Many New Candidate Health- and Caries-Associated Bacterial Species Identified by 16S Pyrosequencing

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Erin Leigh Gross, D.D.S., Ph.D.

Graduate Program in Dentistry

The Ohio State University

2011

Thesis Committee:

Dr. Ann Griffen, Advisor

Dr. Eugene Leys

Dr. Ashok Kumar
Abstract

Dental caries is the most common chronic disease of childhood and disproportionately affects poor and minority children. It is understood that a combination of cariogenic bacteria, frequent sugar intake, and a susceptible host are required for caries development, and that the best predictor of caries is past caries experience. However, the complex microbial etiology of caries is not completely understood.

The purpose of this study was to compare bacterial species and communities associated with early childhood caries and health using next-generation sequencing techniques. Young children with established and rampant caries of the primary dentition were identified, along with a group of healthy controls, and plaque samples were collected according to a clinical model of caries progression. 16S pyrosequencing was used to identify the bacteria present in the samples.

This technique identified more than 300 species in the oral microbiome of childhood, with many being associated with caries or health status. Overall a greater number of species were found at low levels and associated with health.
Potential caries pathogens included *Streptococcus mutans*, *Veillonella atypica dispar parvula*, *Lactobacillus casei rhamnosus*, *Lactobacillus gasseri*, *Prevotella histicola*, *Streptococcus vestibularis salivarius*, *Olsenella profusa*, *Rothia dentocariosa*, *Lactobacillus fermentum*, *Rothia mucilaginosa*, and *Streptococcus parasanguinis oralis*.

Using 16S pyrosequencing, a better understanding was gained of the complexities of the oral microbiomes in health and disease. This new knowledge will allow investigations into possible screening, prevention, and treatment strategies for caries, decreasing the burden of this disease in children.
Dedicated to Ryan, Cameron, Mom, Daddy, Kari, and Devon for your constant support and unending love.
Acknowledgments

I would like to thank the members of my thesis committee, Dr. Ann Griffen, Dr. Gene Leys, and Dr. Ashok Kumar for their guidance and support on this project, as well as for their friendship throughout my years in the lab. Without them this thesis would not be possible.

I am grateful to Dr. Clifford Beall, Noah Firestone, Jennifer Harris, Danny Dayeh, and George Williams for their invaluable contributions to this work.
Vita

2000...................................................... Salem High School

2005...................................................... B.S. Biology, The Ohio State University

2009...................................................... D.D.S., The Ohio State University

2009...................................................... Ph.D. Oral Biology, The Ohio State University

2009 to present ...................................... Graduate Student and Resident in

Pediatric Dentistry, The Ohio State University and Nationwide Children’s Hospital

Publications


Fields of Study

Major Field: Dentistry
Table of Contents

Abstract ................................................................................................................................. ii
Dedication ............................................................................................................................ iv
Acknowledgments ............................................................................................................... v
Vita ......................................................................................................................................... vi
List of Figures ....................................................................................................................... ix
Introduction ......................................................................................................................... 1
Materials and Methods ........................................................................................................ 4
Results .................................................................................................................................. 11
Discussion ............................................................................................................................ 16
References ............................................................................................................................ 37
List of Figures

Figure 1. Mean levels of the most common phyla by sample type..................27
Figure 2. Mean levels of the most common genera by sample type. ..............28
Figure 3. Mean levels of the most common species by sample type. .............30
Figure 4. Decreasing species diversity as caries severity increased..............31
Figure 5. Species diversity by subject group. ............................................32
Figure 6. Plots of a non-metric multidimensional scaling ordination based on
Bray-Curtis Dissimilarity for all samples from caries and healthy subjects......34
Figure 7. Plots of a non-metric multidimensional scaling ordination based on
Bray-Curtis Dissimilarity for baseline enamel samples from caries and healthy
subjects. ......................................................................................................35
Figure 8. Comparison of measures of community similarity......................36
Introduction

Dental caries is the most common chronic disease of childhood and disproportionately affects poor and minority children (48). Any child under the age of six years with caries is considered to have early childhood caries, which may be severe (reviewed by (59)). For many children, early childhood caries is a source of pain and infection, and treatment requires costly restorative dentistry completed under general anesthesia. It is understood that a combination of cariogenic bacteria, frequent sugar intake, and a susceptible host are required for caries development, and that the best predictor of caries is past caries experience (reviewed by (59)).

Many studies have focused on the oral microbiome of early childhood caries and health, in an attempt to gain a better understanding of this complex disease. *Streptococcus mutans* has been implicated by many as an important caries pathogen (1, 5, 9, 16, 17, 31, 42, 45, 49). However, these studies used culture or molecular techniques, such as 16S cloning and sequencing, that were targeted or had limited sensitivity for detecting species at low levels. Additionally, few studied healthy communities.
Pyrosequencing is a next-generation sequencing technique that provides deeper coverage of communities than traditional approaches, and has recently been utilized to study different aspects of the oral microbiome, including different oral environment and different health and disease statuses (7, 8, 15, 18, 32, 35, 36, 39, 63-65). The overwhelming consensus of these studies is that pyrosequencing has revealed a greater diversity of taxa in the oral microbiome than previously found based on traditional techniques. Li et al (36) directly compared pyrosequencing to Sanger sequencing of endodontic infections and found a 600-fold increase in “depth of coverage” using pyrosequencing, and Ling et al (39) compared pyrosequencing to PCR-DGGE and observed greater diversity using pyrosequencing. And yet, several studies have revealed that still greater depth of coverage is needed for a full appreciation of the species richness of the oral microbiome (35, 39, 64). Few studies have used pyrosequencing to study caries in children.

The purpose of this study was to compare bacterial species and communities associated with early childhood caries and health using next-generation sequencing techniques. Young children with established and rampant caries of the primary dentition were identified, along with a group of healthy controls, and pyrosequencing of the 16S rRNA gene was used for identification of the oral microbiota associated with caries and health. This is the first study of its kind to utilize 16S pyrosequencing and a previously-described clinical model of
caries progression (5). This study corroborated previous reports regarding the importance of *Streptococcus mutans* as a caries pathogen, but importantly, it also identified many new candidates for disease-associated bacterial species, as well as new candidate health-associated species. Bacterial community profiles at different types of caries sites were different than those of health, with a decrease in species diversity as health-associated species were lost as caries became more severe.
Materials and Methods

The purpose of this open-ended study was to use next-generation sequencing techniques to compare bacterial species and communities associated with early childhood caries and health.

Clinical Methods

Subject Recruitment

Subjects were recruited at the Nationwide Children’s Hospital Dental Clinic and Dental Surgery Center in Columbus, Ohio for this institution-approved study. Inclusion criteria for the caries group were age less than six years and cavitated lesions extending into dentin in at least six primary teeth. A healthy control group with no caries and no existing restorations was also recruited. Exclusionary criteria for either group were history of antibiotic use in the previous 30 days or professional dental cleaning in the previous 30 days.

Sampling and Clinical Data Collection

Dental plaque was collected by wiping the surface of the tooth or lesion with a microbrush. Dental plaque was sampled from the intact enamel of healthy
subjects. The facial surfaces of four primary teeth were sampled, including at least one anterior tooth and one posterior tooth. In caries subjects, dental plaque was sampled separately from the surfaces of three different types of sites: (i) healthy, intact enamel, (ii) white spot lesions, and (iii) cavitated lesions. Caries samples were pooled from multiple teeth. Additionally, following removal of the surface biofilm with a microbrush, carious dentin from one pulpally-involved tooth was sampled using a spoon excavator. Samples were placed in 270 μl Buffer ATL and frozen in the laboratory until analysis.

For healthy subjects, demographic data collected included date of birth (month and year only), gender, race, and ethnicity. The same demographic data was collected for caries subjects, as well as the number of fully erupted primary teeth present and the number of teeth with cavitated lesions. The impression of the sampling clinician regarding caries activity was also recorded at the time of sampling. Caries activity was considered to be low if most lesions appeared to be arrested and there were few or no white spot lesions. The lesions were dark in color with a leathery or firm consistency, and were shallow on excavation. Caries activity was considered to be high if most lesions were very soft and deeply penetrating, and there were multiple white spots or incipient lesions. Caries activity was called medium if it were judged to be intermediate between low and high.
Following sampling, healthy subjects received a dental prophylaxis, examination, and fluoride application. Caries subjects received treatment under general anesthesia, including dental prophylaxis, examination, fluoride application, and restorative dentistry.

**Laboratory Methods**

**Sample Preparation and DNA Isolation**

Samples were thawed, and 30 μl Proteinase K was added to the 270 μl Buffer ATL plus microbrush. This was vortexed and incubated at 56°C for two hours, vortexing every 20 minutes. Following incubation, the tip of the microbrush was steriley cut and transferred to a 500 μl microcentrifuge tube with a hole in the bottom. This tube was placed inside a 1.5 ml microcentrifuge tube and centrifuged for one minute at 14,000 rpm. This step was repeated with a new 1.5 ml microcentrifuge tube, the tubes were combined, and transferred to a tube with 0.25 g of 0.1 mm glass beads. The tubes were processed for 60 seconds at 4,800 rpm in a Biospec Products bead beater. The bacterial DNA was purified using the QIAGEN QIAamp® DNA Mini Kit according to the manufacturer’s Tissue Protocol and two elutions were frozen until further analysis.

**PCR Amplification for Pyrosequencing**
Bacterial 16S rRNA genes were amplified by PCR using a reaction mixture containing 19.3 μl water, 2.5 μl 10X AccuPrime™ PCR Buffer II, Accuprime™ Taq DNA Polymerase High Fidelity, 2 μl of primer mixture, and 0.5 μl of sample DNA. The primer mixture contained 2 μM of each of two primers: a barcoded A primer (5’-CCATCTCATCCCTGCGTGTCTCCGACTCAG-X-ATTACCGCGGCTGCTGG, where X denotes the position of a 5- to 9-base pair sample-specific barcode sequence) and a non-barcoded B primer (5’-CCTATCCCCTGTGCTGCTTGGCAGTCTCAGAGAGTTTGATCCTGGCTCAG). These primers amplified the V1-V3 region of the 16S rRNA gene. 25 cycles of PCR were performed using the following conditions: 95°C for 2 minutes (first cycle only), 95°C for 20 seconds (repeating cycles), 56°C for 30 seconds, and 72°C for 5 minutes. Amplification was confirmed by gel electrophoresis on 1% agarose.

Pyrosequencing

Pyrosequencing was performed at the Plant-Microbe Genomics Facility (PMGF) at The Ohio State University in Columbus, Ohio. PCR products were cleaned using Agencourt AmPure Beads and quantified using the Quant-IT ds DNA high sensitivity assay. Samples were pooled so that equal concentrations of each sample were added to the final pool, and the pool was cleaned again using Agencourt AmPure Beads. Emulsion PCR was performed, followed by
enrichment, and pyrosequencing using the GS-FLX+ System on the Titanium platform. The results of the pyrosequencing run were subjected to the Amplicon Filtering Module, which eliminated sequences of insufficient length or quality. The final sequences and a quality file were then returned to our laboratory.

**Bioinformatics Methods**

Sequences were further filtered using the program mothur (Dr. Patrick Schloss, The University of Michigan), based on the quality file provided by PMGF. Mothur also eliminated sequences with more than one difference in the barcode sequence, more than four differences in the primer sequence, or strings of more than 10 of the same base. Mothur was also used to remove the barcode and primer sequences and identify the sample from which the sequence was isolated. Following quality filtering in mothur, only sequences with at least 400 base pairs were considered for further analysis.

Sequences were identified by comparing them to a curated, locally hosted oral microbiome database (25) using Blast (3). Clinical sequences were required to be ≥97.95% similar to species in the database for identification. UCHIME (24) was used to check for chimeric sequences among those with similarity scores less than 97.95%, and chimeric sequences were discarded. Remaining sequences were clustered into groups using UCLUST (23), and a representative sequence from clusters with more than 25 sequences was compared to
GenBank. If this sequence matched a GenBank entry, then the GenBank sequence was added to the oral microbiome database.

**Data Management and Statistical Analysis**

Levels of each species were calculated as a percent of total bacteria for each sample. Mean bacterial levels were determined for the most prevalent phyla, genera, and species using JMP (JMP, Version 9.0, SAS Institute Inc., Cary, NC). Repeated measures analysis was performed using PROC MIXED in SAS (SAS Institute Inc., SAS 9.2, Cary, NC) using the default various components structure. The type of site from which each sample was collected was expressed as a level of severity and used in the PROC MIXED analysis. Using this scale, healthy control samples were assigned a value of 1, intact enamel samples 2, white spot lesions 3, cavitated lesions 4, and dentin lesions 5. The PROC MIXED analysis used a linear model to calculate an estimate of the percent change for each species as caries progressed from severity 1 through 5, and determined the significance of the estimate at $\alpha = 0.05$.

Species abundance matrices were generated based on the BLAST classification of sequences. These matrices were then used to calculate the Bray-Curtis dissimilarity between each pair of samples. The calculation was performed with the `vegdist` function of the `vegan` package of the R programming language (51). Direct comparisons were done on the Bray-Curtis matrix using the
function `t.test` of the `stats` package of R (57), and an ordination was done by non-metric multidimensional scaling (NMDS) (function `metaMDS` of the `vegan` package). This procedure attempts to position the samples in two dimensions so that the graphical distance between any two points reflects the dissimilarity. The centroids of points for the five different severity levels were calculated, and a line connecting the centroids was drawn. ANOSIM, a permutation-based test to measure differences between sample groups, was performed using the PRIMER 6 program (PRIMER-E, Plymouth, UK). Shannon Diversity Index was calculated using function `diversity` in the `vegan` package, and was analyzed by PROC MIXED. Samples contained a varying number of sequences, with 681 being the smallest number for a sample that was included in the analysis. A random selection of 681 sequences was taken and the Shannon Diversity Index was calculated 10 times for each sample, and the average of these 10 indices was used for further analysis.

Exploratory analyses on caries activity were done by ANOVA, computed in JMP. *Post hoc* comparisons were made using *t*-tests and paired *t*-tests and were computed in JMP. The False Discovery Rate correction for multiple comparisons was determined using SAS.
Results

Thirty subjects with caries and 29 healthy controls were recruited for this study. The caries group was divided into 18 subjects with high caries activity, seven subjects with medium caries activity, and four subjects with low caries activity. One caries subject did not have a caries activity designation. There were no differences between the healthy control and caries group by age, gender, race, or ethnicity.

Of the sequences received from PMGF, 609,408 remained following quality filtering in mothur. Of these, 602,728 sequences were at least 400 base pairs long and were submitted for Blast. A total of 401,120 sequences were identified by the Blast with a match ≥97.95%. 201,608 sequences that did not have a match in the CORE oral microbiome database were clustered. A sequence from clusters with at least 25 sequences was compared to GenBank, and based on this curation, 36 taxa were added to the CORE. This allowed an additional 72,218 sequences to be identified. 129,570 sequences were discarded because they were not a part of a cluster of at least 25 sequences or they were identified as chimeras. An additional nine sequences were discarded because they were the only sequences identified in three samples (one sequence from
each of two samples, and seven sequences from one sample). So 473,149 sequences were considered in the statistical analysis. The minimum number of sequences for a single sample was 681, and the maximum was 7,787. The average number of sequences per sample was 3,263. The maximum length of a sequence was 659 base pairs, with an average length of 497 base pairs.

Overall 363 species were identified in this study, which could be assigned to 67 genera and 12 phyla. 14.1% of total clones were uncultivated. Mean bacterial levels by sample type are plotted in Figure 1 for the most abundant phyla (greater than 0.03% of clones), in Figure 2 for the most abundant genera (greater than 0.5% of clones), and in Figure 3 for the most abundant species (greater than 0.3% of clones). Figures 1, 2, and 3 also indicate phyla, genera, and species that significantly increased or decreased as caries progressed, according to their PROC MIXED estimates. In general for all three types of classification, more taxa decreased with caries progression than increased, and more were significantly associated with health than disease. Of the phyla that were submitted for PROC MIXED analysis, seven had significant negative PROC MIXED estimates (levels decreased as caries severity increased). Firmicutes was the only phylum to increase as caries progressed, and its PROC MIXED estimate was significant. Twelve of the 17 genera considered that had negative estimates were significant, and the genera Lactobacillus, Streptococcus, Veillonella, Rothia, and Olsenella significantly increased with caries progression. Negative estimates
were calculated for 45 species and 34 were significant, while 11 of 14 species with positive estimates were significant, including *Streptococcus mutans*, *Veillonella atypica dispar parvula*, *Lactobacillus casei rhamnosus*, *Lactobacillus gasseri*, *Prevotella histicola*, *Streptococcus vestibularis salivarius*, *Olsenella profusa*, *Rothia dentocariosa*, *Lactobacillus fermentum*, *Rothia mucilaginosa*, and *Streptococcus parasanguinis oralis*.

The relationship between caries severity and bacterial community diversity is shown in Figure 4. The PROC MIXED estimate was -0.37, meaning that overall, bacterial diversity decreased significantly as caries progressed from health to dentin lesions. This relationship was highly significant ($p = 1.02 \times 10^{-17}$). *Post hoc* comparisons indicated a significant difference between each sample type pair except two: healthy controls and intact enamel of subjects with caries, and white spot lesions and cavitated lesions. Figure 5 considers bacterial community diversity with regard to caries activity. Subjects with high caries activity had significantly less diverse bacterial communities than subjects with low caries activity at both cavitated and dentin sites.

The top panel of Figure 6 contains a plot of a non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis Dissimilarity for all samples in the study. The centroid point for each level of severity was calculated and is indicated by a number, and a generally rightward progression of severity in the ordination can be visualized. ANOSIM comparisons revealed significant
differences between healthy controls and white spots, healthy controls and cavitated lesions, and healthy controls and dentin lesions. The remaining plots of Figure 6 use the same NMDS ordination, but superimpose the abundance of specific species or genera onto the ordination, highlighting their contributions to their communities. The four health-associated species shown, *Corynebacterium matruchotii, Neisseria flava mucosa pharyngis*, *S. mitis* group, and *Fusobacterium nucleatum* subsp. *polymorphum* had the greatest magnitude of negative change, according to their PROC MIXED estimates. All four of these species were among the most abundant in the study. *Streptococcus mutans*, *Veillonella atypica dispar parvula, Streptococcus vestibularis salivarius* had some of the highest PROC MIXED estimates and were in the top ten most abundant species in the study. The genus *Lactobacillus* was considered as a whole because of the contributions of several *Lactobacillus* species to caries. Though lactobacilli were not a major player with regard to overall abundance (being only the eleventh most common genus in the study), they displayed the greatest increase in abundance as caries progressed.

Exploratory analyses of the contributions of caries activity to bacterial communities were performed. Figure 7 is a plot of a non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis Dissimilarity for all enamel samples in the study. ANOSIM analysis revealed that there were no significant differences among bacterial communities for healthy subjects or those with low,
medium, or high caries activity. Abundance of the ten most common species was compared for each type of caries site by ANOVA, revealing that subjects with low caries activity had higher levels of *Fusobacterium nucleatum* subsp. *polymorphum* at white spot and cavitated sites than subjects with medium or high caries activity. *Veillonella atypica dispar parvula* was higher in the cavitated lesions of subjects with high caries activity than subjects with low or medium caries activity, but this comparison did not remain significant following the false discovery rate correction.

Exploratory analyses using ANOVA also revealed the genera *Actinomyces* and *Campylobacter* to be significantly higher in white spot lesions than other types of caries lesions (as well as intact enamel for *Actinomyces*). When the levels of these genera in white spot lesions were compared in low, medium, and high caries activity subjects, no significant differences were found.
Discussion

This is the first time such a study has been conducted on early childhood caries, and while it confirms much of what is already known regarding caries microbiology, it has also revealed many new candidate caries- and health-associated bacterial species. Gaining a more complete understanding of the bacterial communities associated with early childhood caries and health will make development of screening mechanisms and therapies possible.

Pyrosequencing identified 363 bacterial species as members of the oral microbiome of early childhood caries and health. These species represented 67 genera and 12 phyla. Only the most abundant taxa were analyzed by PROC MIXED, and eight phyla, 17 genera, and 45 species were significant. This is many more taxa identified, as well as found to be significant, than a recent study from this laboratory using the same clinical study model, but with 16S cloning and sequencing (26). Admittedly that study was a different population, as it was permanent teeth in adolescents, but the results of the current study are more likely due to the deeper coverage of pyrosequencing and its greater sensitivity for identifying low-abundance species.
Uncultivated bacteria. Other studies have found at least 50% uncultivated species in the oral microbiome (2, 33, 34, 52, 53), but only 14.1% of the sequences generated for this study were attributed to uncultivated species. This finding corroborates similar evidence from a previous study in this lab, and can be attributed to differences in supragingival and subgingival communities, but more importantly to continuous improvements in sequence identification using local and public databases (26). This study utilized the CORE oral microbiome database (25), a locally-hosted database that is continuously curated to contain accurate sequence information that is representative of the oral microbiome.

Caries-associated species. The distribution of the most abundant caries-associated species is complex and is shown in Figure 6.

Streptococcus mutans. S. mutans was the second most abundant species present in the study and had a highly significant positive PROC MIXED estimate. S. mutans is highly acidogenic and aciduric, and is a long-established caries pathogen (reviewed in (54). It has been associated with caries by many previous clinical studies (1, 5, 16, 17, 31, 49), and is believed to be important in caries initiation (reviewed in (54)).

Streptococcus vestibularis salivarius. S. vestibularis salivarius is a pair of species that cannot be differentiated based on 16S sequence and were combined in the CORE database. Each species has been found in the oral cavity and has been shown to be cariogenic in laboratory studies (salivarius: (22, 27,
S. vestibularis was found to have intermediate cariogenic potential in one laboratory study (13), and produced low levels of caries when compared to S. salivarius in another laboratory study (62). While S. salivarius has been shown to produce caries in the laboratory (see references above), it also has caries-protective properties. S. salivarius produces urease, which hydrolyzes urea to ammonia. S. salivarius urease activity has been shown to increase in response to decreasing pH \textit{in vitro} (11), and recombinant bacteria with S. salivarius urease decreased acidification in the presence of glucose in another laboratory study (12). A recent \textit{in vivo} study did not find an increase in urease activity following a sucrose challenge, unless only subjects with low levels at baseline were considered (58). Studies using a rat model also showed a strain of S. salivarius that did not produce caries could inhibit colonization by S. mutans (55, 56), and in one study, this also inhibited caries (55). S. salivarius has been associated with caries in other clinical studies using DNA-based methods (1, 5), but it has also been associated with health (16). The role of S. salivarius in caries progression may be complex.

\textit{Veillonella atypica dispar parvula}. In the current study, this group of Veillonella species was detected in every sample, and its association with caries was significant by PROC MIXED analysis. Veillonella rely solely on lactate and other organic acids as an energy source, but the results of studies regarding a potential decrease in lactate production are mixed (44, 47). The association of
Veillonella species with caries in the current study seems to support laboratory findings that Veillonella contribute to greater acid production (47), as this could potentially increase caries experience. Veillonella species were also found to be significantly associated with caries in children in previous molecular studies (1, 5, 30, 31).

**Lactobacillus species.** Lactobacillus species were combined and considered as a whole genus for this analysis because of their functional similarity (46). Lactobacillus species have been associated with caries progression (reviewed in (60)), a finding which seems to be corroborated by these data. No Lactobacillus species were detected in healthy controls, and when they were detected at healthy sites in caries subjects, it was never at a level greater than 0.05%. Lactobacilli were found in few white spot lesions, and the highest level detected was only 0.16%. However, in cavitated and dentin lesions, lactobacilli were detected in most samples and levels were increased. In fact, in two dentin samples, levels were approximately 90% of total sequences. Lactobacilli are acidogenic and aciduric, and have been associated with caries in other clinical studies (10, 14, 26, 40).

**Acid-producing species dominate caries communities.** Figure 6 provides an interesting visualization of the contribution of these important acid-producers. Of importance is that not every caries site has high levels of *S. mutans*. Especially interesting is the comparison of the NMDS plots including *S.*
mutans and Lactobacillus abundance. It can be seen that there are distinct subsets of caries lesions that have relatively high levels of S. mutans, and they are not the same as those that are dominated by Lactobacillus. A subset of samples also contains increased levels of both S. mutans and Lactobacillus. This finding was also found in a similar study of permanent teeth in adolescents (26). There are also a limited number of samples that have increased levels of the cariogenic species S. vestibularis salivarius. The bacterial communities at these caries sites, dominated by a limited number of species, are significantly different than healthy control communities by ANOSIM analysis.

Other caries-associated species. Other species associated with caries in this study included O. profusa, which ferments glucose to lactic acid (19) and has been associated with carious dentin (14, 46). R. dentocariosa and R. mucilaginosa were also associated with caries. Rothia species are also lactic-acid producing and R. dentocariosa has been found in carious dentin (14, 46). Prevotella histicola was significantly associated with caries. Its major fermentation products are acetic and succinic acids, but it can also produce lactic acid (21). Other species of Prevotella have been associated with caries in clinical studies (14, 31, 53).

Decreasing species diversity. The current study supports previous investigations that associated caries with a reduction in overall species diversity (4, 26, 29, 37)(Fig 4). It also provides support for the ecologic plaque hypothesis.
which explains caries as the result of an ecologic disruption of the normal healthy community. When carbohydrates are available, acid-producing species lower pH, eliminating acid-sensitive species.

**Health-associated species.** Most species analyzed in this study decreased as caries progressed, with 34 having significant negative PROC MIXED estimates. Most were present at low levels in health, and were eliminated as acid-producing species increased in caries, contributing to the overall decrease in species diversity. Many have been associated with health in other clinical studies, while others are new candidates as protective factors. These species contribute to healthy communities in a variety of ways. For example, members of the *S. mitis* group, as well as *S. gordonii* and *S. cristatus* catabolize arginine to produce ammonia, which can raise pH (20, 38). Another example is *C. matruchotii*, which can utilize lactate (28), potentially raising pH.

**Caries activity.** It was expected that subjects with different levels of caries activity would have differences in bacterial community composition, because of their vastly different clinical presentations. There were no differences in overall community composition of enamel sites (Figure 7). White spot and cavitated lesions had significantly higher levels of *F. nucleatum* subsp. *polymorphum* in subjects with low caries activity, as compared to subjects with medium or high caries activity. *F. nucleatum* subsp. *polymorphum* was associated with health in the current study, and may play a protective role,
limiting the rate of caries progression. *Veillonella atypica dispar parvula* was significantly higher in the cavitated lesions of subjects with high caries activity than in those with medium caries activity. As discussed above, this group is caries-associated. Other work from this laboratory has identified *Veillonella atypica dispar parvula* as a risk factor for caries initiation, as it was increased in subjects who were healthy at baseline but subsequently developed caries (unpublished data). Diversity was lower in high subjects than low subjects at cavitated and dentin sites. This was not surprising since simpler caries communities are dominated by acid-producers (26), which would contribute to a high rate of caries progression.

**White spot lesions and caries initiation.** White spot lesions are the first stage of caries, the initial demineralization of enamel before the surface becomes cavitated. They can progress quickly to cavitated lesions, but can also become arrested. Understanding the microbiology at this surface, where a change may prevent the development of more severe disease, is important for developing prevention strategies. Elimination of initiating species may contribute to the arrest of lesions before the enamel surface is irreversibly broken, eliminating costly restorative treatment.

The genera *Actinomyces* and *Campylobacter* were significantly higher in white spot lesions than in cavitated and dentin lesions (*Actinomyces* were also significantly higher in intact enamel of subjects with caries), which may indicate
their importance in the initiation of caries. *Actinomyces* species are acidogenic and have been associated with the initiation of caries in previous laboratory and clinical studies (reviewed in (6), and were found at high levels in white spot lesions (1, 5). *Campylobacter* species utilize organic acids, like *Veillonella* species (discussed above). *C. showae* has been associated with health and *C. gracilis* has been associate with caries (1). Based on another study from this laboratory, it was suggested that *Campylobacter* may be a part of a functional consortium with acid-producing *Selenomonas* species (26). Further studies are required to understand the role of *Campylobacter* species in caries initiation.

It was expected that levels of bacteria associated with caries initiation would be different in subjects with differing levels of caries activity. However, levels of the genera *Actinomyces* and *Campylobacter* in white spots were not significantly different in subjects with low, medium, or high caries activity.

**Measures of differences among bacterial communities.** One of the aims of this study was to look at bacterial communities as a whole, comparing community composition among different types of sites, as well as different types of subjects. Comparisons presented here included healthy sites versus lesions of different levels of caries severity, and healthy subjects versus subjects with different levels of caries activity (Figures 6 and 7, respectively). One method for comparing bacterial communities is UniFrac and principle coordinate analysis (PCA), which is based on the bacterial sequences, using a measure of the
phylogenetic distance between communities (41). Another is the Bray-Curtis Dissimilarity and NMDS (presented in Figures 6 and 7), which is based on the different species or taxa present. Bray-Curtis Dissimilarity has been widely used in ecology to measure differences between communities and has a number of desirable properties, including giving sensible results for abundance matrices with many zeros and often showing good correlations with underlying environmental factors (50). Phylogenetic methods such as UniFrac may be more informative than species-based methods because they consider evolutionary divergence, rather than the number of shared taxa (41).

We considered both phylogenetic- and species-based methods to analyze the study data, and a PCA based on UniFrac and an NMDS based on Bray-Curtis Dissimilarity are shown in Figure 8. It was determined that the NMDS of Bray-Curtis Dissimilarity better represented the separation of communities of different types of sites, so it was used for further analyses of the data. Though phylogenetic methods for comparisons are often considered to be more informative than species-based methods, it seems that this does not hold true for this dataset. It may be explained by functional differences between taxa that are not considered by phylogenetic analysis. The streptococci are an important example of this. Using UniFrac, communities with many streptococci may appear to be more similar to each other than with Bray-Curtis Dissimilarity, because they are phylogenetically similar. However, we know that streptococci behave very
differently in the oral cavity. A community dominated by a caries-associated streptococci, such as *S. mutans* or *S. vestibular salivarius*, is functionally very different than a community dominated by health-associated streptococci, such as the *S. mitis* group. Species-based methods of community comparison (Bray-Curtis Dissimilarity) would simply consider these to be different taxa, and they may appear to be more different than with the UniFrac measures. While using evolutionary distances may be an advantage in many scenarios, here it is a disadvantage because it does not consider the importance of the function of oral species.

**Limitations and Future Directions.** While advances in 16S pyrosequencing have made it possible to deeply mine bacterial communities, there are still limitations to the technique. In the current study the average sequence length obtained using this technique was fewer than 500 base pairs. This may not be enough sequence information to distinguish between closely-related species, especially oral species such as streptococci. As this technology continues to evolve, however, longer sequence reads will be possible.

It was difficult to explore similarities and differences among caries communities from subjects with different levels of caries. When this study was designed, it was planned to recruit equal numbers of healthy control and caries subjects. However, by subdividing the caries subjects by caries activity, sample sizes became much smaller, making statistical analysis difficult. Not only were
samples sizes small, but they were also different for each group. In future studies, an increased and equal number of subjects should be recruited for each level of caries activity. It should also be noted that the criteria used to define these subgroups were subjective, though samples were collected by a limited number of clinicians who discussed the criteria at length. Interobserver reliability would have to be considered in future studies.

Using this model and 16S pyrosequencing in future clinical studies will develop a more complete understanding of caries communities, and the interplay of species as they increase and decrease in response to environment and community factors. An index may be able to be developed that considers ratios of health- and caries-associated bacteria in determining risk for caries initiation or progression. Applying such an index to clinical samples may identify subjects at need for preventive therapy or other therapeutic intervention.
**Corynebacterium matruchotii**

**Neisseria flava mucosa pharyngis**

**Fusobacterium nucleatum subsp.**

**Leptotrichia wadei**

**Leptotrichia shahii**

**Streptococcus sanguinis**

**Leptotrichia AF189244**

**Prevotella loescheii**

**Porphyromonas CW034**

**Leptotrichia 501D10**

**Actinomyces viscosus naeslundii**

**Porphyromonas**

**Actinobacteria**

**Proteobacteria**

**Bacteroidetes**

**Spirochaetes**

**Fusobacteria**

**OP11**

**TM7**

**Leptotrichia oral taxon 225**

**Actinomyces massiliensis**

**Leptotrichia hofstadii**

**Neisseria meningitidis**

**Lautropia mirabilis**

**Kingella denitrificans**

The discovery rate correction was applied.

Figure 1. Mean levels of the most common phyla by sample type. Phyla found at greater than 0.03% of total sequences are shown. The taxa are sorted by magnitude of PROC MIXED estimates, with the solid black line dividing taxa with negative estimates above and positive estimates below. * indicates taxa with $p<0.05$ and ** indicates taxa with $p<0.01$ after the false discovery rate correction was applied.
Figure 2. Mean levels of the most common genera by sample type. Genera found at greater than 0.5% of total sequences are shown. The taxa are sorted by magnitude of PROC MIXED estimates, with the solid black line dividing taxa with negative estimates above and positive estimates below. * indicates taxa with $p<0.05$ and ** indicates taxa with $p<0.01$ after the false discovery rate correction was applied.
Figure 3. Mean levels of the most common species by sample type. Species found at greater than 0.3% of total sequences are shown. The taxa are sorted by magnitude of PROC MIXED estimates, with the solid black line dividing taxa with negative estimates above and positive estimates below. * indicates taxa with $p<0.05$ and ** indicates taxa with $p<0.01$ after the false discovery rate correction was applied.
**Corynebacterium matruchotii**
**Neisseria flava mucosa pharyngis**
**Streptococcus mitis pneumoniae infantis oralis**
**Fusobacterium nucleatum subsp. polymorphum**
**Leptotrichia wadei**
**Leptotrichia shahii**
**Streptococcus sanguinis**
**Leptotrichia AF189244**
**Prevotella loeschei**
**Capnocytophaga spuriigena**
**Capnocytophaga granulosa**
**Streptococcus cristatus**
**Streptococcus gordoni**
**Capnocytophaga leadbetteri**
*Porphyromonas CW034*
**TM7 oral taxon 348**
**Leptotrichia 501D10**
**Leptotrichia IK040**
**Porphyromonas catoniae**
**Kingella oralis**
**Leptotrichia BU064**
**Actinomyces IO077**
**Actinomyces gerencseriae**
**Propionibacterium propionicus**
**Corynebacterium durum**
*Streptococcus genomosp. C8*
**Porphyromonas AP085**
**Granulicatella adiacens**
**Fusobacterium nucleatum subsp. animalis**
Campylobacter gracilis
*Prevotella melaninogenica*
*Corynebacterium durum*
**Fusobacterium nucleatum subsp. polymorphum**
**TM7 oral taxon 348**
**Leptotrichia 501D10**
**Leptotrichia IK040**
**Porphyromonas catoniae**
**Kingella oralis**
**Leptotrichia BU064**
**Actinomyces IO077**
**Actinomyces johnsonii**
**Abiotrophia defectiva**
**Actinomyces massiliensis**
Selenomonas noxia
**Leptotrichia hofstadi**
*Streptococcus genomosp. C8*
**Neisseria meningitidis**
**Streptococcus mutans**
**Neisseria meningitidis**
**Lactobacillus gasseri**
*Streptococcus vestibularis salivarius**
**Prevotella histicola**
**Actinomyces IP073**
**Lactobacillus gasseri**
**Lactobacillus casei rhamnosus**
**Veillonella atypica dispar parvula**
**Streptococcus mutans**

Figure 3
Figure 4. Decreasing species diversity as caries severity increased. Mean Shannon Diversity Indices at the level of species with 95% confidence intervals are shown for each level of caries severity. The decrease in severity was modeled using PROC MIXED analysis, and is shown as a dashed arrow (estimate = -0.37, $p = 1.02 \times 10^{-17}$). All post hoc comparisons were significant, except for the comparison between healthy controls and intact enamel and the comparison between white spot lesions and cavitated lesions.
Figure 5. Species diversity by subject group. Mean Shannon Diversity Indices at the level of species with 95% confidence intervals are shown for each group of caries subjects by lesion type. ANOVA indicated differences among cavitated and dentin samples, and significant post hoc comparisons are indicated by a red line.
Figure 6. Plots of a non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis Dissimilarity for all samples from caries and healthy subjects. The centroid for each type of sample is indicated by a number. All panels are based on the MDS plot in the top panel, with the points in lower panels being sized by abundance for the common health- and caries-associated species ($p$-values are for their PROC MIXED estimates). Samples without detectable levels are shown as empty points.
Caries-Associated Species

- S. mutans  
  \( p < 0.0001 \)

- V. atypica dispar parvula  
  \( p < 0.0001 \)

Health-Associated Species

- C. matruchotii  
  \( p < 0.0001 \)

- N. flava mucosa pharyngis  
  \( p = 0.0001 \)

- S. mitis group  
  \( p = 0.0002 \)

- F. nucleatum subsp. polymorphum  
  \( p < 0.0001 \)

Figure 6
Figure 7. Plots of a non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis Dissimilarity for baseline enamel samples from caries and healthy subjects. Subjects with caries are divided by whether they had low, medium, or high caries activity. Overall ANOSIM was not significant.
Figure 8. Comparison of measures of community similarity. The left panel is a plot of principal coordinates analysis of the UniFrac measurement. The right panel is a plot of the NMDS or Bray-Curtis Dissimilarity distances.
REFERENCES


