2010

Genotyping of Streptococcus mutans Using Multilocus Sequence Typing

Stephanie Momeni
University of Alabama at Birmingham

MOLECULAR ANALYSIS OF *STREPTOCOCCUS MUTANS* GENETIC DIVERSITY FOR THE EPIDEMIOLOGICAL STUDY OF ORAL DISEASE

by

STEPHANIE STALHBUHK MOMENI

ASIM K. BEJ, COMMITTEE CHAIR
NOEL K. CHILDERS, CO-MENTOR
STEPHEN A. MOSER
R. DOUGLAS WATSON
THANE WIBBELS

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2016

MOLECULAR ANALYSIS OF *STREPTOCOCCUS MUTANS* GENETIC DIVERSITY FOR THE EPIDEMIOLOGICAL STUDY OF ORAL DISEASE

STEPHANIE STALHBUHK MOMENI

BIOLOGY

ABSTRACT

Dental caries is chronic global infectious disease. *Streptococcus mutans* are most frequently associated with dental caries. The goal of this work is to use molecular genotyping methods to investigate the genetic diversity and evidence for transmission of *S. mutans* for the epidemiological study of dental caries. It was first necessary to validate methods proposed in this study including choice of selective media, choice of multilocus sequence typing (MLST) scheme, and the frequency of clonality.

Analysis included 13,906 *S. mutans* isolates obtained from oral samples collected from children and their household members in an 8-year longitudinal epidemiological study. Two genotyping methods were used: Repetitive extragenic palindromic PCR (rep-PCR), a cost effective, rapid gel based typing method, differentiated by short, non-coding repetitive elements interspersed throughout the genome; and MLST, a molecular sequence based approach that differentiates strains based on 8 highly conserved housekeeping genes. Both methods are standardized, reproducible within and between laboratories, with web-based analysis tools for global data sharing.

Results indicated that choice of media and MLST typing scheme were valid. Clonality of *S. mutans* was observed to be a characteristic of this population and a reference rep-PCR library for *S. mutans* was established. While 63% of children shared at least 1 genotype with their mothers, 72% of genotypes were not shared with any household family members. Rep-PCR was reasonably accurate at predicting strains of *S. mutans* requiring additional analysis with MLST. MLST analysis of the diverse genotype (G12) suggests that younger children share strains with their mother. Some strains of S*. mutans* appear to be highly transmissible and probable source individuals were identified.

This work contributes to the understanding of *S. mutans* genotype diversity and its association with dental caries through molecular analysis approaches. These findings offer new insights into the genotypic distribution of strains using rep-PCR and MLST. The importance of child-to-child transmission is clearly demonstrated, although more validation with another method, such as MLST or whole genome sequencing is needed. Continued study of this data offers the potential for new understanding of how and which *S. mutans* strains contribute to the initiation and progression of dental caries.

Keywords: genetic diversity, genotyping, multilocus sequence typing, repetitive extragenic palindromic PCR, transmission

DEDICATION

For my mother, Sherry Jane Light, and my father, Gerd Karl Stahlbuhk, who were and will always be the brightest lights of love, knowledge and support in my life. Thank you for all your encouragement and for reminding me there is magic in thinking bigger.

ACKNOWLEDGMENTS

I wish to give my most sincere and deepest gratitude to the following people who have helped me in my quest for knowledge. Drs. Yihong Li and Page Caufield for offering me my first opportunity to participate in research as part of the Specialized Caries Research Center. Dr. Lisa Ann Blankinship, Dr. Cecilia Cheon, and Dr. Jessica Scoffield for always being an inspiration, an invaluable resource, and a sounding board. Dr. John Ruby whose unwavering support over the years has always guided me forward in both my professional and educational aspirations (and for convincing Dr. Childers to hire me for the Uniontown project). Members of my committee: Dr. Watson, Dr. Wibbles, and Dr. Moser for their time, support, thoughtful questions and insights. Dr. Asim Bej for his commitment to my success in this program, as a teaching role model and his guidance on the path to understanding the field of science. Dr. Noel Childers, for whom words are not enough… who always makes time for me, indulging my questions and meandering tangential thoughts on how to be a scientist and a leader.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table *Page*

INTRODUCTION

MUTANS STREPTOCOCCI ENUMERATION AND GENOTYPE SELECTION USING DIFFERENT BACITRACIN-CONTAINING MEDIA

ASSESSMENT OF TWO MULTILOCUS SEQUENCE TYPING (MLST) SCHEMES AVAILABLE FOR *STREPTOCOCCUS MUTANS*

LIST OF TABLES (continued)

Table *Page*

ASSESSMENT OF CLONALITY AND SEROTYPES OF *STREPTOCOCCUS MUTANS* AMONG CHILDREN BY MULTILOCUS SEQUENCE TYPING

GENETIC DIVERSITY AND EVIDENCE FOR TRANSMISSION OF *STREPTOTOCCUS MUTANS* BY DIVERSILAB REP-PCR

LIST OF TABLES (continued)

Table *Page*

TRANSMISSION PATTERNS OF *STREPTOCOCCUS MUTANS* DEMONSTRATED BY A COMBINED REP-PCR AND MLST APPROACH

LIST OF FIGURES

Figure *Page*

INTRODUCTION

MUTANS STREPTOCOCCI ENUMERATION AND GENOTYPE SELECTION USING DIFFERENT BACITRACIN-CONTAINING MEDIA

ASSESSMENT OF TWO MULTILOCUS SEQUENCE TYPING (MLST) SCHEMES AVAILABLE FOR *STREPTOCOCCUS MUTANS*

ASSESSMENT OF CLONALITY AND SEROTYPES OF *STREPTOCOCCUS MUTANS* AMONG CHILDREN BY MULTILOCUS SEQUENCE TYPING

LIST OF FIGURES (continued)

Figure *Page*

GENETIC DIVERSITY AND EVIDENCE FOR TRANSMISSION OF *STREPTOTOCCUS MUTANS* BY DIVERSILAB REP-PCR

LIST OF ABBREVIATIONS

AP-PCR	arbitrarily primed polymerase chain reaction
bp	base pair
CDC	Centers for Disease Control and Prevention
CH1	Uniontown Cohort 1 (children)
CH2	Uniontown Cohort 2 (infants)
DLV	double-locus variant
DMFS/dmfs	decayed, missing, filled surfaces or teeth (permanent/primary)
dN/dS	non-synonymous/synonymous rate ratio
DS	Do MLST typing scheme
F-C	father-child pair
GG	rep-PCR genotype group
GSTB	glucose—sucrose—tellurite—bacitracin media
GT	rep-PCR genotype
I_A	index of association

KEGG	Kyoto Encyclopedia of Genes and Genomes
M-C	mother-child pair
ME	minimum evolution
MEGA	Molecular Evolutionary Genetics Analysis
mL	milliliters
MLST	multilocus sequence typing
MoI	mole ratio of total DNA bases
MS	mutans streptococci
MSB	mitis salivarius media (with bacitracin and sucrose)
MSKB	mitis salivarius—kanamycin—bacitracin media
ng/µl	nanogram per microliter
NGS	next generation sequencing
NS	Nakano MLST typing Scheme
NSyn	non-synonymous changes
OD	optical density
PCR	polymerase chain reaction
PFGE	pulse-field gel electrophoresis

RAPD	randomly amplified polymorphic DNA
rep-PCR	repetitive extragenic palandromic polymerase chain reaction
RGP	rhamnose-glcuose polysaccharide
SID	Simpson’s index of diversity
SLV	single-locus variant
SmTC	*Streptococcus mutans* total plate count
ST	sequence type
START2	Sequence Type Analysis and Recombinational Tests version 2
Syn	synonymous changes
TH	Todd Hewitt
THB	Todd Hewitt Broth
TPC	total plate count
TYCSB	tryptone—yeast—cysteine—sucrose—bacitracin media
TYS20B	trypticase soy—sucrose—bacitracin media
UPGMA	un-weighted pair group method using arithmetic averages
WGS	whole genome sequencing
WHO	World Health Organization

INTRODUCTION

Dental caries is a pandemic disease affecting not only the oral health but also the overall health of over 80% of the human population worldwide (Bagramian et al., 2009; Edelstein, 2006). Dental caries is a complex, multi-factorial disease, the etiology of which is not fully understood. The disease affects both children and adults of all demographics; although individuals of minority, low socioeconomic status, and those with limited access to dental care are typically at higher risk. Other factors are also important in the development of dental caries including oral hygiene practices, diet, and access to fluoride. The development of dental caries requires four main factors: the presence of teeth, presence of microorganisms, a dietary substrate (fermentable carbohydrates, especially sucrose), and time (Keyes and Jordan, 1963).

The Centers for Disease Control and Prevention (CDC) (www.cdc.gov) lists dental caries as the most common chronic disease of children and adolescents. Furthermore, the CDC reports that 22.9% of US children (age 2-19) have untreated dental caries (Sebelius et al., 2010). The World Health Organization (WHO) indicates that 60-90% of school children have dental caries in industrialized countries and predicts that developing nations will see an increase due to consumption of fermentable carbohydrates and lack of proper fluoridation (Petersen et al., 2005).

Dental caries is typically reported as a caries score based on the number of decayed, missing, or filled teeth/decayed, missing, or filled surfaces (called dmfs and DMFS in primary dentition and permanent dentition, respectively) as established by WHO criteria (WHO, 1997). According to the American Academy of Pediatric Dentistry, early childhood caries (ECC) is defined as any child less than 6 years old with a dmfs score equal to 1 or more. A particularly virulent form of the disease is severe early childhood caries (S-ECC), which is characterized by the presence of extensive decay (≥4 at age 3, ≥5 at age 4, or ≥6 at age 5) in primary teeth (American Acacdemy of Pediatric Dentistry, 2015). Its prevalence is epidemic and ranges from 28% to 82% depending on the population being evaluated (Berkowitz, 2003; Leong et al., 2013). Recent data from the National Health and Nutrition Examination Survey 2011-2012 indicates that children age 2-5 and 6-8 years have dental caries at a prevalence of 23% and 56% respectively (Dye et al., 2015). The study further reported that untreated tooth decay was 14% for children 2-8 years old.

A primary focus in understanding the etiology of dental caries has been the search for which microorganisms involved in the initiation and progression of the disease. The primary pathogenic bacteria associated with the initiation of dental caries are the mutans streptococci, which include *Streptococcus mutans* and *Streptococcus sobrinus* (Gross et al., 2012; Loesche, 1986; Tanner et al., 2002). The presence of these organisms can lead to an ecological shift in the pH balance of the oral cavity favoring demineralization according to the Ecological Plaque Hypothesis (Marsh, 1991). The result is microbial dysbiosis conducive to

an environment selective for these and other highly aciduric organisms. While more recent studies have indicated the role of other microorganisms in the progression of dental caries, the mutans streptococci (in particular *S. mutans*) remain the best early indicator of caries risk (Gross et al., 2012).

Studies have shown that *S. mutans* are highly diverse in their phenotypic characteristics and cariogenic potential (Gilbert et al., 2014; Palmer et al., 2013). However, some key questions remain to be answered. Are there strains of *S. mutans* that are more virulent? Are there strains that convey a protective effect against dental caries? Are some strains more transmissible? How and when do children acquire these strains? To answer these questions, the genetic diversity of *S. mutans* has been evaluated using a number of bacterial typing methods. These include serotyping, biotyping, restriction endonuclease analysis, arbitrarily primed polymerase chain reaction (AP-PCR), ribotyping, randomly amplified polymorphic DNA (RAPD), pulse field gel electrophoresis, and repetitive sequence PCR (e.g. rep-PCR) (Alaluusua et al., 1994; Berkowitz and Jones, 1985; Bratthall, 1970; Davey and Rogers, 1984; Kulkarni et al., 1989; Li and Caufield, 1995; Mitchell et al., 2009; Moser et al., 2010; Perch et al., 1974; Redmo Emanuelsson et al., 2003). The primary purpose of using genotyping methods has been to develop libraries of isolate-specific molecular fingerprints for comparison to determine epidemiological relatedness and surveillance. However, the methods mentioned above are gel-based, which can be labor intensive and, in some cases, difficult to reproduce. It is important to remember that when utilizing gel-based techniques, discriminatory power can vary between

techniques and that gel fractionation provides a limited scope (Maiden, 2006; van Belkum et al., 2007). Furthermore, technical interpretation between laboratories varies greatly depending on the criteria established (e. g. are all bands counted or only major bands?) and reproducibility.

Selective Media for Genotype Comparisons

Before beginning any genotyping study, it is important to determine that the microbiological approaches employed to acquire the bacterial strains for study are optimal. A common critique of such studies is directed at the selective media used for the isolation of the bacteria since by its very nature, selective media is designed to enhance the growth of some microorganisms while inhibiting or preventing altogether the growth of other organisms. This means that even within a species, the recovery of certain strains may be impacted by the selective media used for bacterial isolation. Thus it was important to investigate if the selective media being used in our laboratory was impacting the recovery of genotypes.

The first selective media described for the isolation of the mutans streptococci was Gold's media (also known as MSB), which consist of modified mitis salivarius media with added bacitracin and sucrose (Gold et al., 1973). As this is the first and most often used media it was the selected media used for Uniontown study. Over the years additional media have been introduced including: 1) mitis salivarius—kanamycin—bacitracin (MSKB), 2) glucose—

sucrose—tellurite—bacitracin (GSTB), 3) trypticase soy—sucrose—bacitracin (TYS20B), and 4) tryptone—yeast—cysteine—sucrose—bacitracin (TYCSB), each self reported as having improved and superior recovery of *S. mutans* (Kimmel and Tinanoff, 1991; Schaeken, 1986; Tanzer et al., 1984; Van Palenstein Helderman et al., 1983). Wan *et al.* independently investigated the ability of the five selective media available for *S. mutans* for growth and enumeration, reporting that TYCSB was the most sensitive media for the recovery of *S. mutans* (Wan et al., 2002). However no studies were found investigating the recovery of specific *S. mutans* genotypes based on the selective media used. Thus, **Paper 1** discusses a study designed to evaluate the recovery of mutans streptococci on all five media using both prototype and clinical strains to determine if the genotypes recovered varied by the selective media used.

Repetitive Extragenic Palindromic-PCR

Repetitive extragenic palindromic PCR (rep-PCR) has rapidly evolved in recent years as an effective epidemiological tool for analysis of infectious outbreaks of clinically relevant bacteria. For this method, primers are designed to amplify non-coding, highly conserved repetitive elements, typically 30-38 base pairs in length that are interspersed throughout the bacterial genome. While the function of these repetitive elements remains undefined, changes in the larger coding regions between these repetitive elements can be visually compared by using gel

fractionation. DiversiLab™ (bioMérieux, Durham, NC) has standardized this method to improve reproducibility between laboratories and provides web-based software for global analysis and sharing of data (Healy et al., 2004). Analysis of *S. mutans* by this method was originally performed by our laboratory and is currently in use for large-scale epidemiological longitudinal study of *S. mutans* in a high-caries risk population located in Uniontown, AL, USA (Cheon et al., 2011; Moser et al., 2010). This study population is located in a rural, relatively stable community (i.e., families rarely move away). *S. mutans* isolates were collected from African American infants and children as well as their residential household family members over an 8 year period. In total, over 32,000 isolates were collected and 13,906 were analyzed by rep-PCR analysis.

The benefit of using rep-PCR as compared to other available bacterial typing methods is that this approach is standardized, rapid and cost effective, making it very useful for high-throughput epidemiological studies. However, it is recommended that all genotyping methods use alternate methods to assist in determining genotypes and confirm linage assignments (Foley et al., 2006; van Belkum et al., 2007). Therefore, in addition to rep-PCR, analysis with a molecular based typing method for the characterization of *S. mutans* was needed since nucleotide sequences provide greater discriminatory power, reproducibility, and portability for analysis and comparison with other laboratories through internet-based technologies (Maiden, 2006).

Multilocus Sequence Typing

Multilocus sequence typing (MLST) is a sequence based molecular typing method, which uses PCR to amplify highly conserved constitutive (housekeeping) genes. Since this method is sequence based, it provides an effective means for global evolutionary study and epidemiological surveillance of microorganisms in a clinical or environmental setting. A well-designed MLST scheme will have loci selected based on nucleotide variation at a frequency capable of providing sufficient discriminatory power to evaluate phylogenetic or epidemiological concerns. This method is highly reproducible and results can be compared easily between laboratories for global surveillance through registration with the Oral Streptococcus PubMLST database (www.pubmlst.org) (Maiden, 2006).

With MLST, typically PCR is performed on 6-10 housekeeping genes. Resulting forward and reverse amplicons (400-600 base pairs in length) are sequenced and the data is compared to a control reference strain. Any mutations occurring in the isolate's sequence for a given gene results in the assignment of a new allele number for that gene. This generates a 6-10 digit sequence type (ST) ID number. Comparison of bacterial isolates may then be performed in one of two ways, based on either the allelic designations of the ST or on the concatenated nucleotide sequences depending on the population structure (clonal vs. non-clonal) (Maiden, 2006).

Two MLST schemes are currently available for *S. mutans*. The first scheme was introduced by Nakano (Nakano et al., 2007a) in an evolutionary study and consisted of 8 genes. Another scheme was later introduced by Do (Do et al., 2010) using 6 constitutive genes and 2 punitive virulence-associated genes. It is suggested that when multiple MLST schemes exist for a single bacteria, if the schemes are well designed, then the results should be compatible. However, one of the concerns in using MLST is determining which genes to include in the MLST scheme so that the scheme provides sufficient discriminatory power (Maiden, 2006). For this reason, **Paper 2** compares the two available schemes to determine which scheme may be best for the investigation of *S. mutans* genetic diversity and transmission in the Uniontown Study.

Other studies using MLST have reported that *S. mutans* ST are highly diverse, which would seem to indicate that *S. mutans* are not clonal by MLST analysis. (Do et al., 2010; Nakano et al., 2007a). However, a previous study conducted in our laboratory indicated clonality of *S. mutans* ST among a small number of isolates in the Uniontown Study (Momeni et al., 2013). Furthermore, the previous studies of others found no relationships supporting the geographical distributions of ST (i.e., ST are not localized), while our study indicated that many of the Uniontown ST were unique only to this population and that the same ST could occurred in unrelated individuals. Since the other studies evaluated *S. mutans* strains from broader populations with subjects having no known interactions, is it possible that the results observed in the Uniontown study were

because of sample selection bias (pre-selected using rep-PCR genotype) or is clonality of *S. mutans* sequence types a result of using a localized, interactive population? **Paper 3** investigates the clonality of 100 randomly selected *S. mutans* isolates and their geographic distribution as compared to other MLST studies.

Serotyping

An earlier form of bacterial typing included a method known as serotyping, which characterized *S. mutans* bacteria based on the antigenic properties of cell wall rhamnose-glcuose polysaccharides (RGP). The serotypes of *S. mutans* differ in the arrangements of the alternating glucose side chains on the rhamnose backbones resulting in 3 clinically relevant *S. mutans* serotypes: serotype *c* (α 1,2-linkage), serotype *e* (β 1,2-linkage), and serotype *f* (α 1,3- linkage). *S. sobrinus* was assigned serotypes *d* and *g (Bratthall, 1970; Perch et al., 1974)*. In 2004, a fourth *S. mutans* serotype was discovered in Japan, serotype *k* (no glucose side chains) *(Nakano et al., 2004a)*. The prevalence of the most commonly reported serotypes, serotypes *c* and *e,* are respectively about 70-80% and 20-30% of all *S. mutans* isolates recovered from the oral cavity (Hamada and Slade, 1980; Hirasawa and Takada, 2003). Serotypes *f* and *k* are considered minor oral serotypes as these strains are recovered in only about 5% of isolates (Nakano et al., 2004b). Traditionally, determination of *S. mutans* serotypes was performed by an immune-diffusion method. However advances in

molecular based PCR techniques and sequencing of the RPG operon lead to a more cost effective multiplex PCR based approach for serotypes *c, e, f* and then later a conventional PCR approach for serotype *k (Nakano et al., 2004b; Shibata et al., 2003)*.

In tandem with serotype data, two collagen-binding proteins, Cnm and Cbm encoded by genes *cnm* and *cbm* repectively, are often investigated (Nomura et al., 2012; Sato et al., 2004). These proteins, particularly when associated with the minor *S. mutans* oral serotypes *f* and *k,* have been reported to invade human endothelial cells, avoid phagocytosis, and to interfere with platelet aggregation linked to bacteremias, infective endocarditis, hemorrhagic stroke, as well as other cardiovascular and inflammatory conditions (Kojima et al., 2014; Lapirattanakul et al., 2011; Lapirattanakul et al., 2013; Nakano et al., 2006; Nakano et al., 2007c; Nakano et al., 2008; Nakano et al., 2009; Nakano et al., 2010; Nakano et al., 2011; Nomura et al., 2013). Current literature suggests that serotypes *f* and *k* often have defects in the protein antigen (PA) gene preventing these strains from binding to teeth thus accounting for their classification as minor oral serotypes (Lapirattanakul et al., 2015; Nakano et al., 2008).

Previous studies using MLST have included serotype and collagen-binding protein data so inclusion of this data within Paper 3 is appropriate to determine the prevalence of these serotypes and collagen-binding proteins within the Uniontown population.

Genetic Diversity

As mentioned earlier, many genotypic studies of *S. mutans* have been performed using a variety of approaches. In those studies genotypic diversity is often compared with age, race, sex, tooth eruption and caries status of the children. Li and Caufield reported that female children were more likely to share genotypes with their mothers; however male children who acquired *S. mutans* strains from their mother were 13 times more likely to develop caries (Li and Caufield, 1995). Some studies have shown no significant correlation between genotype diversity (i.e., number of genotypes) and caries using AP-PCR (Kreulen et al., 1997; Lembo et al., 2007; Liu et al., 2007; Mattos-Graner et al., 2001) while other have reported significance using multiple methods including AP-PCR, MLEE, and ribotyping (Alaluusua and Matto, 1996; Gamboa et al., 2010; Napimoga et al., 2004; Pieralisi et al., 2010). Mattos-Graner *et al.* looked at genotype diversity and age but found no correlation (Mattos-Graner et al., 2001). Since our laboratory is the first to evaluate *S. mutans* genotypic diversity using rep-PCR, it is logical to consider how the genotypic diversity of the Uniontown population compares by sex, age, and caries status for comparison with other previously reported methodologies.

Transmission of *S. mutans*

Publications focused on the genetic diversity of *S. mutans* are often accompanied by a discussion on the transmission of strain types. The prevailing theory in the literature is that mother-to-child transmission (sometimes called vertical transmission) is the primary route of infection (Berkowitz and Jones, 1985; Carletto-Korber et al., 2010; Davey and Rogers, 1984; Douglass et al., 2008; Klein et al., 2004; Lapirattanakul et al., 2008; Li and Caufield, 1995; Lynch et al., 2015; Teanpaisan et al., 2012). However, studies have shown strains can be shared with fathers and siblings (interfamilial transmission) (Douglass et al., 2008; Kozai et al., 1999; Redmo Emanuelsson and Wang, 1998; Tedjosasongko and Kozai, 2002). Others have evaluated child-to-child transmission (often called horizontal or extra-familial transmission) by demonstrating the commonality of strains shared among children suggesting that other children may also be potential reservoirs for transmission. (Alves et al., 2009; Baca et al., 2012; Domejean et al., 2010; Liu et al., 2007; Mattos-Graner et al., 2001; Tedjosasongko and Kozai, 2002). However, current knowledge is incomplete since studies evaluating all sources of infection (mother-to-child, interfamilial transmission, and child-to-child transmission) in a cohesive study are limited, and those available typically have small sample sizes (20 or fewer families). A key limitation of some studies of genetic diversity is that the children are often recruited at a clinic or hospital meaning the sample population frequently have no interaction and are therefore, less likely to have strains in common. This sampling approach has a confounding effect on the results, with a bias toward

mother-to-child transmission and limiting the likelihood of observing horizontal transmission. In addition, many studies of interfamilial strains typically look only at a nuclear family (mother, father, 2 or less siblings) and not at larger or extended households whose individuals can have regular interactions with the children. In contrast to previous publications, many of the households in the Uniontown study include grandparents, aunts, uncles, and cousins. **Paper 4** evaluates the genetic diversity and transmission of *S. mutans* using rep-PCR within an isolated, interactive population (Uniontown) with the goal of providing insights into the sharing of genotypes at a level not previously evaluated.

Rep-PCR and MLST for Epidemiological Study

While the cost of molecular based typing methods, such as MLST or whole genome sequencing (WGS) have been decreasing in recent years, the cost of doing such application remains significantly higher than using a gel-based typing method such as rep-PCR. When investigating strain type diversity, investigators are faced with the dilemma of selecting strains that will provide sufficient diversity at reasonable costs without biasing the findings. The typical approach has been to randomly analyze bacterial isolates using these methods only to find that often, many strains provide similar information. In a study such as the Uniontown study where commonality of strains is more frequently observed, it would be more logical to pre-screen isolates for costly downstream applications in order to optimize genetic diversity of the results.

A method for using MLST and rep-PCR together for large-scale epidemiological study of *S. mutans* has been suggested but not tested (Momeni et al., 2013). Using an arbitrary cut-off of 98% on a rep-PCR percent similarity dendrogram, analysis of the isolates corresponding to the 2 most extreme branches and further analysis of paired internal branches for each rep-PCR genotype may be used to predict which isolates will result in matching or non-matching MLST ST. Such an approach could possibly be expanded to ensure that diverse isolates are selected for analysis with other methods such as WGS.

Another application of rep-PCR pre-screening may be to use a combined rep-PCR and MLST approach to address specific clinical questions. For instance, MLST is often considered a more discriminative typing method than rep-PCR since a single base change can result in a new ST. In a previous study, it was observed that MLST ST within a group of isolates of the same genotype often clustered together on the rep-PCR dendrograms (Momeni et al., 2013). Based on these preliminary data, is it possible to use a rep-PCR dendrogram to predict the MLST ST within a rep-PCR genotype, thus further discriminating rep-PCR genotype? Such an approach would be helpful in transmission studies especially given the commonality of many of the rep-PCR genotypes in the Uniontown study.

Paper 5 assesses the rep-PCR predictor approach as an initial screening tool, with the aim that it may be possible to limit the number of isolates requiring analysis with MLST or WGS, thereby improving throughput and reducing costs in large-scale epidemiological genotyping studies. It further evaluates the use of

this approach in a clinical application for a transmission study of a specific rep-PCR genotype (G-12).

Summary

The first aim of the following papers is to address three of the major concerns affecting genotyping with rep-PCR and MLST: choice of selective media; choice of MLST scheme; clonality and geographic distribution. The second aim is to report on the genetic diversity of *S. mutans* and its transmission within a localized population using rep-PCR and MLST. Through combined analysis with rep-PCR and MLST it may be possible to expand our understanding of the role of specific *S. mutans* genotypes in the epidemiology of dental caries and how these genotypes are shared within a localized, interactive population.

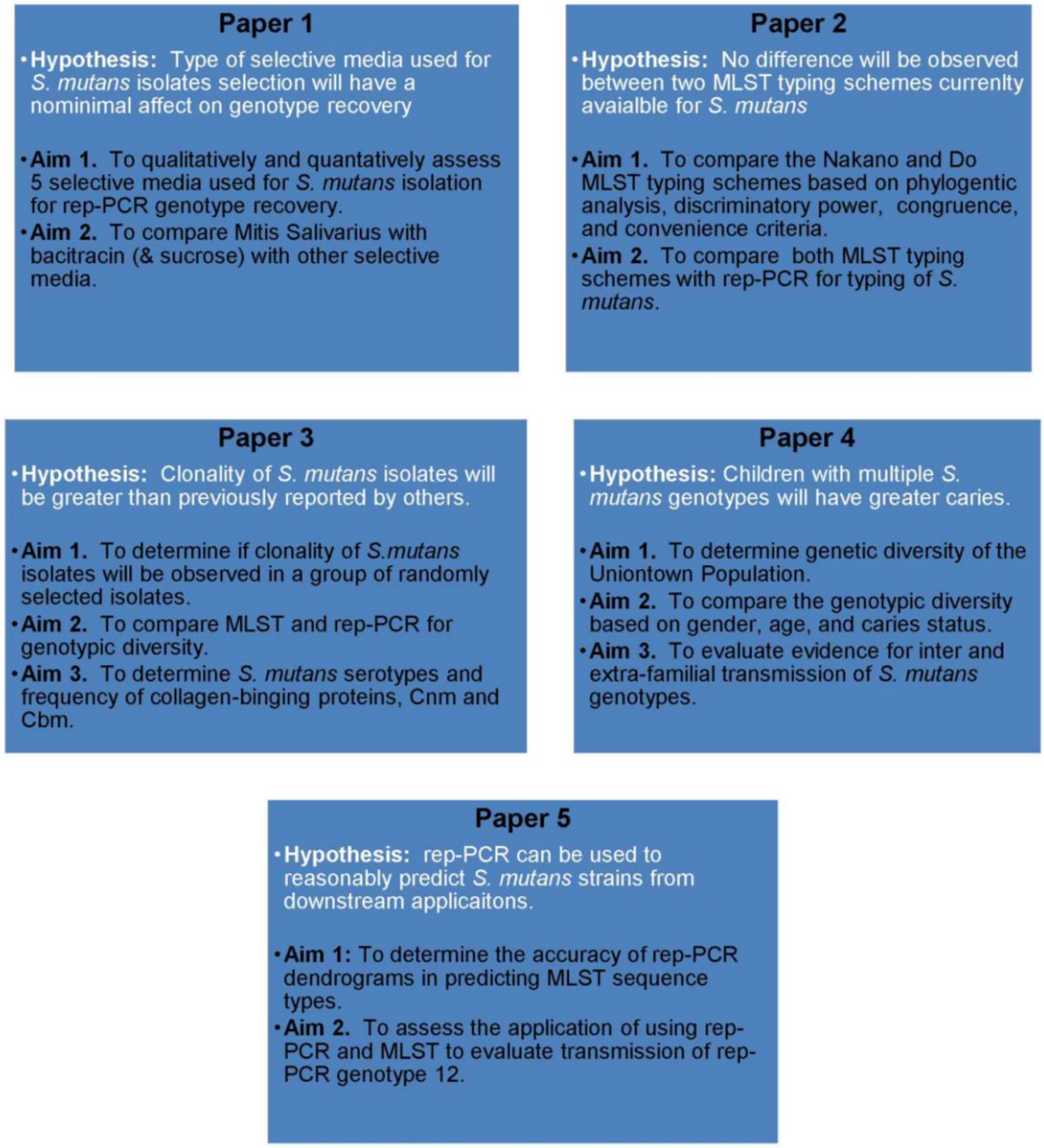


Figure 1. Flowchart of papers presented with hypotheses and specific aims.

MUTANS STREPTOCOCCI ENUMERATION AND GENOTYPE SELECTION USING DIFFERENT BACITRACIN-CONTAINING MEDIA

by

STEPHANIE S. MOMENI, PAUL PATRICK, HOWARD W. WIENER, GARY R. CUTTER, JOHN D. RUBY, KYOUNGA CHEON, JENNIFER WHIDDON, STEPHEN A. MOSER, NOEL K. CHILDERS

Journal of Microbiological Methods Aug;103:53-57

Abstract

The primary etiological agents associated with dental caries include the mutans streptococci (MS) comprised of *Streptococcus mutans* and *Streptococcus sobrinus.* The effective cultivation and isolation of MS is necessary for the study of MS, including their proper clinical assessment in the epidemiological study of dental caries. Several selective media have been developed for the isolation, enumeration, and characterization of MS. However, inhibition of MS may occur, reducing counts and perhaps limiting selection of some strains. The purpose of this study was to compare five culture media containing bacitracin recommended for the isolation of MS.

Five commonly used bacitracin-containing media (MSB, MSKB, GTSB, TYS20B, and TYCSB) used for MS isolation were quantitatively evaluated. Standard plate counts were performed in duplicate for 2 prototype MS strains (*S. mutans* UA159 and *S. sobrinus* 6715) and for MS isolates from clinical saliva samples obtained from 16 children (approximate age 5 years) to determine total plate counts, and total *S. mutans* count. Selected isolates (n=249) from all of the five media from 5 saliva samples were further confirmed as *S. mutans* with real-time PCR then subsequently evaluated qualitatively with rep-PCR for genotype determination.

All media resulted in variable enumeration with no significant difference in

MS counts. MS prototype strains grew well on all five media; clinical isolates demonstrated more variability in counts but no overall significant differences were found. MSB demonstrated comparable ability to grow *S. mutans* but allowed for more non-*S.mutans* growth. All 5 media identified a consistent predominant genotype by rep-PCR. Recovery of minor genotypes was not inhibited by media type.

1. INTRODUCTION

The mutans streptococci (MS) are primary etiological agents associated with the initiation of dental caries (Loesche, 1986). *Streptococcus mutans* and *Streptococcus sobrinus* are the two clinically relevant species of MS found in the oral cavity. These two species are typically identified and quantified in caries epidemiological and early intervention studies (Saravia et al., 2013).

Numerous selective media have been introduced for the isolation, quantitation, and characterization of MS. Typically these media are supplemented with bacitracin and sucrose as selective agents for clinically relevant streptococci (Gold et al., 1973, Hildebrandt and Bretz, 2006). However, unintentional inhibition of MS may occur, reducing counts and perhaps limiting the quantity of detection and isolation of some strains (Hildebrandt and Bretz, 2006, Tanzer et al., 1984). Initially, mitis salivarius with bacitracin (MSB) medium was considered a reliable selective media for isolation of MS (Gold, Jordan and Van Houte, 1973). As such, MSB medium was selected as the primary selective medium for a large-scale epidemiological study of *S. mutans* by our laboratory. However, several studies have suggested that MSB medium results in lower MS counts than other available selective media and can possibly result in false-negatives (Saravia, Nelson-Filho, Silva, De Rossi, Faria, Silva and

Emilson, 2013, Schaeken, 1986, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman et al., 1983, Wan et al., 2002).

In this study, five selective media are evaluated: mitis salivarius-bacitracin (MSB), mitis salivarius-kanamycin-bacitracin (MSKB), glucose-sucrose-tellurite-bacitracin (GSTB), trypticase soy-sucrose-bacitracin (TYS20B), and tryptone-yeast-cysteine-sucrose-bacitracin (TYCSB) (Gold, Jordan and Van Houte, 1973, Kimmel and Tinanoff, 1991, Schaeken, 1986, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman, Ijsseldijk and Huis in 't Veld, 1983). Detailed descriptions of these media can be found in the literature, but briefly, these media typically are supplemented with bacitracin and high concentrations of sucrose that act as selective agents for clinically relevant MS. Sucrose allows MS to be distinguished while inhibiting other organisms and bacitracin is known to inhibit other oral streptococci (Gold, Jordan and Van Houte, 1973, Hildebrandt and Bretz, 2006, Schaeken, 1986). Mitis salivarius media (in MSB, MSKB) is commonly used as a base medium since it is selective for streptococci and because MS colony morphology is easily distinguished on this media (Gold, Jordan and Van Houte, 1973). Sucrose is replaced by sorbitol in MSKB and the addition of kanamycin was incorporated to enhance the activity of bacitracin (Kimmel and Tinanoff, 1991).

The purpose of this study is to evaluate these five media quantitatively and qualitatively using MS prototype strains and clinical isolates. In addition, evaluation of MSB in comparison to the other four media will be assessed.

2. MATERIAL AND METHODS

2.1. MS Prototype Strains

S. mutans UA159 and *S. sobrinus* 6715 were selected as the MS prototype strains.

2.2. S. mutans Clinical Strains

For clinical isolates, saliva was collected from pre-school children (n=16, approximate age 5 years) as part of an ongoing longitudinal epidemiological study of a high caries risk population located in Uniontown, AL, a small rural city. Sample collections were performed at a community health center during a scheduled follow-up visit. Inclusion criteria were that children were free of systemic disease. Human use approval was obtained from the University of Alabama at Birmingham (UAB) Institutional Review Board, with parents of participants providing informed consent, while children gave assent. Whole saliva samples were collected using sterile cotton swabs, which were then stored in 4.5 ml of reduced transport fluid (Syed and Loesche, 1972). Samples were shipped on ice to our laboratory at the University of Alabama at Birmingham School of Dentistry and stored at 4^{o}C until processed (within 24 h of collection).

2.3. Media & Plating

All five media were prepared as originally describe in original publications (Gold, Jordan and Van Houte, 1973, Kimmel and Tinanoff, 1991, Schaeken, 1986,

Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman, Ijsseldijk and Huis in 't Veld, 1983).

Prototype strains were inoculated from -80^{o}C frozen cultures into 5 ml Todd Hewitt Broth (THB) (Becton Dickinson, Sparks, MD, USA) and incubated anaerobically (10% CO_2, 10% H_2, and 80% N_2) at 37^{o}C for 24 h. A 500 µl aliquot of fresh culture was transferred to 4.5 ml of THB and grown to late log phase. Optical density was determined at 600 nm (OD_{600}) using a Bio-Rad SmartSpec Plus spectrophotometer (Bio-Rad, Hercules, CA, USA), then adjusted to an OD_{600} of approximately 1.0 with THB, which corresponds to 1 x 10^9 cells/ml using electronic enumeration (Childers et al., 2011). Bacterial samples were serially diluted to 1 x 10^{-5} and plated in duplicate on each of the five media using a spiral plater (Spiral Systems Inc., Cincinnati, Ohio, USA) dispensing a 50 µl volume. This process was repeated in four independent tests.

Clinical saliva samples in reduced transport fluid were vortexed for 20 s on medium speed prior to plating. Undiluted samples were plated in duplicate on each of the five media using the spiral plater. All plates were incubated anaerobically at 37^{o}C for 48 h.

2.4. Enumeration

After incubation, all plates were counted using a stereomicroscope for total plate counts (TPC) and total *S. mutans* counts (SmTC) by manual counting according to the instructions for the spiral plater (Spiral Systems Inc., Cincinnati, Ohio, USA) to enumerate the colony-forming units per ml (CFUs/ml). Briefly, a

quadrant sector-count method was used to enumerate bacteria using a minimum count of 20 colonies within a quadrant. Opposite quadrants in the same sector were counted and the bacterial density was found by adding the counts of the two sectors and dividing by the sample volume corresponding to the innermost sector counted.

Plates were randomly selected to perform counts without knowledge of counts on other media. The mean plate counts were calculated using duplicate plates. For the clinical isolates, SmTC were recorded based on colony morphology of the prototype strains (Fig. S1) (McGhee et al., 1982, Schaeken, 1986, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman, Ijsseldijk and Huis in 't Veld, 1983).

2.5. Genotype Evaluation

For clinical samples that demonstrated sufficient *S. mutans* growth on all five selected media (n = 5), 10 presumptive *S. mutans* colonies were selected for genotype identification using repetitive extragenic palindromic polymerase chain reaction (rep-PCR) as described previously (Moser et al., 2010). Briefly, colonies were inoculated in 5 ml THB and grown anaerobically at 37^{o}C for 48 hours. DNA was extracted and checked for purity. All isolates were verified as *S. mutans* using SYBR Green^{TM} PCR with sequence specific Yoshida primers before performing rep-PCR (Yoshida et al., 2003). Genotypes were determined using the DiversiLab^{TM} (DL) system and cross-referenced with the library of known genotypes from the larger epidemiological study (Cheon et al., 2013).

2.6. Statistical Analysis

Raw counts were transformed to approximate normality by taking logarithms of counts (count plus one, i.e. for counts of zero). Ratios of counts of *S. mutans* to total counts were based on raw counts, not on transformed counts. Statistical significance was ascertained using the GLM procedure in SAS version 9.2 using a significance cut-off value of $p < 0.05$.

3. RESULTS

3.1. MS Prototype Strains

S. mutans UA159 and *S. sobrinus* 6715 grew equally well on all five media with no significant differences (Tables 1 & S1). For *S. mutans* UA159, the mean log recovery was marginally better on TYS20B (9.18 ± 0.10) than on the other media; MSB (9.16 ± 0.14), GSTB (9.08 ± 0.19), TYCSB (9.09 ± 0.12), and MSKB (9.05 ± 0.16) (p = 0.69). In terms of frequency, the highest counts were most often obtained on MSB than other media (2/4 replicates). When MSB was compared to the other media the difference was also non-significant (p = 0.44).

S. sobrinus 6715 resulted in comparable counts on all five media although at a slightly decreased level as compared to *S. mutans* UA159. The highest counts were most often observed on TYS20B (8.75 ± 0.41) than other media (2/4); MSB (8.79 ± 0.30), TYCSB (8.67 ± 0.29), MSKB (8.63 ± 0.25), and GSTB (8.64 ± 0.35) but not significantly different (p = 0.94) when all five media were

compared. In addition, no significant difference was found when comparing MSB with all other media (p = 0.51).

3.2. *S. mutans* Clinical Strains

Clinical saliva samples from 16 children were processed in order to compare clinical TPC and SmTC. It is characteristic of this study population that *S. sobrinus* is rarely detected either by standard plate count or quantitative PCR as part of the larger ongoing study. Thus, since *S. sobrinus* was not detected, only SmTC are reported in this study.

Four clinical samples were not included in the final analysis since growth was either not observed (n = 1) or minimally observed on only one media (n = 3 for GSTB <2 colonies). The Log mean plate counts observed for TPC and SmTC are reported in Tables 1 and S2, S3 respectively. TPC are reported for 12 clinical samples but SmTC are reported for only 9 clinical samples (i.e., protocol omission in counting colonies for SmTC). The highest TPC were observed on TYS20B (7/16) followed by MSB (4/16). ANOVA analysis of the log TPC for all five media yielded no statistically significant difference (p = 0.14). When MSB was compared to the other media no significant difference was observed (p = 0.32).

Concentrations of *S. mutans,* reported as SmTC, were most frequently highest on GSTB (6/9) compared with all other media types. It was observed on MSB that 22% (2/9) of the samples did not have detectable *S. mutans* when growth was observed on other media. ANOVA analysis of the log SmTC also

resulted in no significant difference between the five media for recovery of clinical *S. mutans* (p = 0.86). The MSB comparison was also found to be insignificant (p = 0.59).

The percentage of *S. mutans* isolated to TPC was used as a measure of the specificity of each media (Tables 1 and S4). The ratio was highest for TYCSB (72% ± 26%), followed by MSKB (70% ± 35%) and GSTB (52% ± 38%). TYS20B (41% ± 37%) and MSB (29% ± 31%) had the lowest percentage of *S. mutans* colonies. Overall, TYCSB and MSKB (both 4/9) most frequently reported the highest percentage of *S. mutans* specificity followed by GSTB (3/9) and MSB (1/9). The difference in the ratio of *S. mutans* to TPC was not significant when comparing all five media (p = 0.07). However, when comparing MSB to the other media the difference is significant (p = 0.03).

The highest counts of non-S. *mutans* were most frequently observed on MSB (5/9), TYS20B (3/9), and MSKB (1/9).

3.3. Rep-PCR Genotypes

All isolates (n = 250) except one (from ID 571, GSTB) were confirmed to be *S. mutans* using SYBR Green™ PCR with Yoshida primers (Yoshida, Suzuki, Nakano, Kawada, Oho and Koga, 2003). Predominate genotypes were defined as “more than 50%” of the isolates being of one genotype (Cheon, Moser, Wiener, Whiddon, Momeni, Ruby, Cutter and Childers, 2013) and were found to be consistent for all samples regardless of media used (Table 2). At least one minor genotype was observed for four of the five clinical samples and one

subject had 2 minor genotypes (ID 561). Minor genotypes were found most frequently with TYCSB (5/5) followed by TYS20B (4/5), MSKB and GSTB (both 3/5). MSB reported minor genotypes less frequently (2/5).

4. DISCUSSION

Direct comparison with other published data is particularly difficult since most media comparison studies do not provide raw data but rather the summary of statistical data and these statistical approaches vary considerably by study. In addition, the selection of samples for analysis also varies by study. For instance, some studies use the non-human *S. mutans* isolates *S. ratti* (serotype a) and *S. criceti* (serotype b) which potentially skew findings when analyzed collectively with the clinically relevant serotypes (Gold, Jordan and Van Houte, 1973, Hildebrandt and Bretz, 2006, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman, Ijsseldijk and Huis in 't Veld, 1983). This is particularly important in the case of MSB since growth of serotype a is not supported.

Some studies provide data only on clinical isolates without clinical controls which is a concern since other factors such as variability of ingredient quality of the culture media, methodology (plating method, incubation conditions), and technique (plate counting, plate randomization) can affect final counts (Dasanayake et al., 1995, Sanchez-Perez and Acosta-Gio, 2001, Schaeken, 1986). The use of prototype strains provides a baseline for comparison that is

essential when so many variables can affect the outcome. In addition, the method for validation of isolates of *S. mutans* was typically biochemical, which is accepted practice but not infallible, particularly in the case of clinical isolates (Petti et al., 2005, Ruoff et al., 1982). It should be noted that none of the studies reviewed reported on accuracy of colony selection or if counts were adjusted to account for isolates that failed confirmation testing. All isolates in the present study were confirmed by PCR using *S. mutans* specific primers instead of biochemical testing.

4.1. MS Prototype Strains

All five media were comparable for enumeration of prototype MS strains *S. mutans* UA159 and *S. sobrinus* 6715 ($p = 0.69$ and 0.94, respectively). When MSB is compared with all other media for the prototype strains no statistically significant difference was observed ($p = 0.51$ and 0.44, respectively). The observed decrease counts for *S. sobrinus* may be due to a difference in cell densities between *S. mutans* and *S. sobrinus,* possibly due to agglutination. Both samples were grown to an $OD_{600} \approx 1.0$, which for *S. mutans* equals approximately 1.0×10^9 cells using electronic enumeration (Childers, Osgood, Hsu, Manmontri, Momeni, Mahtani, Cutter and Ruby, 2011). However, the cell density for *S. sobrinus* has not been determined using electronic enumeration. Selectivity of the media used is not suspected to be the cause of the decreased counts, since lower counts were observed on all the media evaluated.

The finding that *S. sobrinus* grows equally well is noteworthy since some literature suggests that recovery of *S. sobrinus* (serotype d) is inhibited on MSB media (Schaeken, 1986, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984). However, a review of the literature indicates these statements are broad, confusing MSB with another media containing sulfasoxazole. The media containing sulfasoxazole inhibits growth of serotype d (Little et al., 1977). MSB media does not contain sulfasoxazole and will support the growth of serotype d. Several studies present data that supports the current finding that *S. sobrinus* grows well on MSB (Gold, Jordan and Van Houte, 1973, Kimmel and Tinanoff, 1991, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984). It should also be noted that *S. sobrinus* is not reported in the present study for clinical samples because the Uniontown study sample population rarely resulted in recovery of *S. sobrinus* (<0.003%). Although MSB is the selective media used in the Uniontown study, a very rare recovery of *S. sobrinus* in this population has been confirmed using whole saliva samples analyzed by quantitative PCR, a more sensitive detection method, in addition to standard plate counting (unpublished).

4.2. *S. mutans* Clinical Strains

Mean log counts were variable for clinical samples on all five media but variation was not statistically significant indicating all five media may be comparable for enumeration of SmTC for clinical isolates. The highest level of non-*S. mutans* were reported most frequently on MSB (5/9) and TYS20B (3/9), which may indicate these media are less specific for selection of *S. mutans*. It appears that

MSB is as good at supporting *S. mutans* growth as any of the other media tested, but not quite as good at suppressing non-*S. mutans*. This may be an issue if confluent background organisms limit resources available or produce byproducts that may inhibit *S. mutans* growth. In addition, MSB demonstrated potential to report false negatives (2/9 samples). However, it is noteworthy that in these cases the SmTC for other the media were remarkable low.

In evaluating the sensitivity of each media using *S. mutans* to TPC ratios, it was observed that differences were approaching significance ($p = 0.07$) (Table 1 and S4). This value is slightly higher than the *p*-value reported by Wan ($p < 0.01$) (Wan, Seow, Walsh and Bird, 2002). However, when MSB was compared to the other media, the MSB effect was negative and there was a significant difference ($p = 0.03$) indicating that MSB is less effective at distinguishing *S. mutans* from other oral streptococci (i. e. less selective). However, in practice, when growth of non-*S. mutans* is observed on MSB, is not a major issue in that the morphology of MS colonies is distinctive on MSB, and can be successfully managed by confirming selected colonies by SYBR Green™ PCR or biochemical analysis.

4.3. Rep-PCR Genotypes

This is the first study to evaluate recovered MS strains using rep-PCR to evaluate genotype strains recovered. The use of rep-PCR adds a qualitative aspect to this study, the intent of which was to determine whether inhibition/selection of certain genotypes occurs with different media. Using rep-

PCR it was found that the same predominate genotype was observed on all five media for each of the five clinical strains isolates (n = 249) selected for genotyping. This suggests that isolation of a predominate genotype would be successful on any of these media. No evidence of a selection bias for any of the media tested was observed for predominate genotypes.

Isolation of minor genotypes occurred less frequently or not at all depending on media used. For instance, the second most common genotype observed was recovered on at least 4 of the five media tested in 3 of the 4 cases. It is possible that the recovery of minor genotypes occur in less frequency due to limited number of isolates selected (10 colonies per plate) or selection bias. Growth for all 7 genotypes reported in this study was observed with all five media when subsequently plated (data not shown). These findings indicate that the failure to recover some genotypes on some media in the present study was not due to differential selectivity of the media.

It should be noted, that while selection of isolates from MSB resulted in the lowest recovery of minor genotypes in this study, this was not due to inhibition of these genotypes on MSB media. In addition to the finding that all 7 genotypes grew on all 5 media, all of the reported minor genotypes in the present study have been detected from samples of the same subjects in the larger Uniontown epidemiological study using MSB. This suggests that the differential observed in the recovery of minor genotypes on MSB maybe due to the probability of selection of rare genotypes or the limited number of isolates selected (n = 10). The selection of 10 isolates is based on earlier analysis that

determined that 10 isolates had a 95% probability of identifying up to 3 genotypes (Cheon et al., 2011). It is possible that either of these reasons may account for the variable recovery of minor genotypes for the other media as well. In addition, there appears to be no correlation between the SmTC of the five media and the minor genotypes observed indicating that the bias is not likely attributed to lower overall SmTC. Since the predominate genotype recovered was consistent regardless of media used, this would seem to suggest that variability of minor genotypes recovered is not due to media selectivity (Cheon, Moser, Wiener, Whiddon, Momeni, Ruby, Cutter and Childers, 2013). The results found with MSB as compared to the larger Uniontown study further support this possibility. However, sample size was small for this study and further study is required to determine if recovery of minor genotypes is media dependent or a result of sampling efficiency.

A review of the literature evaluating the selective media available for the isolation of MS suggests there are variations in colony counts depending on the media employed and the results of which media is best vary between different studies. None of the media was observed to result in statistically significant recovery of MS while simultaneously inhibiting the growth of background organisms. It is suggested for accurate quantitation of MS in epidemiological studies and caries assessments that multiple media should be used if traditional culture methods are to be employed (Hildebrandt and Bretz, 2006). Alternately, traditional culture methodology may be supplemented or replaced with molecular

testing methods such as quantitative PCR (Childers, Osgood, Hsu, Manmontri, Momeni, Mahtani, Cutter and Ruby, 2011, Hildebrandt and Bretz, 2006).

4.4. Limitations

It is possible that the SmTC may under represent the actual number of *S. mutans* since these counts were based on UA159 prototype strain morphology. However, confirmation of *S. mutans* was performed using SYBR Green™ PCR for selected isolates with 249 of 250 isolates being correctly identified representing a 99.6% accuracy rate in selection of *S. mutans* colonies based on morphology alone. Another possible limitation is the use of PCR to confirm identify *S. mutans,* which is subject to accuracy of the primer selection. A BLAST search of the primers used in this study was performed (October 2013) and were specific for only *S. mutans* at that time.

The robustness of the clinical data reported in this study may have been improved by performing replicates of plated samples in triplicate instead of duplicate. However several studies have used the approach of duplicate plating (Gold, Jordan and Van Houte, 1973, Tanzer, Borjesson, Laskowski, Kurasz and Testa, 1984, Van Palenstein Helderman, Ijsseldijk and Huis in 't Veld, 1983). Duplicate plating was selected for this study because it is the standard method by which all clinical samples are currently processed in the larger longitudinal study in Uniontown. Additionally, the clinical sample size of the current study is small but comparable to another published study (Wan, Seow, Walsh and Bird, 2002).

5. Conclusions

In this study, the five selective media currently available for the enumeration of mutans streptococci media resulted in variable enumeration with no statistically significant differences. Quantitative analysis by standard plate count using both prototype control strains and clinical strains grew comparably on all five media. MSB, when compared to other media, demonstrated comparable ability to grow but not select for MS; although it is noteworthy that colony selection based on morphology remained an effective method for isolates/genotype selection. Qualitative analysis by rep-PCR determined that the predominate genotype grew equally well on all media tested. Minor genotypes varied in recovery rates but this effect may be due to selection probability of rare genotypes.

Additional supporting information may be found in the online version of this article: at http://dx.doi.org/10.1016/j.mimet.2014.05.010.

ACKNOWLEDGEMENTS

The authors wish to thank all the clinical and laboratory participants of this study: Ms. Stephanie McLean, Dr. Stephen Mitchell, Dr. Sonia Makhija, Dr. Rosalyn Bassett, Ms. Mary Slater, Ms. Frances Jackson, Ms. Tonya Wiley and the pediatric dental residents of the UAB School of Dentistry. This investigation was supported by Research Grant DE016684 from the National Institute of Dental and Craniofacial Research (NIDCR), National Institutes of Health, Bethesda, MD

20892. This research is investigator initiated therefore the sponsor had no involvement in the scientific study or preparation for this manuscript. Ms. Momeni is a Dental Academic Research Training (DART) Pre-doctoral Fellow under NIDCR Institutional Grant#T-90 DE022736. The authors declare no conflicts of interest.

REFERENCES

Cheon, K., Moser, S.A., Whiddon, J., Osgood, R.C., Momeni, S., Ruby, J.D., Cutter, G.R., Allison, D.B., Childers, N.K., 2011. Genetic Diversity of Plaque Mutans Streptococci with rep-PCR. J Dent Res. 90, 331-335.

Cheon, K., Moser, S.A., Wiener, H.W., Whiddon, J., Momeni, S.S., Ruby, J.D., Cutter, G.R., Childers, N.K., 2013. Characteristics of *Streptococcus mutans* genotypes and dental caries in children. Eur J Oral Sci. 121, 148-155.

Childers, N.K., Osgood, R.C., Hsu, K.L., Manmontri, C., Momeni, S.S., Mahtani, H.K., Cutter, G.R., Ruby, J.D., 2011. Real-time quantitative polymerase chain reaction for enumeration of *Streptococcus mutans* from oral samples. Eur J Oral Sci. 119, 447-454.

Dasanayake, A.P., Caufield, P.W., Cutter, G.R., Roseman, J.M., Kohler, B., 1995. Differences in the detection and enumeration of mutans streptococci due to differences in methods. Arch Oral Biol. 40, 345-351.

Gold, O.G., Jordan, H.V., Van Houte, J., 1973. A selective medium for *Streptococcus mutans*. Arch Oral Biol. 18, 1357-1364.

Hildebrandt, G.H., Bretz, W.A., 2006. Comparison of culture media and chairside assays for enumerating mutans streptococci. J Appl Microbiol. 100, 1339-1347.

Kimmel, L., Tinanoff, N., 1991. A modified mitis salivarius medium for a caries diagnostic test. Oral Microbiol Immunol. 6, 275-279.

Little, W.A., Korts, D.C., Thomson, L.A., Bowen, W.H., 1977. Comparative recovery of *Streptococcus mutans* on ten isolation media. J Clin Microbiol. 5, 578-583.

Loesche, W., 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol Rev. 50, 353-380.

McGhee, J.R., Michalek, S.M., Cassell, G.H., 1982. Dental Microbiology, Harper & Row Philadelphia.

Moser, S.A., Mitchell, S.C., Ruby, J.D., Momeni, S., Osgood, R.C., Whiddon, J., Childers, N.K., 2010. Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. J Clin Microbiol. 48, 599-602.

Petti, C.A., Polage, C.R., Schreckenberger, P., 2005. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. J Clin Microbiol. 43, 6123-6125.

Ruoff, K.L., Ferraro, M.J., Jerz, M.E., Kissling, J., 1982. Automated identification of gram-positive bacteria. J Clin Microbiol. 16, 1091-1095.

Sanchez-Perez, L., Acosta-Gio, A.E., 2001. Caries risk assessment from dental plaque and salivary *Streptococcus mutans* counts on two culture media. Arch Oral Biol. 46, 49-55.

Saravia, M.E., Nelson-Filho, P., Silva, R.A., De Rossi, A., Faria, G., Silva, L.A., Emilson, C.G., 2013. Recovery of mutans streptococci on MSB, SB-20 and SB-20M agar media. Arch Oral Biol. 58, 311-316.

Schaeken, M.J.M., 1986. Comparative Recovery of *Streptococcus mutans* on Five Isolation Media,Including a New Simple Selective Medium. *J DENT RES* 65, 906.

Syed, S.A., Loesche, W.J., 1972. Survival of human dental plaque flora in various transport media. Appl Microbiol. 24, 638-644.

Tanzer, J.M., Borjesson, A.C., Laskowski, L., Kurasz, A.B., Testa, M., 1984. Glucose-sucrose-potassium tellurite-bacitracin agar, an alternative to mitis salivarius-bacitracin agar for enumeration of *Streptococcus mutans*. J Clin Microbiol. 20, 653-659.

Van Palenstein Helderman, W.H., Ijsseldijk, M., Huis in 't Veld, J.H., 1983. A selective medium for the two major subgroups of the bacterium *Streptococcus mutans* isolated from human dental plaque and saliva. Arch Oral Biol. 28, 599-603.

Wan, A.K., Seow, W.K., Walsh, L.J., Bird, P.S., 2002. Comparison of five selective media for the growth and enumeration of *Streptococcus mutans*. Australian dental journal. 47, 21-26.

Yoshida, A., Suzuki, N., Nakano, Y., Kawada, M., Oho, T., Koga, T., 2003. 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. 41, 4438-4441.

Table 1. Summary statistics for log mean bacterial enumeration of mutans streptococci on five media. UA159 is *S. mutans* prototype strain. 6715 is *S. sobrinus* prototype strain.

Group	MSB	MSKB	GSTB	TYS20B	TYCSB	*p value*[a]	*p value*[b]
Control UA159	9.16 ± 0.14	9.05 ± 0.16	9.08 ± 0.19	9.18 ± 0.10	9.09 ± 0.12	0.69	0.44
Control 6715	8.79 ± 0.30	8.63 ± 0.25	8.64 ± 0.35	8.75 ± 0.41	8.67 ± 0.29	0.94	0.51
Clinical TPC	3.71 ± 0.97	3.06 ± 1.45	3.64 ± 0.93	3.80 ± 0.90	2.67 ± 1.80	0.14	0.32
Clinical SmTC	2.76 ± 1.78	3.01 ± 1.64	3.36 ± 1.14	3.22 ± 1.23	2.67 ± 1.79	0.86	0.59
Clinical Ratio	29% ± 31%	70% ± 35%	52% ± 38%	41% ± 37%	72% ± 26%	0.07	0.03[c]

TPC is total plate count. SmTC is *S. mutans* total count.
[a] *p*-value for all 5 media.
[b] *p*-value for MSB media compared to other media.
[c] Significance. Standard deviation is for 4 replicates for control strains and 2 replicates for clinical strains. Clinical samples did not detect *S. sobrinus*.

Table 2. *S. mutans* rep-PCR genotype counts recovered from clinical isolates on five different media.

		Rep-PCR Genotypes						
ID	**Media**	**G-1**	**G1a**	**G-7**	**G-9**	**G-13**	**G-18**	**G-22**
553	**GSTB**		2				8	
	MSKB		1				9	
	MSB		4				6	
	TYS20B		2				8	
	TYCSB		1				9	
561	**GSTB**		8					2
	MSKB		9					1
	MSB		8					2
	TYS20B		7		3			
	TYCSB		6		1			3
562	**GSTB**	10						
	MSKB	10						
	MSB	10						
	TYS20B	9		1				
	TYCSB	8		2				
571	**GSTB**					9		
	MSKB					10		
	MSB					10		
	TYS20B					10		
	TYCSB					10		
590	**GSTB**				1		9	
	MSKB				1		9	
	MSB						10	
	TYS20B				3		7	
	TYCSB				2		8	

G1a is a subtype of G1.

Table S1. MS Prototype Mean Total Plate Counts for 4 Trials.

		TPC (mean log CFUs/mL ± SD)				
	Test #	**MSB**	**MSKB**	**GSTB**	**TYS20B**	**TYCSB**
UA159	1	9.00 ± 0.40	9.01 ± 0.02	9.21 ± 0.03	9.23 ± 0.22	8.98 ± 0.38
	2	9.23 ± 0.01	9.02 ± 0.10	8.87 ± 0.23	9.20 ± 0.01	8.88 ± 0.32
	3	9.01 ± 0.12	8.88 ± 0.04	8.91 ± 0.23	9.02 ± 0.01	9.13 ± 0.00
	4	9.31 ± 0.02	9.27 ± 0.03	9.27 ± 0.01	9.23 ± 0.02	9.23 ± 0.01
6715	1	8.83 ± 0.10	8.66 ± 0.03	8.43 ± 0.11	9.06 ± 0.18	8.57 ± 0.45
	2	9.01 ± 0.05	8.56 ± 0.17	8.76 ± 0.45	8.59 ± 0.29	8.56 ± 0.03
	3	8.92 ± 0.22	8.94 ± 0.07	9.00 ± 0.01	9.08 ± 0.01	9.06 ± 0.04
	4	8.35 ± 0.09	8.32 ± 0.12	8.26 ± 0.03	8.22 ± 0.06	8.36 ± 0.10

Highest count for each sampling is highlighted**.** Values calculated using mean log CFUs/mL values ± standard deviation. (*S. mutans* UA159 TPC $p = 0.69$, MSB $p = 0.51$, *S. sobrinus* TPC $p = 0.94$, MSB $p = 0.44$)

Table S2. Clinical Samples Total Mean Plate Counts (TPC).

CHILD	**MSB**	**MSKB**	**GSTB**	**TYS20B**	**TYCSB**
505	3.00 ± 0.12	1.15 ± 1.63	2.97 ± 0.26	3.12 ± 0.05	2.79 ± 0.00
552	3.28 ± 0.10	3.75 ± 0.34	3.09 ± 0.43	3.82 ± 0.05	2.76 ± 0.64
553	4.49 ± 0.05	4.05 ± 0.24	4.33 ± 0.01	4.36 ± 0.05	4.16 ± 0.12
554	1.58 ± 2.23	0	2.96 ± 0.07	3.38 ± 0.41	2.90 ± 0.16
561	6.08 ± 0.01	5.65 ± 0.89	5.79 ± 0.21	5.72 ± 0.25	5.63 ± 0.01
562	3.36 ± 0.29	3.37 ± 0.51	3.55 ± 0.14	3.41 ± 0.14	3.33 ± 0.03
563	3.23 ± 0.11	3.03 ± 0.17	3.12 ± 0.05	2.73 ± 0.60	0
564	3.57 ± 0.31	1.30 ± 1.84	3.75 ± 0.12	4.07 ± 0.01	0
569	3.32 ± 0.57	2.91 ± 0.43	2.81 ± 0.28	2.91 ± 0.43	2.70 ± 0.12
571	4.52 ± 0.07	3.29 ± 0.03	4.61 ± 0.04	4.77 ± 0.34	3.53 ± 0.02
588	2.46 ± 0.21	1.15 ± 1.63	2.46 ± 0.21	2.61 ± 0.00	0
590	4.08 ± 0.07	3.69 ± 0.05	3.95 ± 0.03	4.18 ± 0.09	4.03 ± 0.05

Highest count for each clinical sample is highlighted. Values calculated using mean log CFUs/mL values ± standard deviation. (TPC $p = 0.14$, MSB $p = 0.32$)

Table S3. Clinical Samples Total *S. mutans* Mean Plate Counts (SmTC).

CHILD	MSB	MSKB	GSTB	TYS20B	TYCSB
552	2.85 ± 0.09	3.52 ± 0.31	3.09 ± 0.43	3.68 ± 0.19	2.73 ± 0.60
553	3.94 ± 0.21	3.99 ± 0.31	4.26 ± 0.03	4.19 ± 0.09	4.03 ± 0.14
554	0	0	1.39 ± 1.97	1.15 ± 1.63	1.39 ± 1.97
561	5.43 ± 0.11	5.65 ± 0.89	5.79 ± 0.21	5.66 ± 0.19	5.63 ± 0.01
562	2.96 ± 0.49	3.36 ± 0.49	3.38 ± 0.10	3.34 ± 0.11	3.33 ± 0.03
563	2.46 ± 0.21	1.39 ± 1.97	2.85 ± 0.09	1.15 ± 1.63	0
564	3.57 ± 0.31	1.30 ± 1.84	2.55 ± 0.34	1.15 ± 1.63	0
569	0	1.15 ± 1.63	1.15 ± 1.63	2.61 ± 0.00	2.46 ± 0.21
571	3.39 ± 0.18	3.17 ± 0.12	3.62 ± 0.05	3.31 ± 0.32	3.15 ± 0.00

Highest count for each clinical sample is highlighted. Values calculated using mean log CFUs/mL values ± standard deviation. (SmTC $p = 0.14$, MSB $p = 0.32$)

Table S4. Ratio of *S. mutans* to Total Plate Counts.

CHILD	MSB	MSKB	GSTB	TYS20B	TYCSB
552	37%	58%	100%	75%	89%
553	29%	92%	86%	69%	74%
554	0%	0%	33%	3%	38%
561	23%	100%	100%	84%	100%
562	48%	97%	67%	85%	100%
563	18%	27%	54%	13%	0%
564	100%	100%	7%	1%	0%
569	0%	10%	14%	40%	60%
571	8%	79%	10%	3%	42%

Highest percentage for each clinical sample is highlighted. Values calculated on raw count data. ($p = 0.07$)

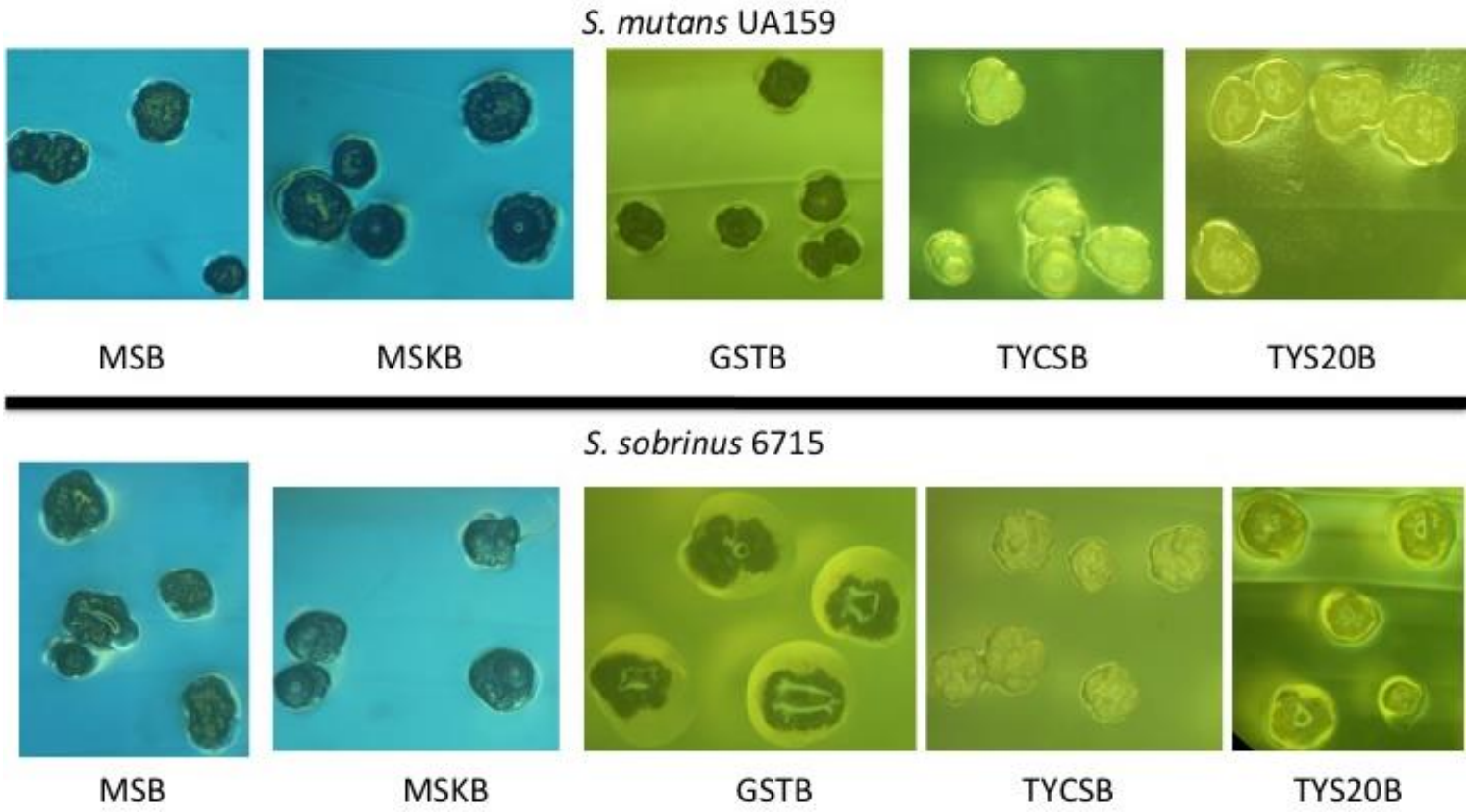


Fig. S1. Comparison of Colony Morphology for Prototype Strains on Five Media. Top row is representative of *S. mutans* UA159 colony morphology observed on each media. Bottom row is representative of *S. sobrinus* 6715 colony morphology observed on each media.

ASSESSMENT OF TWO MULTILOCUS SEQUENCE TYPING (MLST) SCHEMES AVAILABLE FOR *STREPTOCOCCUS MUTANS*

by

STEPHANIE S. MOMENI, JENNIFER WHIDDON, KYOUNGA CHEON, STEPHEN A. MOSER, NOEL K. CHILDERS

Archives of Oral Biology Dec;60(12):1769-1776

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.

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, Gueidan, de Hoog, Lozniewski, & Machouart, 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 30,000 *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 10,310 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 s*paP* (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. (Nakano et al., 2007). MLST analysis for the DS required a modified amplification protocol from the NS, using a lower annealing temperature of 49ºC (NS annealing temp was 55ºC). 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 (3,366 bp Nakano, 3,774 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 using the Maynard-Smith approach using 1,000 trials to evaluate recombination (Jolley, Feil, Chan, & Maiden, 2001; Smith, Smith, O'Rourke, & Spratt, 1993).

Phylogenetic analysis was performed using the Minimum Evolution bootstrap method (1,000 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 $p \leq 0.05$. Discriminatory power and concordance were evaluated by Simpson index of diversity (SID), Adjusted Rand, and Adjusted Wallace coefficients calculated using Comparing Partitions (http://darwin.phyloviz.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 no 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 methods 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 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^{o}C for the NS and 49^{o}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 (3,366 bp vs. 3,351 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 schemes and rep-PCR may 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 results 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 500-700 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^{o}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 *tkt* 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 3,366-bp concatenated sequence verses the 3,351-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 (*tkt, 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.

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.

Acknowledgments

We especially appreciate all the clinical and laboratory participants of this study. Special gratitude to Dr. Jinthana Lapirattanakul 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.

REFERENCES

Ahmed, A., Thaipadungpanit, J., Boonsilp, S., Wuthiekanun, V., Nalam, K., Spratt, B. G., . . . Peacock, S. J. (2011). Comparison of two multilocus sequence based genotyping schemes for *Leptospira* species. [Comparative Study Evaluation Studies Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *PLoS Negl Trop Dis, 5*(11), e1374. doi: 10.1371/journal.pntd.0001374

Carrico, J. A., Silva-Costa, C., Melo-Cristino, J., Pinto, F. R., de Lencastre, H., Almeida, J. S., & Ramirez, M. (2006). Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. [Research Support, Non-U.S. Gov't]. *J Clin Microbiol, 44*(7), 2524-2532. doi: 10.1128/JCM.02536-05

Cheon, K., Moser, S. A., Whiddon, J., Osgood, R. C., Momeni, S., Ruby, J. D., . . . Childers, N. K. (2011). Genetic Diversity of Plaque Mutans Streptococci with rep-PCR. *J Dent Res, 90*(3), 331-335. doi: 90/3/331 [pii] 10.1177/0022034510386375

Cheon, K., Moser, S. A., Wiener, H. W., Whiddon, J., Momeni, S. S., Ruby, J. D., . . . Childers, N. K. (2013). Characteristics of *Streptococcus mutans* genotypes and dental caries in children. *Eur J Oral Sci, 121*(3 Pt 1), 148-155. doi: 10.1111/eos.12044

Childers, N. K., Osgood, R. C., Hsu, K. L., Manmontri, C., Momeni, S. S., Mahtani, H. K., . . . Ruby, J. D. (2011). Real-time quantitative polymerase chain reaction for enumeration of *Streptococcus mutans* from oral samples. *Eur J Oral Sci, 119*(6), 447-454. doi: 10.1111/j.1600-0722.2011.00888.x

Debourgogne, A., Gueidan, C., de Hoog, S., Lozniewski, A., & Machouart, M. (2012). Comparison of two DNA sequence-based typing schemes for the *Fusarium solani* Species Complex and proposal of a new consensus method. [Comparative Study Evaluation Studies]. *J Microbiol Methods, 91*(1), 65-72. doi: 10.1016/j.mimet.2012.07.012

Deng, W., Maust, B. S., Nickle, D. C., Learn, G. H., Liu, Y., Heath, L., . . . Mullins, J. I. (2010). DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. [Research Support, N.I.H., Extramural]. *Biotechniques, 48*(5), 405-408. doi: 10.2144/000113370

Do, T., Gilbert, S. C., Clark, D., Ali, F., Fatturi Parolo, C. C., Maltz, M., . . . Beighton, D. (2010). Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. *PLoS One, 5*(2), e9073. doi: 10.1371/journal.pone.0009073

Foley, S. L., White, D. G., McDermott, P. F., Walker, R. D., Rhodes, B., Fedorka-Cray, P. J., . . . Zhao, S. (2006). Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *J Clin Microbiol, 44*(10), 3569-3577. doi: 44/10/3569 [pii] 10.1128/JCM.00745-06

Francisco, A. P., Bugalho, M., Ramirez, M., & Carrico, J. A. (2009). Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. [Research Support, Non-U.S. Gov't]. *BMC Bioinformatics, 10*, 152. doi: 10.1186/1471-2105-10-152

Healy, M., Huong, J., Bittner, T., Lising, M., Frye, S., Raza, S., . . . Lupski, J. R. (2005). Microbial DNA typing by automated repetitive-sequence-based PCR. *J Clin Microbiol, 43*(1), 199-207. doi: 43/1/199 [pii] 10.1128/JCM.43.1.199-207.2005

Hunter, P. R., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. [Comparative Study]. *J Clin Microbiol, 26*(11), 2465-2466.

Jolley, K. A., Feil, E. J., Chan, M. S., & Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics, 17*(12), 1230-1231.

Kilian, M., Scholz, C. F., & Lomholt, H. B. (2012). Multilocus sequence typing and phylogenetic analysis of *Propionibacterium acnes*. [Comparative Study Research Support, Non-U.S. Gov't]. *J Clin Microbiol, 50*(4), 1158-1165. doi: 10.1128/JCM.r06129-11

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. [Research Support, Non-U.S. Gov't]. *Bioinformatics, 25*(11), 1451-1452. doi: 10.1093/bioinformatics/btp187

Loesche, W. (1986). Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev, 50*(4), 353-380.

Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol, 60*, 561-588. doi: 10.1146/annurev.micro.59.030804.121325

Momeni, S. S., Whiddon, J., Moser, S. A., Cheon, K., Ruby, J. D., & Childers, N. K. (2013). Comparative genotyping of *Streptococcus mutans* by repetitive extragenic palindromic polymerase chain reaction and multilocus sequence typing. *Mol Oral Microbiol, 28*(1), 18-27. doi: 10.1111/omi.12002

Moser, S. A., Mitchell, S. C., Ruby, J. D., Momeni, S., Osgood, R. C., Whiddon, J., & Childers, N. K. (2010). Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. *J Clin Microbiol, 48*(2), 599-602. doi: JCM.01828-09 [pii]10.1128/JCM.01828-09

Nakano, K., Lapirattanakul, J., Nomura, R., Nemoto, H., Alaluusua, S., Gronroos, L., . . . Nakagawa, I. (2007). *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. *J Clin Microbiol, 45*(8), 2616-2625. doi: JCM.02343-06 [pii] 10.1128/JCM.02343-06

Smith, J. M., Smith, N. H., O'Rourke, M., & Spratt, B. G. (1993). How clonal are bacteria? [Research Support, Non-U.S. Gov't]. *Proc Natl Acad Sci U S A, 90*(10), 4384-4388.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H., & Swaminathan, B. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. [Review]. *J Clin Microbiol, 33*(9), 2233-2239.

van Belkum, A., Tassios, P. T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N. K., . . . Struelens, M. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. [Guideline]. *Clin Microbiol Infect, 13 Suppl 3*, 1-46. doi: 10.1111/j.1469-0691.2007.01786.x

Table 1. New alleles and sequence types (ST) registered with PubMLST for Nakano *S. mutans* MLST scheme in this study.

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

[a] New alleles reported in this study are in bold.
[b] L: library isolate, G: genotype group isolate.

Table 2. New alleles and sequence types (ST) registered with PubMLST for Do *S. mutans* MLST scheme in this study.

PubMLST ST	Isolate ID[b]	Allelic profiles[a]							
		accC	*gki*	*lepA*	*recP*	*sodA*	*tyrS*	*gtfB*	*spaP*
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

[a] New alleles reported in this study are in bold.
[b] L: 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
murI	425	8	0.393	11 (2.59)	8	3	0.054	0.302
glnA	460	8	0.378	7 (1.52)	7	0	0.000	-0.491
tkt	435	10	0.449	11 (2.53)	10	1	0.043	-0.976
gyrA	435	6	0.425	6 (1.38)	4	2	0.389	-1.488
gltA	389	10	0.397	12 (3.08)	7	5	0.148	-0.269
aroE	397	10	0.361	10 (2.52)	5	5	0.377	-0.913
glk	405	8	0.409	8 (1.98)	6	2	0.173	-0.613
lepC	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
accC	462	8	0.402	7 (1.52)	5	2	0.146	-1.637
gki	426	9	0.431	10 (2.35)	8	2	0.044	-1.079
lepA	441	7	0.414	8 (1.81)	7	1	0.014	-0.924
recP	474	13	0.403	11 (2.32)	6	5	0.300	-0.613
sodA	492	11	0.412	13 (2.64)	6	7	0.152	-1.767
tyrS	513	13	0.382	16 (3.12)	12	4	0.063	0.404
gtfB	453	11	0.340	10 (2.21)	2	8	0.536	0.351
spaP	513	8	0.395	13 (2.53)	9	4	0.071	-1.360

[a] Number of *S. mutans* isolates used were Nakano scheme (n=33) and Do scheme (n=32). Data generated using DnaSP software v.5.10.1.

[b] Fragment size reported in this study for Nakano scheme is slightly different than originally reported by Nakano et al. (2007).
[c] Fragment size for Do scheme were trimmed to PubMLST length.

Table 4. Diversity indices and congruence values for MLST schemes and rep-PCR using 33 *S. mutans* isolates.[a]

			Adjusted Rand (95% CI[b])		Wallace Coefficient (95% CI)		
	# Partitions	SID[c] (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.

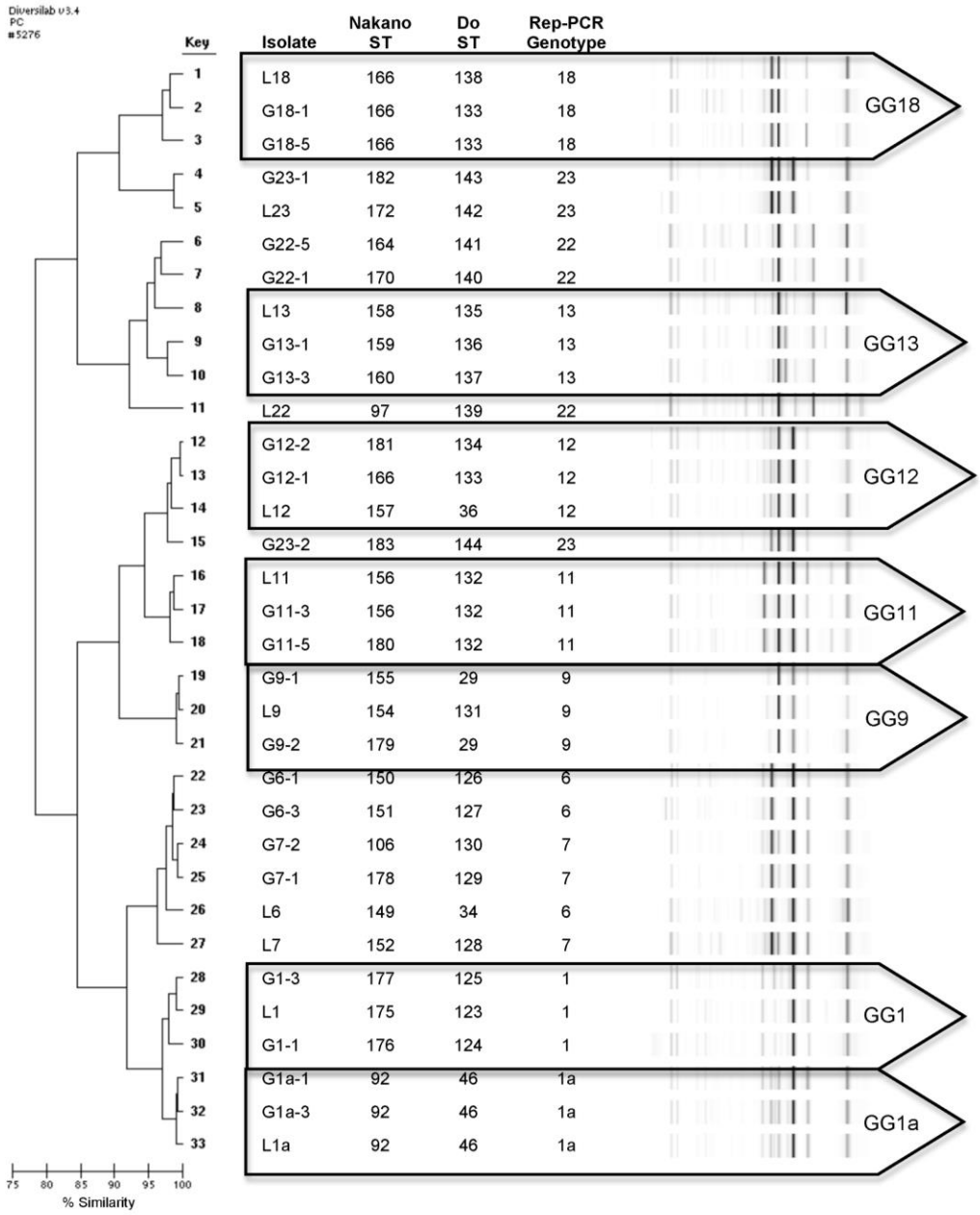


Fig. 1. Rep-PCR dendrogram by Percent Similarity with MLST Sequence Type (ST) assignments for Nakano and Do MLST Schemes for *S. mutans* Genotype Groups. GG = genotype groups. Isolates labeled with "L" prefix are Library representative strains. Isolates Labeled with "G" prefix are additional rep-PCR 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.

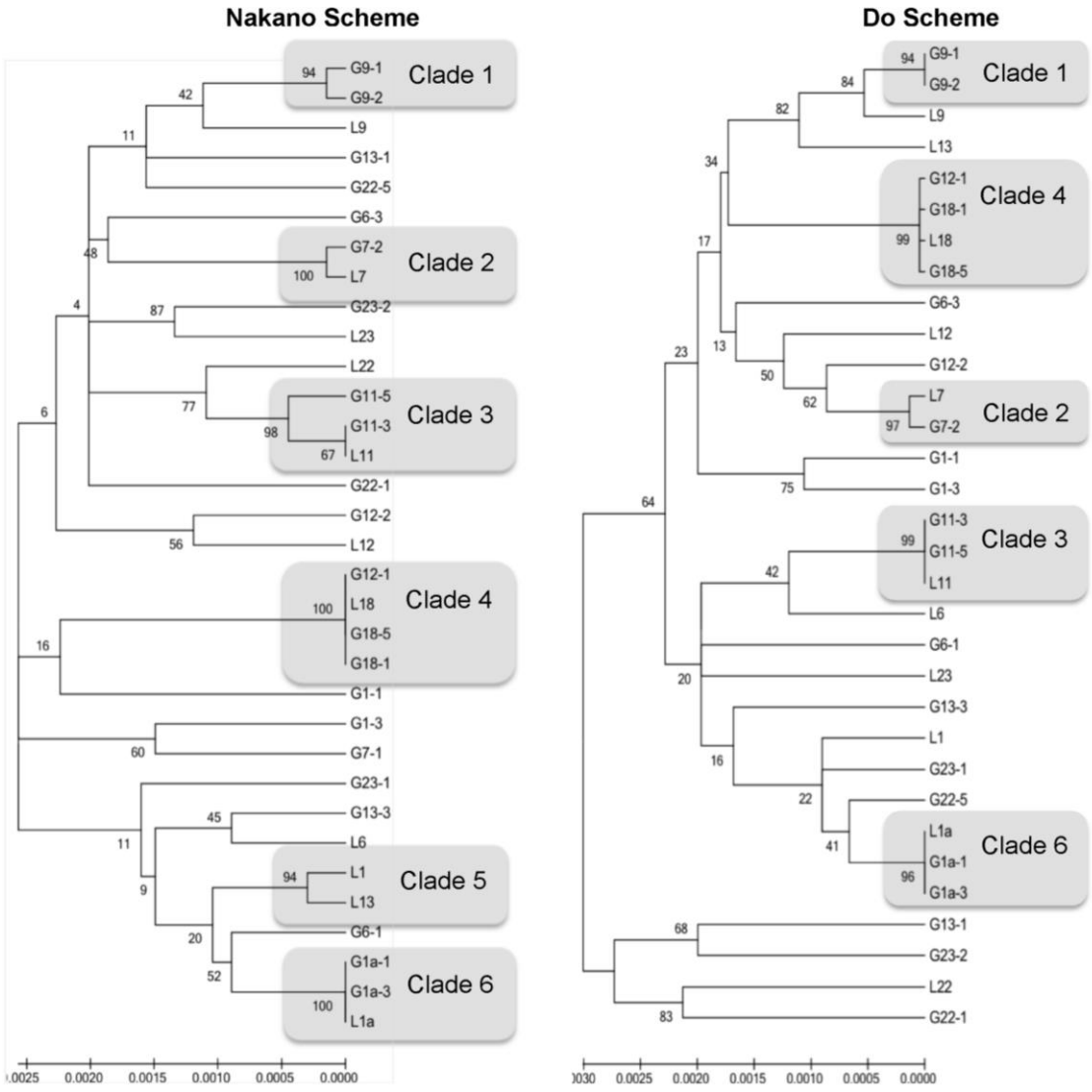


Fig. 2. Minimum Evolution Phylogenetic Tree with Bootstrap (1,000 replicates) for Nakano and Do MLST Typing Schemes. Bootstrap values are indicated on branches. Topology is supported if bootstrap ≥95%. Nakano Clade 5 is further distinguished with Do Scheme. Topology for other clades is supported by both schemes.

ASSESSMENT OF CLONALITY AND SEROTYPES OF *STREPTOCOCCUS MUTANS* AMONG CHILDREN BY MULTILOCUS SEQUENCE TYPING

by

STEPHANIE S. MOMENI, JENNIFER WHIDDON, KYOUNGA CHEON, STEPHEN A. MOSER, NOEL K. CHILDERS

European Journal of Oral Sciences Dec;123(6):416-424

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 (CBPs) encoded by *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 were identified, of which 27 were unique to this population. Seventy-five per cent 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 United States population.

Introduction

Streptococcus mutans is a common colonizer of the oral cavity and has been widely associated with initiation and progression of dental caries (1, 2). Although 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.* (7), using partial fragments of eight conserved housekeeping genes. In a previous study, we reported three out of six repetitive extragenic palindromic PCR (rep-PCR) genotypes groups were identical following MLST (8). This clonality was surprising given that in other studies of *S. mutans* using MLST, sequence types (ST) are rarely shared between unrelated individuals (7, 9). However, the number of isolates in our previous study was small (only three clonal groups) and we could not rule out selection bias because isolates in that study were selected based upon the rep-PCR genotypes (8).

S. mutans has four 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 (encoded by the *cnm* gene) and Cbm protein (encoded by the *cbm* gene), which, when present, are suspected to promote the adherence to and invasion of human endothelial cells by *S. mutans* (11, 13, 14). Approximately 85% of serotype *k* isolates from Japan, Finland, and Thailand are reported to contain either the *cnm* gene or *cbm* gene (11). Serotype *k* strains with the *cbm* gene are those most frequently associated with high collagen-binding activity and invasive properties, whereas 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, following 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-caries risk community in

Uniontown, Alabama, USA. This sample population is considered high-risk because of its low socioeconomic status, limited access to dental care, and markedly high caries prevalence in children under 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 and decayed, missing or filled surfaces (DMFS+dmfs) scores of secondary and primary dentition, respectively 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-age children (5-6 yr of age at initial enrollment) and Cohort 2 (CH2) infants (6-18 mo years 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 children with bacterial isolates that had

been confirmed as *S. mutans* positive by PCR using *gftB*-specific primers and had a rep-PCR genotype assigned (see below). 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

Repetitive extragenic palindromic PCR was performed using DIVERSILAB, as previously reported (18). Briefly, *S. mutans* isolates were confirmed, based on colony morphology, on Gold's Medium (modified mitis salivarius media [Becton, Dickinson and Company, 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* using a SYBR Green PCR approach with *gftB* sequence-specific primers (20, 21). Repetitive extragenic palindromic PCR was performed using the *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 three minor band, one 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, eight PCR amplifications were performed by NAKANO *et al.* This typing scheme for *S. mutans* including primers for partial gene fragments from *murI, tkt, glnA, gyrA, glk, glt*, *glk,* and *lepC* and uses 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, UK) 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, whereas clonal complexes are defined as isolates with similar STs (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 isolate) 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 children either with caries or no caries. 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 STs), 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, and 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 coefficient were calculated using Comparing Partitions (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Home) (28, 29).

Serotypes & Collagen Binding Proteins

Serotyping and identification of the presence of the genes encoding collagen-binding proteins, *cnm* and *cbm* were performed for all 100 isolates using a SYBR Green PCR multiplex approach with previously reported primers (11, 30, 31). PCR was performed on a IQ5 Real-time thermocycler (Bio-Rad Laboratories, Hercules, CA, USA) using a 25-µl reaction mix, containing 2 *µ*l of DNA template (20 ng/*µ*l) and 2x Maxima SYBR Green qPCR master mix (Thermo Scientific, Lafayette, CO, USA), using the following parameters: one cycle of 95^{o}C for 10 min, followed by 30 cycles of 95^{o}C for 20 s, 60^{o}C for 30 s, 72^{o}C for 45 s. Melt curve analysis was performed from 60^{o}C to 95^{o}C, with interval of 0.5^{o}C. Separate multiplex PCR amplifications were performed to identify the serotype and the presence/absence of genes encoding 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 amplifications for the detection of collagen-binding proteins and undetermined serotypes were performed in duplicate. To confirm amplicon size, all amplicons were electrophoresed on 1.3% agarose gels containing 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 (47%) of the 57 children had more than one isolate (range, two to five isolates) (Table 2). Eighteen children exhibited a single rep-PCR genotype and MLST sequence type for all randomly selected isolates (one GT/one ST). For stability analysis, 24 children had isolates from multiple periods of collection; three were excluded because 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 sequence type from more than one period, indicating that stability was 63% (15/24).

MLST

Table 2 shows five children with multiple genotypes and STs (two GTs/two STs) that agreed between methods. Two children (241, 531) had the same rep-PCR genotype but different MLST sequence types, indicating that MLST was more discriminative. Repetitive extragenic palindromic PCR was more discriminative

than MLST for two other children (150, 596) who had different rep-PCR genotypes but fewer MLST sequence types. 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 STs 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, whilst use of the double-locus variant setting resulted in 21 groups. Three new alleles and 16 new STs were identified in this study and were added to the PubMLST database (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 STs identified, seven matched STs available in the PubMLST database; however, 27 were unique to the Uniontown population. Forty isolates were classified into one of the 16 new STs found in this study. Sixty matched previously published STs (19 isolates matched STs from the PubMLST database for Japanese, Thai, or Finnish isolates, and the remaining 41 isolates matched STs previously reported from the Uniontown population) (8, 32).

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 sequence types 92, 192, and 193) (Table 1, Fig. S1). Eight rep-PCR genotype groups (44%; G07, G10, G11, G14, G15, G18, G27, and G50) were further supported by MLST analysis. Five MLST sequence types (156, 179,

191, 192, and 202) represented more than one rep-PCR genotype (e. g. MLST sequence type 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 and S2). DMFS+dmfs (collection) indicates the score at the time the isolate evaluated was obtained, whereas DMFS+dmfs (final) indicates the child's score at last period on record. Final score data should be interpreted with caution as these scores decrease because children in CH1 lose baby teeth and increase in children in either cohort as new teeth erupt.

General allele characteristics are listed in Table 3. The highest and lowest frequencies of polymorphisms were *gltA* (3.6%) *and gyrA* (1.2%) respectively. The number of alleles reported ranged from five (*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 SID, because rep-PCR (0.918) and MLST (0.948) had overlapping confidence intervals. Multilocus sequence typing and rep-PCR were found to be 68% congruent when analyzed using Adjusted Rand. According to Adjusted Wallace, MLST was more likely to predict rep-PCR GTs (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). For all 100 isolates the I_A was 1.943; for 67 representative individuals it was 1.561; and for 34 single ST it 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^{o}C and a single peak at 76-78^{o}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 probably a result of 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^{o}C and were indistinguishable until run on gels. Six of the 10 serotype *k* isolates grouped together into two clonal clusters (clonal clusters 5 and 6, ST205 & ST106), according to MLST analysis, and corresponded to rep-PCR genotypes G14 and G07, respectively (Supporting Information Figs. S1 and S2).

Evidence for the presence of collagen-binding proteins encoded by *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^{o}C and 81^{o}C respectively. Only control strains were observed on gels.

Discussion

Initial analysis used all 100 randomly selected isolates and found that 27 (47%) of the 57 of children 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, nine other children had more than one genotype or ST, indicating that variability of isolates within a child is about 33%. Although the majority multiple isolates in children 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 (33).

It is not surprising that 56% of rep-PCR genotypes were further differentiated by MLST because rep-PCR is based on similarity, whereas a single base-pair change can result in a new ST, according to MLST. Nonetheless, eight (44%) genotype groups were confirmed by MLST analysis, demonstrating these isolates are identical to 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 indicating that 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 United States 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 five cases (Table 1, Fig. S1) (clonal groups 1, 4, 9, 10, and 11), MLST STs were further differentiated by rep-PCR. Although it has been reported that rep-PCR can be more discriminative than MLST for other organisms (34-37), 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 (38, 39). Furthermore, the data support the finding that typing assignments between rep-PCR and MLST do not always correlate. These results suggest that whilst these isolates have identical sequences for 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 (40, 41).

The number of polymorphic sites reported here (range: 5-14 sites) is lower than previously reported by NAKANO *et al.* (range: 15-21 sites) but consistent with our previous study (range: 6-12 sites) (Table 3). The number of variable nucleotide sites was highest in *gltA* in all three studies but lowest in *gyrA* in both Uniontown studies (7, 32). In the present study, *gyrA* was found to have a dN/dS of >1, which indicates positive selection for this gene in the isolates evaluated in this study. The differences observed may be explained by population bias; NAKANO *et al.* (7) evaluated regional *S. mutans* isolates from a variety of sources, whereas the present study used *S. mutans* plaque samples from a relatively localized 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 LAPIRATTANAKUL *et al.* (42) using MLST; however, that study used mother-child

pairs, whereas the current study used randomly selected, unrelated subjects (42). The percentage of shared sequence types among unrelated subjects was 49% (34 STs in 67 samples) in the present study, which is notably higher than previously found by NAKANO *et al.* (7) (9.8%; 92 STs in 102 samples) and DO *et al.* (9) (9.6%; 122 STs in 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 of 0.7721, which is still higher than the I_A values previously reported for *S. mutans* by NAKANO *et al.* (0.0931) and DO *et al.* (0.4379) (7, 9), indicating that the present group of isolates are more clonal than was reported in these studies.

Only three 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 STs indicated sequence diversity typical of other MLST studies of *S. mutans* (7, 9). Other studies using MLST have reported that STs do not appear to be geographically distributed (7, 9). While a few isolates (19%) in the current study had STs available in the global database (seven STs), a majority of the isolates (81%) demonstrated STs unique to the Uniontown population (27 STs). 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. Whilst the use of localized populations is an acknowledged bias, it provides an alternate perspective for these types of studies, especially in cases where broader 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, 43). A study of subjects from southern India (44) reported a higher prevalence (26%) of serotype *k* than in those studies. However, in contrast to the Indian study (44), which 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 a higher prevalence of serotype k, comparable to that of the Indian 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 therefore the results obtained may not be representative of all African-American children.

Seven of the 10 serotype *k* isolates grouped together (G07, G14, and G27) by phylogenetic analysis and corresponded to ST106, ST205, and ST161, respectively (Fig. S1). ST106 was originally reported, by LAPIRATTANAKUL *et al.*

(42), to be serotype *c* in a single isolate from an adult woman, 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, 45).

The absence of any serotype *f* isolates is not surprising because 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 Gold's medium (20) and by PCR using *gftB*-specific primers (21). Furthermore, these two isolates match rep-PCR genotypes (G10 and G12) and MLST STs (ST179 and ST157) of other *S. mutans* strains. The DNA template for PCR was 20 ng/*µ*l, 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 before 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 salvia reported 23% of isolates as undetermined serotypes; however, this study did not include serotype *c* in the analysis (44). In 2007, NAKANO *et al.* (46) reported the recovery of undetermined serotypes, using the PCR approach, in relative abundance in cardiovascular samples and suggested these may be a new, minor oral serotype. Further investigations are needed to evaluate if these

undetermined serotypes are a limitation of detection by PCR approach or a new serotype.

Interestingly, none of the isolates tested positive for collagen-binding proteins Cnm or Cbm, according to 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* (47, 48). The Cbm protein was only recently identified in 2012 by NOMURA *et al.* (11), 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 the sample size was much larger (n = 580, n = 478) than the 100 isolates in the present study (11, 48). 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, significantly different from the 85% reported by NOMURA *et al.* (11) for serotype *k* strains. There are a number of potential explanations for this outcome. NOMURA *et al.* (48) postulated that detection of Cnm proteins may be age or geographically dependent. 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 vs. whole saliva) is not always clearly noted in some references, which can make direct comparison difficult because, if present, *cnm* or *cbm* 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 from 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).

This study demonstrates a high degree of clonality of *S. mutans* isolates randomly selected from children in a relatively isolated population, 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 are 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 at a notably high prevalence in a United States (specifically, African-American) population.

Acknowledgements

This work was supported by Research Grant DE016684 (N.K.C.) from the National Institute of Dental and Craniofacial Research. S.S.M. 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 Lapirattanakul 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.

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

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.

2. TANZER JM, LIVINGSTON J , THOMPSON AM. The microbiology of primary dental caries in humans. *J Dent Educ* 2001; **65**: 1028-1037.

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.

4. HAMADA S , SLADE HD. Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol Rev* 1980; **44**: 331-384.

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.

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.

7. NAKANO K, LAPIRATTANAKUL 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.

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.

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.

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.

11. NOMURA R, NAKANO K, NAKA S, NEMOTO H, MASUDA K, LAPIRATTANAKUL 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.

12. LAPIRATTANAKUL J, NAKANO K, NOMURA R, LEELATAWEEWUD P, CHALERMSARP N, KLAOPHIMAI A, SRISATJALUK 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.

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.

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.

15. LAPIRATTANAKUL 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.

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.

17. WHO. Oral Health surveys: basic methods, 4th ed. 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.

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.

20. GOLD OG, JORDAN HV , VAN HOUTE J. A selective medium for *Streptococcus mutans*. *Arch Oral Biol* 1973; **18**: 1357-1364.

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.

22. JOLLEY KA, FEIL EJ, CHAN MS , MAIDEN MC. Sequence type analysis and recombinational tests (START). *Bioinformatics* 2001; **17**: 1230-1231.

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.

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.

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.

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.

27. LIBRADO P , ROZAS J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; **25**: 1451-1452.

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.

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.

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.

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.

32. MOMENI SS, WHIDDON J, CHEON K, MOSER SA , CHILDERS NK. Assessment of two multilocus sequence typing (MLST) schemes available for Streptococcus mutans. *Arch Oral Biol* 2015; **60**: 1769-1776.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. MAIDEN MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 2006; **60**: 561-588.

42. LAPIRATTANAKUL 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.

43. LAPIRATTANAKUL 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.

44. 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.

45. NAKANO K, NOMURA R, NEMOTO H, LAPIRATTANAKUL 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.

46. 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.

47. NAKANO K, NOMURA R, TANIGUCHI N, LAPIRATTANAKUL 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.

48. NOMURA R, NAKANO K, TANIGUCHI N, LAPIRATTANAKUL 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.

Table 1. Distribution of *Streptococcus mutans* repetitive extragenic palindromic PCR (rep-PCR) genotypes (n = 18) from randomly selected isolates (n = 100) for multilocus sequence typing (MLST) analysis.

Rep-PCR GT	No. of isolates	No. of individuals	MLST ST (no. of 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	

ST, sequence type
The number of individuals with a given rep-PCR genotype (GT) (n = 67) is higher here than reported in text (n = 57) because some children had more than one genotype represented (e.g. one child had up to three different GTs, see Table 2).

Table 2. Stability of *Streptococus mutans* isolates within 27 children for whom more than one isolate was randomly selected.

Index Child Family ID	No. 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*
173	3	G10	179	1 GT/ 1 ST	
185	2	G18	166	1 GT/ 1 ST	
193‡	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‡	2	G50	**202**	1 GT/ 1 ST	
232	2	G09	179	1 GT/ 1 ST	
241	3	G12	132, 181		1 GT/ 2 ST†
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†
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‡	2	G05	**195**	1 GT/ 1 ST	
596	5	G01a, G12	**192**		2 GT/ 1 ST*

New sequence types, identified by multilocus sequence typing (MLST), are shown in bold text.
For Child 150, three genotypes (GTs) were observed following rep-PCR and the sequence types (STs) are noted in corresponding order.

* Isolates were further distinguished with repetitive extragenic palindromic PCR (rep-PCR).
† Isolates were further distinguished by MLST.
‡ Isolates excluded from stability analysis becasuse they were from same collection period.

Table 3. Characteristics of alleles in evaluated in this study.

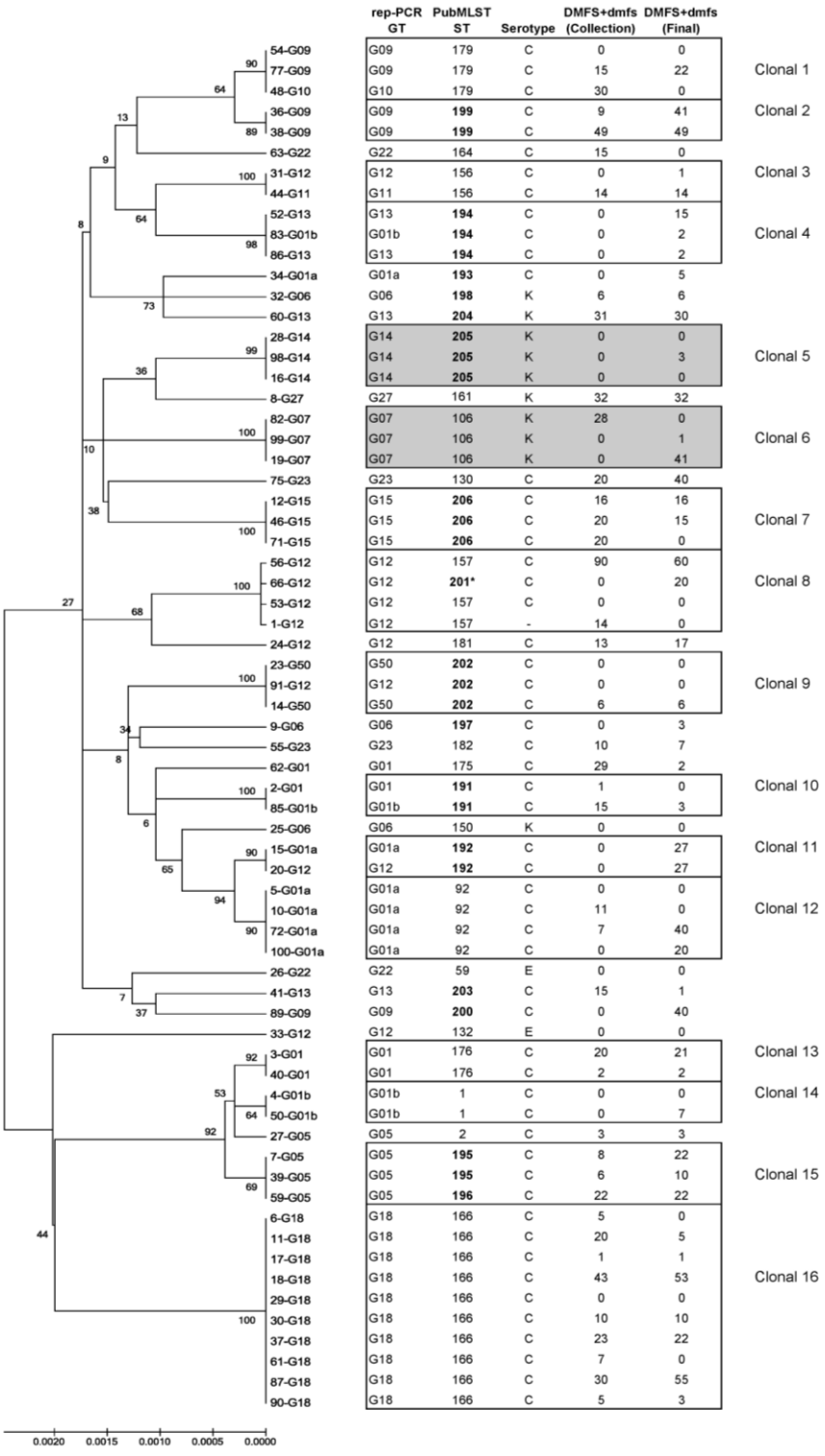
Locus	Fragment size (bp)	No. of alleles	G+C Mol	No. polymorphic sites (%)	Syn	NSyn	dN/dS	Tajima's D test
murI	425	11	0.394	12 (2.82)	8	4	0.0647	0.72634
glnA	460	8	0.378	6 (1.30)	6	0	0.0000	0.38093
tkt	435	8	0.449	10 (2.30)	7	3	0.1379	-0.37665
gyrA	435	5	0.425	5 (1.15)	3	2	1.8375	-0.87632
gltA	389	12	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

dN/dS, non-synonymous/synonymous rate ratio; Mol, mole ratio of total DNA bases; NSyn, nonsynonymous changes Syn, synonymous changes.

Table S1. New *S. mutans* MLST Alleles and Sequence Types added to PubMLST database from this study.

PubMLST ST	Isolate ID	*tkt*	*glnA*	*gltA*	*glk*	*aroE*	*murI*	*lepC*	*gyrA*
191	2-G01	1	1	8	8	**31**	3	1	1
192	15-G01a	14	1	1	1	2	3	1	1
193	34-G01a	1	2	1	1	2	3	11	1
194	83-G01b	1	2	15	1	13	11	24	1
195	7-G05	1	1	1	13	2	1	1	1
196	59-G05	1	1	1	5	2	1	1	1
197	9-G06	4	1	1	8	2	3	1	1
198	32-G06	1	2	7	1	14	7	19	1
199	36-G09	1	2	15	3	1	3	11	1
200	89-G09	1	2	15	13	1	11	1	3
201	66-G12	**24**	2	4	5	4	5	1	1
202	91-G12	3	2	1	1	4	11	1	1
203	41-G13	1	2	15	3	1	3	1	4
204	60-G13	18	23	**35**	23	14	21	11	21
205	16-G14	1	3	15	3	4	2	19	1
206	12-G15	1	1	32	3	4	19	5	4

New alleles are bold text.



rep-PCR GT	PubMLST ST	Serotype	DMFS+dmfs (Collection)	DMFS+dmfs (Final)	
G09	179	C	0	0	
G09	179	C	15	22	Clonal 1
G10	179	C	30	0	
G09	**199**	C	9	41	Clonal 2
G09	**199**	C	49	49	
G22	164	C	15	0	
G12	156	C	0	1	Clonal 3
G11	156	C	14	14	
G13	**194**	C	0	15	
G01b	**194**	C	0	2	Clonal 4
G13	**194**	C	0	2	
G01a	**193**	C	0	5	
G06	**198**	K	6	6	
G13	**204**	K	31	30	
G14	**205**	K	0	0	
G14	**205**	K	0	3	Clonal 5
G14	**205**	K	0	0	
G27	161	K	32	32	
G07	106	K	28	0	
G07	106	K	0	1	Clonal 6
G07	106	K	0	41	
G23	130	C	20	40	
G15	**206**	C	16	16	
G15	**206**	C	20	15	Clonal 7
G15	**206**	C	20	0	
G12	157	C	90	60	
G12	**201***	C	0	20	Clonal 8
G12	157	C	0	0	
G12	157	-	14	0	
G12	181	C	13	17	
G50	**202**	C	0	0	
G12	**202**	C	0	0	Clonal 9
G50	**202**	C	6	6	
G06	**197**	C	0	3	
G23	182	C	10	7	
G01	175	C	29	2	
G01	**191**	C	1	0	Clonal 10
G01b	**191**	C	15	3	
G06	150	K	0	0	
G01a	**192**	C	0	27	Clonal 11
G12	**192**	C	0	27	
G01a	92	C	0	0	
G01a	92	C	11	0	Clonal 12
G01a	92	C	7	40	
G01a	92	C	0	20	
G22	59	E	0	0	
G13	**203**	C	15	1	
G09	**200**	C	0	40	
G12	132	E	0	0	
G01	176	C	20	21	Clonal 13
G01	176	C	2	2	
G01b	1	C	0	0	Clonal 14
G01b	1	C	0	7	
G05	2	C	3	3	
G05	**195**	C	8	22	
G05	**195**	C	6	10	Clonal 15
G05	**196**	C	22	22	
G18	166	C	5	0	
G18	166	C	20	5	
G18	166	C	1	1	
G18	166	C	43	53	Clonal 16
G18	166	C	0	0	
G18	166	C	10	10	
G18	166	C	23	22	
G18	166	C	7	0	
G18	166	C	30	55	
G18	166	C	5	3	

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. Clonal groups are boxed. Grouped serotype *k* isolates are highlighted. GT = genotype, ST = Sequence Type. DMFS+dmfs (collection) =

score at time isolate evaluated was obtained. DMFS+dmfs (final) = score at last period on record. *ST varied by 1 base pair, which was not significant by this method. Bootstrap values are indicated on branches, topology is supported if >95%. Serotypes are capitalized for better visualization.

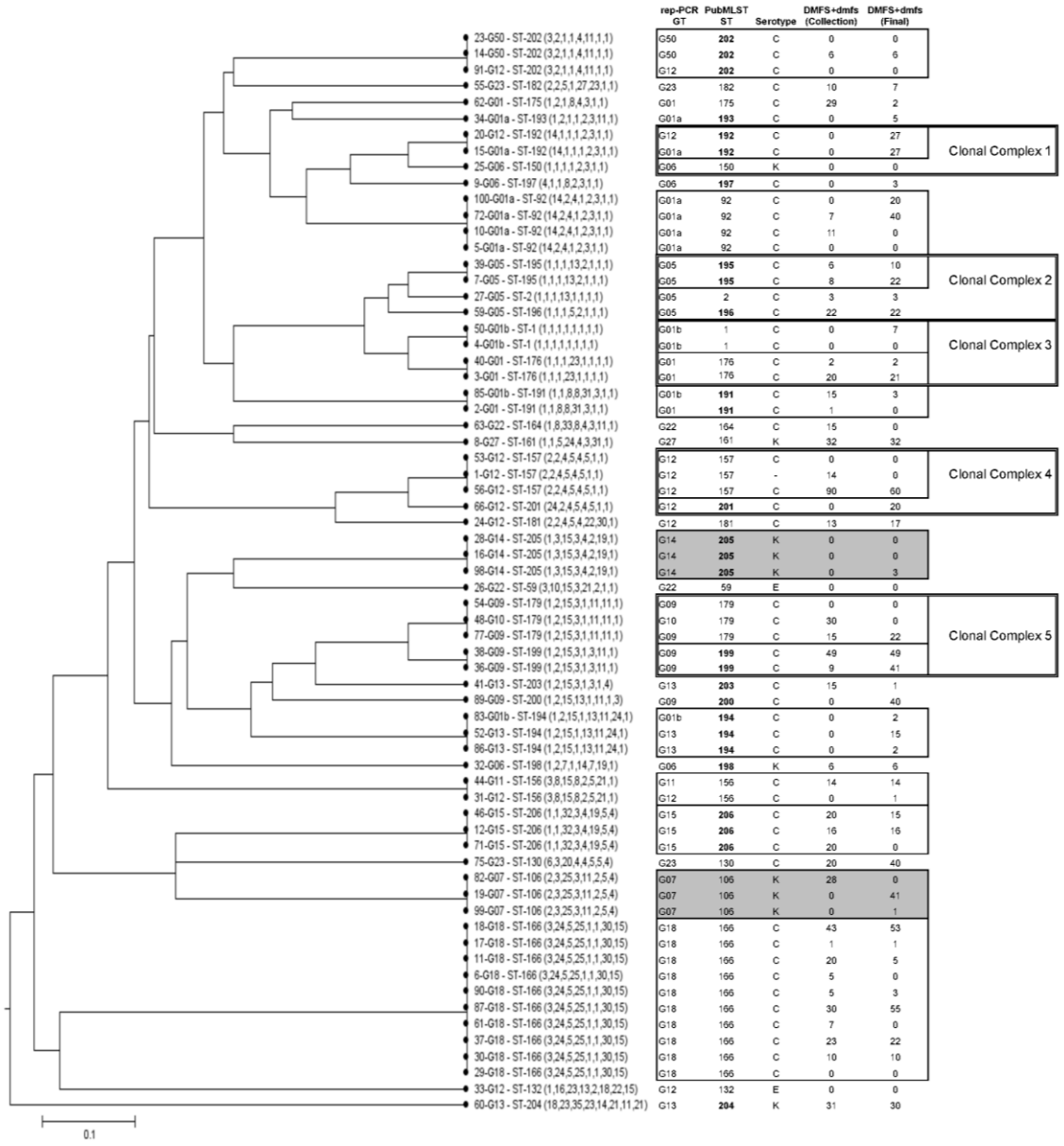


Fig. S2. Phylogenetic tree of 67 *S. mutans* isolates generated using START2 using Un-weighted Pair Group Method using Arithmetic averages (UPGMA). Clonal groups (single lines) and clonal complexes (allowing up to 2 allelic differences, double lines) are boxed. Grouped serotype k isolates are highlighted. Sample labels include sample ID-MLST Sequence type (allelic profile). DMFS+dmfs (collection) = score at time isolate evaluated was obtained. DMFS+dmfs (final) = score at last period on record.

GENETIC DIVERSITY AND EVIDENCE FOR TRANSMISSION OF *STREPTOTOCCUS MUTANS* BY DIVERSILAB REP-PCR

by

STEPHANIE S. MOMENI, JENNIFER WHIDDON, KYOUNGA CHEON, TARIQ GHAZAL, STEPHEN A. MOSER, NOEL K. CHILDERS

Journal of Microbiological Methods Sept;128:108-117

ABSTRACT

This two-part study investigated the genetic diversity and transmission of *Streptococcus mutans* using the DiversiLab repetitive extragenic palindromic PCR (rep-PCR) approach. For children with *S. mutans* and participating household members, analysis for evidence of unrelated child-to-child as well as intra-familial transmission was evaluated based on commonality of genotypes.

A total of 169 index children and 425 household family members from Uniontown, Alabama were evaluated for genetic diversity using rep-PCR. Thirty-four unique rep-PCR genotypes were observed for 13,906 *S. mutans* isolates. For transmission, 117 child and household isolates were evaluated for shared genotype (by child and by genotype cases, multiple matches possible for each child). Overall, children had 1-9 genotypes and those with multiple genotypes were 2.3 times more likely to have caries experience (decayed, missing and filled teeth/surfaces>0). Only 28% of children shared all genotypes within the household, while 72% had at least 1 genotype not shared with anyone in the household. Children had genotype(s) not shared with any household members in 157 cases. In 158 cases children and household members shared a genotype in which 55% (87/158 cases) were shared with more than one family member. Children most frequently shared genotypes with their mothers (54%; 85/158), siblings (46%; 72/158) and cousins (23%; 37/158).

A reference library for *S. mutans* for epidemiological surveillance using the DiversiLab rep-PCR approach is detailed. The genetic diversity of *S. mutans* in this population demonstrated frequent commonality of genotypes. Evidence for both child-to-child and intra-familial transmission of *S. mutans* was observed by rep-PCR.

Key Words: genetic diversity; genotyping; multilocus sequence typing; rep-PCR; *Streptococcus mutans;* transmission

Abbreviations:

AP-PCR = arbitrarily primed PCR

PFGE = pulse field gel electrophoresis

rep-PCR = repetitive extragenic palindromic PCR

MLST = multilocus sequence typing

ST = sequence type

1. INTRODUCTION

Streptococcus mutans are bacteria commonly linked with the initiation and progression of dental caries (Gross et al., 2012, Loesche, 1986, Tanner et al., 2002). The genetic diversity of *S. mutans* strains has been studied using continually evolving gel based typing methods including randomly amplified polymorphic DNA (RAPD) fingerprinting (Redmo Emanuelsson et al., 2003), arbitrarily primed PCR (AP-PCR) (Baca et al., 2012, Cogulu et al., 2006, Domejean et al., 2010, Gamboa et al., 2010, Gilbert et al., 2014, Hu et al., 2014, Lembo et al., 2007, Li and Caufield, 1998, Liu et al., 2007, Lynch et al., 2015, Napimoga et al., 2004, Palmer et al., 2012, Pieralisi et al., 2010, Teanpaisan et al., 2012, Zhao et al., 2014, Zhou et al., 2011), pulse field gel electrophoresis (PFGE) (Mineyama et al., 2004, Mineyama et al., 2007, Roberts et al., 2002) and repetitive extragenic palindromic PCR (rep-PCR) (Cheon et al., 2011, Moser et al., 2010). Molecular based typing methods used to evaluate genetic diversity include multilocus enzyme electrophoresis (MLEE) (Napimoga, Kamiya, Rosa, Rosa, Hofling, Mattos-Graner and Goncalves, 2004) and more recently, multilocus sequence typing (MLST) (Do et al., 2010, Lapirattanakul et al., 2008, Momeni et al., 2013, Nakano et al., 2007). The overall consensus of these studies has indicated that the measure of accuracy for *S. mutans* genotypic

diversity is variable depending on the methods used and sample populations evaluated. In general, the number of genotypes (amplitypes, sequence types, etc.) observed indicate limited genetic variation except in the case of MLST, which can be more discriminative than traditional gel based typing methods since a single nucleotide changes can result in a new sequence type (ST). The two most frequently used methods in the literature are AP-PCR and rep-PCR, both of which are currently being used in large-scale epidemiological studies of *S. mutans* in high-risk communities (Cheon et al., 2013, Lynch, Villhauer, Warren, Marshall, Dawson, Blanchette, Phipps, Starr and Drake, 2015). Genotyping is a valuable tool for epidemiological surveillance of clinical microorganisms, allowing for the efficient investigation of microbial diversity, micro-evolutionary shifts, and transmission tracking.

Many studies have addressed the question of horizontal and vertical transmission of *S. mutans*. Vertical transmission has been defined as the sharing of microbes from caregiver to child (typically mother-to-child). Horizontal transmission has been defined as the sharing of microbes between members of a similar age group (e.g., child-to-child, sibling-to-sibling) although horizontal transmission may also occur between other family members (e.g. spouse-to-spouse) (Berkowitz, 2006). For clarification purposes, this paper uses the terms mother-to-child, intra-familial (i.e., sharing of genotypes within a family, regardless of age, beyond mother-to-child), and child-to-child (i.e., sharing of genotypes among the unrelated originally recruited children) when discussing transmission. Index child, or when used broadly, children will be used to refer to

the originally recruited child in the study and not other children within the home (i.e., siblings or cousins).

Mother-to-child transmission is widely accepted as the primary route of infection with *S. mutans*, although other sources of shared genotypes have been reported (i.e., father, nursery and school classmates) (Alves et al., 2009, Binks and Duane, 2015, Domejean, Zhan, DenBesten, Stamper, Boyce and Featherstone, 2010, Kozai et al., 1999, Liu, Zou, Shang and Zhou, 2007, Redmo Emanuelsson and Wang, 1998, Tedjosasongko and Kozai, 2002). Studies looking at data for transmission by all three routes together in the same study are lacking. Publications considering other household family members (e.g. extended family members within the residence) are rare. Research investigating intra-familial transmission typically evaluates transmission only within families and not between. Furthermore, in many cases, the children recruited for such studies have no known interactions, thus limiting the potential to observe child-to-child transmission. Studies that focus only on one relationship (i.e., mother-to-child or father-to-child) may assume the member compared is the source when other individuals can be possible sources. The dynamic nature of transmission makes it important to evaluate all possible routes (mother-to-child, intra-familial, and child-to-child) of infection when evaluating transmission since children are typically exposed to all of these source types.

Over the last 8 years, this laboratory has utilized the DiversiLab system, an automated rep-PCR method (Healy et al., 2005), to investigate *S. mutans* strain diversity in two cohorts of children and their household family members in

Uniontown, AL, USA (Cheon, Moser, Whiddon, Osgood, Momeni, Ruby, Cutter, Allison and Childers, 2011, Cheon, Moser, Wiener, Whiddon, Momeni, Ruby, Cutter and Childers, 2013, Moser, Mitchell, Ruby, Momeni, Osgood, Whiddon and Childers, 2010). The advantages of this approach are that the method is kit-based and standardized, has been shown to be reproducible over time and between laboratories, and generates data that are easily shared between laboratories through web-based software (Moser, Mitchell, Ruby, Momeni, Osgood, Whiddon and Childers, 2010, Voets et al., 2013). DiversiLab offers a number of bacterial reference strain libraries for clinical researchers but there currently is no reference library available for *S. mutans*. Therefore, the development of a *S. mutans* library would be beneficial for the global epidemiological study of *S. mutans* associated with dental caries and other diseases (e.g. infective endocarditis).

In addition to rep-PCR analysis, this laboratory also utilizes MLST to validate new representative library rep-PCR genotypes. This approach is consistent with the recommendation that genotypes should be evaluated by an alternate method to validate linage assignments (Foley et al., 2006, van Belkum et al., 2007). Previously, 22 representative rep-PCR library strains were reported for *S. mutans* using MLST (Momeni, Whiddon, Moser, Cheon, Ruby and Childers, 2013). Additional rep-PCR genotypes have been discovered and are evaluated here with MLST to validate strain uniqueness.

The present study has two aims. The first is to determine the overall genetic diversity of *S. mutans* isolates obtained from an 8-year longitudinal study

of children and their household family members using rep-PCR. The second aim is to provide an overview of evidence for mother-to-child, intra-familial, and child-to-child transmission of *S. mutans* using commonality of rep-PCR genotypes for children and their household members.

2. MATERIALS AND METHODS

2.1. Sample collection

The University of Alabama at Birmingham Institutional Review Board approved this study. Adult household family members provided informed consent and waiver of assent for participation of children. Two cohorts of children and their household family members were enrolled over a two-year period as part of a longitudinal epidemiological study of *S. mutans* in rural Uniontown, Alabama, USA. This minority (all index children were African-American) sample population was determined to be high risk for early childhood caries due to low socioeconomic status and high caries prevalence (68.5% of children have caries by 4 years of age) (Ghazal et al., 2014). Cohort 1 (CH1) consisted of 91 schoolchildren (age 5-6 years at enrollment) and their household family members recruited from an elementary school. Cohort 2 (CH2) consisted of 90 infants (age 6-18 months initially) and their household family members recruited at a local community center. Inclusion criteria were that household members must reside in the same residence as the child and that all index children were unrelated. Generally, most family members elected to participate; however,

enrollment was not required. Subjects were provided compensation for their participation along with oral health education and dental hygiene supplies (toothbrushes, toothpaste). In total 594 individuals were evaluated. Household members included mother, father and siblings as well as extended family members (i.e., grandparents, aunts, uncles, cousins). The decayed, missing and filled teeth/surfaces (i.e., dmfs/DMFS) were recorded according to the World Health Organization (WHO) criteria (WHO, 1997) for evaluation of caries activity (i.e., unrestored cavitation) and caries experience.

Periodic (i.e., 6-12 month) samples were collected as previously described (Moser, Mitchell, Ruby, Momeni, Osgood, Whiddon and Childers, 2010). In this study plaque, saliva and tongue scraped samples were evaluated and the preferred sample type was plaque samples if *S. mutans* isolates were available. Plaque samples from children were generally collected with a sterile toothpick from all surfaces of specific permanent or primary molar teeth for Cohort 1 and 2, respectively. Household family members had plaque collected from multiple molar teeth with sterile toothpick for a pooled sample. Caries status of specific teeth were recorded but not used as the criteria for specific teeth collected. Samples were inoculated on mitis salivarius media supplemented with bacitracin and sucrose (MSB, Gold's Media) and incubated anaerobically for 48 hours after which up to 10 presumptive *S. mutans* isolated colonies were selected based on morphological appearance (Gold et al., 1973). Sub-cultured isolates were stored in 80% glycerol (20% culture) at -80˚C until processed with rep-PCR. In total, 30,822 putative *S. mutans* isolates were

collected over 8 years of which 13,906 samples were submitted for DNA extraction and were of sufficient quality to be processed with rep-PCR. Confirmation of *S. mutans* was performed using SYBR Green PCR with primers specific to the *gtfB* gene to screen for *S. mutans* (Moser, Mitchell, Ruby, Momeni, Osgood, Whiddon and Childers, 2010, Yoshida et al., 2003).

2.2. Rep-PCR

Rep-PCR analysis was performed as previously described (Moser, Mitchell, Ruby, Momeni, Osgood, Whiddon and Childers, 2010). In addition to the clinical isolates, various prototype (control) strains of *S. mutans* (Ingbritt C, UA96, OMZ-175, GS-5, kpsp2, LM7, OR22P1, 10449, YT1, NG8, UA159, UA130 & FT1) were also evaluated. Briefly, DNA was isolated using the UltraClean microbial DNA isolation kit (bioMérieux, Inc., Durham, NC). Quality and quantity of DNA was measured to determine acceptable purity using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Walthan, MA). Rep-PCR was performed using the DiversiLab system with the *Streptococcus* DNA fingerprinting kit. PCR was performed on a GeneAmp 9700 with gold heating block (Applied Biosystems). Amplicons were then analyzed with microfluidics LabChip (bioMérieux, Inc.) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The DiversiLab web-based software was used to generate reports including similarity matrix, dendrograms, and electorpherograms. Genotypes were determined using a three minor or 1 major band difference guideline. Percent similarity was not used for determining genotypes as percent similarity

becomes increasing unreliable when larger numbers of isolates are analyzed (Momeni et al., 2015). Major and minor bands were differentiated using a cut-off of 100 fluorescence units.

2.3. Genetic Diversity

The genetic diversity of *S. mutans* was evaluated using rep-PCR genotype data for the 13,906 isolates from children and their household family members. For each child, the number of genotypes, sex, age group (i.e., cohorts), number of isolates, and caries experience were included in the analysis. For household members, only the number of genotypes was evaluated. Analysis per child was performed to determine if sharing of genotypes with mother, father or any household members was associated with children's sex, age group, or caries experience.

2.4. Evaluation of Commonality for Possible Transmission

To investigate potential transmission, data was evaluated two ways: by child and by rep-PCR genotype. Only children with 1 or more household members with *S. mutans* isolates were considered in the analysis (n = 118). One child had no *S. mutans* detected and was therefore excluded from analysis since genotypes were required to be included. Seven mothers had no *S. mutans* detected but were included in statistical analysis because their child did have genotype information. Thus a total of 117 children and their household family members were available for analysis. For each rep-PCR genotype, each child was

evaluated as having either 1) no match with any household member or 2) matching with one or more household member(s). Children with no match to any household member were counted as likely child-to-child transmission since the source is probably not within the household. Children in this group were further evaluated by dividing into three subgroups based on the number of isolates with the genotype (groups: children with only 1 isolate termed “minor genotypes”, children with 2-9 isolates termed “borderline genotypes”, and children with ≥10 isolates termed “established genotypes”). Since the focus of this analysis was based on the genotype perspective, in this paper the term “cases” is used since some children had more than one genotype. For instance, a child may have 3 genotypes that did not match any household member. These would be counted at 3 cases for only 1 child.

Children with a genotype matching a household member were counted as evidence for interfamilial transmission. This group was further evaluated as having either 1) only one household member with the genotype or 2) more than 1 household member with the genotype. Children with more than one household member with the same genotype were considered stronger candidates for interfamilial transmission.

2.5. MLST

In addition to previously reported ST, MLST was performed on 13 newly identified rep-PCR genotype representative Library strains (Momeni, Whiddon, Moser, Cheon, Ruby and Childers, 2013). Briefly, independent PCR reactions

were performed using primer sets specific to the eight housekeeping gene fragments of the MLST Scheme for *Streptococcus mutans* as published by Nakano et al. (Nakano, Lapirattanakul, Nomura, Nemoto, Alaluusua, Gronroos, Vaara, Hamada, Ooshima and Nakagawa, 2007). Resulting amplicons were purified using the UltraClean PCR Clean-up kit (MoBio Laboratories, Inc., Carlsbad, CA) and sequenced by Sanger sequencing using the UAB Heflin Genomic Core Laboratory. Sequence data analysis was performed using the CLC DNA workbench with MLST module version 5.7.1 (CLC bio USA, Cambridge, MA). Sequences were aligned with a reference strain *Streptococcus mutans* UA159. Alleles and ST were compared with the PubMLST database (www.pubmlst.org) for numerical assignment. New alleles were repeated from PCR through sequencing to confirm new sequences prior to submitting to PubMLST. All new alleles and ST were submitted to the PubMLST database and added to GenBank.

2.6. Statistical Analysis

Statistically analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). The maximum caries score observed for all periods were used for caries experience except for children with less than 7 periods of rep-PCR data or for whom caries exam data was unavailable, in which cases the outcome value was imputed by using last observation carried forward (e.g., if a child in CH2 was last seen at age 2 and was caries free, this was recorded as caries free at age 4). Logistic Regression was used for both bivariate and multivariable analysis. For

bivariate analyses, logistic regression modeling was used to assess individually the relationship between each covariate and the outcome. In the multivariable model building, only variables with an overall p-value <0.25 in the bivariate analyses were included. Backward selection was used to eliminate consecutively the variables with p-values >0.1 from the models. The final model contained only variables with p-values <0.1. Changes in Akaike Information Criterion (AIC) were not considered when evaluating the fitted model generated from the preceding step, because AIC is sensitive to changes in sample size associated with adding variables. The outcome variable was defined as the presence of more than one *S. mutans* genotypes (yes/no) across all study visits (e.g., if a child had only one specific genotype at all periods evaluated, this was recorded as a child with a single genotype; or if a child had more than one genotype at any one period, or over different periods, then this was recorded as a child with multiple genotypes). The main explanatory variables were caries at ages 12 and 4 years (for CH1 and CH2, respectively), cohort, sex, matching the genotype with household member and number of isolates group (isolates group were defined as group 1 (0-9 isolates), group 2 (10-29 isolates), group 3 (30-82 isolates), and group 4 (more than 83 isolates)

3. RESULTS

3.1. Genetic Diversity

Thirty-four unique rep-PCR genotypes were observed for 13,906 *S. mutans* isolates from 594 individuals (Fig. 1). The total number of individuals in Fig. 1 is higher than the total from CH1 and CH2 because individuals can have more than one genotype. *S. mutans* genotypes were reported for 82 children and 191 household family members (mean 3.58, range 0-11 members) in CH1 and 87 children and 234 household family members (mean 3.13, range 0-10 members) in CH2. Of these children, 30 CH1 and 13 CH2 children had no household members evaluated. The most frequently occurring genotypes were G18 (2,110 isolates), G09 (1,686 isolates), and G01a (1,363 isolates). The least reported genotypes were G21 (3), G42 (4) and G38 (5). G09 and G18 were the most prevalent genotypes, being found in 174 and 133 different individual subjects respectively. Four genotypes (G27, G10, G42, and G21) occurred in only one individual each, although the number of isolates with these genotypes found in these individuals ranged from 3-112. For the more common *S. mutans* prototype strains, UA159 demonstrated a rep-PCR genotype similar to G01b, Ingbritt C was most like G15, LM7 was similar to G23, and OMZ-175 and GS-5 matched G18.

A summary of children with single versus multiple genotypes is compared based on sex, age group, and caries experience in Table 1. Although more males were included in the study (91 males versus 77 females), the ratio of

children with multiple genotypes, as evaluated by sex, were comparable. A comparable number of children were included in each age group (CH1 and CH2); however, a greater percentage of younger children had multiple genotypes (72% for CH2 versus 58% in CH1). A larger number of children in this population had caries experience (127 children versus 41 children in the caries free group). As would be expected, children in the caries free group had far fewer *S. mutans* isolates available and were less likely to have multiple genotypes (60% presented only 1 genotype).

Table 2 provides the results of the bivariate and multivariable statistical analysis for having multiple genotypes based on sex, age group (cohorts), caries status, and number of isolates analyzed. Younger children were significantly more likely to have multiple genotypes than older children ($p= 0.053$) and having multiple genotypes was highly associated with dental caries experience ($p <0.001$). The probability for having multiple genotypes was significantly higher for children having great than 10 isolates evaluated ($p <0.001$). When applying multivariable analysis, the younger age group remained significant, as did having greater than 10 isolates. Accounting for the number of isolates evaluated between the caries groups for the multivariable analysis indicated that the p-value for multiple genotypes was approaching significance with an odds ratio of 2.6. Having multiple genotypes was not significantly different by gender. No significant associations were observed for sharing of genotypes with mother, father or any household members based on sex, age group, or caries experience.

Interestingly, having multiple genotypes did not improve the odds ratio of matching any household members (data not shown, p=0.2070, OR 0.548).

3.2. Commonality for Possible Transmission

Evidence for transmission is presented first as an overall basic summary for the children and then as individual genotype cases (since children may have multiple genotypes).

Overall, 117 children and their household members were evaluated in this part of the study representing 13,124 *S. mutans* isolates. Index children with more than one household member were evaluated for 76% (89/117, mean 3.24, range 1-11) of families. This analysis included 50 children from CH1 and 67 from CH2 (Table S1). Of these, 27 children (23%, 27/117) shared no genotypes with any household members. Eighty-four children (72%, 84/117) had at least one genotype not shared with any household members (genotypes range 1-5 per child); however this data does not consider if genotypes are minor or established. Analysis also found 33 children (28%, 33/117) that shared all genotypes (range 1-4 genotypes) within the household. The number of genotypes per child ranged from 1-9 genotypes with an average of 2.72 genotypes per child. There were 75 children (63%, 75/105; total number adjusted for children with mothers) that shared at least 1 genotype with their mothers.

Table 3 (Table S2) lists 157 cases among 21 genotypes where a genotype occurred in a child, but did not occur in any of the corresponding household members. The number of cases is higher than the number of families evaluated

(n=117) because each genotype was considered independently and thus a child may have multiple genotypes that did not match any household members. Thirteen genotypes were excluded from the table because 1 genotype (G19) did not match the profile for child-to-child transmission (intra-familial only), 4 genotypes (G10, G20, G24, G45) had only 1 child with no matching household members and did not match other children, and 8 genotypes did not have any *S. mutans* isolates from children (household member only; G03, G17, G21, G25, G27, G30, G38, G42). Three genotypes (G01d, G02, G11) had only one child in Table S2 but these matched other children that were excluded from this study due to the lack of household members for comparison but were included in the counts for child-to-child transmission.

For the children that did not share a genotype with a household family member, further analysis was done to evaluate minor, undetermined, and established genotypes (Fig. 2). Fifty-three cases (34%, 53/157) were considered minor or transient genotypes since only one isolate of these genotypes were found in one or more children. Cases where a child had 2-9 isolates of a genotype were considered borderline, thus, 57 cases (36%, 57/157) were treated as having undetermined significance. However, there were 47 cases (30%, 47/157) in which the child presented with ≥10 isolates of the genotype (established genotypes); 37 (24% 37/157) of these cases had more than one household family member evaluated.

Table 3 (and Table S3) lists the 158 cases among 21 genotypes where children and one or more household members shared a genotype. In 55%

(87/158) of these cases more than one household member shared the genotype with the child. Thirteen genotypes were excluded from the table. Of these 4 genotypes, (G20, G24, G45, and G49) had profiles that did not match possible intra-familial transmission (child-to-child only); 6 genotypes (G03, G17, G21, G25, G30, and G38) had no children to compare with household members and 3 genotypes (G10, G27, G42) were excluded because these genotypes had no household members to compare with the child. Of these 87 cases, 44 were cases where the child shared the genotype with 2 household members, 21 with 3 household members, 12 with 4 household members, and 10 with ≥ 5 household members.

The distribution of the shared genotypes among family members can been seen in Fig. 3 (and Table S3). It is possible that a child shared a genotype with multiple siblings, grandmothers, aunts, uncles, and/or cousins. Children sharing a genotype with any respective household member were counted only once. For instance, if a child had genotype G01 in common with 5 siblings and 3 cousins, they were counted as only 1 sibling case and 1 cousin case. Note that totals for the individual household members will not add to the total number of cases since a child may share the genotype with multiple family members. However, totals for each household member type are listed to indicate the number of cases in which each household member type was shared.

Fig. 4 compares the number of children with evidence for child-to-child and intra-familial transmission by rep-PCR. Six genotypes (G01b, G05, G14, G15, G22 and G26) appear to be predominately shared among the children

(child-to-child transmission). In contrast, 4 genotypes (G09, G11, G18, and G50) were predominately shared within households (intra-familial transmission). However, these differences were not statistically significant by multiple Chi Square tests (data not shown).

3.3. Multilocus Sequence Typing

Following MLST for 13 newly identified genotypes, 7 new alleles and 7 new ST were identified (ST 184-190) in this study for the rep-PCR representative Library strains (Table 4). The 7 new alleles were added to Genbank (Accession numbers KU216130-KU216136). All new ST and alleles were added to the PubMLST database. Nine of the 13 new ST were found to be unique. The other 4 strains matched previously reported ST for other Library or clinical strains (NL01b matched L5; NL03 matched L18, and NL26 matched L22, and NL27 matched G13-4).

4. DISCUSSION

4.1. Genetic Diversity

Several studies using a variety of methods have previously evaluated the genetic diversity of *S. mutans*. AP-PCR has been employed most frequently; however, this methodology raises concerns of reproducibility between laboratories and variability of interpretation approaches (Bidet et al., 2000, van Belkum, Tassios, Dijkshoorn, Haeggman, Cookson, Fry, Fussing, Green, Feil, Gerner-Smidt,

Brisse and Struelens, 2007). In the present study, the genotypic diversity of *S. mutans* using the DiversiLab rep-PCR approach, which has been found to be highly reproducible between laboratories, is summarized.

In this study 34 unique *S. mutans* rep-PCR genotypes from 13,906 isolates are reported (Fig. 1). The diversity of genotypes observed in this study is generally similar to other large-scale studies of *S. mutans* using AP-PCR (Lynch, Villhauer, Warren, Marshall, Dawson, Blanchette, Phipps, Starr and Drake, 2015, Palmer, Nielsen, Peirano, Nguyen, Vo, Nguyen, Jackson, Finlayson, Sauerwein, Marsh, Edwards, Wilmot, Engle, Peterson, Maier and Machida, 2012). While some studies have reported more amplitypes with AP-PCR from smaller sample pools, it should be noted in all gel based genotyping studies there is a degree of variability due to interpretation, techniques and conditions employed. This inconsistency makes direct comparison between studies difficult.

The present study is unique in its scale with over 13,000 isolates evaluated. However, due to improved reproducibility with the rep-PCR method, only 34 genotypes were identified, which resulted in extensive commonality of isolates among children and household family members that may not be seen with other methods. This commonality, although improving consistency, complicated the ability to evaluate if transmission was exclusive for genotypic matching of individuals, whether intra- or extra-familial. It is important to note that the results of rep-PCR here are based on similar banding patterns and not percent similarity. For instance, 2,110 isolates are reported for G18 and all of these isolates are uniform in their rep-PCR gel images. In addition the same ST

(ST166) has been found for all isolates tested for G18 with MLST in other studies (Momeni et al., 2015). However, due to the allowance for up to 2 minor band differences, other genotypes, such as G12, demonstrate subtle genetic drift as more isolates are evaluated. In order to gain meaningful data, the 3 minor bands, 1 major band guideline was necessary to allow for some grouping of isolates. Without this guideline, most isolates would have been unique genotypes and no valuable conclusions could have been made. Thus the balance of commonality and diversity of genotypes should be considered.

Some studies have reported that having multiple genotypes is not significantly associated with dental caries using AP-PCR and random fragment length polymorphism (Kreulen et al., 1997, Lembo, Longo, Ota-Tsuzuki, Rodrigues and Mayer, 2007, Liu, Zou, Shang and Zhou, 2007, Mattos-Graner et al., 2001). However, our finding that children with multiple genotypes were about 4.5 times more likely to have caries than children with a single genotype (Table 2) is consistent with more recent reports using AP-PCR and multilocus enzyme electrophoresis (Alaluusua and Matto, 1996, Gamboa, Chaves and Valdivieso, 2010, Napimoga, Kamiya, Rosa, Rosa, Hofling, Mattos-Graner and Goncalves, 2004, Pieralisi, Rodrigues, Segura, Maciel, Ferreira, Garcia and Poli-Frederico, 2010, Zhao, Li, Lin, Chen and Yu, 2014, Zhou, Qin, Qin and Ge, 2011). This is the first report of this association using rep-PCR and combined with the findings of the recent studies would seem to suggest that having multiple *S. mutans* genotypes may be an important factor in dental caries development. Even when multivariable analysis was performed to account for the number of isolates

analyzed per child and sex, children with multiple genotypes remained 2.3 times more like to have dental caries. However, the results of the present study are only an association. Further study is planned to evaluate this association in more detail using a longitudinal approach to the data, as well as, in vitro and in vivo studies evaluating the potential for additive or synergistic effects of multiple genotypes (e. g., biofilm and metabolomics studies).

No significant difference was observed between the number of genotypes and sex, which is consistent with other studies (Liu, Zou, Shang and Zhou, 2007, Pieralisi, Rodrigues, Segura, Maciel, Ferreira, Garcia and Poli-Frederico, 2010). Li and Caufield previously reported that female infants are more likely to share genotypes with their mothers than male children (Li and Caufield, 1995). However, in the present study, no significant difference was observed between sharing of genotypes with the mother and sex. This study also reports a significant difference in the number of genotypes observed between age groups, differing from a previous report (Mattos-Graner, Li, Caufield, Duncan and Smith, 2001). However this difference may be because the age groups evaluated by Mattos-Graner *et al.* (2001) were all under 2.5 years, whereas the current study evaluated infants and children with different age ranges (infants and children).

4.2. Evidence for Transmission

Evidence for transmission by rep-PCR was evaluated by querying whether a given genotype was shared with household family members. (Tables 3, S1, S2) The goal was to estimate, in broad terms, the number of cases of potential extra-

familial and intra-familial transmission as estimated by rep-PCR analysis. Generally, no weight was given to the number of household members (except as described in the results) or time of acquisition. The results of this study are intended as an overview to determine if a more detailed longitudinal analysis is warranted and to provide potential sample pools for additional transmission analysis. Further study relating the rep-PCR genotypes found in this study with transmission on an individual child basis for evaluation of longitudinal (multiple collection periods) with clinical features (e.g. genotypes related to caries) and detailed household associations are beyond the scope of the current manuscript but are planned as future studies.

The data in this study is presented in two ways: by child and by genotype. The purpose of adding the analysis by genotype is an important addition since other studies make a potentially faulty assumption that when multiple genotypes are found in a child with caries that all the genotypes isolated from that child are associated with caries. Analysis using only the child also biases the data to underreport the number of genotypes not shared, thus failing to consider other sources. For instance, many studies base their analysis on a simple match/no match with mother. By this approach if any genotype is shared with the mother, the case is consider mother-to-child transmission when the child may have other genotypes not shared with the mother that are not represented in the analysis.

As reported in Table S1, a total of 117 children and their household family members were evaluated for transmission. The total number of children was adjusted for comparison with mothers (n=105) since not all children had mothers

participating in the study and 1 child was eliminated because no *S. mutans* were recovered. The number of genotypes in this study was 1-9 per child whereas others have reported that children harbor between 1-6 genotypes (Alves, Nogueira, Stipp, Pampolini, Moraes, Goncalves, Hofling, Li and Mattos-Graner, 2009, Cheon, Moser, Whiddon, Osgood, Momeni, Ruby, Cutter, Allison and Childers, 2011, Cheon, Moser, Wiener, Whiddon, Momeni, Ruby, Cutter and Childers, 2013, Domejean, Zhan, DenBesten, Stamper, Boyce and Featherstone, 2010, Liu, Zou, Shang and Zhou, 2007, Lynch, Villhauer, Warren, Marshall, Dawson, Blanchette, Phipps, Starr and Drake, 2015, Pieralisi, Rodrigues, Segura, Maciel, Ferreira, Garcia and Poli-Frederico, 2010). The average number of genotypes per child in this study was 2.7 and is notably higher than what others have reported for children (range 1.5-2.0) (Cheon, Moser, Whiddon, Osgood, Momeni, Ruby, Cutter, Allison and Childers, 2011, Cheon, Moser, Wiener, Whiddon, Momeni, Ruby, Cutter and Childers, 2013, Domejean, Zhan, DenBesten, Stamper, Boyce and Featherstone, 2010, Liu, Zou, Shang and Zhou, 2007, Lynch, Villhauer, Warren, Marshall, Dawson, Blanchette, Phipps, Starr and Drake, 2015). This finding is significant since the presence of more genotypes in children is linked with higher caries scores possibly due to inter-species competition or augmentation. (Alaluusua and Matto, 1996, Baca, Castillo, Liebana, Castillo, Martin-Platero and Liebana, 2012, Hirose et al., 1993, Napimoga, Kamiya, Rosa, Rosa, Hofling, Mattos-Graner and Goncalves, 2004, Pieralisi, Rodrigues, Segura, Maciel, Ferreira, Garcia and Poli-Frederico, 2010, Zhao, Li, Lin, Chen and Yu, 2014).

Analysis of genotypes resulted in a total of 157 cases where a child presented a genotype with no matching household family members (Table S2). This is the largest known reported number of extra-familial horizontal cases. Other studies have reported horizontal transmission between children at lower rates, and these studies typically focused on cohorts of children (in nurseries or classes) without looking at households (Alves, Nogueira, Stipp, Pampolini, Moraes, Goncalves, Hofling, Li and Mattos-Graner, 2009, Domejean, Zhan, DenBesten, Stamper, Boyce and Featherstone, 2010, Liu, Zou, Shang and Zhou, 2007, Mattos-Graner, Li, Caufield, Duncan and Smith, 2001). For instance, Alves et al. reported 14 pairs of children in 7 nurseries shared an AP-PCR amplitype, although only 4 were confirmed by RFLP. In contrast, the present data is more compelling since it evaluates the commonality of the genotypes among children with more than 1 household family member for 76% of children (89/117).

The 27 children (23%, 27/117) and 157 cases of genotypes where the child had a genotype not shared with any household members indicates that the source of transmission is probably not within the household (Tables S1 and S2). Thus, given the high frequency of commonality of genotypes among children in this study, child-to-child transmission with classmates was considered as a probable alternative source. Considering minor versus established genotypes, the 53 cases of minor genotypes (1 isolate only) may be less significant as these bacteria occur in low numbers, indicating these isolates are likely transient and may not be clinically relevant to established caries infections. Although

longitudinal analysis was not a part of the current analysis, transience was assumed since only 1 isolate was observed, though as many as 7 periods were collected for most children. Assuming these strains are not clinically relevant suggest that at least 33% of strains (potentially more if considering the borderline group) may be confounding the search for strains related to disease. Typical analysis in genotyping studies includes all isolates, however the current findings seem to indicate that an adjusted analysis should also be performed. It further demonstrates that *S. mutans* strains are highly transmissible but may not become established strains due to bacterial competition or host immune factors.

In contrast to the transient isolates observed above, the 47 cases where ≥10 isolates were found for children with no matching household members seems to be strong evidence consistent with child-to-child transmission. These genotypes were considered established genotypes since more than 10 isolates indicates these genotypes were generally observed for at least 2 periods since only 10 isolates were evaluated per period. The strength of this data is further supported by the evaluation of multiple household members, in particular for those children with ≥10 isolates where 37 of the 47 children had more than 1 household member tested. The 4 cases where children had genotypes that did not match household members or other children are considered independent cases that would suggest another reason (i.e., source of genotype was beyond the household members or classmates evaluated in this study).

For children that shared a genotype with their household, it was found that these children most commonly shared genotypes with their mothers (54%;

85/158) and their siblings (46%; 72/158) (Table S3, Fig. 2). The next highest group was 23% (37/158) who also shared a genotype with cousins. Interestingly, of the 73 cases that did not share the genotype with the mother, 37 (51%, 37/73 or 23%, 37/158) shared the genotype only with sibling(s) and/or cousin(s). This data is consistent with the 58% strain homology for children and siblings reported by Tedjosasongko and Kozai (2002). It also suggests that children may be acquiring genotypes from other children within the household who likely acquire the genotype through their classmates or other outside interactions since no adults within the household demonstrated the genotypes. However, further analysis is needed to determine which direction the genotype is shared within households since the addition of a time of acquisition component would be required.

The high number of cases (58%; 87/158) in which a child shared a genotype with more than one household family member is consistent with intra-familial transmission by rep-PCR. Since the commonality of genotypes is frequent in this study, child-to-child transmission for the these children cannot be ruled out entirely; however for the purpose of this paper, it is logical to hypothesize that when a child shares the genotype with more than one family member it is more probable that the child acquired the genotype within the home.

4.3. Limitations and Future Directions

While every effort was made to recruit all household family members residing in the home, in rare instances, some household members choose to not to

participate. In particular; fathers, grandparents, aunts and uncles were available less often typically because they were either absent from the household or less frequently, because they were unavailable for sampling. *S. mutans* were collected by standard plate count method, which did not always result in recovery of *S. mutans* isolates from every household member. Analysis of standard plate counts are beyond the scope of the current study but partial data has been published elsewhere and a more comprehensive report is planned (Childers et al., 2011). The purpose of the current analysis was to provide a broad overview of genotype diversity and possible transmission so as to determine if a more detailed longitudinal analysis of children and household members is warranted. Fewer isolates were typically collected for household members (average isolates per child 47, median 31; average per family member 13, median 13). Also, rep-PCR is based on banding similarity, which allows a degree of variability within rep-PCR genotypes. However, closer examination of genotypes shows sub-groupings within the genotypes that may be further distinguished using an alternate typing method such as MLST. Thus, the 157 cases of possible child-to-child transmission observed by rep-PCR analysis in this study may be overestimated in some cases. Results will be better refined through future additional analysis with MLST.

4.4. MLST Library Update

Previously, we used MLST to validate the uniqueness of rep-PCR genotypes reported for 22 representative Library strains (Momeni, Whiddon, Moser, Cheon, Ruby and Childers, 2013). In that study, all 22 stains were found to have unique MLST ST, which was anticipated since MLST is typically a more discriminative method. Here another 13 different representative Library rep-PCR genotypes have been identified (Table 4). As in the first report, 9 additional Library strains were confirmed as unique by MLST. However, in contrast to the previous findings, 4 of the new Library strains (NL01b, NL03, NL26 and NL27) in this study matched previously reported ST, 3 of which were previous library strains. This finding indicates that these strains have identical housekeeping genes but differ within the repetitive extragenic elements. This finding offers more support of the recent report that rep-PCR can be more discriminative than MLST in some cases when evaluating *S. mutans* (Momeni, Whiddon, Cheon, Moser and Childers, 2015). It should also be noted that one of the previous library genotypes (L16) was reassigned to L23 due to similarities in rep-PCR profiles. Thus, the current total of Library strains evaluated with MLST agrees with the number of rep-PCR genotypes at 34.

5. CONCLUSIONS

In summary, this study presents a DiversiLab rep-PCR reference library for *S. mutans* for epidemiological surveillance. This study highlights the effective use of the DiversiLab rep-PCR approach as a cost-effective screening tool for large-scale epidemiological study of bacterial isolates. This is the first known study to evaluate transmission by all three possible routes of transmission (unrelated child-to-child, mother-to-child, and interfamilial including extended family) within the same study. Although this study's findings support mother-to-child transmission, it also provides compelling evidence that child-to-child transmission is observed in 37% of children and for 50% of genotypes of *S. mutans* in this population. The high degree of child-to-child transmission reported warrants further study with an alternate method such as MLST to validate possible routes of transmission. The significance of age group and caries status related to genotype diversity reported here indicates that a more through, longitudinal analysis is justified. These findings advocate the importance of considering child-to-child, as well as intra-familial, acquisition of *S. mutans* in risk assessments and prevention strategies for dental caries.

ACKNOWLEDGEMENTS

This work was supported by National Institute of Dental and Craniofacial Research (NIDCR) Grant DE016684 (NKC). Ms. Momeni is a Dental Academic Research Training (DART) Pre-doctoral Fellow under NIDCR Institutional

Grant#T-90 DE022736. This research is investigator initiated therefore the sponsor had no involvement in the scientific study or preparation for this manuscript. We especially appreciate all the clinical and laboratory participants of this study. Dr. Mei Han of the UAB Heflin Genomic Core provided sequencing data; Dr. Jinthana Lapirattanakul directed the confirmation and registration of isolates with PubMLST; Dr. Ping Zhang and Dr. Kazuhiko Nakano generously donated some controls strains. The authors declare no conflicts of interest.

REFERENCES

Alaluusua, S., Matto, J., 1996. Oral colonization by more than one clonal type of mutans streptococcus in children with nursing-bottle dental caries. *Arch Oral Biol*. 41, 167-173.

Alves, A.C., Nogueira, R.D., Stipp, R.N., Pampolini, F., Moraes, A.B., Goncalves, R.B., Hofling, J.F., Li, Y., Mattos-Graner, R.O., 2009. Prospective study of potential sources of *Streptococcus mutans* transmission in nursery school children. J Med Microbiol. 58, 476-481.

Baca, P., Castillo, A.M., Liebana, M.J., Castillo, F., Martin-Platero, A., Liebana, J., 2012. Horizontal transmission of *Streptococcus mutans* in schoolchildren. Med Oral Patol Oral Cir Bucal. 17, e495-500.

Berkowitz, R.J., 2006. Mutans streptococci: acquisition and transmission. Pediatr Dent. 28, 106-109; discussion 192-108.

Bidet, P., Lalande, V., Salauze, B., Burghoffer, B., Avesani, V., Delmee, M., Rossier, A., Barbut, F., Petit, J.C., 2000. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. J Clin Microbiol. 38, 2484-2487.

Binks, C., Duane, B., 2015. Mother-to-child transmission of *Streptococcus mutans*. Evid Based Dent. 16, 39-40.

Cheon, K., Moser, S.A., Whiddon, J., Osgood, R.C., Momeni, S., Ruby, J.D., Cutter, G.R., Allison, D.B., Childers, N.K., 2011. Genetic Diversity of Plaque Mutans Streptococci with rep-PCR. J Dent Res. 90, 331-335.

Cheon, K., Moser, S.A., Wiener, H.W., Whiddon, J., Momeni, S.S., Ruby, J.D., Cutter, G.R., Childers, N.K., 2013. Characteristics of *Streptococcus mutans* genotypes and dental caries in children. Eur J Oral Sci. 121, 148-155.

Childers, N.K., Osgood, R.C., Hsu, K.L., Manmontri, C., Momeni, S.S., Mahtani, H.K., Cutter, G.R., Ruby, J.D., 2011. Real-time quantitative polymerase chain reaction for enumeration of *Streptococcus mutans* from oral samples. Eur J Oral Sci. 119, 447-454.

Cogulu, D., Sabah, E., Uzel, A., Ozkinay, F., 2006. Genotyping of *Streptococcus mutans* by using arbitrarily primed polymerase chain reaction in children with Down Syndrome. Arch Oral Biol. 51, 177-182.

Do, T., Gilbert, S.C., Clark, D., Ali, F., Fatturi Parolo, C.C., Maltz, M., Russell, R.R., Holbrook, P., Wade, W.G., Beighton, D., 2010. Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. PLoS One. 5, e9073.

Domejean, S., Zhan, L., DenBesten, P.K., Stamper, J., Boyce, W.T., Featherstone, J.D., 2010. Horizontal transmission of mutans streptococci in children. J DENT RES. 89, 51-55.

Foley, S.L., White, D.G., McDermott, P.F., Walker, R.D., Rhodes, B., Fedorka-Cray, P.J., Simjee, S., Zhao, S., 2006. Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. J Clin Microbiol. 44, 3569-3577.

Gamboa, F., Chaves, M., Valdivieso, C., 2010. Genotypic profiles by AP-PCR of *Streptococcus mutans* in caries-active and caries-free preschoolers. Acta odontologica latinoamericana : AOL. 23, 143-149.

Ghazal, T., Levy, S.M., Childers, N.K., Broffitt, B., Cutter, G., Wiener, H.W., Kempf, M., Warren, J., Cavanaugh, J., 2014. Prevalence and incidence of early childhood caries among African-American children in Alabama. J Public Health Dent.

Gilbert, K., Joseph, R., Vo, A., Patel, T., Chaudhry, S., Nguyen, U., Trevor, A., Robinson, E., Campbell, M., McLennan, J., Houran, F., Wong, T., Flann, K., Wages, M., Palmer, E.A., Peterson, J., Engle, J., Maier, T., Machida, C.A., 2014. Children with severe early childhood caries: streptococci genetic strains within carious and white spot lesions. J Oral Microbiol. 6.

Gold, O.G., Jordan, H.V., Van Houte, J., 1973. A selective medium for *Streptococcus mutans*. Arch Oral Biol. 18, 1357-1364.

Gross, E.L., Beall, C.J., Kutsch, S.R., Firestone, N.D., Leys, E.J., Griffen, A.L., 2012. Beyond Streptococcus mutans: dental caries onset linked to multiple species by 16S rRNA community analysis. PLoS One. 7, e47722.

Healy, M., Huong, J., Bittner, T., Lising, M., Frye, S., Raza, S., Schrock, R., Manry, J., Renwick, A., Nieto, R., Woods, C., Versalovic, J., Lupski, J.R., 2005. Microbial DNA typing by automated repetitive-sequence-based PCR. J Clin Microbiol. 43, 199-207.

Hirose, H., Hirose, K., Isogai, E., Miura, H., Ueda, I., 1993. Close association between *Streptococcus sobrinus* in the saliva of young children and smooth-surface caries increment. Caries Res. 27, 292-297.

Hu, D., Cui, W., Luo, Y., Yang, J., Deng, B., Xu, J., Chu, B., Wang, C., 2014. [Horizontal transmission of *Streptococcus mutans* in caries-active preschool children]. Nan Fang Yi Ke Da Xue Xue Bao. 34, 636-640.

Kozai, K., Nakayama, R., Tedjosasongko, U., Kuwahara, S., Suzuki, J., Okada, M., Nagasaka, N., 1999. Intrafamilial distribution of mutans streptococci in Japanese families and possibility of father-to-child transmission. Microbiol Immunol. 43, 99-106.

Kreulen, C.M., de Soet, H.J., Hogeveen, R., Veerkamp, J.S., 1997. *Streptococcus mutans* in children using nursing bottles. ASDC J Dent Child. 64, 107-111.

Lapirattanakul, J., Nakano, K., Nomura, R., Hamada, S., Nakagawa, I., Ooshima, T., 2008. Demonstration of mother-to-child transmission of *Streptococcus mutans* using multilocus sequence typing. Caries Res. 42, 466-474.

Lembo, F.L., Longo, P.L., Ota-Tsuzuki, C., Rodrigues, C.R., Mayer, M.P., 2007. Genotypic and phenotypic analysis of *Streptococcus mutans* from different oral cavity sites of caries-free and caries-active children. Oral Microbiol Immunol. 22, 313-319.

Li, Y., Caufield, P.W., 1995. The fidelity of initial acqusition of mutans streptococci by infants from their mothers. J Dent Res. 74, 681-685.

Li, Y., Caufield, P.W., 1998. Arbitrarily primed polymerase chain reaction fingerprinting for the genotypic identification of mutans streptococci from humans. Oral Microbiol Immunol. 13, 17-22.

Liu, Y., Zou, J., Shang, R., Zhou, X.D., 2007. Genotypic diversity of *Streptococcus mutans* in 3- to 4-year-old Chinese nursery children suggests horizontal transmission. Arch Oral Biol. 52, 876-881.

Loesche, W., 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol Rev. 50, 353-380.

Lynch, D.J., Villhauer, A.L., Warren, J.J., Marshall, T.A., Dawson, D.V., Blanchette, D.R., Phipps, K.R., Starr, D.E., Drake, D.R., 2015. Genotypic characterization of initial acquisition of *Streptococcus mutans* in American Indian children. J Oral Microbiol. 7, 27182.

Mattos-Graner, R.O., Li, Y., Caufield, P.W., Duncan, M., Smith, D.J., 2001. Genotypic diversity of mutans streptococci in Brazilian nursery children suggests horizontal transmission. J Clin Microbiol. 39, 2313-2316.

Mineyama, R., Yoshino, S., Fukushima, K., 2004. Genotypic analysis of strains of mutans streptococci by pulsed-field gel electrophoresis. Microbiol Res. 159, 181-186.

Mineyama, R., Yoshino, S., Maeda, N., 2007. DNA fingerprinting of isolates of *Streptococcus mutans* by pulsed-field gel electrophoresis. Microbiol Res. 162, 244-249.

Momeni, S.S., Whiddon, J., Cheon, K., Moser, S.A., Childers, N.K., 2015. Assessment of clonality and serotypes of *Streptococcus mutans* among children by multilocus sequence typing. Eur J Oral Sci.

Momeni, S.S., Whiddon, J., Cheon, K., Moser, S.A., Childers, N.K., 2015. Assessment of two multilocus sequence typing (MLST) schemes available for *Streptococcus mutans*. Arch Oral Biol. 60, 1769-1776.

Momeni, S.S., Whiddon, J., Moser, S.A., Cheon, K., Ruby, J.D., Childers, N.K., 2013. Comparative genotyping of *Streptococcus mutans* by repetitive extragenic palindromic polymerase chain reaction and multilocus sequence typing. Mol Oral Microbiol. 28, 18-27.

Moser, S.A., Mitchell, S.C., Ruby, J.D., Momeni, S., Osgood, R.C., Whiddon, J., Childers, N.K., 2010. Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. J Clin Microbiol. 48, 599-602.

Nakano, K., Lapirattanakul, J., Nomura, R., Nemoto, H., Alaluusua, S., Gronroos, L., Vaara, M., Hamada, S., Ooshima, T., Nakagawa, I., 2007. *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. J Clin Microbiol. 45, 2616-2625.

Napimoga, M.H., Kamiya, R.U., Rosa, R.T., Rosa, E.A., Hofling, J.F., Mattos-Graner, R., Goncalves, R.B., 2004. Genotypic diversity and virulence traits of *Streptococcus mutans* in caries-free and caries-active individuals. J Med Microbiol. 53, 697-703.

Palmer, E.A., Nielsen, T., Peirano, P., Nguyen, A.T., Vo, A., Nguyen, A., Jackson, S., Finlayson, T., Sauerwein, R., Marsh, K., Edwards, I., Wilmot, B., Engle, J., Peterson, J., Maier, T., Machida, C.A., 2012. Children with severe early childhood caries: pilot study examining mutans streptococci genotypic strains after full-mouth caries restorative therapy. Pediatr Dent. 34, e1-10.

Pieralisi, F.J., Rodrigues, M.R., Segura, V.G., Maciel, S.M., Ferreira, F.B., Garcia, J.E., Poli-Frederico, R.C., 2010. Genotypic Diversity of *Streptococcus mutans* in Caries-Free and Caries-Active Preschool Children. Int J Dent. 2010, 824976.

Redmo Emanuelsson, I.M., Carlsson, P., Hamberg, K., Bratthall, D., 2003. Tracing genotypes of mutans streptococci on tooth sites by random amplified polymorphic DNA (RAPD) analysis. Oral Microbiol Immunol. 18, 24-29.

Redmo Emanuelsson, I.M., Wang, X.M., 1998. Demonstration of identical strains of mutans streptococci within Chinese families by genotyping. Eur J Oral Sci. 106, 788-794.

Roberts, M.C., Riedy, C.A., Coldwell, S.E., Nagahama, S., Judge, K., Lam, M., Kaakko, T., Castillo, J.L., Milgrom, P., 2002. How xylitol-containing products affect cariogenic bacteria. J Am Dent Assoc. 133, 435-441; quiz 492-433.

Tanner, A.C., Milgrom, P.M., Kent, R., Jr., Mokeem, S.A., Page, R.C., Riedy, C.A., Weinstein, P., Bruss, J., 2002. The microbiota of young children from tooth and tongue samples. J Dent Res. 81, 53-57.

Teanpaisan, R., Chaethong, W., Piwat, S., Thitasomakul, S., 2012. Vertical transmission of mutans streptococci and lactobacillus in Thai families. Pediatr Dent. 34, e24-29.

Tedjosasongko, U., Kozai, K., 2002. Initial acquisition and transmission of mutans streptococci in children at day nursery. ASDC J Dent Child. 69, 284-288, 234-285.

van Belkum, A., Tassios, P.T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N.K., Fussing, V., Green, J., Feil, E., Gerner-Smidt, P., Brisse, S., Struelens, M., 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 13 Suppl 3, 1-46.

Voets, G.M., Leverstein-van Hall, M.A., Kolbe-Busch, S., van der Zanden, A., Church, D., Kaase, M., Grisold, A., Upton, M., Cloutman-Green, E., Canton, R., Friedrich, A.W., Fluit, A.C., 2013. International multicenter evaluation of the DiversiLab bacterial typing system for *Escherichia coli* and *Klebsiella* spp. J Clin Microbiol. 51, 3944-3949.

WHO, 1997. Oral Health surveys: basic methods, World Health Organization, Geneva.

Yoshida, A., Suzuki, N., Nakano, Y., Kawada, M., Oho, T., Koga, T., 2003. 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. 41, 4438-4441.

Zhao, W., Li, W., Lin, J., Chen, Z., Yu, D., 2014. Effect of sucrose concentration on sucrose-dependent adhesion and glucosyltransferase expression of *S.*

mutans in children with severe early-childhood caries (S-ECC). Nutrients. 6, 3572-3586.

Zhou, Q., Qin, X., Qin, M., Ge, L., 2011. Genotypic diversity of *Streptococcus mutans* and *Streptococcus sobrinus* in 3-4-year-old children with severe caries or without caries. Int J Paediatr Dent. 21, 422-431.

Table 1. Distribution of children with single or multiple *S. mutans* rep-PCR genotypes by sex, age group, and caries experience.

	No. of isolates tested	No. of genotypes detected	1 genotype	>1 genotype
Sex				
Male (N = 91)	4024	24	29 (31.9%)	62 (68.1%)
Females (N = 77)	4014	26	29 (37.7%)	48 (62.3%)
Age Group				
Infants (Cohort 2, N = 87)	3108	24	24 (27.6%)	63 (72.4%)
Children (Cohort 1, N = 81)	4930	27	34 (42.0%)	47 (58.0%)
Caries Experience				
Caries Free (N = 41)	684	17	25 (60.1%)	16 (39.0%)
Caries (N = 127)	7354	28	33 (26.0%)	94 (74.0%)

Table 2. Results of bivariate and multivariable logistic regression modeling the probability of children (N = 168) having multiple genotypes.

A. BIVARIATE ANALYSIS				
Variable: Reference	**Level (% with multiple genotype)**	**Odds Ratio**	**95%CI**	**Overall p-value**
Age: Cohort 2	Cohort 1 (58.02%) Cohort 2 (72.41%)	1.899	0.996-3.618	0.053
Sex: Male	Female (62.34) Male (68.13)	1.292	0.683-2.444	0.431
HH genotype (N =125): Match	No (79.38%) Yes (67.86%)	0.548	0.216-1.394	0.207
	Group 1 (48.78%)			
Isolates group 2 vs. 1	Group 2 (35.71%)	0.583	0.242-1.406	<0.001
Isolates group 3 vs. 1	Group 3 (90.00%)	9.45	2.844-31.402	
Isolates group 4 vs. 1	Group 4 (86.67%)	6.825	2.376-19.608	
Caries: Yes	No (39.02%) Yes (74.02%)	4.451	2.119-9.350	<0.001

B. MULTIVARIABLE ANALYSIS			
Variable Reference	**Odds Ratio**	**95% CI**	**Overall p-value**
Cohort 2	2.905	1.271-6.638	0.0114
Isolates group 2 vs. 1	0.53	0.201-1.397	
Isolates group 3 vs. 1	6.317	1.730-23.070	<0.001
Isolates group 4 vs. 1	7.397	2.129-25.695	
Caries (yes)	2.315	0.938-5.712	0.0685

Isolate groups defined as: group 1 (0-9 isolates), group 2 (10-29 isolates), group 3 (30-82 isolates), group 4 (more than 83 isolates)

Table 3. Summary of rep-PCR genotypes observed in index children shared and not shared with household family members.

Genotype	Total Children with Genotype*	Only Child has Genotype (%)	Child and Household Share Genotype (%)
G01	22	9 (41)	13 (59)
G01a	25	11 (44)	14 (56)
G01b	15	13 (87)	2 (13)
G01d	2	1 (50)	1 (50)
G02	2	1 (50)	1 (50)
G05	9	7 (78)	2 (22)
G06	19	8 (42)	11 (58)
G07	19	10 (53)	9 (47)
G08	4	2 (50)	2 (50)
G09	31	9 (29)	22 (71)
G11	6	1 (17)	5 (83)
G12	21	10 (48)	11 (52)
G13	21	10 (48)	11 (52)
G14	12	9 (75)	3 (25)
G15	10	7 (70)	3 (30)
G18	39	16 (41)	23 (59)
G19	1	-	1 (100)
G22	18	12 (67)	6 (33)
G23	20	10 (50)	10 (50)
G26	7	6 (86)	1 (14)
G49	2	2 (100)	-
G50	10	3 (30)	7 (70)
Total Cases		157	158

*Total represent children evaluated in this study (N = 117) and excludes children for whom no household members were evaluated or no *S. mutans* were observed. Children can have more than one genotype represented. Prefix “G”; rep-PCR genotype.

Table 4. New rep-PCR Library Strains with MLST sequence types.

Isolate ID	MLST ST	*tkt*	*glnA*	*gltA*	*glk*	*aroE*	*murI*	*lepC*	*gyrA*
NL01	175	1	2	1	8	4	3	1	1
NL01b	1	1	1	1	1	1	1	1	1
NL01d	**184**	1	1	1	2	2	3	1	1
NL03	166	3	24	5	25	1	1	30	15
NL25	**185**	1	2	1	1	19	3	24	1
NL26	97	3	2	1	1	21	11	24	1
NL27	161	1	1	5	24	4	3	31	1
NL30	**186**	1	2	**34**	4	2	**24**	1	21
NL38	**187**	**22**	2	31	4	4	3	11	1
NL42	179	1	2	15	3	1	11	11	1
NL45	**188**	3	2	1	23	**30**	3	**34**	1
NL49	**189**	2	3	20	4	5	5	1	**22**
NL50	**190**	**23**	8	15	8	2	5	21	1

ST= MLST sequence types.
NL = New Library strains.
New Sequence types and alleles identified in this study are in bold.

Table S1. Summary of genotypes observed for children and all participating household family members.

	Children Cohort 1	Children Cohort 2	Total
Total children evaluated for transmission	50	67	117
Total children with no *S. mutans* detected	1	0	1
Total children with mothers *S. mutans* available	36	62	98
Total mothers with no *S. mutans* detected	5	2	7
Total No. of genotypes per child			
Average (median) †	2.92 (3)	2.57 (2)	2.72 (2)
No. of Genotypes			
1	12	15	27
2	12	29	41
3	10	9	19
4	7	5	12
5	5	4	9
6	3	3	6
7	-	2	2
9	1	-	1
No. of genotypes per child not shared with any household members			
Average (median)	1.36 (2)	1.39 (2)	1.38 (2)
No. of Genotypes (not shared per child)			
1	17	22	39
2	10	14	24
3	6	7	13
4	2	3	5
5	1	2	3
Other			
Children having at least 1 genotype not shared with any household members	36 (72%)	48 (72%)	84 (72%)
Children sharing all genotypes within the household	14 (27%)	19 (28%)	33 (28%)
Children sharing no genotypes with any household members	9 (18%)	18 (27%)	27 (23%)
Children sharing at least 1 genotype with mother*	28 (68%)	38 (59%)	75 (63%)

* Includes mothers that were tested but no *S. mutans* isolates were recovered (CH1 = 5, CH2 = 2); excludes mothers with no isolates tested (CH1 = 1).
† Number of genotypes are not normally distributed.

Table S2. Commonality of cases of genotypes that occur only in children (no household family members with matching genotype) indicating child-to-child transmission of *S. mutans*.

Genotype	Total Children with Genotype*	Child Only has Genotype (%)	Total Isolates per Child (mean) ‡	No. of Family Members Evaluated	Range no. of Family Members (median)
Sm01	22	9 (41)	142 (16)	34	2-9 (3)
Sm01a	25	11 (44)	30 (3)	29	1-7 (2)
Sm01b	15	13 (87)	142 (11)	44	1-11 (3)
Sm01d	2	1 (50)	1 (1)	2	2
Sm02	2	1 (50)	2 (1)	1	1
Sm05	9	7 (78)	211 (30)	19	1-4 (3)
Sm06	19	8 (42)	69 (9)	22	1-5 (3)
Sm07	19	10 (53)	173 (17)	31	1-6 (2.5)
Sm08	4	2 (50)	4 (2)	7	3-4 (3.5)
Sm09	31	9 (29)	101 (11)	26	1-5 (3)
Sm11	6	1 (17)	7 (7)	1	1
Sm12	21	10 (48)	276 (28)	34	1-7 (3)
Sm13	21	10 (48)	163 (16)	41	1-10 (4)
Sm14	12	9 (82)	175 (19)	25	1-5 (3)
Sm15	10	7 (70)	77 (11)	24	1-7 (3)
Sm18	39	16 (41)	292 (18)	45	1-7 (2)
Sm22	18	12 (67)	194 (16)	40	1-8 (4)
Sm23	20	10 (50)	53 (5)	32	1-7 (3)
Sm26	7	6 (86)	50 (8)	24	1-8 (4)
Sm49	2	2 (100)	3 (2)	11	1-10 (5.5)

Genotype	Total Children with Genotype*	Child Only has Genotype (%)	Total Isolates per Child (mean) ‡	No. of Family Members Evaluated	Range no. of Family Members (median)
Sm50	10	3 (30)	9 (3)	4	1-2 (1)
Totals		157	2174	496	

* Totals represent children evaluated in this study. Children that had no household members analyzed with rep-PCR were excluded from this analysis.
† Special cases where the child matched another child that was excluded because the other child had no household members evalutated with rep-PCR.
‡ "Total Isolates per Child" is for "Child Only has Genotype" (column 3)
§ "No. of Family Members Evaluated" is the number of household members analyzed by rep-PCR with no match to child.
Prefix "G"; rep-PCR genotype.

Table S3. Commonality of cases of genotypes that occur in children and their household family members indicating probable intra-familial transmission of *S. mutans*.

Genotype	Total Children with Genotype*	Child + Family Member(s) Share Genotype (%)	Total Isolates Child (mean) †	Total Isolates Family Members‡	Match 1 Family Member	Match >1 Family Member (%)	Children with genotype shared with Household Members						
							M	F	S	G	A	U	C
Sm01	22	13 (59)	408 (31)	179	6	7 (54)	9	1	4	2	2	1	1
Sm01a	25	14 (56)	720 (51)	328	5	9 (64)	11	3	2	1	3	3	6
Sm01b	15	2 (13)	17 (2)	23	1	1 (50)	1	1	1	1	-	-	-
Sm01d	2	1 (50)	22 (22)	23	-	1 (100)	1	-	1	-	-	-	-
Sm02	2	1 (50)	1 (1)	34	-	1 (100)	-	-	1	-	-	-	-
Sm05	9	2 (22)	68 (34)	28	1	1 (50)	2	-	1	-	-	-	-
Sm06	19	11 (58)	292 (27)	130	4	7 (64)	6	1	5	-	2	-	2
Sm07	19	9 (47)	172 (19)	73	7	2 (22)	2	-	6	1	1	1	-
Sm08	4	2 (50)	6 (3)	5	1	1 (50)	-	-	2	-	-	-	1
Sm09	31	22 (71)	634 (29)	344	12	10 (45)	11	1	9	4	4	4	4
Sm11	6	5 (83)	242 (48)	109	1	4 (80)	2	-	3	1	1	-	1
Sm12	21	11 (52)	371 (34)	134	3	8 (73)	8	1	5	1	2	-	2
Sm13	21	11 (52)	363 (33)	202	5	6 (55)	8	-	6	2	2	-	2
Sm14	12	3 (27)	114 (38)	85	1	2 (67)	1	-	1	-	-	-	1
Sm15	10	3 (30)	72 (24)	23	2	1 (33)	1	-	2	1	-	-	-
Sm18	39	23 (59)	969 (42)	509	12	11 (48)	8	-	8	5	9	1	8
Sm19	1	1 (100)	18 (18)	27	-	1 (100)	-	-	1	-	-	-	1
Sm22	18	6 (33)	141 (24)	93	3	3 (50)	3	1	2	2	1	-	2
Sm23	20	10 (50)	247 (25)	266	3	7 (70)	7	2	6	-	3	2	5

Genotype	Total Children with Genotype*	Child + Family Member(s) Share Genotype (%)	Total Isolates Child (mean) †	Total Isolates Family Members‡	Match 1 Family Member	Match >1 Family Member (%)	Children with genotype shared with Household Members						
							M	F	S	G	A	U	C
Sm26	7	1 (14)	10 (10)	7	-	1 (100)	1	-	1	1	-	-	-
Sm50	10	7 (70)	295 (42)	51	4	3 (43)	3	-	5	-	-	-	1
Totals		158	5182	2673	71	87	85	11	72	22	30	12	37

* Totals represent children evaluated in this study and excluded children where no household members analyzed with rep-PCR.
† Total Isolates Child is for “Child and Family Member Share GT” (column 3)
‡ Total Isolates Family Members is for “Child and Family Member Share GT” (column 3)
Note: Not all children have all family members.
Abbreviations: M = mother, F = father, S = sibling(s), G = grandmother(s), A = aunt (s), U = uncle(s), C = cousin(s).
Prefix “G”; rep-PCR genotype.

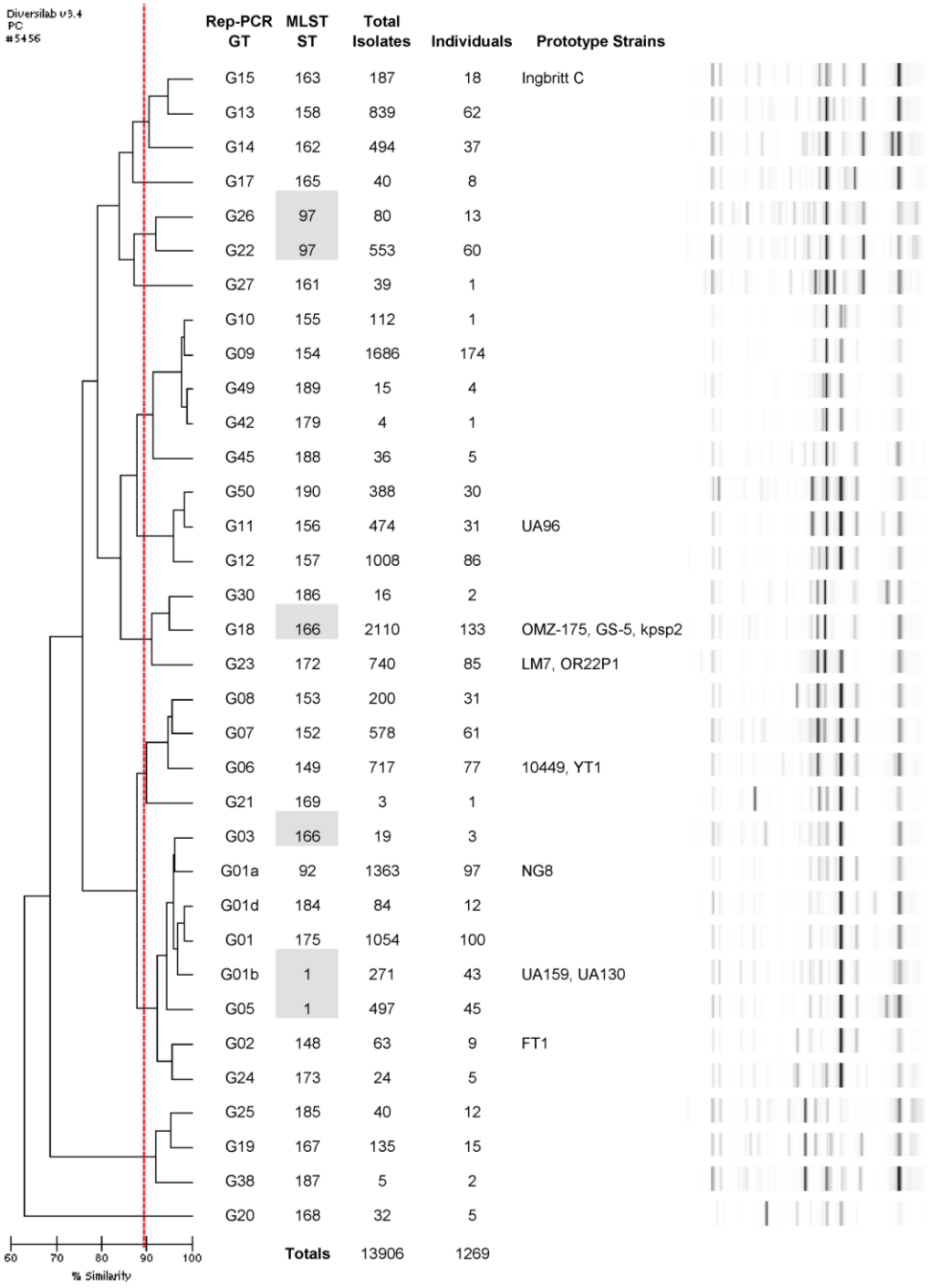


Fig. 1. Summary rep-PCR dendrogram of 34 representative *S. mutans* genotypes. Total number of isolates and individuals with similar rep-PCR genotypes are indicated. MLST ST is for representative isolate only (other isolates with the same rep-PCR genotype may have different ST). Best match rep-PCR genotypes for prototype strains are indicated. Prefix “G”; rep-PCR genotype.

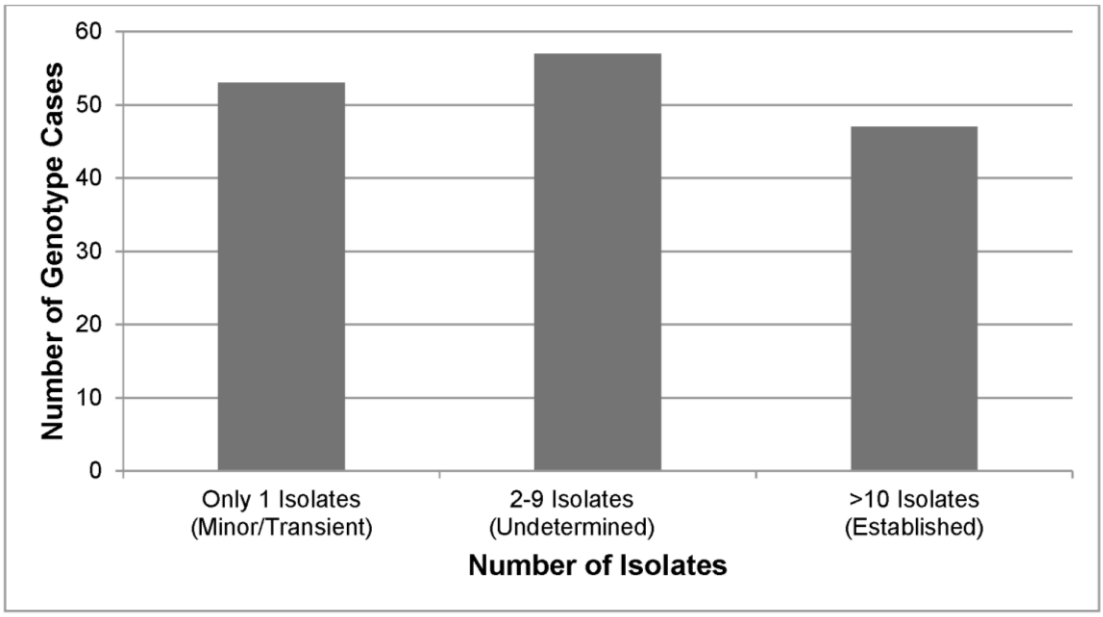


Fig. 2. Distribution of minor, undetermined, and established genotypes among children who did not share the genotype with any household family members evaluated.

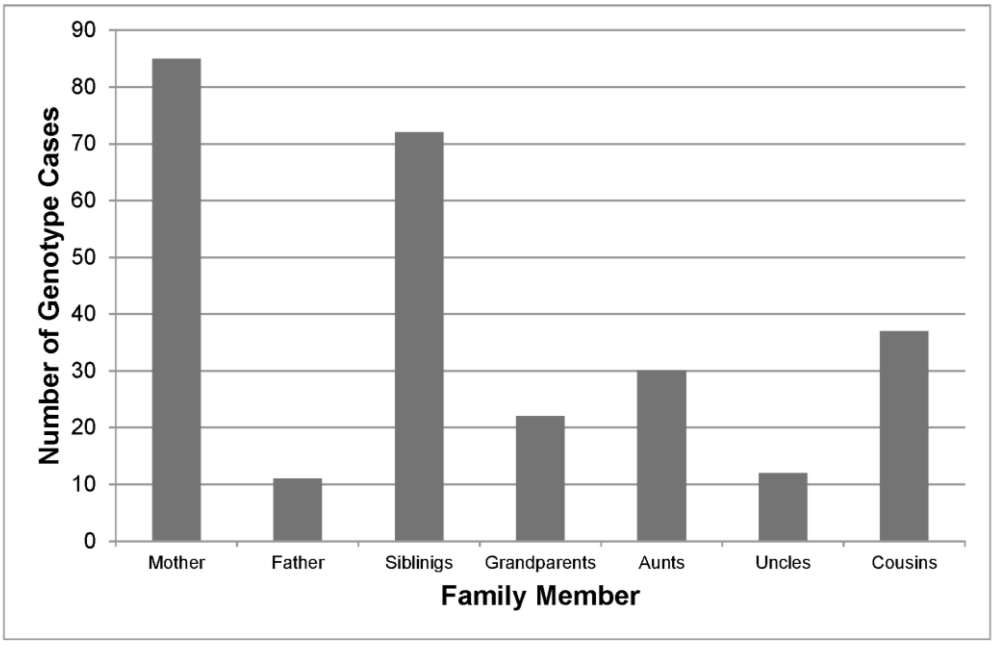


Fig. 3. Number of genotype cases for intra-familial transmission (N=158) by family members where a child shared a genotype with at least one household member**.** Totals will not equal 158 since children may share genotypes with multiple family members.

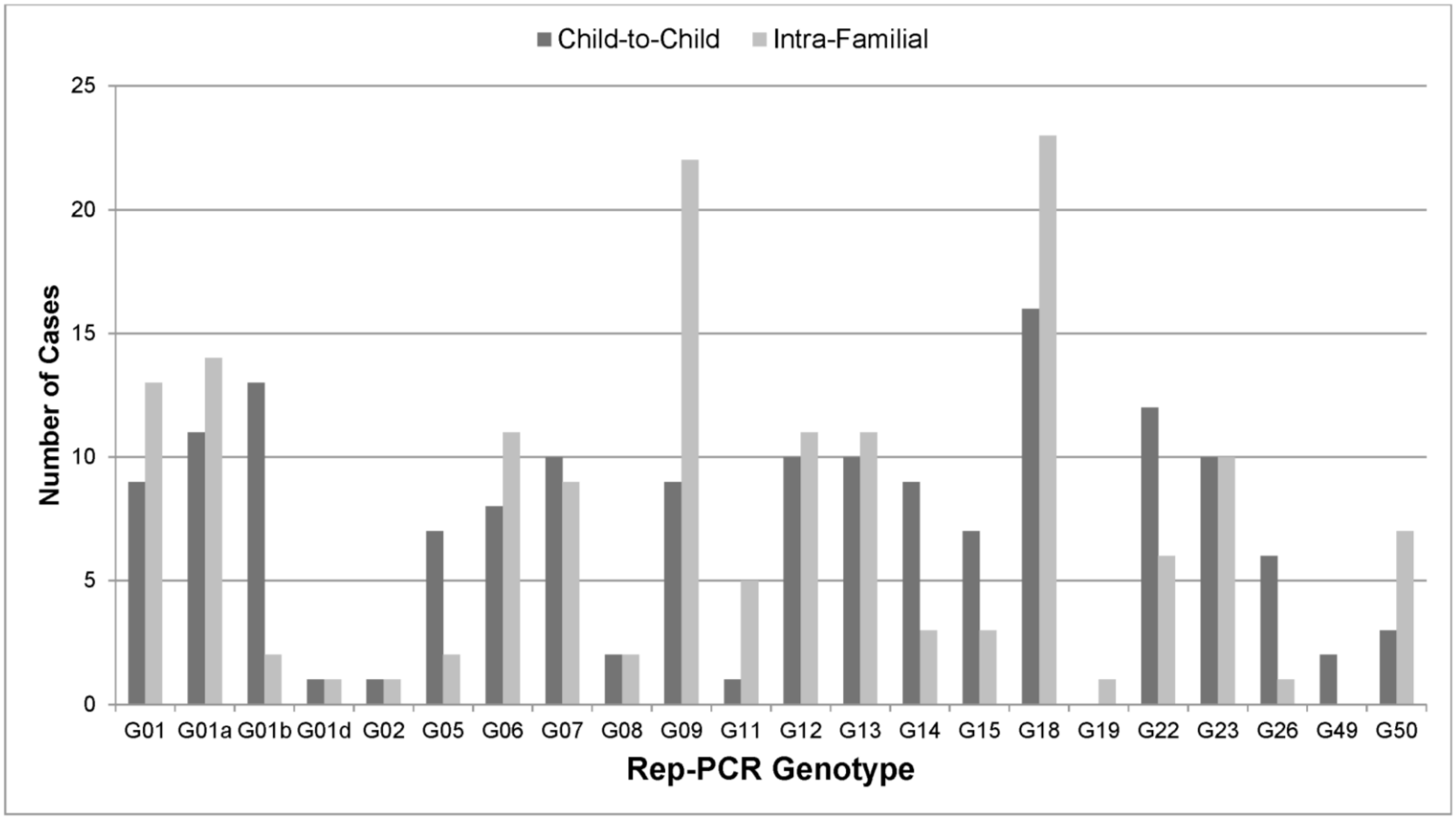


Fig. 4. Comparison of child-to-child and intra-familial transmission among children evaluated by rep-PCR genotype. Prefix “G”; rep-PCR genotype.

TRANSMISSION PATTERNS OF *STREPTOCOCCUS MUTANS* DEMONSTRATED BY A COMBINED REP-PCR AND MLST APPROACH

by

STEPHANIE S. MOMENI, JENNIFER WHIDDON, STEPHEN A. MOSER, NOEL K. CHILDERS

In preparation for *Archives of Oral Biology*

Format adapted for dissertation

ABSTRACT

Streptococcus mutans is widely associated with early childhood dental caries. Molecular analysis of the genetic diversity of *S. mutans* using approaches such as multilocus sequence typing (MLST) and whole genome sequencing (WGS) are typically more specific, but also more expensive. Repetitive extragenic palindromic PCR (rep-PCR) is relatively inexpensive gel-based bacterial genotyping method that offers a pre-screening alternative for selection for more expensive approaches. This two-part study evaluated whether rep-PCR dendrograms could be used to accurately predict MLST sequence type diversity and applied this approach to a sub-group of clinical isolates to assess evidence for horizontal transmission.

In part one of the study, *S. mutans* isolates previously analyzed with rep-PCR to determine genotypes were used to generate rep-PCR dendrograms for the 10 most common rep-PCR genotypes. Twenty isolates were randomly selected for each report. Isolates with greater than 98% similarity were predicted to have an MLST sequence type (ST) match. Prediction accuracy was calculated as the number of correctly predicted ST divided by the total evaluated. In part two, this approach was applied to evaluate transmission of genotypes of

children and their family members with a rep-PCR genotype that was found to be more diverse (G12) by MLST.

Overall, rep-PCR was 73% effective at determining matching MLST ST when single- and double-locus variants were included and 100% effective when applied to related individuals. For transmission, infants shared ST with their mothers but 50% of older children had ST not identified in their mother. Six ST were shared between different families and probable source members were identified.

This study confirms that rep-PCR offers an affordable option to predict diverse isolates for downstream applications. Using a combined rep-PCR and MLST approach, it is possible to track probable transmission and determine strain sources for some *S. mutans* genotypes. Further studies using other rep-PCR genotypes are needed to confirm these findings.

INTRODUCTION

Dental caries is a prevalent global chronic disease affecting over 80% of humans worldwide and is the most common chronic disease among children and adolescents (1). Although many organisms are suspected in the progression of dental caries, the mutans streptococci (*Streptococcus mutans* and *Streptococcus sobrinus*) are the primary bacteria most frequently associated with dental caries initiation. Studies evaluating the genotypic diversity of the mutans streptococci, in particular *S. mutans*, could be important for the improvement of caries risk assessments and treatment options.

Bacterial genotyping is the foundation of investigations of bacterial strain genetic diversity, microevolution, and epidemiological surveillance studies. Traditionally, genotyping of *S. mutans* has been performed using gel-based typing methods including pulse field gel electrophoresis (PFGE), arbitrarily primed PCR (AP-PCR), and repetitive extragenic palindromic PCR (rep-PCR) (2-7). Current advances in technology are increasing the application of molecular based typing applications such as multilocus sequence typing (MLST) and whole genome sequencing (WGS) for *S. mutans* strain comparisons (8-12). Although costs for using these methods continue to decrease, they are significantly more expensive than traditional gel-based typing methods and may not be financially feasible for all laboratories. Alternately, it may be more economical to prescreen

isolates prior to employing more expensive testing approaches when appropriate to the research question. For instance, a study using rep-PCR can be utilized to select for isolates demonstrating strain diversity, creating a sample pool from which it would be possible to select isolates that are far more likely to provide diverse data for a comparative genomics study (13-15). This approach is potentially more efficient than using a random selection of isolates for more expensive, discriminative analysis, as using random isolates could result in multiple genomic profiles that are similar or identical, yielding little or no new insights for the research question.

Therefore, the DiversiLab System is an economical option for the investigation of genetic diversity of bacterial isolates using a standardized rep-PCR approach (16). After PCR, micro fluidics chip technology allows the samples to be visualized using dendrograms, percent similarity and graphic overlays. This method is rapid and economical for high-throughput epidemiological studies (our average cost is $22/sample). It has been widely used for epidemiological surveillance of clinically relevant bacteria but its application for typing oral bacteria is limited and unique to our laboratory (3, 17-21).

MLST is a molecular genotyping method that characterizes bacteria based on changes in the molecular sequence of select conserved housekeeping gene loci (22). This method is considered more discriminative and reproducible than gel based methods because interpretation of gel based typing approaches are typically based on similarity where as with MLST a single nucleotide change can

result in a new sequence type (ST). Processing and data analysis for MLST analysis is more time consuming and more expensive than for rep-PCR (our average base cost is $96/sample).

Studies investigating the genotypic diversity of *S. mutans* frequently include transmission analysis. The current literature offers several possible sources of transmission. Mother-to-child transmission of *S. mutans* has been described as the primary route of transmission (2, 23-28). Other studies have suggested that *S. mutans* strains are often shared with other household family members including fathers and siblings indicating intra-familial transmission may be possible (25, 26, 29, 30). Furthermore, extra-familial sharing of *S. mutans* genotypes has been reported for unrelated children within nurseries or school classes (6, 31-33). There are several limitations to these studies, which include using only one genotyping approach (typically gel based); limited numbers of bacterial isolates (small sample sizes, thus decreased statistical power); and a limited focus (e.g., many studies focus only on mother-to-child without looking at other potential sources).

We have reported that rep-PCR is beneficial for large scale epidemiological studies of *S. mutans* (3) and we predict that it may also effectively be used to determine which isolates are most likely to yield diverse data from additional, more specific but also more expensive techniques such as MLST or whole genome sequencing. The current study used a two-part study design. The first aim was to determine if rep-PCR could be used to predict *S. mutans* MLST sequence types. The second aim was to assess the practical

application of this approach for a specific research question using rep-PCR to predict sequence types for an *S. mutans* rep-PCR genotype (genotype G12) to determine likely transmission of this genotype among and between children and their families.

METHODS

Sample Selection

Isolates used in this study were collected as part of an ongoing longitudinal epidemiological study of *S. mutans* in Uniontown, Alabama, United States. This population is primarily African-American, low socioeconomic status with limited access to dental care. This University of Alabama at Birmingham IRB approved study used informed consent from the parent/guardian and wavier of assent due to the young age of children. *S. mutans* isolates used in this study were from 2 cohorts of children. Cohort 1 (CH1) consisted of school aged children (age 5-6 years at initial recruitment, family ID# less than 300) and Cohort 2 (CH2) were infants (age 12-18 months at initial collection, family ID# 500 and over). Specifics on sample selection for each of the two-part study design are described in more detail below.

Repetitive extragenic palindromic PCR (rep-PCR)

Rep-PCR was performed as part of the larger Uniontown epidemiological study as previously described (3). Briefly, DNA was extracted using the UltraClean

microbial DNA isolation kit (bioMérieux, Inc., Durham, NC) from *S. mutans* isolates previously confirmed as *S. mutans* using *gftB*-specific primers. Quality of DNA was assessed and rep-PCR performed using the DiversiLab system with the Streptococcus DNA fingerprinting kit. Amplicons were evaluated using an Agilent 2100 Bio-analyzer (Agilent Technologies, Santa Clara, CA) with microfluidics LabChip (bioMérieux, Inc.). DiversiLab web-based software was used to determine genotypes using dendrograms and electropherogram overlays. An isolate was classified as a new genotype when an isolate differed by 1 major band or 3 minor bands. Major and minor bands were determined based on a 100-fluorescence units cut-off.

In part one of the current study, from rep-PCR data obtained from ~13,000 *S. mutans* isolates analyzed, 20 plaque isolates from 20 individual unrelated children were randomly selected for each of the 10 most frequently reported rep-PCR genotypes. All available isolates from children for a rep-PCR genotype were complied into a list and randomized in Excel using the Rand function. From the randomized list, the first 20 isolates from different children for which sufficient DNA and a strong rep-PCR pattern were available were selected. Genotypes were previously determined as part of the larger epidemiological study by rep-PCR using methods described elsewhere (3). In rare cases when plaque isolates were not available for all 20 children, then saliva or tongue isolates were used. These isolates were then used to generate a rep-PCR dendrogram using the DiversiLab web-based software. The dendrogram consist of a percent similarity tree, sample identification, and virtual gel images. Using

an arbitrarily assigned 98% percent similarity cut-off, isolates were predicted to either have a matching or non-match sequence type. This cut-off was selected based on previous observations from studies in our laboratory using MLST indicating that isolates with <98% percent similarity tended to have different MLST sequence types (data not shown). If the node connecting the two isolates was less than 98% similar (major branch), then it was predicted that the two isolates would have a non-matching sequence type. If the connecting node had a greater than 98% similarity (minor branch), then it was predicted that the two isolates would have a matching sequence type. This process was initially performed using two isolates selected from the most extremes branches of the dendrogram, then repeated for 7-9 isolates for various internal branches (i.e., branches within the tree). In rare cases, a single isolate occurred on a single branch. For these cases the isolate was compared with the closest isolates on a neighboring branch. Prediction accuracy was calculated as the number of correctly predicted sequence types divided by the total sequence types evaluated. A total of 97 isolates were selected for MLST analysis.

In part two of the study, a pool of all index children and their mothers and fathers with a specific rep-PCR genotype (G12) were selected for analysis. Genotype G12 was selected based on previous observation in our laboratory that found this common genotype group to contain diverse sequence types by MLST analysis (13, 14). A total of 9 mother-child pairs (M-C), 1 father-child pair (F-C), and 14 children only (mother and father did not have G12) groups were used to generate rep-PCR dendrograms. Dendrograms were generated for all G12

isolates for each parent-child group or for each individual child (when the child had G12 but the parents did not have this genotype). These reports included a total of 701 isolates from children and their parents. Some children had less than twenty G12 isolates available for a report. These typically had only minor branches (above 98%) so only 1 isolate per individual was evaluated for these children. For reports with 20 or more isolates, extreme isolates and at least 1 internal isolate were selected for analysis. Additional isolates were selected from very large reports consistent with the major branches observed. For available mothers and fathers, 1 isolate for each major branch on which an isolate from a mother or father was observed was selected. Using the method described in part one, sequence types were predicted and a total of 69 *S. mutans* isolates were selected from the reports for MLST analysis.

Multilocus Sequence Typing

MLST was performed as previously described (13). Briefly, primers described by Nakano et al. for 8 housekeeping partial gene fragments from *murI, tkt, glnA, aroE, gyrA, glk, glt*, and *lepC* were amplified using a SYBR green real-time PCR approach (10). Amplicons were purified and sequenced. Manual alignments were performed using CLC DNA Workbench version 5.7.1 to obtain consensus sequences that were cross-referenced with the oral streptococci www.pubmlst.org database for allele and sequence type assignments. New alleles and sequence types were confirmed and added to the PubMLST database. New alleles were also submitted to Genbank.

Following MLST analysis, sequence type data was added to the rep-PCR dendrograms to determine the prediction accuracy of rep-PCR percent similarity trees in predicted MLST sequence type. First the sequence types of the two extreme isolates were compared. Isolates with the same sequence type were considered a match while isolates with different sequence types were considered a no match. Next, internal branches were evaluated. This approach focused on prediction accuracy between two selected isolates but not between different groups. The number of instances that the prediction agreed or disagreed was recorded to determine prediction accuracy. Given that rep-PCR is based on similarity, further consideration was given to isolates that differed by either a single- or double-locus variants (SLV and DLV, respectively) because up to two allelic differences are considered clonal complexes (i.e., recently diverged, similar strains) (10, 22).

Transmission

A subset of the rep-PCR G12 clinical isolates were used to investigate evidence for transmission patterns based on MLST sequence types. The commonality of *S. mutans* strains by MLST sequence types for G12 isolates for mother-child, father-child and children only (children with no parent with G12 genotype) were assessed to track transmission. Individuals sharing the G12 rep-PCR genotype and the same MLST sequence type were considered as mother-to-child transmission, father-to-child or extra-familial transmission (between unrelated children). Subsequently, it was possible to identify potential sources because a

number of strains shared among different children appeared to be shared with only one adult in the study. To further provide evidence for the source, an extended analysis of other household family member was performed. Other family members with the G12 genotype were analyzed with MLST and sequence types compared. In cases where the child shared the sequence type with the mother, siblings or cousins with the G12 genotype were excluded from the analysis because regardless of the sequence type of the sibling, they would be considered as unlikely sources since the mother already had the shared sequence type. Siblings and cousins were evaluated for cases where the mother and child had different sequence types.

Statistics

Statistically analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). A Fisher's exact test was performed to evaluate the ability to correctly predict match or no-match sequence types from the rep-PCR dendrograms (i.e., sensitivity/specificity and predictive values). This analysis included both extreme and internal branches for the top 10 genotypes.

RESULTS

Overall, this study identified three new alleles and 6 new ST were added to the PubMLST Database (Table 1). Genbank accession numbers KX383950-52 were assigned for the three new alleles.

Prediction Accuracy

In the initial part of the study, using rep-PCR of the 10 most common genotypes to predict MLST sequence types from the two extreme branches resulted in 2 genotypes (G18 and G09) that were connected by branches with greater than 98% (predicted match) while 8 were non-matching (Table 2). Based on this, it was predicted that any internally selected isolates would match the sequence type for the two extreme branch isolates for G18 and G09. While G18 did match this prediction, G09 has other sequence types that were different, but only by SLV and DLV. Thus, using extreme isolates only resulted in 90% accuracy in predicting if internal branch sequence types would or would not match.

Analysis of 97 isolates on 52 branch groups (5-7 groups per genotype) resulted in 62% (32/52) prediction accuracy (Table 2). Allowing for SLV and DLV increased the accuracy of prediction to 73% (38/52). Statistical analysis of 62 groups (52 internal and 10 extreme branches combinations) indicated there was a significant predictive relationship using rep-PCR for identifying matching vs. non-matching MLST sequence types ($p < 0.001$) both with and without SLV and DLV included in the analysis (Table S1). Rep-PCR had a greater negative predictive value (91%) vs. the positive predictive value of (55%). Sensitivity was 92% and specificity was 53%. The negative predictive value remained the same when SLV and DLV were considered; however the positive predictive value increased to 70%. Sensitivity remained high at 93% and specificity increased to 63%.

When evaluating G12 isolates for prediction accuracy, M-C and F-C pairs were evaluated separately from the children only group (Table 3). Comparing the extreme isolates in predicting sequence type was only 50% accurate for the parent-child pairs and 20% for individual children. Cases where the actual extreme match results were able to predict internal isolates sequence type was 82% (9/11) overall. Rep-PCR dendrograms were also used to predict if M-C and F-C would match with 88% (7/8) accuracy. Statistical analysis was not performed on the data from this portion of the study due to small sample size and because the data would be biased since isolates from children and their parents were more likely to have the same sequence type.

Transmission

Table 4 summarizes the isolate and MLST data for rep-PCR genotype G12. For the 9 M-C pairs, 6 children shared the same rep-PCR GT and MLST ST with the mother (107, 180, 182, 518, 596, 600); 3 children had different MLST sequence types from their mothers although these children shared the rep-PCR genotype. Sequence types were consistent within individuals except for 170-C which had 2 sequence types (ST157, ST202). The one F-C pair did not share any MLST sequence types. For the 14 children only (no mother or father with G12), sequence types were consistent within children except for 241-C who had two sequence types (ST181, ST132). ST157, ST201 and ST202 were shared among multiple individuals within different families. ST119, ST132, and ST181 were each only shared between 2 families. ST119 (Thailand), ST132 (Thailand), and

ST92 (Japan) were isolates matching strains identified in published studies from other countries while all other isolates were unique to the Uniontown population (United States) (10, 13, 14, 34, 35).

For the extended family analysis (Table S2), 8 of the 9 M-C pairs had additional household family members with G12. MLST analysis was performed for 6 additional family members from 5 families; however, only 3 households (107, 180, and 596) had another adult (assumed a potential source) besides the mother with a G12. In the 14 children only (no parents with G12), 3 of the children (159, 241, and 601) did not have a mother participating in the study for evaluation. Fathers participate less often in the study overall. MLST analysis was performed for 5 additional family members from 2 families (219 and 509). Twelve of the children only group had no other family members with G12 observed although the mean number of family members evaluated by rep-PCR was 4. The one F-C pair had 3 additional family member evaluated by rep-PCR but none with G12.

DISCUSSION

This study evaluates whether rep-PCR dendrograms can be used to predict MLST sequence types. Successfully application of this approach would be cost-beneficial for other downstream applications such as WGS.

Initial analysis using only the two extreme isolates from a dendrogram was 90% accurate in predicting which genotypes needed further analysis (Table 2). This was expected as unrelated individuals were expected to be more likely to

have different sequence type. For the M-C and F-C pairs as well as the individual children evaluated in the second part of the study (i.e., G12), the prediction accuracy of the two extreme isolates was much lower at 50% and 20% respectively. For M-C and individual isolates it is expected that these isolates are more likely to have the same ST within a M-C pair or individual. The observed results were low because of low specificity; the extreme isolates often predicted no match when analysis revealed a match.

In part one of study, the overall prediction accuracy (extreme and internal branches) was only 62% and the specificity was 53%, which was lower than anticipated (Table 2). However, it is important to note the rep-PCR branching is based largely on percent similarity and there can be minor variations of isolates classified as the same rep-PCR genotype (3). In contrast, MLST is a more sensitive and discriminative method because a single base pair change results in a new MLST sequence type. Thus, the analysis of MLST sequence types was expanded to include SLV and DLV so as to be more comparable to the rep-PCR approach. This modification resulted in an increase to 73% prediction accuracy however even when SLV and DLV were considered specificity remained low at 63% (Table 4, Table S1). These findings suggest that rep-PCR is very effective at predicting MLST sequence types that will not match. Furthermore, it suggests that when isolates shared a minor branch (>98% similarity) and were predicted to match, the isolates were sometimes further differentiated by MLST. In part two of the study, the prediction accuracy for G12 M-C and F-C pairs was 100% and 75% for G12 individual children. It is noteworthy that for the individual children

the reduced predictability was due to false no matches (cases where the sequence type were predicted to not-match but actually matched). Together, these findings suggest that rep-PCR is reasonably accurate in predicting MLST sequence types and can provide an affordable pre-screening approach for isolates to minimize the samples necessary for more expensive downstream applications.

Transmission of rep-PCR genotype G12 was evaluated using the combined rep-PCR and MLST approach described in this paper (Table 3). Overall, transmission analysis included 24 children and 20 of their family members resulting in 13 different MLST sequence types for the rep-PCR genotype G12 (Table 4). It was observed that for the 15 individuals with more than one isolate evaluated, only 2 had more than one sequence type (13%, 2/15), indicating that sequence types are typically stable, at least within these individuals, for G12 isolates.

For M-C pairs including infants, sequence types were the same (ID 518, 596, 600). In contrast, M-C pairs for older, school-aged children had 2 pairs for which the child has sequence types that were different from the mother (ID 170, 227). This finding suggest that younger children are more likely to share strains with their mother, whereas older children are more likely to have other genotype not shared with their mothers. However, further study with a larger sample size is needed to confirm these results. Although these children and their mother share a rep-PCR genotype (G12), the isolates were further distinguished with MLST. This may be an indication that the mother was not the original source of

the *S. mutans* or that longer habitation time within the child lead to minor genetic mutations. Given that the sequence types observed in these two children (ST 157 and 202) appear to be commonly shared in this population it is also possible these strains were acquired through child-to-child transmission.

Fourteen children had no rep-PCR G12 match with a parent. However, some of these children did have sequence type in common with some of the M-C pairs. ST132, ST157, and ST202 were most commonly shared, indicating these strains to be more easily transmissible. Although it is very difficult to prove transmission conclusively, it is interesting to note that is appears that 180-M may be the initial source for strain ST157. Although 227-M also shares this sequence type, her child does not carry the sequence type. Both 180 mother and child carry the strain. Thus, this mother and her child seem the most likely candidates for sharing the strain with five other children (159-C, 223-C, 597-C, 608-C, 226-C). The same conclusion may be drawn for ST202, for which 182 mother and child share the sequence type. No other mothers carry this strain yet 4 other children have this strain (170-C, 227-C, 218-C, 509-C). Based on this data, it is possible to track transmission and the source of commonly shared *S. mutans* strains using rep-PCR to initially screen isolates for further differentiation with MLST. This finding is further strengthened through the overall family analysis (Table S2). For the 24 children evaluated in this study, a total of 91 household members were evaluated with rep-PCR but only 24 had rep-PCR genotype G12. Subsequent MLST analysis for 22 of the family members confirmed the source was not another participating family member.

There are a few limitations to this study. For the prediction accuracy, the data in this study suggest that individuals within the same family are more likely to have the same sequence types so it is probable that the higher prediction accuracies observed in the second part of the study are partly biased. For both parts of the study, not all isolates on a rep-PCR dendrogram were evaluated so it is possible that these results could have varied if more isolates all isolates were tested. For transmission, not all household members chose to participate in the study and thus other sources not evaluated here may be possible. For instance, most children in this study did not have a father evaluated either because the father was not in the household or chose not to participate. Also, other individuals in the larger Uniontown Study had G12 strains but were not evaluated in this study since these individuals did not reside with children in this study and possible interactions ere unknown. Furthermore, the approach described may not be applicable to all *S. mutans* strains, such as rep-PCR G18, which consistently have the same rep-PCR and MLST sequence type. Therefore, for some genotypes, alternate approaches would be required.

This study supports that rep-PCR can be used to reasonably predict strain diversity for downstream application such as MLST, providing a more cost effective approach to large scale epidemiological studies. As regards transmission, some strains of *S. mutans* genotype G12 are commonly shared while other are family specific. Children may acquire additional sequence types from sources, other than their mothers, as they get older. Using a combined

rep-PCR and MLST approach, it is possible to track possible transmission and identify potential sources for some strains of *S. mutans*.

ACKNOWLEDGEMENTS

This work was supported by National Institute of Dental and Craniofacial Research (NIDCR) Grant DE016684 (NKC). Ms. Momeni is a Dental Academic Research Training (DART) Pre-doctoral Fellow under NIDCR Institutional Grant#T-90 DE022736. We especially appreciate all the clinical and laboratory participants of this study. Dr. Mei Han of the UAB Heflin Genomic Core provided sequencing data; Dr. Jinthana Lapirattanakul directed the confirmation and registration of isolates with PubMLST. The authors declare no conflicts of interest.

REFERENCES

1. Bagramian RA, Garcia-Godoy F, Volpe AR. The global increase in dental caries. A pending public health crisis. *Am J Dent* 2009;**22**(1):3-8.

2. Li Y, Caufield PW. The fidelity of initial acqusition of mutans streptococci by infants from their mothers. *J Dent Res* 1995;**74**(2):681-685.

3. 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. *J Clin Microbiol* 2010;**48**(2):599-602.

4. Lynch DJ, Villhauer AL, Warren JJ, Marshall TA, Dawson DV, Blanchette DR, et al. Genotypic characterization of initial acquisition of *Streptococcus mutans* in American Indian children. *J Oral Microbiol* 2015;**7**:27182.

5. Mitchell SC, Ruby JD, Moser S, Momeni S, Smith A, Osgood R, et al. Maternal transmission of mutans Streptococci in severe-early childhood caries. *Pediatr Dent* 2009;**31**(3):193-201.

6. Baca P, Castillo AM, Liebana MJ, Castillo F, Martin-Platero A, Liebana J. Horizontal transmission of *Streptococcus mutans* in schoolchildren. *Med Oral Patol Oral Cir Bucal* 2012;**17**(3):e495-500.

7. Gilbert K, Joseph R, Vo A, Patel T, Chaudhry S, Nguyen U, et al. Children with severe early childhood caries: streptococci genetic strains within carious and white spot lesions. *J Oral Microbiol* 2014;**6**.

8. 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. *PLoS One* 2010;**5**(2):e9073.

9. Lapirattanakul 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**(6):466-474.

10. Nakano K, Lapirattanakul J, Nomura R, Nemoto H, Alaluusua S, Gronroos L, et al. *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. *J Clin Microbiol* 2007;**45**(8):2616-2625.

11. Song L, Wang W, Conrads G, Rheinberg A, Sztajer H, Reck M, et al. Genetic variability of mutans streptococci revealed by wide whole-genome sequencing. *BMC Genomics* 2013;**14**:430.

12. Cornejo OE, Lefebure T, Pavinski Bitar PD, Lang P, Richards VP, Eilertson K, et al. Evolutionary and Population Genomics of the Cavity Causing Bacteria Streptococcus mutans. *Mol Biol Evol* 2013;**30**(4):881-893.

13. 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**(1):18-27.

14. Momeni SS, Whiddon J, Cheon K, Moser SA, Childers NK. Assessment of two multilocus sequence typing (MLST) schemes available for *Streptococcus mutans*. *Arch Oral Biol* 2015;**60**(12):1769-1776.

15. 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. *Clin Microbiol Infect* 2007;**13 Suppl 3**:1-46.

16. Healy M, Huong J, Bittner T, Lising M, Frye S, Raza S, et al. Microbial DNA typing by automated repetitive-sequence-based PCR. *J Clin Microbiol* 2005;**43**(1):199-207.

17. Kim GH, Breidt F, Fratamico P, Oh DH. Acid Resistance and Molecular Characterization of *Escherichia coli* O157:H7 and Different Non-O157 Shiga Toxin-Producing *E. coli* Serogroups. *J Food Sci* 2015;**80**(10):M2257-2264.

18. Mutlu Sariguzel F, Berk E, Koc AN, Sav H, Demir G. Investigation of the relationship between virulence factors and genotype of Candida spp. isolated from blood cultures. *J Infect Dev Ctries* 2015;**9**(8):857-864.

19. Werner G, Fleige C, Neumann B, Bender JK, Layer F, Klare I. Evaluation of DiversiLab(R), MLST and PFGE typing for discriminating clinical *Enterococcus faecium* isolates. *J Microbiol Methods* 2015;**118**:81-84.

20. Ford CD, Lopansri BK, Haydoura S, Snow G, Dascomb KK, Asch J, et al. Frequency, risk factors, and outcomes of vancomycin-resistant Enterococcus colonization and infection in patients with newly diagnosed acute leukemia: different patterns in patients with acute myelogenous and acute lymphoblastic leukemia. *Infect Control Hosp Epidemiol* 2015;**36**(1):47-53.

21. Aguadero V, Gonzalez Velasco C, Vindel A, Gonzalez Velasco M, Moreno JJ. Evaluation of rep-PCR/DiversiLab versus PFGE and spa typing in genotyping methicillin-resistant *Staphylococcus aureus* (MRSA). *Br J Biomed Sci* 2015;**72**(3):120-127.

22. Maiden MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 2006;**60**:561-588.

23. Teanpaisan R, Chaethong W, Piwat S, Thitasomakul S. Vertical transmission of mutans streptococci and lactobacillus in Thai families. *Pediatr Dent* 2012;**34**(2):e24-29.

24. Klein MI, Florio FM, Pereira AC, Hofling JF, Goncalves RB. Longitudinal study of transmission, diversity, and stability of *Streptococcus mutans* and *Streptococcus sobrinus* genotypes in Brazilian nursery children. *J Clin Microbiol* 2004;**42**(10):4620-4626.

25. Kozai K, Nakayama R, Tedjosasongko U, Kuwahara S, Suzuki J, Okada M, et al. Intrafamilial distribution of mutans streptococci in Japanese families and possibility of father-to-child transmission. *Microbiol Immunol* 1999;**43**(2):99-106.

26. Douglass JM, Li Y, Tinanoff N. Association of mutans streptococci between caregivers and their children. *Pediatr Dent* 2008;**30**(5):375-387.

27. da Silva Bastos Vde A, Freitas-Fernandes LB, Fidalgo TK, Martins C, Mattos CT, de Souza IP, et al. Mother-to-child transmission of Streptococcus mutans: a systematic review and meta-analysis. *J Dent* 2015;**43**(2):181-191.

28. Binks C, Duane B. Mother-to-child transmission of *Streptococcus mutans*. *Evid Based Dent* 2015;**16**(2):39-40.

29. Tedjosasongko U, Kozai K. Initial acquisition and transmission of mutans streptococci in children at day nursery. *ASDC J Dent Child* 2002;**69**(3):284-288, 234-285.

30. Redmo Emanuelsson IM, Wang XM. Demonstration of identical strains of mutans streptococci within Chinese families by genotyping. *Eur J Oral Sci* 1998;**106**(3):788-794.

31. Domejean S, Zhan L, DenBesten PK, Stamper J, Boyce WT, Featherstone JD. Horizontal transmission of mutans streptococci in children. *J DENT RES* 2010;**89**(1):51-55.

32. Liu Y, Zou J, Shang R, Zhou XD. Genotypic diversity of *Streptococcus mutans* in 3- to 4-year-old Chinese nursery children suggests horizontal transmission. *Arch Oral Biol* 2007;**52**(9):876-881.

33. Alves AC, Nogueira RD, Stipp RN, Pampolini F, Moraes AB, Goncalves RB, et al. Prospective study of potential sources of *Streptococcus mutans* transmission in nursery school children. *J Med Microbiol* 2009;**58**(Pt 4):476-481.

34. Lapirattanakul J, Nakano K, Nomura R, Leelataweewud P, Chalermsarp N, Klaophimai A, et al. Multilocus sequence typing analysis of *Streptococcus mutans* strains with the cnm gene encoding collagen-binding adhesin. *J Med Microbiol* 2011;**60**(Pt 11):1677-1684.

35. Momeni SS, Whiddon J, Cheon K, Moser SA, Childers NK. Assessment of clonality and serotypes of *Streptococcus mutans* among children by multilocus sequence typing. *Eur J Oral Sci* 2015.

Table 1. Summary of new alleles* and sequence types identified by MLST analysis in this study.

PubMLST ST	*tkt*	*glnA*	*gltA*	*glk*	*aroE*	*murI*	*lepC*	*gyrA*
207	20	2	24	3	4	3	1	1
208	19	2	10	3	8	3	5	1
209	2	2	4	3	4	**25**	1	4
210	1	2	5	3	14	3	1	**23**
211	1	16	8	1	2	3	1	1
212	14	2	4	1	2	3	1	**24**

*New alleles are **bold**.

Table 2. Prediction accuracy for two extreme isolates and internal branch groups for the ten most common rep-PCR genotypes.

	Extreme Branches (2 isolates)		Internal Branches (2-3 isolates)				
Genotype	**Predicted match**	**Actual match***	**Predicted match**	**No. of isolates**	**Groups**	**Predicted correctly (%)**	**SLV, DLV predicted correctly (%)**
G18	Yes	**Yes**	Yes	10	5	5 (100)	5 (100)
G09	Yes	No	No	9	5	2 (40)	4 (80)
G01a	No	**No**	No	10	5	3 (60)	4 (80)
G12	No	**No**	No	10	5	3 (60)	4 (80)
G01	No	**No**	No	11	5	3 (60)	4 (80)
G13	No	**No**	No	11	7	4 (57)	4 (57)
G06	No	**No**	No	9	5	2 (40)	2 (40)
G07	No	**No**	No	9	5	4 (80)	5 (100)
G23	No	**No**	No	9	5	2 (40)	2 (40)
G22	No	**No**	No	9	5	4 (80)	4 (80)
Predicted accurately		9 of 10	Totals	97	52	32 (62)	38 (73)

* Correct predictions are bold.
SLV; single locus variant
DLV; double locus variant
Total combinations represented is 62 (52 from internal branch groups and 10 from extremes for each genotype)

Table 3. Prediction accuracy for children only and child-parent pairs with rep-PCR genotype G12.

A. Parent-child Pairs

Family member ID*	No. of Isolates	Predicted extreme ST match	Actual extremes ST match	Internal isolates ST match extreme ST	Predicted parent-child match	Actual parent-child match
107-C	1	-	-	-	Yes	Yes
107-M	1	-	-	-		
170-C	4	No	No	Yes	No	No
170-M	1	-	-	-		
180-C	1	-	-	-	Yes	Yes
180-M	1	-	-	-		
182-C	6	No	Yes	Yes	Yes	Yes
182-M	2	No	Yes	-		
223-C	1	-	-	-	Yes	Yes
223-M	1	-	-	-		
227-C	3	Yes	Yes	Yes	No	No
227-M	1	-	-	-		
518-C	5	Yes	Yes	Yes	Yes	Yes
518-M	1	-	-	-		
596-C	4	No	Yes	Yes	Yes	Yes
596-M	1	-	-	-		
600-C	6	No	Yes	Yes	Yes	Yes
600-M	3	No	Yes	Yes		
226-C	2	Yes	Yes	-	Yes	No
226-F	2	No	Yes	-		
	Predicted Correctly (%)		50.0%	100.0%		100% M-C 0% F-C

B. Child Only**

Family Member ID*	No. of isolates	Predicted extreme ST match	Actual extreme ST match	Internal isolates ST match extreme ST
159-C	4	No	Yes	Yes
219-C	2	Yes	Yes	-
241-C	6	No	Yes	No
597-C	3	No	Yes	Yes
608-C	4	No	Yes	Yes
	Predicted Correctly (%)		20.0%	75.0%

* C; Child, M; mother, F; Father
**For child only table, only 5 of the 14 children are shown because all others had only one isolate evaluated with MLST.
ST; MLST sequence type

Table 4. Commonality of MLST sequence types for children only and parent-child pairs with rep-PCR genotype G12.

A. Mother-Child Pairs with G12

ID	Total G12 Isolates	Isolates Tested	ST
107-C	1	1	181
107-M	2	1	181
170-C	24	4	157 (3), 202 (1)
170-M	13	1	212
180-C	15	1	157
180-M	1	1	157
182-C	74	6	202
182-M	4	2	202
223-C	10	1	157
223-M	5	1	119*
227-C	42	3	202
227-M	11	1	157
518-C	35	5	156
518-M	1	1	156
518-B	13	1	156
596-C	32	4	192
596-M	7	1	192
600-C	92	6	201
600-M	4	3	201
Totals	383	42	

B. Child Only with G12

ID	Total G12 Isolates	Isolates Tested	ST
150-C	1	1	175
159-C	113	4	157
218-C	3	1	202
219-C	16	2	119*
222-C	1	1	166
241-C	106	6	181 (2), 132* (4)
228-C	1	1	150
502-C	1	1	92**
509-C	12	1	202
528-C	1	1	201
588-C	1	1	201
597-C	24	3	157
601-C	1	1	201
608-C	22	4	157

C. Father-Child Pairs with G12

226-C	15	2	157
226-F	9	2	132*
Totals	327	32	

Sequence types (ST) shared between unrelated individuals are colored.
* Matches a sequence type previously reported from Thailand
** Matches a sequence type previously reported from Japan

Table S1. Summary statistics of using rep-PCR to predict matching and non-matching MLST sequence types

Actual Match/No-Match

	Actual Match	Actual No-match	*p*-value
Predicted Match	22	18	0.0004
Predicted No Match	2	20	
Sensitivity	91.7%		
Specificity	52.6%		
Positive Predictive Value	55.0%		
Negative Predictive Value	90.9%		

Match/No-Match Including SLV and DLV

	Actual Match	Actual No-match	*p*-value
Predicted Match	28	12	<0.0001
Predicted No Match	2	20	
Sensitivity	93.3%		
Specificity	62.5%		
Positive Predictive Value	70.0%		
Negative Predictive Value	90.9%		

Table S2. Listing of other family members for all children with rep-PCR genotype G12 indicating genotype match and MLST sequence types for transmission source tracking.

A. Mother-Child Match

Child ID	Other Family Members	No. of Isolates analyzed with rep-PCR	Rep-PCR G12	MLST ST (Child)	MLST ST (Family)
107	Mother	17	yes	181	181
	Grandmother	19	yes		TBD
170	Mother	28	yes	157, 202	212
	Aunt	12	no		
	Cousin 1	14	no		
	Cousin 2	13	no		
	Cousin 3	6	no		
180	Mother	11	yes	157	157
	Sister	-	-		
	Brother 1	-	-		
	Brother 2	-	-		
	Brother 3	-	-		
	Aunt	12	yes		TBD
182	Mother	23	yes	202	202
	Sister***	20	yes		-
223	Mother	8	yes	157	119
	Sister	7	yes		TBD
	Brother	6	no		
227	Mother	14	yes	202	157
	Brother 1	21	no		
	Brother 2	24	no		
	Cousin 1	27	yes		TBD
	Cousin 2	24	no		
	Cousin 3	10	no		
	Cousin 4	25	no		
	Cousin 5	27	yes		TBD
518	Mother	3	yes	156	156
	Father	4	no		
	Brother	20	yes		156
596	Mother	20	yes	192	192
	Aunt	24	yes		TBD
600	Mother	24	yes	201	201
	Sister***	12	yes		-

B. Child Only

Child ID	Other Family Members	No. of Isolates analyzed with rep-PCR	Rep-PCR G12	MLST ST (Child)	MLST ST (Family)
150	Mother	13	no	175	
	Brother	18	no		
	Brother 2	-	-		
159	Grandmother	22	no	157	
218	Mother	19	no	202	
	Father	10	no		
	Sister	13	no		
	Brother	13	no		
	Grandmother	20	no		
	Aunt	25	no		
	Cousin	20	no		
219	Mother	25	no	119	
	Father	5	no		
	Sister 1	15	yes		TBD
	Sister 2	14	yes		TBD
	Sister 3	3	yes		TBD
	Brother	9	no		
	Aunt	17	no		
	Cousin 1	14	no		
	Cousin 2	14	no		
	Other	7	yes		TBD
222	Mother	25	no	166	
	Brother 1*	-	no		
	Brother 2	13	no		
	Brother 3	11	no		
228	Mother*	-	no	150	
	Father	16	no		
	Brother	3	no		
	Brother	13	no		
241	Sister 1	7	no	181, 132	
	Sister 2	24	no		
	Brother	15	no		
502	Mother	15	no	92	
	Sister 1	1	no		
	Sister 2	1	no		
	Sister 3	14	no		
	Brother	11	no		

B. Child Only (continued)

Child ID	Other Family Members	No. of Isolates analyzed with rep-PCR	Rep-PCR G12	MLST ST (Child)	MLST ST (Family)
509	Mother	6	no	202	
	Sister	23	yes		TBD
	Aunt	24	no		
	Cousin 1	9	no		
	Cousin 2	24	no		
528	Mother	10	no	201	
	Father	7	no		
	Brother 1	10	no		
	Brother 2	10	no		
588	Mother	3	no	201	
	Sister	10	no		
	Grandmother	6	no		
	Uncle	22	no		
597	Mother	22	no	157	
601	Brother*	-	no	201	
	Cousin	4	no		
608	Mother	10	no	157	
	Aunt	1	no		
	Cousin 1	23	no		
	Cousin 2	7	no		
	Cousin 3**	-	-		
	Cousin 4	2	no		

C. Father-Child Match

Child ID	Other Family Members	No. of Isolates analyzed with rep-PCR	Rep-PCR G12	MLST ST (Child)	MLST ST (Family)
226	Father	12	yes	157	132
	Brother	10	no		
	Aunt	7	no		
	Cousin	1	no		

* Isolates analyzed were not *S. mutans*
** No *S. mutans* available
*** Only siblings shared genotype, thus no MLST analysis was performed.
TBD; MLST sequence type to be determined.

SUMMARY

The current body of work the aim was to investigate the genetic diversity of *S. mutans* by molecular typing with rep-PCR and MLST. Therefore, a main focus was to evaluate evidence for transmission of strains, especially upon the initial colonization. Additionally, the ability to track *S. mutans* strains is particularly important for the large epidemiological study if more caries virulent strains are to be identified. Understanding how these strains are transmitted and shared within a population will be very beneficial assessment of caries risk and to design targeted treatment/prevention strategies.

Initially, it was important to confirm that the methods to be used were valid. In paper 1, the use of Mitis Salivarius supplemented with bacitracin (MSB or Gold's) media was justified by demonstrating this media was effective by comparison to other selective medias available for cultivation of mutans streptococci. In this study caries related bacteria, *S. mutans* and *S. sobrinus* prototype strains, were found to grow equally well on all five selective media. This study also confirmed that the all media evaluated were comparable in cultivation of predominate rep-PCR genotypes. In addition, the recovery of minor genotypes did vary by media when clinical samples were plated directly; however, this appeared to be sampling variation since these strains were found

to grow successfully when plated directly. These results indicate that recovery of minor genotypes isolation was a factor of abundance rather than inhibition by the selective media used. This study confirmed that the media used (MSB) would not limit the genotypes recovered.

In Paper 2, evaluation of 2 MLST typing schemes for the selection of one for general use in further studies was performed. Although it was postulated that both MLST schemes should provide comparable results, it was necessary to independently evaluate the schemes to determine if any significant difference would be observed before apply either scheme to our research question (Maiden, 2006). The Nakano and Do MLST schemes were each evaluated for discriminatory power, congruence, phylogenetic agreement and convenience criteria (Do et al., 2010; Nakano et al., 2007a). No significant differences were observed between the two schemes and thus, the Nakano scheme was select for consistency since our initial published work was completed using this scheme (Momeni et al., 2013).

Using a sample size of 33 isolates, we found the results to be representative by selecting the most common rep-PCR genotypes. Other researchers planning to employ MLST may need to validate the schemes for consistency for their typing method (e.g. AP-PCR). An important future study would be a global analysis of the MLST ST in the www.PubMLST.org data for comparison to unique isolates from this Uniontown cohort. Using the findings in Paper 2, the importance of registering the sequence type data, in conjunction with those already reported to the www.PubMLST.org database for global

analysis, is exemplified by the realization that concatenated sequences reported by our laboratory were different in length that originally reported by Nakano (Momeni et al., 2013; Nakano et al., 2007a). This global analysis will be used to determine if *S. mutans* sequence types from the Uniontown study are unique. Depending on the similarity of our strains to those reported elsewhere, an ancestral source might be identified, contributing significantly to the understanding of the evolutionary history of *S. mutans*.

A previous report from our laboratory found clonality of *S. mutans* sequence types but selection bias could not be ruled out because rep-PCR was used to pre-select the isolates for MLST analysis (Momeni et al., 2013). It was therefore necessary in Study 3 to investigate if the clonality of *S. mutans* MLST sequence types previously observed was a true characteristic of the Uniontown study population. Using a randomly selected pool of *S. mutans* isolates for this localized population, clonality of *S. mutans* sequence types was confirmed. This may be because the population was geographically localized and ethnically similar (the subjects were all African-American).

This manuscript included serotype data by PCR identification to be consistent with other *S. mutans* MLST publications. This additional analysis resulted in the discovery of *S. mutans* serotype *k* for the first time in an United States population. Furthermore, the prevalence of serotype *k* was notably higher (17.5%) than reported by others (with the exception of a study in India) (Lapirattanakul et al., 2009; Nakano et al., 2004b; Rao and Austin, 2014). This unexpected finding may have significant clinical impact since the rarely found

oral *S. mutans* serotype *k* has been associated with systemic disease such as cardiovascular health (in particular, infective endocarditis), bacteremia, hemorrhagic stroke and inflammatory conditions. The studies reporting this association have found key collagen-binding proteins (Cnm, Cbm) associated with this serotype, which have been suggested to make serotype *k* strains highly invasive to human endothelial cells, and therefore may be important in the etiology of systemic complications. While the *S. mutans* evaluated in this study did not have these collagen-binding proteins, the sample size and the use of individual bacterial isolates may not have been sufficient to adequately screen this population.

The discovery of serotype *k* also opened a new avenue for investigations for the *S. mutans* isolates from this population. Future studies are underway investigating the prevalence of all *S. mutans* serotypes, especially serotype *k*, using a both a multiplex and conventional PCR based approach. Using whole saliva and plaque samples, as well as individual bacterial isolates, to determine the serotype profile for all the index children will allow us to evaluate for difference of serotype profiles more comprehensively. Using the whole saliva and plaque samples, rather than just isolates will result in a more comprehensive screening that may identify an even higher prevalence of serotype *k* than already observed. Further study is needed to determine if increased serotype *k* prevalence is unique to this population. It would also be beneficial to further investigate the frequency diseases associated with serotype *k* in the Uniontown

population, possibly initially with a questionnaire to previously participating familes.

Another important finding in Study 3 was the finding of 2 strains for which no serotype was identified by PCR analysis. These strains were confirmed as *S. mutans* both by 16s identification and with *gtfB*-specific primers. Nakano *et al.* identified 11 *S. mutans* strains that also could not be identified by the PCR approach in isolates obtained from cardiovascular tissue samples and they theorized this maybe a new minor oral serotype (Nakano et al., 2007b). Our laboratory is in the process of performing whole genome sequencing on the two strains identified in Study 3 to confirm if there are significant differences in the RPG operon consistent with these strains being a new *S. mutans* serotype. If confirmed, additional analysis may be needed to determine if these isolates are new serotype(s). Once established as a new serotype, a PCR identification approach will need to be designed, tested against the available strains through international collaboration, and applied to the Uniontown samples to determine prevalence. If the two strains identified in Study 3 match those of the Japanese group's cardiovascular samples, then additional studies will need to be conducted to determine this serotype's role in human cardiovascular health. This is an example of the significance and opportunity for further international collaborative study initiated by the findings of the studies reported.

Paper 4 had two primary goals: 1) to summarize the overall genetic diversity of more than 13,000 *S. mutans* isolates evaluated thus far by rep-PCR from the Uniontown study providing a DiversiLab reference library to be made

available for the global epidemiological surveillance of *S. mutans* and 2) to provide a broad overview of evidence for transmission of *S. mutans* among the children and their family members as well as with extra-familial subjects in the study by evaluating the commonality of the genotypes. The genetic diversity of *S. mutans* by rep-PCR analysis was found to be minimal given the 34 representative genotypes. This finding is primarily due to the banding pattern criteria used to define rep-PCR genotypes (i.e., based upon similarity which allowed up to two minor bands to be different) (Moser et al., 2010). Subgroups of rep-PCR genotypes are clearly discernable; therefore it was possible for some rep-PCR genotype to be further differentiation with an alternate typing method, such as MLST (Momeni et al., 2013). The choice to include up to two bands variation was important to classifications, since making every band count would have resulted in far too many rep-PCR genotypes to provide meaningful analysis.

Paper 4 further established the association of multiple genotypes in the risk of children experiencing dental caries. Further analysis on a child-by-child based will be required to provide more insight into the caries etiology. It has been suggested that intra-species microbial interactions of multiple *S. mutans*, and competitive survival interactions may increase caries risk (Alaluusua and Matto, 1996; Pieralisi et al., 2010; Zhao et al., 2014). Given that younger children in this study were almost 2.3 times more likely to have multiple genotypes and the caries prevalence by age 4 is 66% seems to provide some support for this hypothesis (Ghazal et al., 2015). Often multiple genotypes were found to be transient (only found once during the longitudinal sampling). The

probable reason that younger children have more genotypes is likely do to common child sharing behaviors (e.g., a shared ice cream cone) and a lack concern by children regarding the sharing items tainted with saliva (e.g., toys to the mouth) (Emilson et al., 1989; Liu et al., 2007).

The future direction of this part of Paper 4 is rich with possibilities and opportunities for further study of the etiology of dental caries. For example, a study using a rodent caries model might be beneficial to investigating the individual genotypes identified to be associated with or protective from caries for experimental disease outcomes. It would also be important to evaluate the synergistic cariogenic effects of different combinations of multiple *S. mutans* genotypes for differences caries outcomes. Taken one step further, cross-species microbial interactions could provide valuable insights into the ability of different *S. mutans* stains to inhibit or to be inhibited by commensal bacteria such as the mitis group of streptococci, which are high producers of hydrogen peroxide, an inhibitor of *S. mutans*.

In the second aim of Paper 4, evidence for the transmission of *S. mutans* was broadly evaluated. Whilst the findings do support that the mother is the primary source of *S. mutans* among 63% of children, the added benefit of performing analysis by both child and then by genotype indicated that 72% of children had at least 1 genotype not shared with any one evaluated from their household. This finding highlights potential of the high level of child-to-child transmission. Given the high number of transient genotypes in this study, it is possible that some of the unshared genotypes may have no clinical significance;

however, a more detailed analysis of the data is warranted to determine the role these genotypes play in the development of dental caries. Further study is also needed to determine the role of the host immune system in influencing which genotypes become established (and lead to disease). For instance, questions such as: Is the established predominate genotype the same as mothers or from another source, such as interfamilial or child-to-child transmission; and How often do these genotypes from different sources result in caries compared to maternal genotypes?

Another important finding in Paper 4 is the discovery of so many minor (or transient) genotypes, which may have no relevance to disease and may be confounding the search for strains with greater virulence. Previous studies have typically not made corrections for minor genotypes because the samples sizes have been to small to determine minor genotypes. This finding provides a unique advantage for larger scale studies to perform analysis that include and exclude these isolates to determine which strains may be associated with disease. Previous studies have also not included another valuable contribution of the discovery of these minor genotypes, which is understanding of the importance of the addition of a time component in the search for strains associated with disease. Dental caries is a chronic disease and it can take time after an *S. mutans* genotype is acquired before dental caries is observed. For instance, initial analysis may indicate a genotype is associated with disease for 80% of children analyzed. The addition of the time component would easily eliminate minor genotypes, while simultaneously eliminating children who have

not had the strain long enough to cause disease, greatly altering the percentages and bringing values closer to meaningful conclusions.

The clinically significant contribution of Study 4 is that when strains are identified as more virulent strains (therefore, cause more aggressive early childhood caries), then it will be possible to develop rapid screening tools, possibly PCR based, that can tell clinicians if a child is at higher risk of severe dental caries (cavities) because they harbor a high risk stain of *S. mutans*. This analysis may even be extended to pre-natal mothers or the primary caregivers. Based on these caries risk screenings the clinicians would be able to choose more appropriate treatment plans (e.g. which children are at higher risk and would best benefit from fluoride varnish treatments. If a child or their mother has a high risk *S. mutans* strain, then justification based on scientific evidence can be made to insurers to cover early, preventative treatments.)

The final study, represented in Study 5, sought to determine how accurately rep-PCR could determine MLST sequence types. The overall goal of testing this new method was to offer investigators a more cost-beneficial approach to large-scale epidemiological studies of clinically relevant bacteria. By pre-screening the isolates with the less expensive rep-PCR method it could be possible to focus which isolates should be evaluated with more specific, and more expensive, downstream application such as MLST or whole genome sequencing. The results in Paper 5, indicate that rep-PCR is reasonably capable of predicting MLST sequence types, with 73% predication accuracy for unrelated and 100% accuracy for related isolates, for *S. mutans*. This approach would

need to be tested with other clinically relevant bacterial strains to confirm its application for other bacteria because some studies have indicated that rep-PCR does not always have a comparable discriminatory power as other typing methods, such as MLST, for some clinical bacteria (Al Nakib et al., 2011; Behringer et al., 2011).

The secondary goal of Study 5 was to evaluate the transmission of rep-PCR genotype 12 among children and their household members. Initially this study focused on children with G12 and their mothers and fathers. This analysis suggested potential sources for commonly shared G12 isolates with the same sequence types. Thus other family members for these children that had G12 were also evaluate to confirm the source. This data indicated that younger children had the same genotypes as their mothers, but older children were more likely to have distinct genotypes from the mothers suggesting that these children acquire new strains from sources other than their mothers later in life.

The data in Paper 5 is intended as preliminary data to illustrate the potential of a combined rep-PCR and MLST analysis for epidemiological study of *S. mutans*. This study shows that MLST sequence types were typically shared within a family but that some other sequence types are more transmissible and shared within the community (e.g. ST 202). The probable sources of many of these strains were identified among those that participated in the study. Although, other sources are possible, this study clearly demonstrates how effective this combined approach could be in tracking *S. mutans strains* related to dental caries.

Study 5 also provides some insights into the scale of analysis that is required to see significant observations in a large-scale epidemiological study. Consider that overall 13,906 *S. mutans* isolates were analyzed by rep-PCR of which only 1,008 were genotype G12. Of these, there were only 24 children with 23 household family members in this study that had G12. This narrowing sample pool explains the need to analyze large numbers of isolates to observe these results. Furthermore, it indicates that that this approach will not be useful for all the rep-PCR genotypes observed in the study. Only those with a larger number of isolates would benefit from such analysis. An additional challenge is that some genotypes, such a G18 that has consistently shown the same rep-PCR genotype and MLST ST (ST166), could require an additional genotyping method to further differentiate the groups. Although, an alternate possibility is that it is that genotypes like G18 may be highly common and transmissible across populations. These strains may be early colonizers that, by themselves, are not leading to disease but that in aggregation with other strains acquired later lead to caries or that host factors may cause these strains to mutate leading to dysbiosis of the oral cavity. Further analysis of the data is needed to determine how such strains may contribute to dental caries.

In conclusion, the current body of work has sought to demonstrate the importance of using molecular based genotyping techniques for investigating the genetic diversity of *S. mutans* in the epidemiological study of dental caries. These findings have contributed to the dental and epidemiological fields' understanding of *S. mutans* strain type diversity using a method currently unique

to our laboratory (rep-PCR), as well as a commonly accepted and well-utilized sequence based method (MLST). Furthermore, this work has contributed significantly to information available regarding the evidence for child-to-child transmission and probable strain source tracking. Future directions for continued study and a deeper analysis of this rich data mine offers the potential for new understanding of how and which *S. mutans* strains contribute to the initiation and progression of dental caries, including early childhood caries.

GENERAL LIST OF REFERENCES

Al Nakib M, Longo M, Tazi A, Billoet A, Raymond J, Trieu-Cuot P *et al.* (2011). Comparison of the Diversilab(R) system with multi-locus sequence typing and pulsed-field gel electrophoresis for the characterization of Streptococcus agalactiae invasive strains. *J Microbiol Methods* 85(2)*:*137-142.

Alaluusua S, Alaluusua SJ, Karjalainen J, Saarela M, Holttinen T, Kallio M *et al.* (1994). The demonstration by ribotyping of the stability of oral *Streptococcus mutans* infection over 5 to 7 years in children. *Arch Oral Biol* 39(6)*:*467-471.

Alaluusua S, Matto J (1996). Oral colonization by more than one clonal type of mutans streptococcus in children with nursing-bottle dental caries. *Arch Oral Biol* 41(2)*:*167-173.

Alves AC, Nogueira RD, Stipp RN, Pampolini F, Moraes AB, Goncalves RB *et al.* (2009). Prospective study of potential sources of *Streptococcus mutans* transmission in nursery school children. *J Med Microbiol* 58(Pt 4)*:*476-481.

American Acacdemy of Pediatric Dentistry (2015). Policy on Early Childhood Caries (ECC): Classifications, Consequences, and Preventive Strategies. *Pediatr Dent* 37(6 (suppl))*:*50-52.

Baca P, Castillo AM, Liebana MJ, Castillo F, Martin-Platero A, Liebana J (2012). Horizontal transmission of *Streptococcus mutans* in schoolchildren. *Med Oral Patol Oral Cir Bucal* 17(3)*:*e495-500.

Bagramian RA, Garcia-Godoy F, Volpe AR (2009). The global increase in dental caries. A pending public health crisis. *Am J Dent* 22(1)*:*3-8.

Behringer M, Miller WG, Oyarzabal OA (2011). Typing of Campylobacter jejuni and Campylobacter coli isolated from live broilers and retail broiler meat by flaA-RFLP, MLST, PFGE and REP-PCR. *J Microbiol Methods* 84(2)*:*194-201.

Berkowitz RJ, Jones P (1985). Mouth-to-mouth transmission of the bacterium *Streptococcus mutans* between mother and child. *Arch Oral Biol* 30(4)*:*377-379.

Berkowitz RJ (2003). Causes, treatment and prevention of early childhood caries: a microbiologic perspective. *J Can Dent Assoc* 69(5)*:*304-307.

Bratthall D (1970). Demonstration of five serological groups of streptococcal strains resembling *Streptococcus mutans*. *Odontol Revy* 21(2)*:*143-152.

Carletto-Korber FP, Gonzalez-Ittig RE, Jimenez MG, Cornejo LS (2010). Initial acquisition and genetic identity of *Streptococcus mutans* of mother-child Pairs. *Pediatr Dent* 32(3)*:*205-211.

Cheon K, Moser SA, Whiddon J, Osgood RC, Momeni S, Ruby JD *et al.* (2011). Genetic Diversity of Plaque Mutans Streptococci with rep-PCR. *J Dent Res* 90(3)*:*331-335.

Davey AL, Rogers AH (1984). Multiple types of the bacterium *Streptococcus mutans* in the human mouth and their intra-family transmission. *Arch Oral Biol* 29(6)*:*453-460.

Do T, Gilbert SC, Clark D, Ali F, Fatturi Parolo CC, Maltz M *et al.* (2010). Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. *PLoS One* 5(2)*:*e9073.

Domejean S, Zhan L, DenBesten PK, Stamper J, Boyce WT, Featherstone JD (2010). Horizontal transmission of mutans streptococci in children. *J DENT RES* 89(1)*:*51-55.

Douglass JM, Li Y, Tinanoff N (2008). Association of mutans streptococci between caregivers and their children. *Pediatr Dent* 30(5)*:*375-387.

Dye BA, Thornton-Evans G, Li X, Iafolla TJ (2015). Dental caries and sealant prevalence in children and adolescents in the United States, 2011-2012. *NCHS Data Brief* 191)*:*1-8.

Edelstein BL (2006). The dental caries pandemic and disparities problem. *BMC oral health* 6 Suppl 1(S2.

Emilson CG, Ciardi JE, Olsson J, Bowen WH (1989). The influence of saliva on infection of the human mouth by mutans streptococci. *Arch Oral Biol* 34(5)*:*335-340.

Foley SL, White DG, McDermott PF, Walker RD, Rhodes B, Fedorka-Cray PJ *et al.* (2006). Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *J Clin Microbiol* 44(10)*:*3569-3577.

Gamboa F, Chaves M, Valdivieso C (2010). Genotypic profiles by AP-PCR of *Streptococcus mutans* in caries-active and caries-free preschoolers. *Acta odontologica latinoamericana : AOL* 23(2)*:*143-149.

Ghazal T, Levy SM, Childers NK, Broffitt B, Cutter G, Wiener HW *et al.* (2015). Prevalence and incidence of early childhood caries among African-American children in Alabama. *J Public Health Dent* 75(1)*:*42-48.

Gilbert K, Joseph R, Vo A, Patel T, Chaudhry S, Nguyen U *et al.* (2014). Children with severe early childhood caries: streptococci genetic strains within carious and white spot lesions. *J Oral Microbiol* 6(

Gold OG, Jordan HV, Van Houte J (1973). A selective medium for *Streptococcus mutans*. *Arch Oral Biol* 18(11)*:*1357-1364.

Gross EL, Beall CJ, Kutsch SR, Firestone ND, Leys EJ, Griffen AL (2012). Beyond Streptococcus mutans: dental caries onset linked to multiple species by 16S rRNA community analysis. *PLoS One* 7(10)*:*e47722.

Hamada S, Slade HD (1980). Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol Rev* 44(2)*:*331-384.

Healy M, Reece K, Walton D, Huong J, Shah K, Kontoyiannis DP (2004). Identification to the species level and differentiation between strains of Aspergillus clinical isolates by automated repetitive-sequence-based PCR. *J Clin Microbiol* 42(9)*:*4016-4024.

Hirasawa M, Takada K (2003). A new selective medium for *Streptococcus mutans* and the distribution of *S. mutans* and *S. sobrinus* and their serotypes in dental plaque. *Caries Res* 37(3)*:*212-217.

Keyes PH, Jordan HV (1963). Factors influencing the initiation, transmission and inhibition of dental caries. In: Mechanisms of hard tissue destruction. S RF editor. New York, New York: American Association for the Advancement of Science.

Kimmel L, Tinanoff N (1991). A modified mitis salivarius medium for a caries diagnostic test. *Oral Microbiol Immunol* 6(5)*:*275-279.

Klein MI, Florio FM, Pereira AC, Hofling JF, Goncalves RB (2004). Longitudinal study of transmission, diversity, and stability of *Streptococcus mutans* and *Streptococcus sobrinus* genotypes in Brazilian nursery children. *J Clin Microbiol* 42(10)*:*4620-4626.

Kojima A, Nomura R, Naka S, Okawa R, Ooshima T, Nakano K (2014). Aggravation of inflammatory bowel diseases by oral streptococci. *Oral Dis* 20(4)*:*359-366.

Kozai K, Nakayama R, Tedjosasongko U, Kuwahara S, Suzuki J, Okada M *et al.* (1999). Intrafamilial distribution of mutans streptococci in Japanese families and possibility of father-to-child transmission. *Microbiol Immunol* 43(2)*:*99-106.

Kreulen CM, de Soet HJ, Hogeveen R, Veerkamp JS (1997). *Streptococcus mutans* in children using nursing bottles. *ASDC J Dent Child* 64(2)*:*107-111.

Kulkarni GV, Chan KH, Sandham HJ (1989). An investigation into the use of restriction endonuclease analysis for the study of transmission of mutans streptococci. *J Dent Res* 68(7)*:*1155-1161.

Lapirattanakul J, Nakano K, Nomura R, Hamada S, Nakagawa I, Ooshima T (2008). Demonstration of mother-to-child transmission of *Streptococcus mutans* using multilocus sequence typing. *Caries Res* 42(6)*:*466-474.

Lapirattanakul J, Nakano K, Nomura R, Nemoto H, Kojima A, Senawongse P *et al.* (2009). Detection of serotype *k Streptococcus mutans* in Thai subjects. *Oral Microbiol Immunol* 24(5)*:*431-433.

Lapirattanakul J, Nakano K, Nomura R, Leelataweewud P, Chalermsarp N, Klaophimai A *et al.* (2011). Multilocus sequence typing analysis of *Streptococcus mutans* strains with the cnm gene encoding collagen-binding adhesin. *J Med Microbiol* 60(Pt 11)*:*1677-1684.

Lapirattanakul J, Nomura R, Nemoto H, Naka S, Ooshima T, Nakano K (2013). Multilocus sequence typing of *Streptococcus mutans* strains with the *cbm* gene encoding a novel collagen-binding protein. *Arch Oral Biol* 58(8)*:*989-996.

Lapirattanakul J, Nomura R, Matsumoto-Nakano M, Srisatjaluk R, Ooshima T, Nakano K (2015). Variation of expression defects in cell surface 190-kDa protein antigen of *Streptococcus mutans*. *Int J Med Microbiol*.

Lembo FL, Longo PL, Ota-Tsuzuki C, Rodrigues CR, Mayer MP (2007). Genotypic and phenotypic analysis of *Streptococcus mutans* from different oral cavity sites of caries-free and caries-active children. *Oral Microbiol Immunol* 22(5)*:*313-319.

Leong PM, Gussy MG, Barrow SY, de Silva-Sanigorski A, Waters E (2013). A systematic review of risk factors during first year of life for early childhood caries. *Int J Paediatr Dent* 23(4)*:*235-250.

Li Y, Caufield PW (1995). The fidelity of initial acqusition of mutans streptococci by infants from their mothers. *J Dent Res* 74(2)*:*681-685.

Liu Y, Zou J, Shang R, Zhou XD (2007). Genotypic diversity of *Streptococcus mutans* in 3- to 4-year-old Chinese nursery children suggests horizontal transmission. *Arch Oral Biol* 52(9)*:*876-881.

Loesche W (1986). Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* 50(4)*:*353-380.

Lynch DJ, Villhauer AL, Warren JJ, Marshall TA, Dawson DV, Blanchette DR *et al.* (2015). Genotypic characterization of initial acquisition of *Streptococcus mutans* in American Indian children. *J Oral Microbiol* 7(27182.

Maiden MC (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 60(561-588.

Marsh PD (1991). Sugar, fluoride, pH and microbial homeostasis in dental plaque. *Proc Finn Dent Soc* 87(4)*:*515-525.

Mattos-Graner RO, Li Y, Caufield PW, Duncan M, Smith DJ (2001). Genotypic diversity of mutans streptococci in Brazilian nursery children suggests horizontal transmission. *J Clin Microbiol* 39(6)*:*2313-2316.

Mitchell SC, Ruby JD, Moser S, Momeni S, Smith A, Osgood R *et al.* (2009). Maternal transmission of mutans Streptococci in severe-early childhood caries. *Pediatr Dent* 31(3)*:*193-201.

Momeni SS, Whiddon J, Moser SA, Cheon K, Ruby JD, Childers NK (2013). Comparative genotyping of *Streptococcus mutans* by repetitive extragenic palindromic polymerase chain reaction and multilocus sequence typing. *Mol Oral Microbiol* 28(1)*:*18-27.

Moser SA, Mitchell SC, Ruby JD, Momeni S, Osgood RC, Whiddon J *et al.* (2010). Repetitive extragenic palindromic PCR for study of *Streptococcus mutans* diversity and transmission in human populations. *J Clin Microbiol* 48(2)*:*599-602.

Nakano K, Nomura R, Nakagawa I, Hamada S, Ooshima T (2004a). Demonstration of *Streptococcus mutans* with a cell wall polysaccharide specific to a new serotype, *k*, in the human oral cavity. *J Clin Microbiol* 42(1)*:*198-202.

Nakano K, Nomura R, Shimizu N, Nakagawa I, Hamada S, Ooshima T (2004b). Development of a PCR method for rapid identification of new *Streptococcus mutans* serotype *k* strains. *J Clin Microbiol* 42(11)*:*4925-4930.

Nakano K, Tsuji M, Nishimura K, Nomura R, Ooshima T (2006). Contribution of cell surface protein antigen PAc of *Streptococcus mutans* to bacteremia. *Microbes Infect* 8(1)*:*114-121.

Nakano K, Lapirattanakul J, Nomura R, Nemoto H, Alaluusua S, Gronroos L *et al.* (2007a). *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. *J Clin Microbiol* 45(8)*:*2616-2625.

Nakano K, Nemoto H, Nomura R, Homma H, Yoshioka H, Shudo Y *et al.* (2007b). Serotype distribution of Streptococcus mutans a pathogen of dental caries in cardiovascular specimens from Japanese patients. *J Med Microbiol* 56(Pt 4)*:*551-556.

Nakano K, Nomura R, Nemoto H, Mukai T, Yoshioka H, Shudo Y *et al.* (2007c). Detection of novel serotype *k Streptococcus mutans* in infective endocarditis patients. *J Med Microbiol* 56(Pt 10)*:*1413-1415.

Nakano K, Nomura R, Nemoto H, Lapirattanakul J, Taniguchi N, Gronroos L *et al.* (2008). Protein antigen in serotype *k Streptococcus mutans* clinical isolates. *J Dent Res* 87(10)*:*964-968.

Nakano K, Nemoto H, Nomura R, Inaba H, Yoshioka H, Taniguchi K *et al.* (2009). Detection of oral bacteria in cardiovascular specimens. *Oral Microbiol Immunol* 24(1)*:*64-68.

Nakano K, Nomura R, Taniguchi N, Lapirattanakul J, Kojima A, Naka S *et al.* (2010). Molecular characterization of *Streptococcus mutans* strains containing the *cnm* gene encoding a collagen-binding adhesin. *Arch Oral Biol* 55(1)*:*34-39.

Nakano K, Hokamura K, Taniguchi N, Wada K, Kudo C, Nomura R *et al.* (2011). The collagen-binding protein of *Streptococcus mutans* is involved in haemorrhagic stroke. *Nat Commun* 2(485.

Napimoga MH, Kamiya RU, Rosa RT, Rosa EA, Hofling JF, Mattos-Graner R *et al.* (2004). Genotypic diversity and virulence traits of *Streptococcus mutans* in caries-free and caries-active individuals. *J Med Microbiol* 53(Pt 7)*:*697-703.

Nomura R, Nakano K, Naka S, Nemoto H, Masuda K, Lapirattanakul J *et al.* (2012). Identification and characterization of a collagen-binding protein, Cbm, in *Streptococcus mutans*. *Mol Oral Microbiol* 27(4)*:*308-323.

Nomura R, Naka S, Nemoto H, Inagaki S, Taniguchi K, Ooshima T *et al.* (2013). Potential involvement of collagen-binding proteins of *Streptococcus mutans* in infective endocarditis. *Oral Dis* 19(4)*:*387-393.

Palmer SR, Miller JH, Abranches J, Zeng L, Lefebure T, Richards VP *et al.* (2013). Phenotypic heterogeneity of genomically-diverse isolates of Streptococcus mutans. *PLoS One* 8(4)*:*e61358.

Perch B, Kjems E, Ravn T (1974). Biochemical and serological properties of *Streptococcus mutans* from various human and animal sources. *Acta Pathol Microbiol Scand B Microbiol Immunol* 82(3)*:*357-370.

Petersen PE, Bourgeois D, Ogawa H, Estupinan-Day S, Ndiaye C (2005). The burden of oral disease: challenges to improving oral health in the 21st century. *Bulletin of the World Health Organization* 83(1)*:*3.

Pieralisi FJ, Rodrigues MR, Segura VG, Maciel SM, Ferreira FB, Garcia JE *et al.* (2010). Genotypic Diversity of *Streptococcus mutans* in Caries-Free and Caries-Active Preschool Children. *Int J Dent* 2010(824976.

Rao AP, Austin RD (2014). 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* 5(3)*:*296-301.

Redmo Emanuelsson IM, Wang XM (1998). Demonstration of identical strains of mutans streptococci within Chinese families by genotyping. *Eur J Oral Sci* 106(3)*:*788-794.

Redmo Emanuelsson IM, Carlsson P, Hamberg K, Bratthall D (2003). Tracing genotypes of mutans streptococci on tooth sites by random amplified polymorphic DNA (RAPD) analysis. *Oral Microbiol Immunol* 18(1)*:*24-29.

Sato Y, Okamoto K, Kagami A, Yamamoto Y, Igarashi T, Kizaki H (2004). *Streptococcus mutans* strains harboring collagen-binding adhesin. *J Dent Res* 83(7)*:*534-539.

Schaeken MJM (1986). Comparative Recovery of *Streptococcus mutans* on Five Isolation Media,Including a New Simple Selective Medium. *J Dent Res* 65(906.

Sebelius K, Frieden TR, Sondik EJ (2010). Health, United States, 2009: With Special Feature on Medical Technology. In: NCfH Statistics editor. Hyattsville, MD: U. S. Government Printing Office, Washington, DC, pp. 306-307.

Shibata Y, Ozaki K, Seki M, Kawato T, Tanaka H, Nakano Y *et al.* (2003). Analysis of loci required for determination of serotype antigenicity in *Streptococcus mutans* and its clinical utilization. *J Clin Microbiol* 41(9)*:*4107-4112.

Tanner AC, Milgrom PM, Kent R, Jr., Mokeem SA, Page RC, Riedy CA *et al.* (2002). The microbiota of young children from tooth and tongue samples. *J Dent Res* 81(1)*:*53-57.

Tanzer JM, Borjesson AC, Laskowski L, Kurasz AB, Testa M (1984). Glucose-sucrose-potassium tellurite-bacitracin agar, an alternative to mitis salivarius-bacitracin agar for enumeration of *Streptococcus mutans*. *J Clin Microbiol* 20(4)*:*653-659.

Teanpaisan R, Chaethong W, Piwat S, Thitasomakul S (2012). Vertical transmission of mutans streptococci and lactobacillus in Thai families. *Pediatr Dent* 34(2)*:*e24-29.

Tedjosasongko U, Kozai K (2002). Initial acquisition and transmission of mutans streptococci in children at day nursery. *ASDC J Dent Child* 69(3)*:*284-288, 234-285.

van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK *et al.* (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* 13 Suppl 3(1-46.

Van Palenstein Helderman WH, Ijsseldijk M, Huis in 't Veld JH (1983). A selective medium for the two major subgroups of the bacterium *Streptococcus mutans* isolated from human dental plaque and saliva. *Arch Oral Biol* 28(7)*:*599-603.

Wan AK, Seow WK, Walsh LJ, Bird PS (2002). Comparison of five selective media for the growth and enumeration of *Streptococcus mutans*. *Australian dental journal* 47(1)*:*21-26.

WHO (1997). Oral Health surveys: basic methods. 4th ed. Geneva: World Health Organization.

Zhao W, Li W, Lin J, Chen Z, Yu D (2014). Effect of sucrose concentration on sucrose-dependent adhesion and glucosyltransferase expression of *S. mutans* in children with severe early-childhood caries (S-ECC). *Nutrients* 6(9)*:*3572-3586.

APPENDIX

IRB APPROVAL FORMS

Project Revision/Amendment Form

Today's Date: December 8, 2009

1. Contact Information

Principal Investigator's Name: Noel Childers, DDS BlazerID: nkc E-mail: nkc@uab.edu
Contact Person's Name: Stephanie McLean BlazerID: smclean E-mail: smclean@uab.edu
Telephone: 996-6628 Fax: 934-7013
Campus Address: SDB 304B

RECEIVED DEC 14 2009

2. Protocol Identification

Protocol Title: Epidemiology of Dental Caries and Immunity in Children (Alabama)
IRB Protocol Number: F060328001

Current Status of Project (check only one):

- ☒ Currently in Progress (Number of participants entered: 868)
- ☐ Study has not yet begun (No participants entered)
- ☐ Closed to participant enrollment (remains active)—
  Number of participants on therapy/intervention:_____
  Number of participants in long-term follow-up only:_____
- ☐ Closed to participant enrollment (data analysis only)—
  Total number of participants enrolled:_____

☑ Approved Expedited ☐ To Convened IRB

Chair or Vice-Char MD Date Dec 17, 2009

This submission changes the status of this study in the following manner (check all that apply):

- ☐ Protocol Revision
- ☒ Protocol Amendment
- ☐ Study Closed to participant entry
- ☐ Study Closure
- ☐ Other, (specify) _____
- ☐ Revised Consent Form
- ☐ Addendum (new) consent form
- ☐ Enrollment temporarily suspended by sponsor
- ☒ Change in protocol personnel

3. Reason for change

We are adding Stephanie Momeni to our list of study personnel. Ms. Momeni will be conducting experiments with our plaque samples that were identified only as "bacteriologic analysis" in our original HSP. These experiments consist of determining the number of *Streptococcus mutans* genotypes, recovered from patient plaque samples, using multilocus sequence typing (MLST) analysis. This is an additional method to evaluate for accuracy of the repetitive extragenic palindromic polymerase chain reaction (rep-PCR) method that is being used for *S. mutans* isolates. MLST will include an additional means of evaluation of commonality and diversity of these individual bacterial genotypes using both rep-PCR and MLST analysis. We are also evaluating similar methods using PCR and Microarrays to serve to confirm bacterial counts are *S. mutans and* as a substitute for the culture methods for counting bacteria in our project. The plaque samples used to cultivate these bacterial isolates are coded and do not contain any patient indentifying information. We have attached a copy of our current ICF for you to evaluate if this proposed change in protocol will require a change in our ICF.

4. Does this change revise or add a genetic or storage of samples component? ☐Yes ☒No

If yes, please see the Guidebook to assist you in revising or preparing your submission, or call the IRB office at 934-3789.

5. Does the change affect subject participation (e.g., procedures, risks, costs, location of services, etc.)? ☐Yes ☒No

If yes, Fiscal Approval Process (FAP)-designated units complete a FAP submission and send to fap@uab.edu. For more on the UAB FAP, see www.uab.edu/ohr.

6. Does the change affect the consent document(s)? ☐Yes ☒No

If yes, briefly discuss the changes. _____
Include the revised consent document with the changes highlighted.
Will any participants need to be reconsented as a result of the changes? ☐Yes ☒No
If yes, when will participants be reconsented?_____

Signature of Principal Investigator Noel K Childers Date 12/11/09
DOLA 12-11-09

OMB No. 0990-0263
Approved for use through 1/31/2012

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule. Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[ ] GRANT [ ] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
Epidemiology of Dental Caries and Immunity in Children (Alabama)	CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960, the expiration date 10/26/2010 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 11/11/2009 or [ ] Expedited Review on (date) _____
[ ] If less than one year approval, provide expiration date _____

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments
Protocol subject to Annual continuing review.

Title F060328001
Epidemiology of Dental Caries and Immunity in Children (Alabama)

IRB Approval Issued: 12/08/09

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: smoore@uab.edu	University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
14. Name of Official Ferdinand Urthaler, M.D.	15. Title Chairman, IRB
16. Signature Ferdinand Urthaler, MD	17. Date 12/08/09

Authorized for local Reproduction — Sponsored by HHS

Public reporting burden for this collection of information is estimated to average less than an hour per response. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: OS Reports Clearance Officer, Room 503 200 Independence Avenue, S.W., Washington, DC 20201. *Do not return the completed form to this address.*

OMB No. 0990-0263
Approved for use through 3/31/2015

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule. Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[ ] GRANT [X] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
Epidemiology of Dental Caries and Immunity in Children (Alabama)	CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960, the expiration date 01/24/2017 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 8/29/2012 or [ ] Expedited Review on (date)_____
[ ] If less than one year approval, provide expiration date _____

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments	Title F060328001
Protocol subject to Annual continuing review.	Epidemiology of Dental Caries and Immunity in Children (Alabama)

IRB Approval Issued: 09-12-12

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: irb@uab.edu	University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
14. Name of Official Ferdinand Urthaler, M.D.	15. Title Chairman, IRB
16. Signature Ferdinand Urthaler, MD/db	17. Date 09-12-12

Authorized for local Reproduction — Sponsored by HHS

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0990-0263. The time required to complete this information collection is estimated to average 30 minutes per response. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Health & Human Services, OS/OCIO/PRA, 200 Independence Ave., S.W., Suite 336-E, Washington D.C. 20201, Attention: PRA Reports Clearance Officer.

OMB No. 0990-0263
Approved for use through 3/31/2015

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule.

Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[ ] GRANT [ ] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
Epidemiology of Dental Caries and Immunity in Children (Alabama)	CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960 , the expiration date 01/24/2017 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 8/7/2013 or [ ] Expedited Review on (date) _____
[ ] If less than one year approval, provide expiration date _____

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments	Title F060328001
Protocol subject to Annual continuing review. HIPAA Waiver Approved?: No Partial HIPAA Waiver Approved?: No	Epidemiology of Dental Caries and Immunity in Children (Alabama)
IRB Approval Issued: 1/21/14	IRB Approval No Longer Valid On: August 7, 2014

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: irb@uab.edu	University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
14. Name of Official Albert Oberman, M.D., MPH	15. Title Vice Chair, IRB
16. Signature Albert Oberman, MD, MPH	17. Date 1/21/14

Authorized for local Reproduction — Sponsored by HHS

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0990-0263. The time required to complete this information collection is estimated to average 30 minutes per response. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Health & Human Services. OS/OCIO/PRA. 200 Independence Ave., S.W., Suite 336-E. Washington D.C. 20201. Attention: PRA Reports Clearance Officer.

OMB No. 0990-0263
Approved for use through 3/31/2015

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule.

Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[X] GRANT [ ] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	
4. Title of Application or Activity Epidemiology of Dental Caries and Immunity in Children (Alabama)		5. Name of Principal Investigator, Program Director, Fellow, or Other CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960 , the expiration date 01/24/2017 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 7/23/2014 or [ ] Expedited Review on (date) _____
[ ] If less than one year approval, provide expiration date _____

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments	Title F060328001
Protocol subject to Annual continuing review. HIPAA Waiver Approved?: No Partial HIPAA Waiver Approved?: No	Epidemiology of Dental Caries and Immunity in Children (Alabama)
IRB Approval Issued: 8/11/2014	IRB Approval No Longer Valid On: 7/23/2015

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: irb@uab.edu	University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
14. Name of Official Ferdinand Urthaler, M.D.	15. Title Chairman, IRB
16. Signature Ferdinand Urthaler MD MS	17. Date 8/11/2014

Authorized for local Reproduction — Sponsored by HHS

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0990-0263. The time required to complete this information collection is estimated to average 30 minutes per response. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Health & Human Services, OS/OCIO/PRA, 200 Independence Ave., S.W., Suite 336-E, Washington D.C. 20201, Attention: PRA Reports Clearance Officer.

OMB No. 0990-0263
Approved for use through March 31, 2018

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule. Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[X] GRANT [ ] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
Epidemiology of Dental Caries and Immunity in Children (Alabama)	CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960, the expiration date 01/24/2017 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 6/3/2015 or [ ] Expedited Review on (date)_____
[ ] If less than one year approval, provide expiration date _____
[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments	Title F060328001
Protocol subject to Annual continuing review. HIPAA Waiver Approved?: No Partial HIPAA Waiver Approved?: No	Epidemiology of Dental Caries and Immunity in Children (Alabama)
IRB Approval Issued: 6-3-2015	IRB Approval No Longer Valid On: 6-3-2016

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: irb@uab.edu	University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
14. Name of Official Albert Oberman, M.D., MPH	15. Title Vice Chair, IRB
16. Signature Albert Oberman MD, MPH (FIS)	17. Date 6-3-2015

Authorized for local Reproduction — Sponsored by HHS

OMB No. 0990-0263
Approved for use through March 31, 2018

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule. Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [✓] CONTINUATION [ ] EXEMPTION	[ ] GRANT [ ] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:_____	
4. Title of Application or Activity Epidemiology of Dental Caries and Immunity in Children (Alabama)		5. Name of Principal Investigator, Program Director, Fellow, or Other CHILDERS, NOEL

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960, the expiration date 01/24/2017 IRB Registration No. IRB00000196

[ ] This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance No._____, the expiration date_____ IRB Registration/Identification No._____*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph_____.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [X] Full IRB Review on (date of IRB meeting) 5/11/2016 or [ ] Expedited Review on (date)_____
[ ] If less than one year approval, provide expiration date _____

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments	Title F060328001
Protocol subject to Annual continuing review. HIPAA Waiver Approved?: No Partial HIPAA Waiver Approved?: No	Epidemiology of Dental Caries and Immunity in Children (Alabama)
IRB Approval Issued: 5/11/16	IRB Approval No Longer Valid On: 5/11/17

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.	10. Name and Address of Institution University of Alabama at Birmingham 701 20th Street South Birmingham, AL 35294
11. Phone No. *(with area code)* (205) 934-3789 12. Fax No. *(with area code)* (205) 934-1301 13. Email: irb@uab.edu	
14. Name of Official Ferdinand Urthaler, M.D.	15. Title Chairman, IRB
16. Signature Ferdinand Urthaler, MD	17. Date 5/11/16

Authorized for local Reproduction — Sponsored by HHS

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0990-0263. The time required to complete this information collection is estimated to average 30 minutes per response. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Health & Human Services, OS/OCIO/PRA, 200 Independence Ave., S.W., Suite 336-E, Washington D.C. 20201, Attention: PRA Reports Clearance Officer.