



Published in final edited form as:

Arch Oral Biol. 2015 December ; 60(12): 1769–1776. doi:10.1016/j.archoralbio.2015.09.012.

Assessment of two multilocus sequence typing (MLST) schemes available for *Streptococcus mutans*

Stephanie S. Momeni^a, Jennifer Whiddon^b, Kyounga Cheon^a, Stephen A. Moser^b, and Noel K. Childers^{a,*}

^aDepartment of Pediatric Dentistry, School of Dentistry, University of Alabama at Birmingham, SOD 304, 1720 2nd Avenue South, Birmingham, AL 35294-0007, USA

^bDepartment of Pathology, School of Medicine & Dentistry, University of Alabama at Birmingham, WP 230, 619 19th Street South, Birmingham, AL 35249-7331, USA

Abstract

Objective—Two multilocus sequencing typing (MLST) schemes are currently available for *Streptococcus mutans*. The first, introduced by Nakano et al. in 2007, consists of 8 conserved housekeeping genes. The second, introduced in 2010 by Do et al., includes 6 housekeeping genes and 2 putative virulence genes. The purpose of the current study was to compare the two MLST schemes for use in validating repetitive extragenic palindromic polymerase chain reaction (rep-PCR) genotypes.

Design—Thirty-three *S. mutans* isolates, representing the 11 most commonly occurring rep-PCR genotype groups, were selected for MLST. MLST was performed with SYBR Green™ PCR with published primers for both MLST schemes. Amplicons were purified, sequenced, and data checked against the www.PubMLST.org database for allelic and sequence type (ST) assignment. Discriminatory power, congruence, and convenience criteria were evaluated. Concatenated sequences for each scheme were analyzed using MEGA to generate phylogenetic trees using minimum evolution with bootstrap.

Results—No significant difference in discriminatory power was observed between the two MLST schemes for *S. mutans*. Clonal clusters were consistent for both schemes. Overall, MLST demonstrated marginally greater discriminatory power than rep-PCR; however all methods were found to be congruent. New alleles and ST are reported for each scheme and added to the PubMLST database.

Conclusions—Clonality, supported by both methods and rep-PCR, indicates *S. mutans* genotypes are shared between unrelated subjects. Both Nakano and Do schemes demonstrates similar genotype discrimination for *S. mutans* isolates suggesting each are well designed and may be used to verify rep-PCR genotypes.

*Corresponding author at: Department of Pediatric Dentistry, University of Alabama at Birmingham, School of Dentistry Room 304, 1720 2nd Avenue South Birmingham, AL 35294-0007, USA. Fax: +1 205 934 7013. jennndoc@uab.edu (S.S. Momeni), kcheon@uab.edu (K. Cheon), moser@uab.edu (S.A. Moser), nkc@uab.edu (N.K. Childers).

Keywords

Dental caries; MLST; rep-PCR; Genotyping; *Streptococcus mutans*; Genetic diversity

1. Introduction

Dental caries is a prevalent global infectious disease. *Streptococcus mutans* is suggested as being highly associated with dental caries in humans (Loesche, 1986). Bacterial genotyping of *S. mutans* for epidemiological surveillance and evolutionary studies is important to the understanding of the initiation and progression of dental caries. Repetitive extragenic palindromic polymerase chain reaction (rep-PCR) using the DiversiLab™ system is a standardized rapid and cost effective gel based typing method currently being used for a large-scale epidemiological study of *S. mutans* (Cheon et al., 2013; Healy et al., 2005; Momeni et al., 2013; Moser et al., 2010). However, it is recommended that gel-based typing methods like rep-PCR be verified by alternate methods, preferably a molecular typing method such as multilocus sequence typing (MLST) or next generation sequencing (NGS) (Foley et al., 2006 van Belkum et al., 2007).

Currently, there are two published *S. mutans* MLST typing schemes available through the www.pubMLST.org database. The first scheme was introduced in 2007 by Nakano and consisted of 8 highly conserved housekeeping genes (Nakano et al., 2007). A second scheme was introduced in 2010 by Do using 6 constitutive housekeeping genes and 2 putative virulence-associated genes (Do et al., 2010). Maiden (2006) suggests that multiple MLST schemes for the same organism will be comparable if the schemes are well designed. However, when multiple schemes are available, it is important to compare schemes based on discriminative and phylogenetic resolutions since these may vary between schemes depending on the diversity of the genes selected (Ahmed et al., 2011; Debourgogne et al., 2012; Kilian, Scholz, & Lomholt, 2012; Maiden, 2006). The purpose of this study was to compare the two MLST schemes (Nakano and Do schemes) available for *S. mutans* for use in the verification of rep-PCR genotypes.

2. Materials and methods

2.1. Sample selection and processing

S. mutans isolates for this study were obtained from oral samples collected in Uniontown, Alabama, USA, a low socioeconomic minority community considered a high caries risk population. Informed consent was obtained in accordance with the regulations established by the University of Alabama at Birmingham (UAB) Institutional Review Board. The original samples were collected at an elementary school or a local community health center. Plaque, tongue, and saliva sample collections and processing were performed as previously described (Cheon et al., 2011; Childers et al., 2011). Samples were selected from a bank of more than 30000 *S. mutans* isolates obtained from children and their household family members over an 8-year period. For comparison of the two MLST schemes available for *S. mutans*, 33 isolates were selected from 32 individuals. These isolates represent the 11 most common rep-PCR genotype groups (GG) found in children. Each rep-PCR GG consisted of

the representative Library isolate and two other randomly selected isolates with the same rep-PCR genotype. Nineteen samples were collected from children (age 5–8) and 14 from household family members. In one case, 2 isolates were selected from the same subject, but these isolates had different rep-PCR genotypes (G1 and G6). Six samples were from three related individual pairs (a child and a household member); however, isolates within pairs had different rep-PCR genotypes. Sample selection was not limited to one sample type since the Library strain was a tongue or saliva sample in some cases. This resulted in 28 isolates collected from dental plaque, 4 from tongue scrapings, and 1 from saliva. Overall, these 11 GG represent 10310 isolates from 549 individuals that have been analyzed to date by rep-PCR genotyping.

2.2. PCR for MLST and sequence analysis

The Nakano scheme is based on partial fragments from the following genes: *murI*, *tkt*, *glnA*, *gyrA*, *aroE*, *gltA*, *glk*, and *lepC* (Nakano et al., 2007). The Do scheme is based on *accC*, *gki(glk)*, *lepA*, *recP(tkt)*, *sodA*, *tyrS*, *gtfB*, and *spaP* (Do et al., 2010). For simplicity, the terms NS (Nakano Scheme) and DS (Do Scheme) will be used to represent the two schemes, respectively.

MLST analysis with the NS was performed as previously reported (Momeni et al., 2013). Briefly, purified DNA was extracted from pure cultures, confirmed as *S. mutans* using SYBR Green™ PCR with *gtfB* sequence specific primers, and real-time PCR was performed using the 8 housekeeping gene primer sets originally reported by Nakano et al. (2007). MLST analysis for the DS required a modified amplification protocol from the NS, using a lower annealing temperature of 49 °C (NS annealing temp was 55 °C). The housekeeping and putative virulence gene primers were used as published for the DS (Do et al., 2010) except for the *accC* gene which required a different forward primer (5'-ATTGCCAATCGTGGTG-3') from the originally published primer in order to obtain the fragment needed for submission to the PubMLST database (Do, personal communication).

Primers used for PCR were also used for sequencing for both schemes. Sequence data analysis was performed as previously described (Momeni et al., 2013). New allele and sequence types (ST) were assigned after comparing consensus sequence data with the PubMLST (www.pubMLST.org) database and confirmed by database curators. All new alleles were confirmed by repeating PCR and sequencing steps.

Alignments were created using both allelic variation and concatenated sequences phylogenetic analysis (3366 bp Nakano, 3774 bp Do). Sequence Type Analysis and Recombinational Tests Version 2 (START2; PubMLST.org/University of Oxford, Oxford, England) with Un-weighted Pair Group Method using Arithmetic averages (UPGMA) was used for allelic variation alignment. Up to 2 allelic differences were considered to be a clonal complex. The index of association (I_A) was calculated from both the complete data set (Both schemes, $n = 33$ ST) and a single representative of each ST (NS, $n = 27$ ST; DS, $n = 26$ ST). Calculations were performed by START2 with the Maynard-Smith approach using 1000 trials to evaluate recombination (Jolley, Feil, Chan, & Maiden, 2001; Smith, Smith, O'Rourke, & Spratt, 1993).

Phylogenetic analysis was performed by the Minimum Evolution bootstrap method (1000 replicates) using Molecular Evolutionary Genetics Analysis (MEGA; www.megasoftware.net/The Biodesign Institute, Tempe, AZ, USA) software version 5.2.2. Variable sites data was obtained using DIVEIN website (last accessed 6/13/14) (Deng et al., 2010). Genetic linkages and clonal complexes were investigated using Global Optimal eBURST version 1.2.1 (goeBURST; www.goeburst.phylovis.net) (Francisco, Bugalho, Ramirez, & Carrico, 2009).

2.3. Comparison criteria

The two MLST typing schemes were compared using typing system concordance in addition to comparison with rep-PCR genotypes. Coding sections of allele fragments were determined using BioEdit software v.7.2.5 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>, Ibis Biosciences, Carlsbad, CA). Information for each locus including G + C content, polymorphic sites, average non-synonymous/synonymous ratio (dN/dS) and Tajima's D were calculated using DnaSP software v5.10.1 (DnaSP; www.ub.edu/dnasp/Universitat de Barcelona, Barcelona, Spain) (Librado & Rozas, 2009). Statistical significance of Tajima's D in DnaSP is the result of a two-tailed test assuming beta distribution with significance defined as $p < 0.05$. Discriminatory power and concordance were evaluated by Simpson's index of diversity (SID), Adjusted Rand, and Adjusted Wallace coefficients calculated using Comparing Partitions (<http://darwin.phylovis.net/ComparingPartitions/index.php?link=Home>) (Carrico et al., 2006; Hunter & Gaston, 1988). In addition to statistical methods, convenience criteria were also evaluated.

2.4. Nucleotide sequence accession numbers

Sequences for new allele types generated by MLST were submitted to Genbank (<http://www.ncbi.nlm.nih.gov/genbank>) and assigned accession numbers KM889557–KM889561 (NS) and KM889562–KM889591 (DS).

3. Results

3.1. MLST analysis

In total, 33 *S. mutans* isolates from 11 rep-PCR GG were analyzed with MLST using both MLST schemes. New alleles and ST are reported for the NS (5 alleles, 9 ST) and DS (30 alleles, 22 ST). A total of 27 distinct ST were identified for the NS and 26 ST for the DS with goeBURST. Using single locus variant (SLV) setting, 24 clusters were found for both schemes while a double locus variant (DLV) setting resulted in 21 (NS) and 23 (DS) clusters. Isolate G7-1 was removed for the DS because the *gtfB* gene contained a deletion of 3 nucleotides that prevented alignment in MEGA and START2 resulting in only 32 isolates available for phylogenetic analysis.

Fig. 1 displays the percent similarity rep-PCR dendrogram with the assigned rep-PCR genotypes, for display of the resulting Nakano and Do MLST ST. MLST clones and clonal complexes were supported by both schemes. For instance, in GG-1a all three ST were identical for both schemes. For GG-12, all three ST were different in both schemes. For GG-18 and GG-11, the difference in ST was the result of a single base pair difference

resulting in a new ST so one scheme is clonal and the other is a clonal complex (up to two alleles different). Some GG were not grouped together on the dendrogram (e.g., GG-7 and GG-23). This is because the DiversiLab™ software that assesses the dendrograms is based on percent similarity and can be less accurate when working with greater number of isolates (e.g., GG-7 has 552 and GG-23 has 727 isolates). Further technical analysis (e.g., graphic overlays) showed these to be similar rep-PCR genotypes. For instance, GG-7 has an extra major band (following the first dark band) that differentiates it these isolates for GG-6; however, the percent similarity is mixed for GG-6 and GG-7. Another example is L-23, which the percent similarity groups separately from the other two GG-23 isolates but visually (confirmed by overlays) these isolates are the same. When randomly selected isolates from larger genotype groups are run in smaller reports like the one used in Fig. 1, genotypes do not always appear to group out. This is due to (1) how the DiversiLab algorithm calculates percent similarity, (2) how the rep-PCR genotypes are determined using the 1 major band, 3 minor band guide, and (3) genetic drift as more isolates are added to the larger pool from which these isolates were obtained (Moser et al., 2010).

When allelic data was compared with the PubMLST database 30 new alleles were observed for the DS and 5 new alleles for the NS. New ST are reported for both NS (9 new) and DS (22 new). Tables 1 and 2 list the new ST and alleles for the NS and DS, respectively. New ST resulted from addition of new alleles (NS 5, DS 14 new alleles) or new combinations of previously published alleles (NS 4, DS 8 new combinations). Fewer new alleles and ST are reported for the NS since some were previously reported (Momeni et al., 2013).

The general characteristics for all loci are listed in Table 3. The number of polymorphic sites ranged from 1.38% (*gyrA*) to 3.08% (*gltA*) for NS and 1.52% (*accC*) to 3.12% (*tyrS*) for the DS. Altered amino acid sequences (dN) and silent changes (dS) were determined and the ratio dN/dS for each allele was calculated to be substantially <1. Results for the Tajima's D test were negative in most cases and none of the values were found to be significant. Analysis in DIVEIN resulted in a comparable number, 51 (NS) and 56 (DS), of informative sites (genetic variations occurring in more than one isolate), which supports the topology of phylogenetic trees generated. The number of private sites (or singletons, variations occurring in only one isolate) varied between schemes: 22 (NS) and 32 (DS) and suggest that individual unique mutations are more frequent in the DS.

3.2. Phylogenetic analysis

Fig. 2 shows the phylogenetic analysis based on concatenated sequences evaluated in MEGA resulting in 6 clades for the NS and this clonal structure was maintained in 5 clades for the DS. Clades were either clonal isolates or clonal complexes (up to two alleles different). Clades were supported by both schemes except Clade 5 (L1 and L13), which clustered together in the NS due to a single locus variant but were separated in the DS due to a three locus variants. Phylogenetic analysis using allelic profiles using START2 were similar (data not shown). The I_A for entire data sets were 1.091 (NS, $n = 33$) and 0.8878 (DS, $n = 33$). When adjusted for a single representative of each ST I_A values were 0.5762 (NS, $n = 27$) and 0.4206 (DS, $n = 26$).

3.3. Discriminatory power and congruence

Both schemes demonstrated similar discriminatory power based on overlapping confidence intervals for Simpson's index of diversity (Table 4). When compared with rep-PCR, both MLST schemes were found to have marginally more discriminatory power. However, using phylogenetic analysis both NS and DS were able to discriminate further 52% (17/33) and 56% (18/32) of rep-PCR isolates respectively. Overall concordance, as calculated by Adjusted Rand, found both MLST schemes to be congruent (0.694). The Adjusted Rand values were lower for rep-PCR than both the MLST methods; however, these method may still be congruence since the 95% CI overlap. Both MLST typing schemes (NS 68%, DS 78%) are likely to predict genotypes similarly according to the Adjusted Wallace coefficient since the 95% CI overlap. However, rep-PCR is less likely to predict MLST ST for either scheme (NS 18%, DS 23%).

3.4. Convenience criteria

The annealing temperatures required for PCR amplification were 55 °C for the NS and 49 °C for the DS. Amplicon sequence lengths ranged from 387 to 462 bp (Nakano) and 526 to 970 bp (Do). Two genes-transketolase (*tkt*-NS, *recP*-DS) and glucose kinase (*glk*-NS, *gki*-DS) were shared between the two schemes, with some overlapping of the fragments. The *spaP* gene (surface protein antigen I/II) reverse primer has two binding location at bp 610–627 and 856–873, which can results in double peaks using the SYBR green real-time PCR approach.

Gene fragments generated in this study using a primer-to-primer fragment were found to be slightly different than previously reported by Nakano (Table 3) resulting in a longer concatenated sequence (3366 bp vs. 3351 bp reported by Nakano). The current study used full primer-to-primer fragments while fragments in the PubMLST database have some trimming or extension of the primer regions to create fragments to within the reading frame for translation to amino acids.

4. Discussion

The objective of the current study was to compare two available MLST typing schemes for *S. mutans* and to discuss the application of each for use with rep-PCR to validate emerging rep-PCR genotypes of *S. mutans* in an ongoing, longitudinal epidemiological study. This study provides an independent verification of both the NS and DS designs and their practical use for large-scale study of *S. mutans*.

The NS and DS were found to be comparable indicating the schemes have similar resolutions which is consistent with others observations of multiple schemes for single bacteria (Ahmed et al., 2011; Debourgogne et al., 2012; Maiden, 2006). The two schemes have different gene compositions. The NS contains 8 housekeeping genes of which 5 (*tkt*, *glnA*, *gyrA*, *murI*, *lepC*) are conserved across *Streptococcus* and 3 genes are specific for *S. mutans* (Nakano et al., 2007). In contrast, the composition of the DS utilizes 6 housekeeping genes and 2 extracellular virulence-associated genes (*gtfB* and *spaP*) (Do et al., 2010). The *accC*, *gki*, *gtfB*, and *spaP* gene fragments are specific to *S. mutans*. The genes for

transketolase and glucose kinase are used in both schemes with some sequencing overlap that introduces some redundancy when applying both schemes. Both schemes are publicly accessible through www.pubMLST.org to allow investigators to compare data.

4.1. Genotype groups

The number of polymorphic sites observed is notably lower than those previously reported (Table 3). The maximums reported here (3.08% NS and 3.12% DS) were comparable to the minimums reported in the original studies (3.24 NS and 3.40 DS) (Do et al., 2010; Nakano et al., 2007). Furthermore, the increase of nucleotide changes reported for the DS for the putative virulence genes *gtfB* and *spaP* was not observed in this study (Do et al., 2010). This may be due to the sample size or possible bias of selecting isolates based on rep-PCR genotype groups that may have limited variability. The dN/dS ratios were considerably less than 1 for all loci, and were comparable to those previously reported, supporting the finding that these loci are not under positive selection. Most of the alleles had a negative result for the Tajima's *D* neutrality test indicating a low frequency of polymorphisms.

Phylogenetic trees generated from concatenated sequences indicate that isolates differentiated in the NS are also differentiated in the DS (Fig. 2). Similarly, clonal groups of isolates were supported by both schemes with one exception, Clade 5, which varied by 2 alleles in the NS and 3 alleles in the DS. Trees generated with allele profiles further supported this phylogenetic agreement. Additionally, clonal isolates (Clades 2, 4, and 6) supported by both MLST schemes and rep-PCR indicate that these isolates, from different individuals, are identical and may provide some evidence of transmission.

The SID indicated that rep-PCR and both MLST typing schemes were robust, having SID values between 0.938 and 0.981 (Table 4). Although rep-PCR had the lower discriminatory value, overlapping confidence interval means these methods all have similar discriminatory powers. Concordance as estimated by the Adjusted Wallace coefficient indicates that rep-PCR is a poor predictor of MLST ST (NS = 0.197, DS = 0.228). However, it is important to understand that this calculation uses only the genotype assignment and does not take into account the percent similarity and technical adjustments (1 major band, 3 minor band rules) that can improve the ability of rep-PCR data to predict which genotype can be further distinguished by MLST (Momeni et al., 2013; Moser et al., 2010; Tenover et al., 1995). The DS demonstrated a higher probability of predicting both NS ST (0.694) and rep-PCR genotypes (0.787). Since the NS demonstrated a lower predicative power for rep-PCR (0.680) than the DS (0.787), this suggests that using the NS to validate rep-PCR genotypes will provide more information than using the DS as it is likely to result in more ST. This analysis of the discriminatory power and congruence of rep-PCR and both MLST typing schemes supports their combined use for large-scale epidemiological study.

The I_A values as calculated using START2 indicate significant linkage disequilibrium for both schemes. Values for I_A calculations using all ST as well as representative ST are reported since limiting to a single representative ST does not accurately reflect the clonal structure of this population since many ST were shared between different individuals. The I_A reported for the NS (1.091 all, 0.5762 single ST) was much higher than the I_A reported by Nakano (0.0931) (Nakano et al., 2007). The I_A reported for the DS (0.8878 all, 0.4206 single

ST), is comparable to the I_A reported by Do (0.4379) using only on a single representative of each ST that indicated a clonal population structure (Do et al., 2010). All I_A values in this study support that *S. mutans* is a clonal population. Differences observed in this study for the NS I_A may be due to the sample population being limited to single geographical/ethnic group or because rep-PCR was used to select isolates. The I_A may have reflected more recombination if samples from other populations were included for the NS. However, this data highlights the importance of focused populations studies to understand evolutionary changes and possible transmission on a scale that may be missed in studies where regional or global samples are used.

4.2. Convenience criteria

Overall, both NS and DS were similar in their practical aspects (i.e., ease of use). Both schemes consisted of 8 genes and therefore cost the same to perform. The primer design for each allele resulted in fragment lengths that were comparable lengths in the NS (389–460 bp) but variable for the DS (560–970 bp). Fragment lengths for the DS required trimming to PubMLST lengths listed in Table 3. For instance, the sequenced fragment for the *tyrS* allele produced a 970 bp fragment that required trimming to PubMLST length of 513 bp. This may be a concern since ideal sequencing lengths for some systems range between 500 and 700 bp. Larger fragment sizes are also a concern for using the SYBR Green PCR approach since this system is optimized for amplification of products <300 bp. The lower annealing temperature of 49 °C for the DS may also present an issue as this may allow for PCR artifacts. Although these issues were not observed to be a major concern in the present study, they are noted here for others planning to employ this approach.

It is important to note if a global comparison of *S. mutans* isolates is to be performed, that researchers planning to employ the NS scheme should align sequences with a representative download for each allele (possibly UA159) from the PubMLST database to trim sequences to the proper length since fragments trimmed to the primer regions will have slight variations with the sequence lengths in the database. This is because the gene fragments in the PubMLST database are trimmed to be within a reading frame. For instance, the NS for the *tkl* gene has a gene fragment size of 432-bp in the database, but trimming primer-to-primer produces a 435-bp fragment. As a result, the current study produced a 3366-bp concatenated sequence verses the 3351-bp sequence reported by Nakano. While the variation within gene fragments noted here did not impact the phylogenetic analysis of this study, these variations can be problematic when concatenated sequences are to be compared for global analysis, especially when translating to amino acids. Using primer to primer also requires corrections for submission of new allelic sequences for curation with the PubMLST database. Variations in fragment length can also affect the determination of codons for calculations of synonymous, non-synonymous, and dN/dS calculations. Using primer-to-primer fragments resulted in the following start codons for NS: 1 (*aroE*, *glnA*, *lepC*, *murI*), 2 (*gltA*, *gyrA*), and 3 (*tkl*, *glk*). Aligning fragments with a reference from PubMLST will adjust the start codon for all genes in the NS to 1. Primer design for the DS allows for a start codon of 1 for all genes.

4.3. Conclusions

The data presented in this study independently validates both MLST schemes currently available for analysis of *S. mutans*. Since both the NS and DS demonstrated comparable discriminatory power, congruence, and phylogenetic ability the use of either scheme for validation of rep-PCR genotypes is supported. The practical aspects of the two schemes varied slightly which researchers should consider before employing either scheme. This study contributes to the global epidemiological surveillance of *S. mutans* by adding new alleles and ST to the PubMLST database (NS-5 alleles, 9 ST and DS-30 alleles, 22 ST). The occurrence of 3 clonal groups (Clades 2, 4, 6) supported by rep-PCR and both MLST schemes provides some evidence that these isolates are shared between individuals and may provide some indication of transmission.

Acknowledgments

We especially appreciate all the clinical and laboratory participants of this study. Special gratitude to Dr. Jinthana Lapidattanakul and Dr. Thuy Do for directing the confirmation and registration of isolates with PubMLST; Dr. Mei Han of the UAB Heflin Genomic Core for sequencing data; and Dr. John Ruby for support.

Funding

This work was supported by Research Grant DE016684 (NC) from the National Institute of Dental and Craniofacial Research. SSM is a Dental Academic Research Training (DART) Pre-doctoral Fellow under NIDCR Institutional Grant#T-90 DE022736.

Abbreviations

MLST	multilocus sequence typing
rep-PCR	repetitive extragenic palindromic polymerase chain reaction
NS	Nakano MLST typing scheme
DS	Do MLST typing scheme
bp	base pair
ST	sequence type
GT	genotype
GG	genotype groups
I_A	index of association
MEGA	molecular evolutionary genetics analysis
START2	Sequence Type Analysis and Recombinational Tests Version 2
UPGMA	Un-weighted Pair Group Method using Arithmetic averages
SID	Simpson's index of diversity

References

Ahmed A, Thaipadungpanit J, Boonsilp S, Wuthiekanun V, Nalam K, Spratt BG, et al. Comparison of two multilocus sequence based genotyping schemes for *Leptospira* species. *PLoS Neglected*

- Tropical Diseases. 2011; 5(11):e1374. [Comparative Study Evaluation Studies Research Support, N. I. H., Extramural Research Support, Non-U. S. Gov't]. [PubMed: 22087342]
- Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, et al. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *Journal of Clinical Microbiology*. 2006; 44(7):2524–2532. [Research Support, Non-U.S. Gov't]. [PubMed: 16825375]
- Cheon K, Moser SA, Whiddon J, Osgood RC, Momeni S, Ruby JD, et al. Genetic diversity of plaque mutans streptococci with rep-PCR. *Journal of Dental Research*. 2011; 90(3):331–335. [PubMed: 21297016]
- Cheon K, Moser SA, Wiener HW, Whiddon J, Momeni SS, Ruby JD, et al. Characteristics of *Streptococcus mutans* genotypes and dental caries in children. *European Journal of Oral Sciences*. 2013; 3(Pt. 1):148–155. [PubMed: 23659236]
- Childers NK, Osgood RC, Hsu KL, Manmontri C, Momeni SS, Mahtani HK, et al. Real-time quantitative polymerase chain reaction for enumeration of *Streptococcus mutans* from oral samples. *European Journal of Oral Sciences*. 2011; 119(6):447–454. [PubMed: 22112030]
- Debourgogne A, Gueidan C, de Hoog S, Lozniewski A, Machouart M. Comparison of two DNA sequence-based typing schemes for the *Fusarium solani* Species Complex and proposal of a new consensus method. *Journal of Microbiological Methods*. 2012; 91(1):65–72. [Comparative Study Evaluation Studies]. [PubMed: 22820199]
- Deng W, Maust BS, Nickle DC, Learn GH, Liu Y, Heath L, et al. DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques*. 2010; 48(5):405–408. [Research Support, N.I.H., Extramural]. [PubMed: 20569214]
- Do T, Gilbert SC, Clark D, Ali F, Fatturi Parolo CC, Maltz M, et al. Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. *Public Library of Science*. 2010; 5(2):e9073.
- Foley SL, White DG, McDermott PF, Walker RD, Rhodes B, Fedorka-Cray PJ, et al. Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *Journal of Clinical Microbiology*. 2006; 44(10):3569–3577. [PubMed: 17021084]
- Francisco AP, Bugalho M, Ramirez M, Carrico JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics*. 2009; 10:152. [Research Support, Non-U.S. Gov't]. [PubMed: 19450271]
- Healy M, Huong J, Bittner T, Lising M, Frye S, Raza S, et al. Microbial DNA typing by automated repetitive-sequence-based PCR. *Journal of Clinical Microbiology*. 2005; 43(1):199–207. [PubMed: 15634972]
- Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. [Comparative Study]. *Journal of Clinical Microbiology*. 1988; 26(11):2465–2466. [PubMed: 3069867]
- Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics*. 2001; 17(12):1230–1231. [PubMed: 11751234]
- Kilian M, Scholz CF, Lomholt HB. Multilocus sequence typing and phylogenetic analysis of *Propionibacterium acnes*. *Journal of Clinical Microbiology*. 2012; 50(4):1158–1165. [Comparative Study Research Support, Non-U.S. Gov't]. [PubMed: 22259216]
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25(11):1451–1452. [Research Support, Non-U.S. Gov't]. [PubMed: 19346325]
- Loesche W. Role of *Streptococcus mutans* in human dental decay. *Microbiological Reviews*. 1986; 50(4):353–380. [PubMed: 3540569]
- Maiden MC. Multilocus sequence typing of bacteria. *Annual Review of Microbiology*. 2006; 60:561–588.
- Momeni SS, Whiddon J, Moser SA, Cheon K, Ruby JD, Childers NK. Comparative genotyping of *Streptococcus mutans* by repetitive extragenic palindromic polymerase chain reaction and multilocus sequence typing. *Molecular Oral Microbiology*. 2013; 28(1):18–27. [PubMed: 23194334]

- Moser SA, Mitchell SC, Ruby JD, Momeni S, Osgood RC, Whiddon J, et al. Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. *Journal of Clinical Microbiology*. 2010; 48(2):599–602. [PubMed: 20007400]
- Nakano K, Lapidattanakul J, Nomura R, Nemoto H, Alaluusua S, Gronroos L, et al. *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. *Journal of Clinical Microbiology*. 2007; 45(8):2616–2625. [PubMed: 17567784]
- Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America*. 1993; 90(10):4384–4388. [Research Support, Non-U.S. Gov't]. [PubMed: 8506277]
- Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. [Review]. *Journal of Clinical Microbiology*. 1995; 33(9):2233–2239. [PubMed: 7494007]
- van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. [Guideline]. *Clinical Microbiology and Infection*. 2007; 13(Suppl. 3):1–46. [PubMed: 17716294]

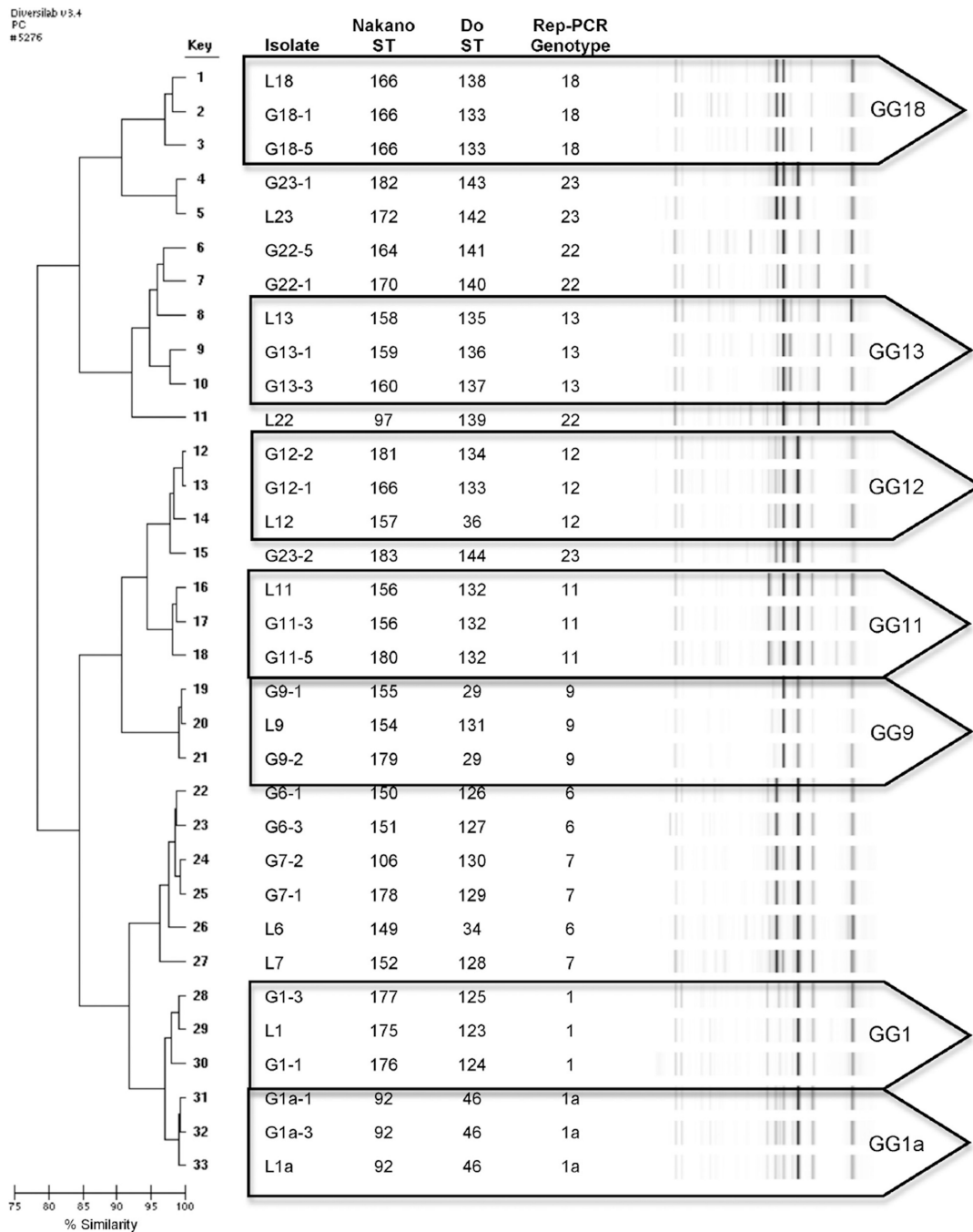


Fig. 1. Rep-PCR dendrogram by Percent Similarity with MLST Sequence Type (ST) assignments for Nakano and Do MLST Schemes for *S. mutans* Genotype Groups. GG = genotype groups. Isolates labeled with an “L” prefix are Library representative strains. Isolates labeled with a “G” prefix are additional rep-PCR genotype group members. Note GG-1a was previously published as GG-1. ST <175 Nakano and <123 Do are previously published. Note that ST assignments for the two schemes do not correspond to each other.

Table 1

New alleles and sequence types (ST) registered with PubMLST for *Nakano S. mutans* MLST scheme in this study.

PubMLST ST	Isolate ID ^b	Allelic profiles ^a										
		<i>tkr</i>	<i>glnA</i>	<i>gltA</i>	<i>glk</i>	<i>aroE</i>	<i>murI</i>	<i>lepC</i>	<i>gyrA</i>			
175	NL01	1	2	1	8	4	3	1	1	1	1	
176	G1-1	1	1	1	23	1	1	1	1	1	1	
177	G1-3	1	25	31	3	1	1	1	1	1	1	
178	G7-1	1	1	31	8	4	5	1	1	6	6	
179	G9-2	1	2	15	3	1	11	11	11	1	1	
180	G11-5	3	8	15	8	29	5	21	21	1	1	
181	G12-2	2	2	4	5	4	22	30	30	1	1	
182	G23-1	2	2	5	1	27	23	1	1	1	1	
183	G23-2	6	3	20	4	4	5	33	33	4	4	

^aNew alleles reported in this study are in bold.

^bL: library isolate, G: genotype group isolate.

Table 2

New alleles and sequence types (ST) registered with PubMLST for *S. mutans* MLST scheme in this study.

PubMLST ST	Isolate ID ^b	Allelic profiles ^a									
		<i>accC</i>	<i>gki</i>	<i>lepA</i>	<i>recP</i>	<i>sodA</i>	<i>tyrS</i>	<i>gfbB</i>	<i>spaP</i>		
123	L1	3	4	1	1	5	5	2	3		
124	G1-1	1	32	1	38	22	1	1	1		
125	G1-3	1	6	1	1	1	1	1	1		
126	G6-1	3	4	1	18	5	33	35	1		
127	G6-3	3	6	3	34	19	5	20	1		
128	L7	1	6	1	27	5	3	1	1		
129	G7-1	3	6	1	39	23	3	36	9		
130	G7-2	1	6	1	27	5	34	1	1		
131	L9	3	4	6	1	5	35	1	9		
132	L11	14	4	21	15	5	1	3	38		
133	G12-1	3	30	1	13	6	32	34	1		
134	G12-2	1	6	22	27	6	3	1	5		
135	L13	3	33	1	40	20	32	1	1		
136	G13-1	23	32	3	41	22	11	10	39		
137	G13-3	2	34	23	1	6	14	37	9		
138	L18	24	30	1	13	6	32	34	1		
139	L22	25	5	24	34	24	31	1	40		
140	G22-1	26	6	1	34	25	31	2	41		
141	G22-5	3	4	1	42	5	36	35	3		
142	L23	3	4	3	17	12	1	15	1		
143	G23-1	3	4	1	27	10	17	3	3		
144	G23-2	3	4	3	17	5	11	14	5		

^aNew alleles reported in this study are in bold.

^bL: library isolate, G: genotype group isolate.

Table 3

Characteristics of loci for *S. mutans* isolates using Nakano and Do MLST typing schemes.^a

Nakano scheme									
Locus	Fragment size (bp) ^b	No. of alleles	G + C Mol	No. polymorphic sites (%)	Synonymous changes	Non-synonymous changes	dN/dS	Tajima's D test	
<i>murI</i>	425	8	0.393	11 (2.59)	8	3	0.054	0.302	
<i>glmA</i>	460	8	0.378	7 (1.52)	7	0	0	-0.491	
<i>tkf</i>	435	10	0.449	11 (2.53)	10	1	0.043	-0.976	
<i>gyrA</i>	435	6	0.425	6 (1.38)	4	2	0.389	-1.488	
<i>glfA</i>	389	10	0.397	12 (3.08)	7	5	0.148	-0.269	
<i>aroE</i>	397	10	0.361	10 (2.52)	5	5	0.377	-0.913	
<i>glk</i>	405	8	0.409	8 (1.98)	6	2	0.173	-0.613	
<i>lepC</i>	420	8	0.392	8 (1.90)	7	1	0.032	1.146	
Do scheme									
Locus	Fragment size (bp) ^c	No. of alleles	G + C Mol	No. polymorphic sites (%)	Synonymous changes	Non-synonymous changes	dN/dS	Tajima's D test	
<i>accC</i>	462	8	0.402	7 (1.52)	5	2	0.146	-1.637	
<i>gki</i>	426	9	0.431	10 (2.35)	8	2	0.044	-1.079	
<i>lepA</i>	441	7	0.414	8 (1.81)	7	1	0.014	-0.924	
<i>recP</i>	474	13	0.403	11 (2.32)	6	5	0.3	-0.613	
<i>sodA</i>	492	11	0.412	13 (2.64)	6	7	0.152	-1.767	
<i>tyrS</i>	513	13	0.382	16 (3.12)	12	4	0.063	0.404	
<i>glfB</i>	453	11	0.34	10 (2.21)	2	8	0.536	0.351	
<i>spaP</i>	513	8	0.395	13 (2.53)	9	4	0.071	-1.360	

^aNumber of *S. mutans* isolates used were Nakano scheme ($n = 33$) and Do scheme ($n = 32$). Data generated using DnaSP software v.5.10.1.

^bFragment size reported in this study for Nakano scheme is slightly different than originally reported by Nakano et al. (2007).

^cFragment size for Do scheme were trimmed to PubMLST length.

Diversity indices and congruence values for MLST schemes and rep-PCR using 33 *S. mutans* isolates.^a

Table 4

	# Partitions	SID ^c (95% CI)	Adjusted Rand (95% CI) ^b		Wallace coefficient (95% CI)		
			MLST Do	MLST Nakano	Rep-PCR	MLST Do	MLST Nakano
Rep-PCR	11	0.938 (0.937–0.938)	0.353 (0.005–0.735)	0.305 (0.000–0.687)	1	0.228 (0.071–0.385)	0.197 (0.043–0.351)
MLST Do	26	0.981 (0.962–1.000)	–	0.694 (0.312–1.000)	0.787 (0.586–0.988)	1	0.694 (0.477–0.912)
MLST Nakano	27	0.981 (0.958–1.000)	–	–	0.680 (0.341–1.000)	0.694 (0.370–1.000)	1

^a Statistics calculated using Comparing Partitions software.

^b _{jackknife} pseudo-values.

^c SID: Simpson's Diversity Index.