, C. C. Li, S. K. Tung, and C. S. Chiou.** 2007. Epidemiology and evolution of genotype and antimicrobial resistance of an imported *Shigella sonnei* clone circulating in central Taiwan. Diagn Microbiol Infect Dis **58:**469-75.