Published in final edited form as:

Eur J Oral Sci. 2015 December ; 123(6): 416–424. doi:10.1111/eos.12221.

Clonality and serotypes of *Streptococcus mutans* among children by multilocus sequence typing

Stephanie S. Momeni^{1,2}, Jennifer Whiddon³, Kyounga Cheon², Stephen A. Moser³, and Noel K. Childers²

¹Department of Biology, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, AL, USA

²Department of Pediatric Dentistry, School of Dentistry, University of Alabama at Birmingham, Birmingham, AL, USA

³Department of Pathology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

Abstract

Studies using multilocus sequence typing (MLST) have demonstrated that *Streptococcus mutans* isolates are genetically diverse. Our laboratory previously demonstrated clonality of *S. mutans* using MLST but could not discount the possibility of sampling bias. In this study, the clonality of randomly selected *S. mutans* plaque isolates from African American children was examined using MLST. Serotype and presence of collagen-binding proteins (CBP) *cnm/cbm* were also assessed. One hundred *S. mutans* isolates were randomly selected for MLST analysis. Sequence analysis was performed and phylogenetic trees were generated using START2 and MEGA. Thirty-four sequence types (ST) were identified of which 27 were unique to this population. Seventy-five percent of the isolates clustered into 16 clonal groups. Serotypes observed were *c* (n=84), *e* (n=3), and *k* (n=11). The prevalence of *S. mutans* isolates serotype *k* was notably high at 17.5%. All isolates were *cnm/cbm* negative. The clonality of *S. mutans* demonstrated in this study illustrates the importance of localized populations studies and are consistent with transmission. The prevalence of serotype *k*, a recently proposed systemic pathogen, observed in this study is higher than reported in most populations and is the first report of *S. mutans* serotype *k* in a US population.

Corresponding Author: Noel K. Childers, Department of Pediatric Dentistry, University of Alabama at Birmingham, School of Dentistry Room 304, 1720 2nd Avenue South, Birmingham, AL 35294-0007, USA.

Conflicts of interest

The authors declare no conflicts of interest related to this study.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. New *S. mutans* MLST alleles and sequence types added to PubMLST database from this study.

Fig. S1. Phylogenetic tree of 67 *S. mutans* isolates generated using MEGA Minimum Evolution with bootstrap (1,000 replicates) including serotype and DMFS+dmfs scores.

Fig. S2. Phylogenetic tree of 67 *S. mutans* isolates generated using START2 using Un-weighted Pair Group Method using Arithmetic averages (UPGMA).

Keywords

multilocus sequence typing; genotyping techniques; molecular epidemiology; dental caries; rep-PCR

Streptococcus mutans is a common colonizer of the oral cavity and has been widely associated with initiation and progression of dental caries (1, 2). While typically associated with the oral cavity, *S. mutans* has been linked to other systemic diseases including infective endocarditis, inflammatory bowel syndrome, aneurysm formation, and hemorrhagic stroke making this organism clinically relevant to overall human health (3–6).

Multilocus sequence typing (MLST) is a molecular based typing method that is considered highly discriminative for typing bacterial strains. The first MLST typing scheme for *S. mutans* was developed by Nakano *et al.* published in 2007 using partial fragments of 8 conserved housekeeping genes (7). In a previous study, we reported 3 out of 6 repetitive extragenic palindromic PCR (rep-PCR) genotypes groups were identical by MLST (8). This clonality was surprising given that in other studies of *S. mutans* with MLST, sequence types (ST) are rarely shared between unrelated individuals (7, 9). However, the number of isolates in our previous study was small (only 3 clonal groups) and we could not rule out selection bias since isolates in that study were selected based upon the rep-PCR genotypes (8).

S. mutans has 4 serotypes (*c*, *e*, *f*, and *k*) with serotype *c* being the most frequently reported (>70% of clinical isolates) and serotype *k* (<5%) the least (4, 10–12). In addition, two collagen-binding proteins have been reported in *S. mutans*, which may play a role in infective endocarditis (IE) as well as other systemic diseases and binding to tooth surfaces. These have been identified as the Cnm protein (*cnm* gene) and Cbm protein (*cbm* gene) and when present are suspected to promote *S. mutans* adherence to and invasion of human endothelial cells (11, 13, 14). Approximately 85% of serotype *k* isolates from Japan, Finland, and Thailand are reported to contain either the *cnm* or *cbm* gene (11). Serotype *k* strains with the *cbm* gene are most frequently associated with high collagen binding activity and invasive properties while serotype *k* isolates that are *cnm/cbm* negative lack or have minimal collagen binding capabilities (11, 15). Multiplex PCR approaches have been designed for rapid identification of *S. mutans* serotypes and collagen binding genes (11).

The purpose of the current study was to evaluate if clonality of *S. mutans* isolates by MLST is observed in a group of randomly selected isolates. Comparative analysis with rep-PCR was performed to evaluate genotypic distribution of isolates between methods. In addition, the isolates were serotyped and tested for the presence of collagen binding genes, *cnm* and *cbm*.

Materials and Methods

Sample selection

S. mutans isolates were obtained as part of an ongoing longitudinal epidemiological study of dental caries in a high-carries risk community in Uniontown, AL, USA. This sample population is considered high-risk due to low socioeconomic status, limited access to dental

care, and markedly high caries prevalence in children less than 5 yr of age (16). This University of Alabama at Birmingham Institutional Review Board approved study used informed consent obtained from parents of participating children and children gave assent for participation. Decayed, missing, or filled teeth/decay, missing or filled surfaces (DMFS +dmfs) scores were determined according to World Health Organization (WHO) criteria at oral examinations performed by three trained and calibrated dentist examiners (16, 17). DMFS+dmfs scores are used as indicators of caries history.

Initial sample collection, sample processing, and isolate selection have been described previously (18, 19). To date, 14,979 *S. mutans* isolates have been isolated and genotyped from unrelated African American index children and their household family members using rep-PCR. For index children, samples were collected every 6 months for the first 36 months then annually over an 8-yr period. The current study focused on Cohort 1(CH1)-school aged children (5–6 yr of age at initial enrollment) and Cohort 2 (CH2)-infants (2–3 yr of age, initially) with the original samples collected at an elementary school (CH1) or a local community health center (CH2). One hundred *S. mutans* plaque isolates were randomly selected from the two cohorts of children. Children in this study demonstrate all levels of caries as indicated by DMFS+dmfs scores and were otherwise healthy. Inclusion criteria were *S. mutans* isolates from children that had been confirmed as *S. mutans* by PCR using *gftB* specific primers and had a rep-PCR genotype assigned. The available pool consisted of 4,693 isolates from 115 children (45 CH1 and 70 CH2) with 26 rep-PCR genotypes represented. Isolates were not limited to one per child so that stability of genotypes within a child with multiple isolates could potentially be evaluated (i.e., if randomly selected). Randomization of samples was performed in an Excel spreadsheet using the RAND function.

Rep-PCR analysis

Rep-PCR was performed using DiversiLab as previously reported (18). Briefly, *S. mutans* isolates were confirmed based on colony morphology on Gold's Media (modified mitis salivarius media [Becton Dickinson, Franklin Lakes, NJ, USA], supplemented with bacitracin and sucrose), DNA was extracted using UltraClean Microbial DNA Isolation Kit (MoBio, Carlsbad, CA, USA), and isolates were confirmed as *S. mutans* with *gftB* sequence specific primers using a SYBR Green PCR approach (20, 21). Rep-PCR was performed using *Streptococcus* DNA fingerprinting kit (bioMérieux, Durham, NC, USA) and a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). Amplicons were visualized using microfluidics LabChip technology (bioMérieux). Data analysis was performed using the DiversiLab web-based software. New rep-PCR genotypes were determined based on similarity of dendrograms and electropherogram overlays using the 3 minor band, 1 major band differences to distinguish unique genotypes (18).

MLST analysis

PCR for MLST analysis was performed on the 100 randomly selected isolates as previously described (8). Briefly, for each sample, 8 PCR reactions were performed by the Nakano *et al.* MLST Typing scheme for *S. mutans* including primers for partial gene fragments from *murI*, *tkl*, *glnA*, *gyrA*, *glk*, *glt*, *glk*, and *lepC* using a SYBR Green PCR approach (7, 8).

Amplicons were purified and sequenced. Sequence data was analyzed using CLC DNA Workbench 5.7.1 with MLST Module (CLC bio USA, Cambridge, MA, USA). Allelic variation was evaluated using 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) approach (22). Phylogenetic analysis was performed using concatenated sequences (3,366 base pairs) analyzed using Molecular Evolutionary Genetics Analysis version 5.2.2 (MEGA; www.megasoftware.net/ The Biodesign Institute, Tempe, AZ, USA) with a Minimum Evolution with Bootstrap (1,000 replicates) (23). Population structure analysis was performed using goeBURST using both the single and double locus variants settings (24). For this study, clonal groups are defined as isolates with the same ST while clonal complexes are defined as isolates with similar ST (up to two allelic differences) (7).

DiveIn, an online tool used for phylogenetic analysis (<http://indra.mullins.microbiol.washington.edu/DIVEIN/>), was used to calculate variable sites including informative sites (mutations occurring in more than one isolates) and private sites (mutations occurring in only one isolate) using *S. mutans* UA159 as a reference (25). Informative sites data was cross-referenced with DMFS+dmfs scores to determine if common mutations occur in either caries or no caries children. For the purpose of this paper, caries refers to any caries history (current or previous) and no caries refers to no previous or currently detectable caries activity.

The degree of clonality was calculated by the Index of Association (I_A) in START2 using the Maynard Smith approach (26). The I_A was calculated using all 100 (all ST), 67 (single representative ST per individual, including special cases), and 34 (single representative of each ST) isolates. The population is considered a clonal population when the value of I_A differs significantly from zero. General characteristics of the 100 strains including number of polymorphic sites, G+C content, synonymous and non-synonymous changes, average non-synonymous/synonymous rate ratio (dN/dS) were calculated using DnaSP v5.10.1 (DnaSP; www.ub.edu/dnasp/ Universitat de Barcelona, Barcelona, Spain) (27). Simpson's Index of Diversity (SID, discriminatory power), Adjusted Rand (congruence) and Adjusted Wallace were calculated using Comparing Partitions (<http://darwin.phylloviz.net/ComparingPartitions/index.php?link=Home>) (28, 29).

Serotypes and collagen binding proteins

Serotyping and presence of collagen binding proteins, *cnm* and *cbm* were performed for all 100 isolates using a SYBR Green PCR multiplex approach with primers previously reported (11, 30, 31). PCR was performed on a IQ5 Real-time thermocycler (Bio-Rad Laboratories, Hercules, CA, USA) using a 25 µl reaction with 2 µl of DNA template (20 ng/µl) and 2x Maxima™ SYBR Green qPCR master mix (Thermo Scientific, Lafayette, CO, USA) using the following parameters: 1 cycle of 95°C for 10 min followed by 30 cycles of 95°C for 20 s, 60°C for 30 s, 72°C for 45 s. Melt curve analysis was performed at 60°C – 95°C with 0.5°C intervals. Separate multiplex PCRs were performed for serotype and collagen binding proteins. Controls included for serotyping were representative prototype *S. mutans* strains UA159 (serotype *c*), LM7 (serotype *e*), OMZ-175 (serotype *f*), and FT1 (serotype *k*).

Controls for collagen binding proteins were *S. mutans* strains OMZ-175 (*cnm+*) and YT1 (*cbm+*). PCR for the detection of collagen binding proteins and undetermined serotypes were performed in duplicate. All amplicons were subjected to electrophoresis on 1.3% agarose gels to confirm amplicon size with a 100 base pair ladder (New England Bio Labs, Ipswich, MA, USA).

Results

For clarification in this manuscript, genotype (GT) refers to rep-PCR genotype, sequence type (ST) refers to MLST sequence type, and serotype (*c,e,f* or *k*) refers to immunological serotypes as determined by PCR methodology.

Stability

The 100 isolates selected were closely distributed between CH1 (n=51) and CH2 (n=49). A total of 57 individual children (CH1 n=29; CH2 n=28) and 18 genotypes were represented (Table 1). Twenty-seven children (27/57, 47%) had more than one isolate (2–5 isolates) (Table 2). Eighteen children exhibited a single rep-PCR genotype and MLST ST for all randomly selected isolates (1GT/ 1 ST). For stability analysis, 24 children had isolates from multiple periods of collection while 3 of the children were excluded since isolates from those children were from the same collection period (193, 228, 579). Fifteen of the children had isolates with the same rep-PCR genotype and MLST ST from more than one period indicating that stability was 63% (15/24).

MLST

Also in Table 2, 5 children demonstrated multiple genotypes and ST (2GT/ 2ST) that agreed between methods. Two children (241, 531) had the same rep-PCR genotype but had different MLST ST indicating MLST was more discriminative. Rep-PCR was more discriminative than MLST for 2 other children (150, 596) who had different rep-PCR genotypes but fewer MLST ST. Isolates with these special cases were also included in the MLST analysis resulting in 67 total isolates (from 57 children).

Overall a total of 34 ST were identified, representing the 18 rep-PCR genotypes from 100 isolates. Using goeBURST with the single locus variant setting, the 67 isolates were divided into 27 groups while using the double locus variant setting resulted in 21 groups. Three new alleles and 16 new ST were identified in this study and were added to the PubMLST database (Supporting Table S1). Sequences for new allele types generated by MLST were submitted to Genbank (<http://www.ncbi.nlm.nih.gov/genbank>) and assigned accession numbers KR995097-KR995099. Of the 34 ST identified, 7 ST matched ST available in the PubMLST database; however, 27 ST were unique to the Uniontown population. Forty isolates were classified into one of the 16 new STs found in this study. Sixty isolates matched previously published ST (19 isolates matched ST from the PubMLST database for Japanese, Thai, or Finnish isolates and the remaining 41 isolates matched ST previously reported from the Uniontown Study) (unpublished findings, S. S. Momeni, J. Whiddon, K. Cheon, S. A. Moser, N. K. Childers) (8).

For 10 of the 18 rep-PCR genotype groups (56%), MLST further differentiated rep-PCR genotypes (e. g. rep-PCR genotype G01a isolates were typed as MLST ST 92, 192, 193) (Table 1, Supporting Fig. S1). Eight rep-PCR genotype groups (44%, G07, G10, G11, G14, G15, G18, G27, G50) were further supported by MLST analysis. Five MLST ST (156, 179, 191, 192, and 202) represented more than one rep-PCR genotype (e. g. MLST ST 156 includes rep-PCR genotypes G11 and G12).

A total of 70 variable sites were observed of which 49 were informative sites (mutations shared in more than one isolate) and 21 were private sites (single mutations occurring in only 1 isolate). No relationship was observed between caries status (DMFS+dmfs score) and base pair changes reported (Figs. S1, S2). DMFS+dmfs (collection) indicates the score at time the isolate evaluated was obtained while (final) indicates the child's score at last period on record. Final score data should be interpreted with caution as these values may vary since children in CH1 tend to decrease as they loose baby teeth and children in CH2 tend to increase as new teeth erupt.

General allele characteristics are listed in Table 3. Genes *gltA* (3.6%) and *gyrA* (1.2%) have the highest and lowest frequencies of polymorphisms respectively. The number of alleles reported ranged from 5 (*gyrA*) to 12 (*gltA*). The dN/dS ratio was <1 for all genes except *gyrA* (1.84). Discriminatory power for rep-PCR and MLST were determined to be comparable as determined by Simpson's Index of Diversity since rep-PCR (0.918) and MLST (0.948) had overlapping confidence intervals. MLST and rep-PCR were found to be 68% congruent by adjusted Rand. According to Adjust Wallace, MLST was more likely to predict rep-PCR genotypes (0.891; CI=0.858–0.924) than rep-PCR was to predict MLST (0.550; CI=0.495–0.606).

Clonality

To eliminate bias for the investigation of clonality, only one isolate per child was included in phylogenetic analysis (termed "focus group"), except if a child had more than one genotype or ST in which case a representative of each genotype or ST was also included for analysis. A total of 67 isolates were evaluated resulting in 16 clonal groups that contained 75% (50/67) of the focus group isolates (Fig. S1). Allowing for up to two allelic difference (clonal complexes) resulted in 79% (53/67) of isolates belonging to 14 clonal groups (single line boxes) or complexes (double line boxes) (Fig. S2). The I_A for all 100 isolates was 1.943; for 67 representative individuals was 1.561; and for 34 single ST was 0.7721.

Serotype and collagen binding proteins

Serotypes for the 100 isolates were *c* (n=84), *e* (n=3), and *k* (n=11). No serotype *f* isolates were found in this study. Two isolates from two children demonstrated no amplification and were noted as undetermined serotypes. Serotype *k* isolates were found in 10 different subjects. Serotype *c* and *k* demonstrated distinctive melt patterns with a trifurcated product centered at 74.5°C and a single peak at 76–78°C, respectively. All serotype *c* samples resulted in a single band during electrophoresis indicating the trifurcated melt peak observed for serotype *c* by real-time PCR is likely due to the size of the amplicon (727 bp), which exceeds the manufacturer's recommended fragment size (<150 bp) for the master mix used.

Both serotypes *e* and *f* demonstrated a single peak at 74.5°C and were indistinguishable until run on gels. Six of the 10 serotype *k* isolates grouped together into 2 clonal clusters (Clonal clusters 5 & 6, ST205 & ST106) by MLST analysis and corresponded to rep-PCR genotypes G14 and G07 respectively (Figs. S1, S2).

Collagen binding proteins *cnm* and *cbm* were not observed in any of isolates examined. Controls OMZ-175 (*cnm*+) and YT1 (*cbm*+) demonstrated single, distinctive melt peaks at 80°C and 81°C respectively. Only control strains were observed on gels.

Discussion

Initial analysis used all 100 randomly selected isolates and found that 27 of the 57 of children (47%) had more than one isolate represented (Table 2). Of the 27 children with more than one isolate, 15 children's isolates (collected at multiple time points) were the same genotype and ST indicating the stability of isolates within individual children was 63%. However, 9 other children had more than one genotype or ST indicating variability of isolates within a child is about 33%. Although the majority of children with multiple isolates demonstrated stability, the number of children with variable genotypes/ST supports the need to test more than one isolate per child. This is further supported by a previous report that the average number of rep-PCR genotypes per child in this population is about two (32).

It is not surprising that 56% of rep-PCR genotypes were further differentiated by MLST, since rep-PCR is based on similarity while a single base-pair change can result in a new ST by MLST. Nonetheless, 8 (44%) genotype groups were confirmed by MLST analysis demonstrating these isolates are identical by both MLST and rep-PCR analysis (Fig. S1) although it should be noted that two GTs had only one or two isolates, i.e., G27 and G11, respectively. The most noteworthy of the genotypes is G18, overall, the most common isolate, which consistently resulted in the same ST (ST 166) for 10 different children (18 isolates). Additionally, a subgroup of G01a (Clonal group 12) was also supported by both rep-PCR and MLST. These cases of clonality provide some data that indicate isolates are shared between unrelated children in this community. This finding is consistent with common transmission among these children, although the source of transmission was beyond the focus of the current study and was therefore not identified. It is worth mentioning, that the children in this study are not immediately related however in a small community such as Uniontown, Alabama (total population 2,539 as of 2013 according to the US census Bureau), the possibility of distant familial relations is both possible and probable. Therefore, the potential of a common generational source will be considered in future studies.

In 5 cases (Table 1, Fig. S1) (clonal groups 1, 4, 9, 10, 11), MLST ST were further differentiated by rep-PCR. Although it has been reported that rep-PCR can be more discriminative than MLST for other organisms (33–36), this study provides the first known evidence that rep-PCR can be more discriminative than MLST for *S. mutans*. This finding further supports the importance of using an alternate genotype method to validate assignments (37, 38). Furthermore, this data supports the finding that typing assignments between rep-PCR and MLST do not always correlate. These results suggest that while these

isolates have identical sequences for the housekeeping genes, the repetitive elements are different. In contrast to the conserved housekeeping genes used in MLST that have known functions, the purpose of repetitive elements evaluated by rep-PCR remain largely unknown (39, 40).

The number of polymorphic sites reported here (range 5–14 sites) is lower than previously reported by Nakano *et al.* (7) (range 15–21) and are consistent with our previous study (range 6–12) (Table 3). Variable nucleotide sites were highest in *gltA* in all three studies but *gyrA* is lowest in both Uniontown studies (unpublished findings, S. S. Momeni, J. Whiddon, K. Cheon, S. A. Moser, N. K. Childers) (7). In the present study, *gyrA* was found to have a $dN/dS > 1$ which indicates positive selection for this gene in the isolates evaluated in this study. Differences observed may be explained by population bias; Nakano *et al.* (7) evaluated regional *S. mutans* isolates from a variety of sources where as the present study used *S. mutans* plaque samples from a focused, relatively isolated population of children.

The finding that a majority of randomly selected *S. mutans* isolates (75%) were clonal supports the previous results reported by our laboratory and indicates that the clonality previously reported was not the result of pre-selection by rep-PCR (8). This clonality is comparable to the 70% reported by Lapiattanakul *et al.* (41) using MLST; however that study used mother-child pairs where as the current study used randomly selected, unrelated subjects. The percentage of shared sequence types among unrelated subjects was 49% (34 ST/67 samples) in the present study, which is notably higher than previously found by Nakano *et al.* (7) (9.8%; 92 ST/102 samples) and Do *et al.* (9) (9.6%; 122 ST/135 samples) using MLST. Furthermore, the degree of clonality as estimated by the I_A (1.5626) for the 67 representative isolates suggests significant linkage disequilibrium indicative of a clonal population. Adjusting for a single representative resulted in an $I_A = 0.7721$ which is still higher than I_A values previously reported for *S. mutans* by Nakano *et al.* (0.0931) and Do *et al.* (0.4379) (7, 9) indicating the present group of isolates are more clonal than was reported in these studies.

Only 3 new alleles were discovered in the current study, indicating that the library of genotypes previously reported is highly representative of this population and that these genes, while diverse, are relatively homogenous as would be expected for conserved genes. However, the discovery of 16 new ST indicated sequence diversity typical of other MLST studies of *S. mutans* (7, 9). Other studies using MLST have reported that ST do not appear to be geographically distributed (7, 9). While a few isolates (19%) in the current study had ST available in the global database (7 ST), a majority of the isolates (81%) demonstrated ST unique to the Uniontown population (27 ST). This finding along with the high degree of clonality may suggest that epidemiological studies of geographically or ethnically isolated populations may provide important information on transmission or recent evolutionary populations shifts that may be overlooked in large regional/global studies. While the use of localized populations is an acknowledged bias, it provides an alternate perspective for these types of studies, especially in cases where sample pools are too diverse to make meaningful connections between isolates. Further studies are needed to determine if these results are common to other smaller high-risk sample populations.

This study is the first to report the discovery of serotype *k* isolates in a US population, specifically an African American population of children. The prevalence of serotype *k* (17.5%) reported here is higher than previously reported for Japanese (1.4%–2%), Finnish (3.6%) and Thai (2%–2.8%) subjects (10, 11, 42). A study of subjects from southern India (43) reported a higher prevalence (26%). However, in contrast to the India study that used saliva, the present study was conducted using individual *S. mutans* isolates. It is possible that analysis of saliva samples in the Uniontown population would result in an even higher or comparable prevalence to the India study. Future studies are planned to investigate the prevalence of serotype *k* in the Uniontown population using this approach. It should be noted that although the subject demographic of this study was African American children, this is a localized population and may not be indicative of all African American children.

Seven of the 10 serotype *k* isolates grouped together (G07, G14, G27) by phylogenetic analysis and corresponded to ST106, ST205, and ST161 respectively (Fig. S1). ST106 was originally reported to be serotype *c* in a single isolate from an adult female by Lapirattanakul *et al.* (41), which demonstrates that not all isolates with ST106 will be serotype *k*. Six of the 10 serotype *k* strains exhibited lower caries scores (DMFS+dmfs <10) or no caries activity, which is consistent with attenuated cariogenic properties reported for serotype *k* strains (15, 44).

The absence of any serotype *f* isolates is not surprising as it is considered a minor serotype reportedly occurring in fewer than 5% of isolates (30, 31). That two isolates did not produce an amplicon by the multiplex PCR methods previously reported for serotypes *c*, *e*, *f*, and *k* is unanticipated (30, 31). These samples were both confirmed as *S. mutans* based on morphological appearance on MSB and by PCR using *gfbB* specific primers (21). Furthermore, these two isolates match rep-PCR genotypes (G10 and G12) and MLST ST (ST179 and ST157) of other *S. mutans* strains. DNA template for PCR was at a concentration of 20 ng/μl, which is within the detectable range and both samples were repeated to confirm the failure. Shibata *et al.* (30) in 2003 originally reported one isolate that was undetermined, but this work was prior to the development of a serotype *k* primer set by Nakano *et al.* (31) in 2004, which is now commonly used in serotyping by multiplex PCR. A recent study in India of whole saliva reported 23% of isolates as undetermined serotypes; however this study did not include serotype *c* in the analysis (43). In 2007, Nakano *et al.* (45) reported the recovery of undetermined serotypes by the PCR approach in relative abundance in cardiovascular samples and suggested these may be a new, minor oral serotype. Further study is needed to evaluate if these undetermined serotypes are due to a limitation of detection by PCR approach or if these isolates are a new serotype.

Interestingly, none of the isolates tested positive for collagen binding proteins Cnm or Cbm by PCR multiplex approach. Overall, the detection rate of *cnm* positive *S. mutans* is reported to be between 10–20% and is most predominately associated with serotypes *f* and *k* (46, 47). The Cbm protein was only recently identified in 2012 by Nomura *et al.* (47), who reported a detection rate of only 2% in mostly serotypes *k*. Other studies using MLST reported that the *cnm* gene occurs in about 17% of oral isolates from stimulated saliva (n=150) and 26% for Japanese isolates (n=102) (7, 12). Other studies using *S. mutans* isolates reported a frequency of 11–16% for *cnm* but sample sizes was much higher (n=580, n=478) than the

100 isolates in the present study (11, 47). Based on the literature noted above, the outcome of the current study was unexpected. Furthermore, the 10 strains of serotype *k* in this study were found to be *cnm* or *cbm* negative strains, which significantly different from the 85% reported by Nomura *et al.* for serotype *k* strains (11). There are a number of reasons that may explain this outcome. Nomura *et al.* postulated that detection of Cnm proteins may be age or geographically dependent (47). It is also possible that the sample size of the current study (n=100) may be too small to detect the collagen binding proteins in this population. Furthermore, sample type (individual isolates versus whole saliva) is not always clearly noted in some references, which can make direct comparison difficult since the presence of *cnm* or *cbm*, if present, is more likely to be detected in a pooled saliva samples than individual isolates. Further study is needed to determine if whole saliva samples from the Uniontown population will yield the same results as those reported in this study using individual *S. mutans* isolates. Additionally, there may be limitations using a PCR-based approach (i.e., primer binding specificity). Nonetheless, the possibility remains that the *cnm* and *cbm* genes may not be present in the isolates evaluated in this study from the Uniontown population. If these genes are in fact absent in this population, it is predicted that these strains would have negative collagen-binding ability and therefore no invasive qualities associated with systemic disease, such as infective endocarditis (11, 14, 15).

In this study, a high degree of clonality of *S. mutans* isolates randomly selected from children in a relatively isolated population is demonstrated suggesting that geographically or ethnically focused studies may provide valuable insights for the study of *S. mutans*. Isolates identified as identical by both MLST and rep-PCR is consistent with transmission. Together these two methods can provide effective tools for epidemiological studies of transmission for future studies. This study provides the first indication that rep-PCR can be more discriminative than MLST for the analysis of *S. mutans* in some cases. Furthermore, this study provides the first report of *S. mutans* serotype *k* found in a US (specifically, African American) population at a notably high prevalence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by Research Grant DE016684 (NKC) 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. Dr. Mei Han of the UAB Heflin Genomic Core provided sequencing data; Dr. Jinthana Lapidattanakul directed the confirmation and registration of isolates with PubMLST; Dr. Ping Zhang and Kazuhiko Nakano generously donated some controls strains, and Dr. Hernan Grenett provided electrophoresis data.

References

1. Tanner AC, Milgrom PM, Kent R Jr, Mokeem SA, Page RC, Riedy CA, Weinstein P, Bruss J. The microbiota of young children from tooth and tongue samples. *J Dent Res.* 2002; 81:53–57. [PubMed: 11824414]
2. Tanzer JM, Livingston J, Thompson AM. The microbiology of primary dental caries in humans. *J Dent Educ.* 2001; 65:1028–1037. [PubMed: 11699974]

3. Nakano K, Nemoto H, Nomura R, Inaba H, Yoshioka H, Taniguchi K, Amano A, Ooshima T. Detection of oral bacteria in cardiovascular specimens. *Oral Microbiol Immunol*. 2009; 24:64–68. [PubMed: 19121072]
4. Hamada S, Slade HD. Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol Rev*. 1980; 44:331–384. [PubMed: 6446023]
5. Nakano K, Hokamura K, Taniguchi N, Wada K, Kudo C, Nomura R, Kojima A, Naka S, Muranaka Y, Thura M, Nakajima A, Masuda K, Nakagawa I, Speziale P, Shimada N, Amano A, Kamisaki Y, Tanaka T, Umemura K, Ooshima T. The collagen-binding protein of *Streptococcus mutans* is involved in haemorrhagic stroke. *Nat Commun*. 2011; 2:485. [PubMed: 21952219]
6. Kojima A, Nomura R, Naka S, Okawa R, Ooshima T, Nakano K. Aggravation of inflammatory bowel diseases by oral streptococci. *Oral Dis*. 2014; 20:359–366. [PubMed: 23679203]
7. Nakano K, Lapidattanakul J, Nomura R, Nemoto H, Alaluusua S, Gronroos L, Vaara M, Hamada S, Ooshima T, Nakagawa I. *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. *J Clin Microbiol*. 2007; 45:2616–2625. [PubMed: 17567784]
8. 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. *Mol Oral Microbiol*. 2013; 28:18–27. [PubMed: 23194334]
9. Do T, Gilbert SC, Clark D, Ali F, Fatturi Parolo CC, Maltz M, Russell RR, Holbrook P, Wade WG, Beighton D. Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. *PLoS One*. 2010; 5:e9073. [PubMed: 20140210]
10. Nakano K, Nomura R, Nakagawa I, Hamada S, Ooshima T. Demonstration of *Streptococcus mutans* with a cell wall polysaccharide specific to a new serotype, *k*, in the human oral cavity. *J Clin Microbiol*. 2004; 42:198–202. [PubMed: 14715753]
11. Nomura R, Nakano K, Naka S, Nemoto H, Masuda K, Lapidattanakul J, Alaluusua S, Matsumoto M, Kawabata S, Ooshima T. Identification and characterization of a collagen-binding protein, Cbm, in *Streptococcus mutans*. *Mol Oral Microbiol*. 2012; 27:308–323. [PubMed: 22759315]
12. Lapidattanakul J, Nakano K, Nomura R, Leelataweewud P, Chalermarp N, Klaophimai A, Srisattaluk R, Hamada S, Ooshima T. Multilocus sequence typing analysis of *Streptococcus mutans* strains with the *cnm* gene encoding collagen-binding adhesin. *J Med Microbiol*. 2011; 60:1677–1684. [PubMed: 21680768]
13. Sato Y, Okamoto K, Kagami A, Yamamoto Y, Igarashi T, Kizaki H. *Streptococcus mutans* strains harboring collagen-binding adhesin. *J Dent Res*. 2004; 83:534–539. [PubMed: 15218042]
14. Abranches J, Miller JH, Martinez AR, Simpson-Haidaris PJ, Burne RA, Lemos JA. The collagen-binding protein Cnm is required for *Streptococcus mutans* adherence to and intracellular invasion of human coronary artery endothelial cells. *Infect Immun*. 2011; 79:2277–2284. [PubMed: 21422186]
15. Lapidattanakul J, Nomura R, Nemoto H, Naka S, Ooshima T, Nakano K. Multilocus sequence typing of *Streptococcus mutans* strains with the *cbm* gene encoding a novel collagen-binding protein. *Arch Oral Biol*. 2013; 58:989–996. [PubMed: 23506778]
16. Ghazal T, Levy SM, Childers NK, Broffitt B, Cutter G, Wiener HW, Kempf M, Warren J, Cavanaugh J. Prevalence and incidence of early childhood caries among African-American children in Alabama. *J Public Health Dent*. 2015; 75:42–48. [PubMed: 25213319]
17. WHO. Oral Health surveys: basic methods. 4. Geneva: World Health Organization; 1997.
18. Moser SA, Mitchell SC, Ruby JD, Momeni S, Osgood RC, Whiddon J, Childers NK. Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. *J Clin Microbiol*. 2010; 48:599–602. [PubMed: 20007400]
19. Cheon K, Moser SA, Whiddon J, Osgood RC, Momeni S, Ruby JD, Cutter GR, Allison DB, Childers NK. Genetic Diversity of Plaque Mutans Streptococci with rep-PCR. *J Dent Res*. 2011; 90:331–335. [PubMed: 21297016]
20. Gold OG, Jordan HV, Van Houte J. A selective medium for *Streptococcus mutans*. *Arch Oral Biol*. 1973; 18:1357–1364. [PubMed: 4518755]
21. Yoshida A, Suzuki N, Nakano Y, Kawada M, Oho T, Koga T. Development of a 5' nuclease-based real-time PCR assay for quantitative detection of cariogenic dental pathogens *Streptococcus mutans* and *Streptococcus sobrinus*. *J Clin Microbiol*. 2003; 41:4438–4441. [PubMed: 12958287]

22. Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics*. 2001; 17:1230–1231. [PubMed: 11751234]
23. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–2739. [PubMed: 21546353]
24. 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. [PubMed: 19450271]
25. Deng W, Maust BS, Nickle DC, Learn GH, Liu Y, Heath L, Kosakovsky Pond SL, Mullins JI. DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *BioTechniques*. 2010; 48:405–408. [PubMed: 20569214]
26. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci U S A*. 1993; 90:4384–4388. [PubMed: 8506277]
27. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25:1451–1452. [PubMed: 19346325]
28. Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, Ramirez M. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol*. 2006; 44:2524–2532. [PubMed: 16825375]
29. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol*. 1988; 26:2465–2466. [PubMed: 3069867]
30. Shibata Y, Ozaki K, Seki M, Kawato T, Tanaka H, Nakano Y, Yamashita Y. Analysis of loci required for determination of serotype antigenicity in *Streptococcus mutans* and its clinical utilization. *J Clin Microbiol*. 2003; 41:4107–4112. [PubMed: 12958233]
31. Nakano K, Nomura R, Shimizu N, Nakagawa I, Hamada S, Ooshima T. Development of a PCR method for rapid identification of new *Streptococcus mutans* serotype *k* strains. *J Clin Microbiol*. 2004; 42:4925–4930. [PubMed: 15528675]
32. Cheon K, Moser SA, Wiener HW, Whiddon J, Momeni SS, Ruby JD, Cutter GR, Childers NK. Characteristics of *Streptococcus mutans* genotypes and dental caries in children. *Eur J Oral Sci*. 2013; 121:148–155. [PubMed: 23659236]
33. Ben-Darif E, De Pinna E, Threlfall EJ, Bolton FJ, Upton M, Fox AJ. Comparison of a semi-automated rep-PCR system and multilocus sequence typing for differentiation of *Salmonella enterica* isolates. *J Microbiol Methods*. 2010; 81:11–16. [PubMed: 20114063]
34. Lau SH, Cheesborough J, Kaufmann ME, Woodford N, Dodgson AR, Dodgson KJ, Bolton EJ, Fox AJ, Upton M. Rapid identification of uropathogenic *Escherichia coli* of the O25:H4-ST131 clonal lineage using the DiversiLab repetitive sequence-based PCR system. *Clin Microbiol Infect*. 2010; 16:232–237. [PubMed: 19416293]
35. Bonacorsi S, Bidet P, Mahjoub F, Mariani-Kurkdjian P, Ait-Ifrane S, Courroux C, Bingen E. Semi-automated rep-PCR for rapid differentiation of major clonal groups of *Escherichia coli* meningitis strains. *Int J Med Microbiol*. 2009; 299:402–409. [PubMed: 19451030]
36. Bourdon N, Lemire A, Fines-Guyon M, Auzou M, Perichon B, Courvalin P, Cattoir V, Leclercq R. Comparison of four methods, including semi-automated rep-PCR, for the typing of vancomycin-resistant *Enterococcus faecium*. *J Microbiol Methods*. 2011; 84:74–80. [PubMed: 21062634]
37. Foley SL, White DG, McDermott PF, Walker RD, Rhodes B, Fedorka-Cray PJ, Simjee S, Zhao S. Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *J Clin Microbiol*. 2006; 44:3569–3577. [PubMed: 17021084]
38. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect*. 2007; 13 (Suppl 3):1–46. [PubMed: 17716294]

39. Ishii S, Sadowsky MJ. Applications of the rep-PCR DNA fingerprinting technique to study microbial diversity, ecology and evolution. *Environ Microbiol.* 2009; 11:733–740. [PubMed: 19207574]
40. Maiden MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol.* 2006; 60:561–588. [PubMed: 16774461]
41. Lapirotanakul J, Nakano K, Nomura R, Hamada S, Nakagawa I, Ooshima T. Demonstration of mother-to-child transmission of *Streptococcus mutans* using multilocus sequence typing. *Caries Res.* 2008; 42:466–474. [PubMed: 18997467]
42. Lapirotanakul J, Nakano K, Nomura R, Nemoto H, Kojima A, Senawongse P, Srisatjaluk R, Ooshima T. Detection of serotype *k* *Streptococcus mutans* in Thai subjects. *Oral Microbiol Immunol.* 2009; 24:431–433. [PubMed: 19702960]
43. Rao AP, Austin RD. Serotype specific polymerase chain reaction identifies a higher prevalence of *streptococcus mutans* serotype *k* and *e* in a random group of children with dental caries from the Southern region of India. *Contemp Clin Dent.* 2014; 5:296–301. [PubMed: 25191062]
44. Nakano K, Nomura R, Nemoto H, Lapirotanakul J, Taniguchi N, Gronroos L, Alaluusua S, Ooshima T. Protein antigen in serotype *k* *Streptococcus mutans* clinical isolates. *J Dent Res.* 2008; 87:964–968. [PubMed: 18809752]
45. Nakano K, Nemoto H, Nomura R, Homma H, Yoshioka H, Shudo Y, Hata H, Toda K, Taniguchi K, Amano A, Ooshima T. Serotype distribution of *Streptococcus mutans* a pathogen of dental caries in cardiovascular specimens from Japanese patients. *J Med Microbiol.* 2007; 56:551–556. [PubMed: 17374899]
46. Nakano K, Nomura R, Taniguchi N, Lapirotanakul J, Kojima A, Naka S, Senawongse P, Srisatjaluk R, Gronroos L, Alaluusua S, Matsumoto M, Ooshima T. Molecular characterization of *Streptococcus mutans* strains containing the *cnm* gene encoding a collagen-binding adhesin. *Arch Oral Biol.* 2010; 55:34–39. [PubMed: 20005510]
47. Nomura R, Nakano K, Taniguchi N, Lapirotanakul J, Nemoto H, Gronroos L, Alaluusua S, Ooshima T. Molecular and clinical analyses of the gene encoding the collagen-binding adhesin of *Streptococcus mutans*. *J Med Microbiol.* 2009; 58:469–475. [PubMed: 19273643]

Table 1

Distribution of *S. mutans* rep-PCR Genotypes (n=18) from randomly selected isolates (n=100) for MLST Analysis.

Rep-PCR GT	# Isolates	# Individuals	MLST ST (# Isolates)
G01	9	4	175 (1), 176 (5), 191 (3)
G01a	13	6	92 (7), 192 (2), 193 (4)
G01b	4	4	1 (2), 191 (1), 194 (1)
G05	6	3	2 (2), 195 (3), 196 (1)
G06	4	3	150 (1), 197 (2), 198 (1)
G07	4	3	106 (4)
G09	6	5	179 (3), 199 (2), 200 (1)
G10	3	3	179 (3)
G11	2	1	156 (2)
G12	12	8	132 (2), 156 (1), 157 (3), 181 (1), 192 (3), 201 (1), 202 (1)
G13	5	4	194 (3), 203 (1), 204 (1)
G14	3	3	205 (3)
G15	3	3	206 (3)
G18	18	10	166 (18)
G22	2	2	59 (1), 164 (1)
G23	2	2	130 (1), 182 (1)
G27	1	1	161 (1)
G50	3	2	202 (3)
Totals	100	67	

Number of Individuals with a given GT (n=67) is higher here than reported in text (n=57) since some children have more than one genotype represented (e.g., one child with up to 3 different GTs, see Table 2). GT = rep-PCR genotype. ST = MLST sequence type. Number of isolates with a given ST is indicated in parenthesis following ST.

Table 2

Stability of *S. mutans* isolates within 27 children that had more than one isolate randomly selected.

Index Child Family ID	# of Isolates	Rep-PCR GT	PubMLST ST	Agreed	Disagree
102	2	G01a, G15	92, 206	2 GT/2 ST	
110	2	G18	166	1 GT/1 ST	
116	2	G01b, G15	191, 206	2 GT/2 ST	
119	2	G18	166	1 GT/1 ST	
149	2	G05	2	1 GT/1 ST	
150	3	G01, G01b, G13	194, 175, 194		3 GT/2 ST ^a
173	3	G10	179	1 GT/1 ST	
185	2	G18	166	1 GT/1 ST	
193 ^c	2	G18	166	1 GT/1 ST	
214	2	G07	106	1 GT/1 ST	
218	2	G13, G22	164, 203	2 GT/2 ST	
219	5	G18	166	1 GT/1 ST	
228 ^c	2	G50	202	1 GT/1 ST	
232	2	G09	179	1 GT/1 ST	
241	3	G12	132, 181		1 GT/2 ST ^b
252	2	G11	156	1 GT/1 ST	
501	4	G01a	92	1 GT/1 ST	
505	3	G06, G14	197, 205	2 GT/2 ST	
521	2	G01	176	1 GT/1 ST	
524	2	G13	194	1 GT/1 ST	
525	3	G01	176	1 GT/1 ST	
531	2	G05	195, 196		1 GT/2 ST ^b
537	4	G01a	193	1 GT/1 ST	
542	2	G01b, G23	1, 182	2 GT/2 ST	
546	3	G01	191	1 GT/1 ST	
579 ^c	2	G05	195	1 GT/1 ST	
596	5	G01a, G12	192		2 GT/1 ST ^a

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

New ST are in **bold** text. ST = MLST sequence type. GT = rep-PCR genotype.

^a Isolates were further distinguished with rep-PCR.

^b Isolates were further distinguished by MLST.

^c Isolates excluded from stability analysis since isolates were from same collection period. For Child 150, 3 GT were observed and the ST are noted in corresponding order.

Table 3

Characteristics of alleles evaluated in this study.

Locus	Fragment size (bp)	No. of alleles	G+C Mol	No. polymorphic sites (%)	Syn	NSyn	dN/dS	Tajima's D test
murI	425	11	0.394	12 (2.82)	8	4	0.0647	0.72634
glnA	460	8	0.378	6 (1.30)	6	0	0.0000	0.38093
tkt	435	8	0.449	10 (2.30)	7	3	0.1379	-0.37665
gyrA	435	5	0.425	5 (1.15)	3	2	1.8375	-0.87632
gluA	389	12	0.398	14 (3.60)	7	7	0.1507	-0.87840
aroE	397	9	0.361	8 (2.02)	3	5	0.4190	-0.37541
glk	405	9	0.407	8 (1.98)	6	2	0.1842	0.18850
lepC	420	9	0.392	7 (1.67)	6	1	0.0205	1.61063

Syn = Synonymous changes, NSyn = Nonsynonymous changes