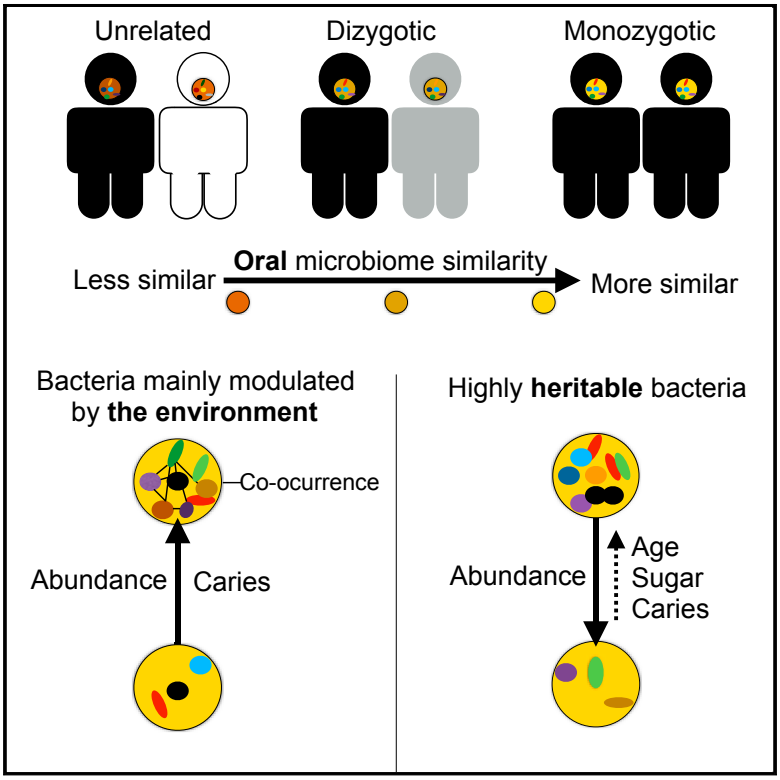
Article

Cell Host & Microbe

Host Genetic Control of the Oral Microbiome in Health and Disease

Graphical Abstract



Authors

Andres Gomez, Josh L. Espinoza, Derek M. Harkins, ..., Marcus B. Jones, Chris L. Dupont, Karen E. Nelson

Correspondence

kenelson@jcvi.org

In Brief

Gomez et al. examine the supragingival plaque microbiome of 5- to 11-year-old twins and find that the early oral microbiome is shaped by both heritable and environmental factors. However, the most heritable bacteria diminish in abundance with age, and potentially cariogenic taxa are not controlled by host genetics.

Highlights

- Heritable components of the oral microbiome identified in a child twin cohort
- Heritable bacteria are not associated with dental caries
- Environmentally derived bacteria tend to co-occur and are potentially cariogenic
- Heritable oral bacteria may be lost over time

Gomez et al., 2017, Cell Host & Microbe 22, 269–278
September 13, 2017
http://dx.doi.org/10.1016/j.chom.2017.08.013

Cell Host & Microbe
Article

Host Genetic Control of the Oral Microbiome in Health and Disease

Andres Gomez,[1,8] Josh L. Espinoza,[2,8] Derek M. Harkins,[3] Pamela Leong,[4] Richard Saffery,[4] Michelle Bockmann,[5] Manolito Torralba,[1] Claire Kuelbs,[1] Rohith Kodukula,[6] Jason Inman,[3] Toby Hughes,[5] Jeffrey M. Craig,[4] Sarah K. Highlander,[1] Marcus B. Jones,[7] Chris L. Dupont,[2] and Karen E. Nelson[1,3,9,*]

[1]Departments of Human Biology and Genomic Medicine, J. Craig Venter Institute, La Jolla, CA 92037, USA
[2]Department of Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA
[3]Departments of Human Biology and Genomic Medicine, J. Craig Venter Institute, Rockville, MD 20850, USA
[4]Murdoch Children's Research Institute and Department of Pediatrics, University of Melbourne, Royal Children's Hospital, Parkville, VIC 3052, Australia
[5]School of Dentistry, The University of Adelaide, Adelaide, SA 5005, Australia
[6]JCVI Summer Intern Program, Torrey Pines High School, San Diego, CA 9213, USA
[7]Human Longevity Institute, La Jolla, CA 92037, USA
[8]These authors contributed equally
[9]Lead Contact
*Correspondence: kenelson@jcvi.org
http://dx.doi.org/10.1016/j.chom.2017.08.013

SUMMARY

Host-associated microbial communities are influenced by both host genetics and environmental factors. However, factors controlling the human oral microbiome and their impact on disease remain to be investigated. To determine the combined and relative effects of host genotype and environment on oral microbiome composition and caries phenotypes, we profiled the supragingival plaque microbiome of 485 dizygotic and monozygotic twins aged 5–11. Oral microbiome similarity always increased with shared host genotype, regardless of caries state. Additionally, although most of the variation in the oral microbiome was determined by environmental factors, highly heritable oral taxa were identified. The most heritable oral bacteria were not associated with caries state, did not tend to co-occur with other taxa, and decreased in abundance with age and sugar consumption frequency. Thus, while the human oral microbiome composition is influenced by host genetic background, potentially cariogenic taxa are likely not controlled by genetic factors.

INTRODUCTION

Although there has been a tremendous expansion in human microbiome research, with hundreds of projects underway globally (Blaser, 2014; Human Microbiome Project Consortium, 2012), the oral microbiome has not received the same level of attention as its gut counterpart. Indeed, this microbial ecosystem is a critical component of oral and systemic human health. For instance, although dental caries, the most common chronic disease in children (Benjamin, 2010), is of a multifactorial nature, it usually occurs when frequent sugar intake is metabolized by a specific bacterial milieu in the oral cavity, resulting in increased acidity and dental demineralization (Takahashi and Nyvad, 2011). In periodontitis, a chronic disease affecting adults, specific bacterial ecology elicits inflammatory responses in the host, leading to the destruction of periodontal tissue, pocket formation, and tooth loss (Loesche, 2011). Likewise, non-plaque-associated bacteria, viruses, and fungi can trigger gingival lesions associated with herpes and candidiasis (Holmstrup, 1999), and there is mounting evidence pointing to a specific microecosystem characterizing cancerous tissue in oral cancer (Schmidt et al., 2014). Interestingly, the connections between oral microbes and health extend beyond the oral cavity, as cardiometabolic, respiratory, and immunological disorders; gastrointestinal cancers; and obstetric complications are thought to have oral microbial associations (Beck et al., 2000; Rubinstein et al., 2013; Seymour et al., 2007).

Consequently, unraveling the forces that shape and define the oral microbiome is crucial for the understanding of both oral and broader systemic health. Research on development and maturation of the human microbiome in the early years of postnatal life has mainly been centered on the gut, pointing at mode of delivery and breastfeeding as important early driving forces (Azad et al., 2013; Dominguez-Bello et al., 2010) and diet and environment as subsequent determinants (Walter and Ley, 2011). Moreover, twin studies have shown that the gut microbiome similarity increases with host genetic background, that some gut taxa are driven by additive genetic effects, and that the abundance of specific gut taxa is linked to genes associated with immune and metabolic functions in the host (Goodrich et al., 2014, 2016). Nonetheless, the available evidence on the forces shaping the oral microbiome is scarce. For example, just as in the gut, oral microbial communities seem to be initially influenced by perinatal factors (Holgerson et al., 2013; Lif Holgerson et al., 2011, 2015). However, reports on the heritable fraction of the oral microbiome are conflicting. For instance, contrary to what has been found in the gut, twin studies on the genetic control of the oral microbiome (saliva and plaque) have shown less or no apparent influence of additive genetic factors (Papapostolou et al., 2011; Stahringer et al., 2012). Yet other twin studies have focused on the abundance

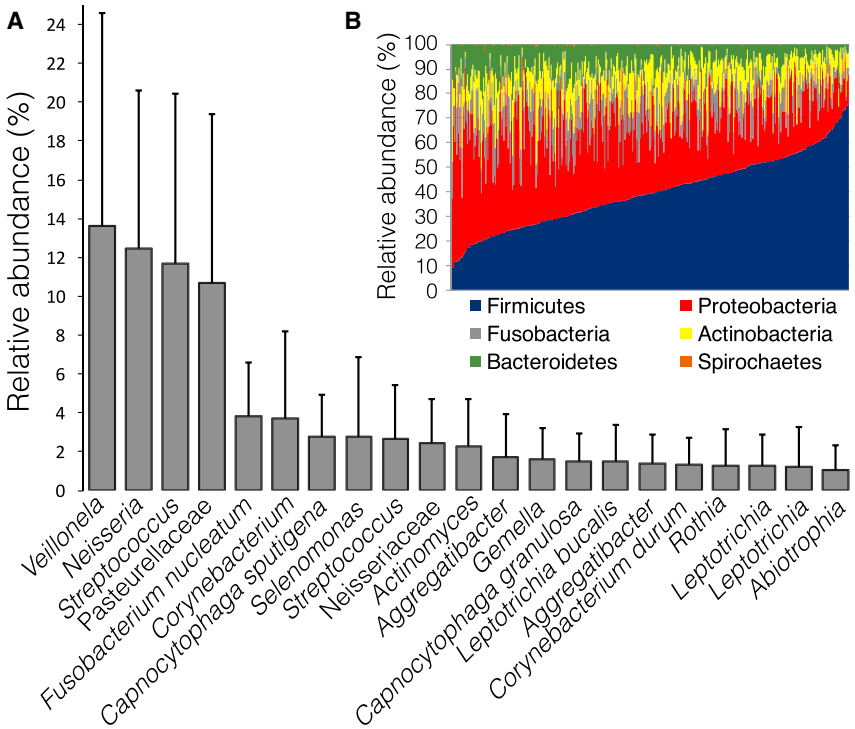


Figure 1. Taxonomic Distribution of the Oral Microbiome of Twin Children

(A and B) The barplot (A) and inset-stacked bar plot (B) show the relative abundances of main OTUs (> 1% in abundance) and phyla, respectively. OTUs in the barplot are sorted in order of mean abundance and standard error. The x axis in the inset-stacked plot represents the 485 twins sampled. Error bars = SEM.

of cariogenic taxa in saliva and plaque, showing that these are highly heritable traits (Corby et al., 2007, 2005).

In this study, we have investigated a large cohort of twin children to shed light on the contributions of host genotype and the early shared environment in shaping the oral microbiome in the context of oral health. As such, we characterized the microbiome in supragingival plaque (biofilm) swabs of monozygotic (MZ) and dizygotic (DZ) twin pairs, including children both with and without dental caries. Specifically, we first sought to establish if caries presence correlates with distinct plaque microbial communities in MZ and DZ twins discordant for dental caries. The presence of dental caries was assessed using two categories based on disease progression: caries confined to the outer enamel layer (enamel) and caries involving both the enamel and the deeper dentinal layer (caries progressing to dentin). Subsequently, based on heritability calculations, we explored the extent to which oral bacterial taxa potentially associated with oral health and disease in this cohort are genetically or environmentally driven.

RESULTS

The microbiome profiles in plaque swabs from 485 DZ (n = 280) and MZ (n = 205) twins between 5 and 11 years old (mean age = 7.8 ± 1.43 years) revealed an oral bacterial community composed of 297 operational taxonomic units (OTUs), clustered at 97% 16S rRNA gene sequence similarity (average number of sequence reads per subject: 22,000 ± 10,800). OTUs across all samples were mainly assigned to even proportions of three main phyla—Firmicutes (39.9% ± 14.5%), Proteobacteria (32.4% ± 14%), and Fusobacteria (9.2% ± 5.7%)—and followed in abundance by Actinobacteria (9.8% ± 6.8%) and Bacteroidetes (8.4% ± 4.8%). Spirochaetes and other minor phyla comprised less than 0.1% of the 16S rRNA gene sequence dataset. The most dominant OTUs were unclassified species affiliated with *Veillonella*, *Neisseria*, *Streptococcus*, and Pasteurellaceae (from ~13% to 10% in abundance, respectively, ~50% of all sequence reads), with all other OTUs not superseding more than 4% in relative abundance (Figure 1).

The Oral Microbiome of Twins Discordant for Oral Caries

An ordination analysis (principal coordinates plot) based on Bray-Curtis distance between twin pairs did not show bacterial community differences in MZ and DZ twin pairs discordant for dental caries in enamel or progressing to dentin (Figures 2A–2D). That is, the overall bacterial community composition of plaque swabs did not predict dental caries in this cohort when controlling for host genotype. These visual patterns were confirmed by permutational multivariate analyses of variance (PERMANOVA) ($p > 0.9$, $R^2 < 0.01$) and random forest analyses (out-of-bag estimate error rate = 60%–70%). However, there was a tendency for the healthy twins to cluster, somewhat consistently, away from diseased twin pairs along PCoA.2, in the MZ ($t = -2.65$, $p = 0.018$) and DZ twins ($t = -2.10$, $p = 0.047$), and only at the dentin level (Figure S1). We did not see the same trend at the enamel level, or along PCoA.1 ($p > 0.1$). This observation indicates that caries that have progressed through dentin may have a compositional effect on the plaque microbiome that is not standard across twins; that is, the taxa involved are not the same for every twin pair. No microbiome composition differences were observed between MZ and DZ twins when considering individuals with past or present evidence of caries, as seen in pairs discordant for caries treatment, in both dentin and enamel (Figure S2). Lastly, no differences were observed in terms of alpha diversity between diseased and healthy twins, regardless of zygosity or tooth surface (Figure S3). Information on caries diagnosis on each individual twin can be seen in Table S1.

Host Genotype as a Driver of Oral Microbiome Composition

Upon closer inspection of the ordination patterns obtained (Figure 2), it was observed that the majority of twin pairs tended to cluster together regardless of caries state. We thus tested the extent to which host genotype shaped oral microbiome

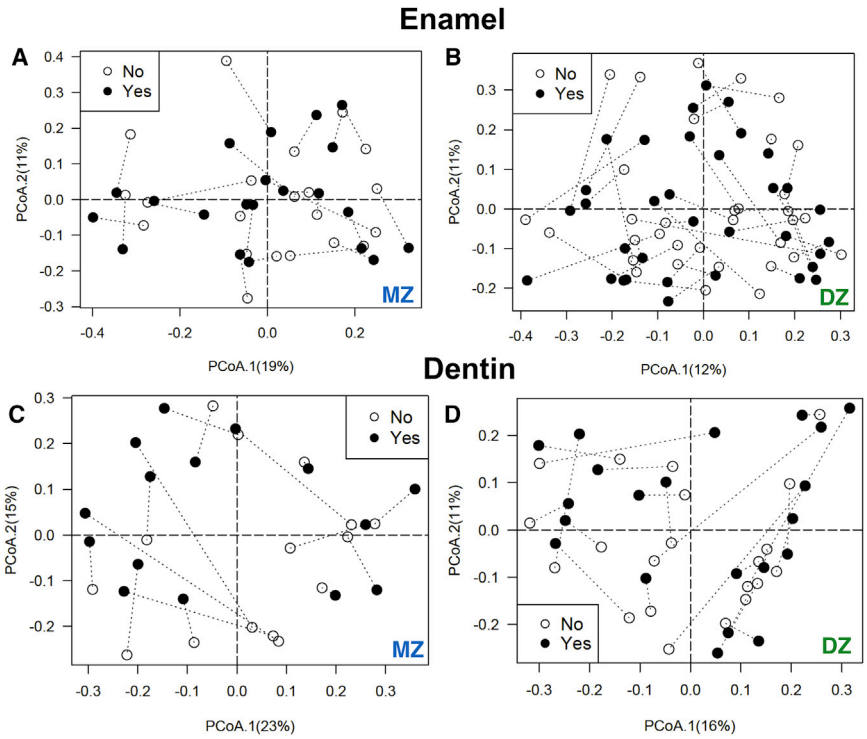


Figure 2. Oral Microbiome Composition of Twins Discordant for Oral Caries

(A–D) Principal coordinates ordination plots show that caries phenotype in dentin (A and B) and enamel (C and D) does not drive plaque microbiome composition in MZ (A and C) and dizygotic (B and D) twins, as demonstrated by PERMANOVA ($p > 0.9$, $R^2 < 0.01$). No, no caries presence; Yes, caries presence.

composition in this twin cohort, independent of caries state. Bray-Curtis distances (expressed as distance between nodes on dendrogram) within the MZ twin group were indeed lower than those observed in the DZ twin set ($p < 0.001$) and in a group of randomly chosen unrelated individuals in both tooth layers (Figures 3A and 3D). Furthermore, a circular dendrogram built on these distances confirmed that MZ twin pairs tended to cluster together in the same nodes (≤ 3 nodes apart) more frequently than DZ twins (Figures 3B, 3C, 3E, and 3F). Randomly selected unrelated individuals did not share nodes in the dendrogram.

Host Genetic and Environmental Drivers of Plaque Microbiome Composition

Heritability (H^2) of a phenotypic trait among individuals in a given population is defined as the amount of the variation in that trait due to genetic variation. In order to determine the amount of variation due to genetic and environmental effects driving specific taxa in the oral microbiome of this twin cohort, two approaches were followed: (1) calculating intraclass correlation coefficients (ICCs) on the abundance of each taxon within MZ and DZ twin pairs, and (2) assessing the additive genetic and environmental factors driving OTU abundance as determined by the ACE model (Eaves et al., 1978), controlling for sex and age. The ACE model, applied here to the full cohort regardless of caries states, assumes that the variability of a given OTU is explained by additive (A) genetic effects, the shared/common (C) environment, and non-shared/unique environmental (E) factors. To these ends, the OTU count data were first normalized (see STAR Methods) and filtered so that all OTUs were present in at least 50% of all individuals. This process yielded a set of 91 OTUs used for all variance partition analyses. Heritability was calculated on the full cohort, independent of oral health state.

Mean ICCs were significantly higher in MZ than in DZ twin pairs and in any kind of twin compared to randomly selected subjects (Kruskal-Wallis test, $p < 0.001$) (Figure 4A). Thus, OTUs in the dental plaque of MZ twin pairs covaried more often than observed in DZ pairs or unrelated individuals. Subsequently, we applied the ACE model to the 91 OTU set, controlling for sex and age across all subjects, revealing that the greatest proportion of the variation in the plaque microbiome of this twin cohort was explained by non-shared (E = 51.5% ± 11.7%) followed by shared factors (C = 29.1% ± 16%). Mean variance due to additive genetic factors (A) in this OTU set was 19.3% ± 17.8%, with 12 OTUs showing the highest additive genetic control (A > 40%, FDR-q < 0.001) and lower variance due to shared environmental or non-shared unique factors (C and E, respectively) (Table S2). Among the highly heritable OTUs, *Prevotella pallens* exhibited the greatest **A** value (A = 65.4% [q < 1e−10], C = 0%, E = 34.5%), followed by *Veillonella* oral taxon (A = 59.6% [q < 1e−10], C = 0%, E = 40%), Pasteurellaceae (A = 58.1% [q < 1e−10], C = 0%, E = 41.8%), and *Corynebacterium durum* (A = 54.3% [q < 4.1e−5], C = 1%, E = 49.3%). Values for other potentially heritable taxa along with their statistical significance, including *Leptotrichia* and *Abiotrophia*, can be seen in Table S2 and Figure 4B.

All OTUs affiliated with the genus *Streptococcus* showed the highest proportion of their variability explained by the twins' shared and unique environment (mean A = 9.5% ± 9.8%, C = 42% ± 18.2%, E = 49% ± 16.3%). For instance, an unclassified *Streptococcus* (likely related to *Streptococcus mitis*, according to a phylogenetic tree constructed with the BLAST tree view) (A = 5%, C = 62% [q < 1e−10], E = 24%) and *Streptococcus salivarius* (A = 0%, C = 62% [q < 1e−10], E = 37%) were the taxa with the highest proportion of their variability explained by shared environment, along with other unclassified streptococci. *Streptococcus mutans*, a known cariogenic taxon, and *Streptococcus anginosus* were shown to be mainly modulated by unique, non-shared environmental forces (Table S2 and Figure 4B). These observations suggest that the dominance of streptococci in the human oral cavity is not primarily driven by genetic factors and that environmental forces may be their main determinants. Other OTUs with a higher proportion of their variation explained by the twins' shared environment (highest C value) include unclassified *Selenomonas*, *Capnocytophaga*, Actinomycetales, and *Neisseria oralis* (Table S2, Figure 4B).

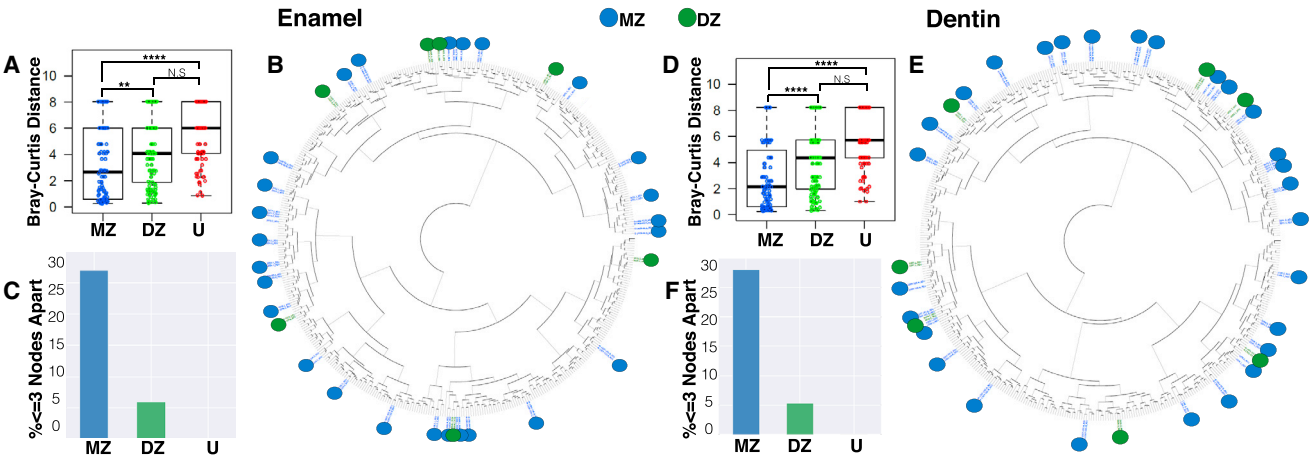


Figure 3. Influence of Host Genotype in Oral Microbiome Composition

(A–F) Both in enamel (A–C) and dentin (D–F), the oral microbiome of MZ twins was more similar than that of DZ twins or unrelated individuals, based on Bray-Curtis distances (expressed as distance between nodes on dendrogram) (A and D). Moreover, MZ twin pairs shared the terminal nodes of a Bray-Curtis distance dendrogram (B and E) more often compared to DZ or unrelated individuals (C and F).

To strengthen our heritability calculations, we also used Falconer's (F) formula (Bodmer, 1961), which determines the heritability of a given trait based on the differences between ICCs in MZ and DZ twins ($F = 2 \times [ICCmz - ICCdz]$). The average F value for this dataset was 0.25 ± 0.23. Providing complementary data to our previous heritability results, Falconer's values correlated strongly with the proportion of the variation explained by additive genetic factors in this dataset (Pearson's rho = 0.8, p = 2.2e−16) (Figure 4C). *P. pallens* and *Veillonella*-oral taxon not only exhibited the highest heritability and Falconer's values but also covaried more strongly among MZ compared with DZ twins (Pearson's rho > 0.5, p = 2.2e−16) (Figure 5). We also observed significant positive correlations between additive genetic effects (A values) and ICCs in MZ twins (Figure S4A). This was in contrast to insignificant or negative correlations between heritability A and ICC values in DZ and unrelated individuals, respectively (Figures S3B and S3C). Likewise, the data showed significantly negative correlations between the proportion of the variation determined by additive genetic effects and that explained by shared environment (Figure S4D).

We also compared the broad sense heritability values obtained herein with those presented by Goodrich et al. for gut microbes (Goodrich et al., 2014). This comparison indicated that, in contrast with the gut microbiome, oral bacteria may be more influenced by additive genetic effects (oral microbiome, A = 19%, Falconer's = 0.25; gut microbiome, A = 12%, Falconer's = 0.14; p < 2e−5); while the gut microbiome exhibits higher influence of shared environmental factors (gut microbiome, C = 29%; oral microbiome, C = 8%; p = 4.1e−51). Likewise, the gut microbiome shows a significantly higher influence of unique non-shared forces (E = 78% versus 51% in the gut and oral cavity, respectively; p = 8.5e−56).

However, these observations may be explained by the fact that the twin cohort in Goodrich et al. is significantly older (23–86 years old) and suggest that the strong interpersonal variation (personalization) of the human microbiome (Schloissnig et al., 2013) increases with age. To test hypotheses about oral microbiome heritability at different ages, we split the cohort into three different age groups (5–7 [n = 181], 7–9 [n = 179], and 9–11 [n = 125]). The results show that mean heritability (A) estimates across the core OTU set do not change with the age ranges examined in this study (p > 0.1, Figure S5A). Nonetheless, heritability estimates for specific OTUs vary significantly among the three different age groups. For instance, heritability estimates for *P. pallens*, the most heritable OTU in the dataset, are high among 5- to 7-year-olds (A = 0.77, p = 2.2e−16), diminish among 7- to 9-year-olds (A = 0.47, p = 1e−10), and disappear among 9- to 11-year-olds (A = 0, p = 1) (Figure S5B). Thus, there was high variability in heritability (A) estimates of specific OTUs across the three age groups, with just one OTU showing relative stability (unclassified *Leptotrichia*) (Figure S5C). Variation in C and E estimates for particular OTUs was also observed; nonetheless, *Streptococcus* spp. seemed to maintain high and constant C estimates across the three age groups (Figure S5C). Additionally, the data show that while heritability estimates remained unchanged among the age groups, C and E estimates were significantly modified; for example, the influence of shared environment (C) decreased from 6 to 11 years old at the expense of individual/unique environmental variation (E) (Figure S5A).

A critical question regarding genetic and environmental determinants of the human microbiome centers on exploring how each of these fractions is associated with health and disease and other phenotypic or lifestyle traits. Therefore, we used heritability information obtained on the core OTU set and explored how oral microbes influenced by additive genetic effects, shared environment, or unique non-shared factors correlated with caries diagnosis, oral health practices, diet, and age in the full cohort. Information on oral health practices, diet, and age of each individual twin can be seen in Table S1.

First, a Wilcoxon rank-sum test was used to determine if OTUs driven by genetic, environmental, and non-shared/unique forces were also associated with caries-positive and caries-negative states in each tooth layer (enamel and dentin) (p < 0.05). These tests were performed in the full cohort, randomly selecting

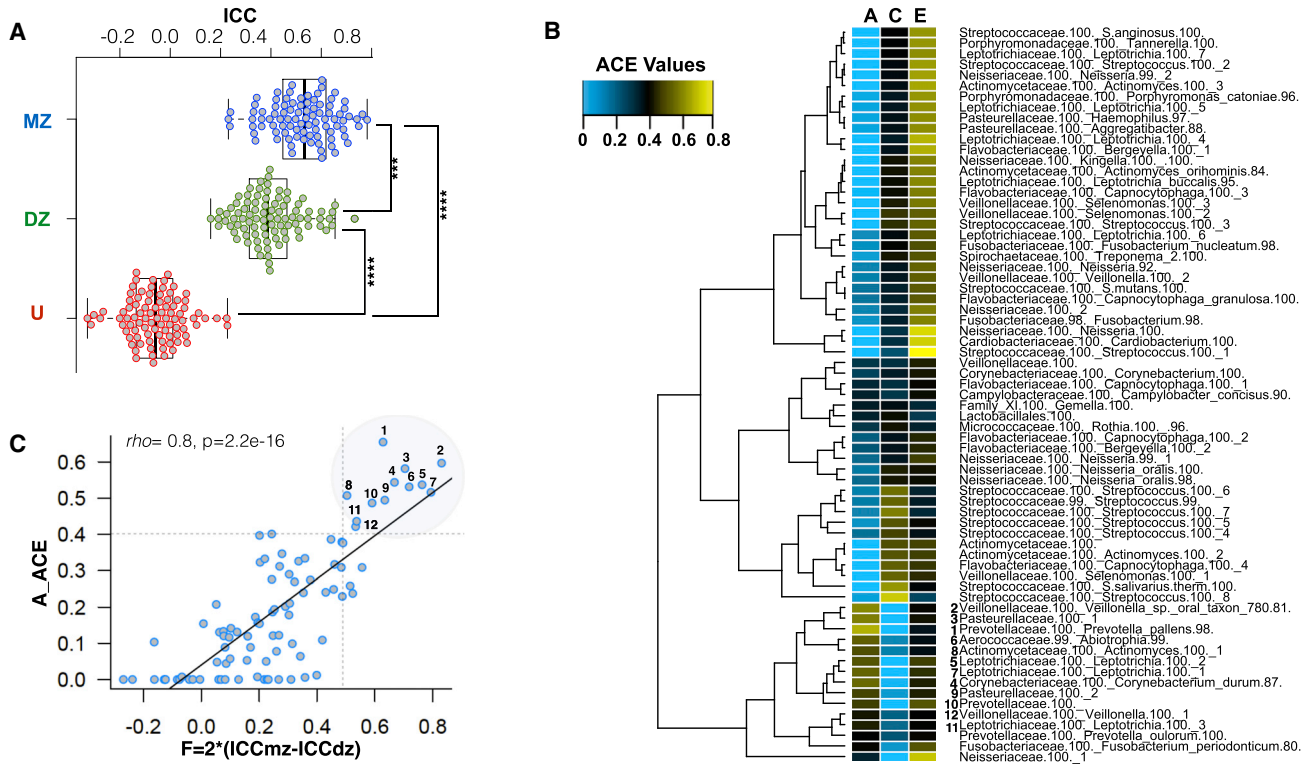


Figure 4. Heritability of Oral Microbes According to ICC and the ACE Model

(A) Intra-class correlation coefficients (ICC) are higher in MZ twins than in DZ twins or unrelated individuals. Plotted are swarm boxplots representing interquartile ranges (IQRs) of ICC medians (dark lines in the boxes), the lowest and highest values within 1.5 times IQR from the first and third quartiles (whiskers above and below the boxes), and outliers (warm symbols beyond the whiskers). Asterisks show statistical significance based on Kruskal-Wallis tests, ***$p < 0.001$, ****$p < 0.0001$.

(B) Hierarchically clustered heatmap (Euclidean distances, complete linkage algorithm) of ACE values showing the taxa that exhibit the highest heritability values (A) (numbered 1–12).

(C) Heritability measures (A) are strongly correlated with values obtained from Falconer's formula ($2*(ICC_{mz}-ICC_{dz})$), according to Pearson's rho = 0.8, $p < 0.0001$. The shaded area on the scatterplot and numbers correspond to taxa in the heatmap with high heritability (A) and high Falconer's values.

All ACE values can be seen in Table S2.

individuals from each pair to assess OTU prevalence in caries-negative and caries-positive groups. The tests revealed that *P. pallens* and unclassified *Veillonella*, two heritable OTUs (A > C and E), were associated with a healthy state in enamel. Other OTUs indicative of a caries-free state in the outer tooth layer were *Streptococcus*, *S. anginosus*, *Cardiobacterium*, Actinomycetaceae, and *Corynebacterium*. These taxa showed both greater shared and unique/non-shared environmental variation (C > A and E or E > A and C). All the taxa associated with dental caries in enamel or progressing into dentin had a greater portion of their variation explained by unique or shared environmental factors (Figures 6A and 6B). In contrast, none of the taxa we identified as heritable showed associations to disease.

Taxa influenced by genetic or environmental effects and associated with oral health or disease were also impacted by age. For instance, *P. pallens,* the most heritable OTU, decreased in abundance as child age increased to 6–11 years old ($q < 0.001$, nonlinear least-squares best fit, locally weighted scatterplot smoother [LOWESS]) (Figure 6C). When other OTUs associated with a caries-free state also covaried with age, they did so in a decreasing manner (e.g., *S. anginosus*, unclassified *Streptococcus*, Actinomycetaceae, and *Corynebacterium* [$Q < 0.001$]) (Figure S6A). These OTUs were influenced by both shared and non-shared environmental forces. *Tannerella*, an OTU significantly modulated by unique environmental variation, was the only caries-associated OTU that also covaried with age and did so in an increasing manner ($q < 0.001$) (Figure 6D).

Some of these taxa also covaried with the frequency of sugar consumption (Figure S6A). For instance, the significantly heritable unclassified *Veillonella* and *Leptotrichia* associated with a caries-free state in enamel and showed lower abundance in twins who always had sugar added to their food and drinks. In contrast, the caries-associated *Aggregatibacter*, an OTU significantly influenced by non-shared environmental forces, increased in abundance with sugar consumption frequency (Kruskal-Wallis test, $p < 0.05$). Although these data do not support significant associations between dental caries and oral health habits (e.g., brushing frequency, toothpaste used, and antibiotic use by the mother or antibiotic use by the infant in the first week postnatal), the study design, cohort selection, and sample collection were primarily focused on heritability rather than lifestyle or oral health habits.

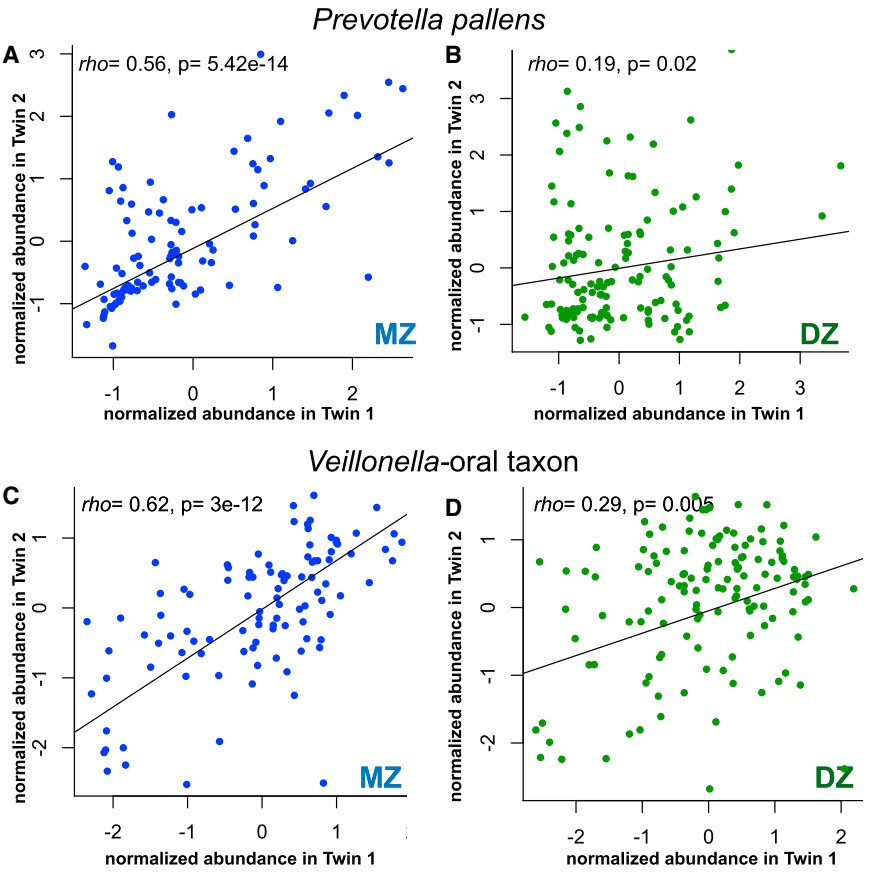


Figure 5. Highly Heritable Taxa Covary Significantly between MZ Twin Pairs

(A–D) Abundances of *P. pallens* (A and B) and unclassified *Veillonella* (C and D). The taxa with the highest heritability values correlated significantly between twin pairs, with stronger associations between MZ (A and C) (Pearson's rho > 0.5, p < 0.001) compared to DZ individuals (B and D).

Ecological Dynamics of Heritable Microbes

Based on the evidence collected, ecological associations of taxa both driven by additive genetic effects or environmental factors and associated with health and disease were investigated. Co-occurrence patterns of each taxon were determined, along with ecological associations predictive of caries phenotype. To that end, we constructed a network based on adjusted multiple correlations of transformed OTU abundances. The amount of variation driven by additive genetic effects (A) was not correlated with degree of connectivity (rho = −0.22, p = 0.15, Figure 7A). Thus, OTUs that had higher heritability estimates (A > C and E), with the sole exception of *Abiotrophia*, did not generally form clusters, represent hubs, or co-occur with multiple taxa (with either high or low heritability). Indeed, the most highly heritable OTUs and the only heritable taxa associated with a caries-free state (*P. pallens* and *Veillonella*) were not significantly associated with any other taxa (based on a minimum correlation threshold of rho = |0.5|) and thus were not seen in the network. These data contrast with results presented for the gut microbiome (Goodrich et al., 2014), where highly heritable taxa tended to co-occur more often.

The low connectivity of highly heritable taxa is in contrast with OTUs showing a higher portion of their variability explained by shared environmental factors, which tended to be highly connected hubs. For instance, shared environmental variation (C) was highly associated with network connectivity (rho = 0.65, p = 2.76e−16) (Figure 7B). Indeed, the OTU with the highest variability explained by the shared environment, an unclassified *Streptococcus*, was the most highly connected node. Interestingly, OTUs driven by non-shared/unique variability (E) did not tend to co-occur with other taxa, confirmed by significantly negative correlation with network connectivity (rho = −0.61, p = 0.15, Figure S7).

OTUs associated with caries presence were always modulated by shared or non-shared environmental factors and, in most cases, co-occurred with another taxon under high environmental control more than one time (degree > 1). For example, *Rothia*, associated with caries prevalence in both enamel and dentin, was part of a highly connected cluster that included *Streptococcus* spp., the genera exhibiting the most shared environmental variation, and the caries-associated *Actinomyces* (Figure 7B). *Selenomonas*, a taxon that also exhibited high control from shared environmental factors and that was predictive of caries in dentin, formed a hub connecting other potentially cariogenic, high-E taxa, such as *Tannerella* and *P. olorum* (Figure S7). Nonetheless, OTUs that characterized a caries-free state (e.g., *Corynebacterium* and *Streptococcus*) also formed direct or indirect associations with caries-indicative taxa, which highlights the ecological complexity of dental caries.

DISCUSSION

This study provides evidence that the composition of the human oral microbiome is influenced by the host genetic background. The data also indicate that shared and non-shared environmental factors are significant drivers of oral microbial communities, especially in connection to oral disease and microbiome maturation. These results add to existing evidence of host genetic control on the human microbiome based on high-throughput sequencing studies. These reports, largely focused on the gut microbiome, have shown heritable microbiome traits involved in food metabolic functions, such as methanogenesis and fermentation (Goodrich et al., 2014; Hansen et al., 2011), and host loci associated with the abundance of specific gut taxa, such as *Bifidobacterium* and *Akkermansia* (Benson et al., 2010; Blekhman et al., 2015; Bonder et al., 2016; Davenport et al., 2015; Goodrich et al., 2016).

Evidence of host genetic control of the human oral microbiome using high-throughput sequencing approaches has been scarcer, especially in the context of oral health and disease.

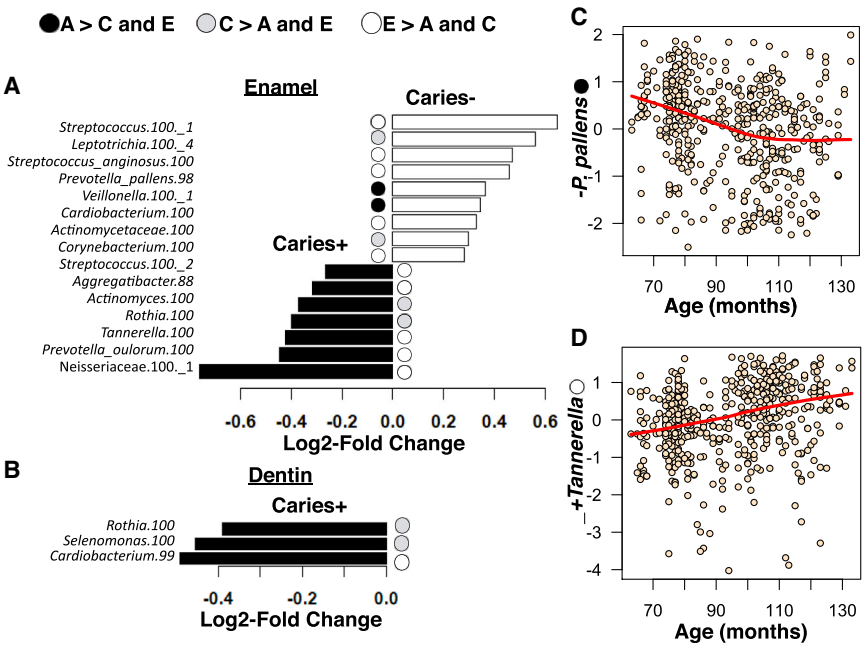


Figure 6. Genetic and Environmental Determinants of the Oral Microbiome in Connection to Oral Health

(A and B) Wilcoxon rank-sum tests identified taxa distinguishing caries-positive versus caries-negative individuals ($p < 0.05$) in enamel (A) and dentin (B), as shown by log2 fold changes of transformed data. Black, gray, and white circles denote OTUs with a greater proportion of their variation explained by additive genetic (A > C and E), shared environmental (C > A and E), and non-shared unique environmental factors (E > A and C), respectively.

(C and D) *P. pallens* (C) associated with a healthy state, decreased in abundance as children aged, while the environmentally driven *Tannerella* (D), predictive of caries, showed the opposite trend. All models are based on nonlinear regression adjusted by sex and zygosity ($Q < 0.01$, locally weighted scatterplot smoother [LOWESS]). Minus and plus signs next to each taxon name indicate taxa associated with caries-negative and caries-positive states, respectively (other taxa following a decreasing trend can be seen in Figure S6A).

Previous twin-based studies on the heritability of oral microbes have relied on culture-based and array-based (checkerboard hybridization) approaches, and contrary to our results, they have suggested that the abundance of the cariogenic *S. mutans* and specific pathogenic attributes such as acid production in dental plaque and saliva are highly heritable traits (Bretz et al., 2005; Corby et al., 2007, 2005).

The data presented herein show no evidence that potentially cariogenic streptococci and other taxa likely associated with caries are significantly driven by genetic factors. This discordance may be due to the fact that checkerboard assays and culture-based approaches hinder resolution of microbial diversity and ecological interactions, which could mask the effects of host-genetic correlates of quantitative microbiome traits (Hansen et al., 2011; Spor et al., 2011). On the contrary, we show that taxa possibly playing a role in caries pathology are mainly modulated by shared and unique environmental forces, and that the heritable fraction of the oral microbiome is not significantly associated with oral disease.

These data also show that some taxa associated with health and disease also covaried with age and sugar consumption patterns and exhibited distinct ecological interactions. Thus, potentially cariogenic and health-associated taxa may emerge or decrease as infants transition into more mature stages and become more exposed to the external environment (Cephas et al., 2011; Lif Holgerson et al., 2015; Rôças et al., 2016). Specifically, it is noteworthy that the only taxon that was simultaneously associated with caries and covaried with age (*Tannerella*) did so in an increasing manner, while bacteria associated with a healthy state, including the most heritable *P. pallens*, were eroded from the oral community over time. These observations, along with the significant variation in heritability estimates for individual taxa and the increased influence of unique environmental forces over time, raise questions about how heritable and environmentally derived fractions of the microbiome influence disease.

For instance, *P. pallens*, the taxon most strongly driven by additive genetic factors, has not previously been reported to be associated with caries. Also, Pasteurellaceae, another taxon with high heritability values, is reported to influence the oral acid-base homeostasis to inhibit aciduric and caries-associated taxa in children's saliva (Morou-Bermudez et al., 2015). Similarly, abundances of *Prevotella*, Pasteurellaceae, and *Leptotrichia* have been previously associated with single-nucleotide polymorphisms (SNPs) in host genes coding for ATP-binding cassettes, protein synthesis, cell division, and tumor suppression (Blekhman et al., 2015). Moreover, some of the most abundant taxa in these dataset, *Veillonella*, *Neisseria*, and Pasteurellaceae (~37% of all reads), were more significantly influenced by additive genetic factors, with no evidence of associations with caries.

Functional genomic data at the host and microbe levels are necessary to make assumptions on microbiome seeding in the oral cavity and its influence on disease phenotypes later in life. The data presented here generate additional questions about the extent to which protective or neutral microbiome features are innately acquired and significantly modulated by the host genotype. Thus, these data could indicate that there is a "wild-type" plaque ecology that exists in the absence of specific environmental stimuli over time and to which the human genome is naturally adapted. Consequently, some oral microbial phenotypes, particularly those characterizing oral disease, may be a consequence of an environmentally derived bacterial milieu to which the host genome is not adapted. This potential evolutionary discordance may be supported by evidence that cariogenic taxa such as *S. mutans* arose after humans transitioned to agricultural and industrialized lifestyles (Adler et al., 2013).

As such, these results may indicate that focusing on modulating specific environments in a given individual is key for oral disease prevention, diagnosis, and treatment. Thus, mining for

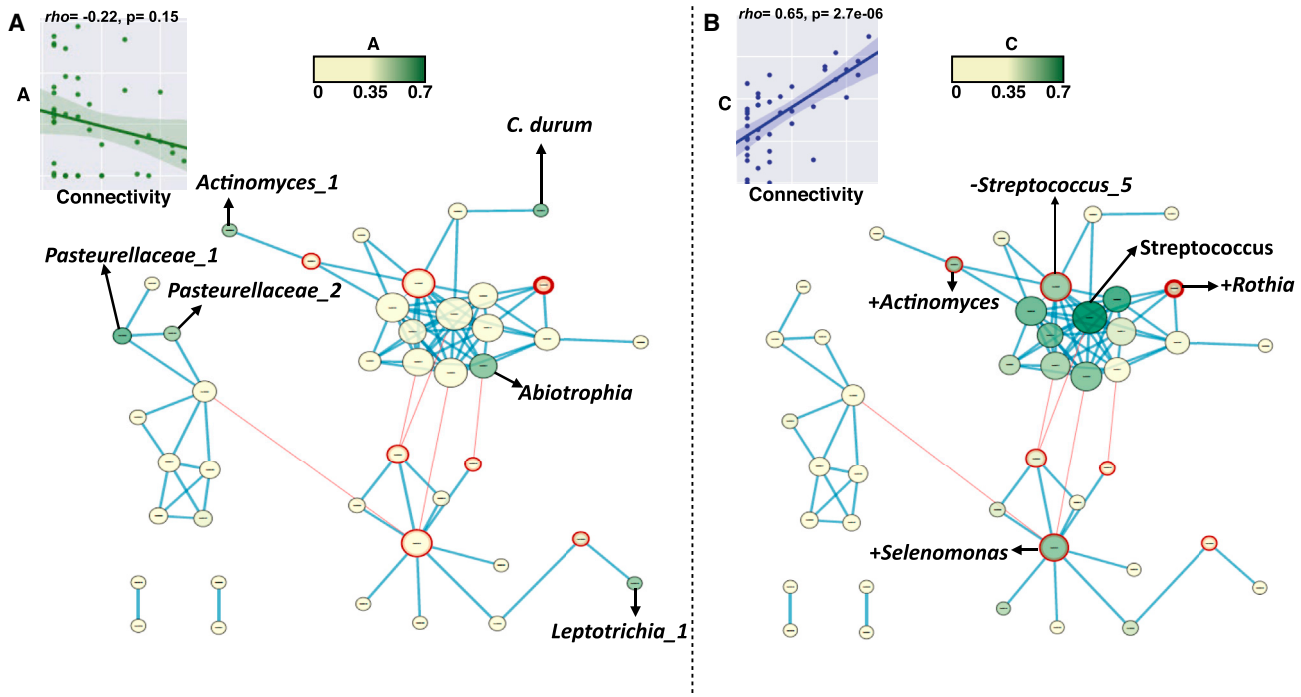


Figure 7. Ecological Dynamics of the Oral Microbiome in the Context of Health and Heritability

(A and B) Network view of co-occurrence patterns between OTUs (nodes) based on heritability estimates (A > C and E) (A) and shared environmental variation (C > A and E) (B). Inset scatterplot shows relationships between degree of connectivity and A or C estimates. Blue and red edges show positive and negative correlation coefficients, respectively (*SparCC*, rho > |0.5|). Yellow and green colors in the nodes indicate taxa from low to high A or C estimates (0–0.7), as shown in key. Node size corresponds to connectivity (degree of associations a given taxon has with other taxa). A red border indicates taxa significantly associated with caries state, and a thicker red border indicates that a taxon is significantly predictive of caries in both enamel and dentin (e.g., *Rothia*). Network view based on taxa mainly driven by higher non-shared environmental variation (E > A and C) can be seen in Figure S6B.

the abundance or expression of particular oral taxa and functions in both saliva and plaque, in the context of environmental variation, presents a window of opportunity to predict oral disease risk at individual and population-wide levels. Moreover, the realization that there remain wild-type protective taxa that may be lost with time and environmental influences offers clues for developing potential prevention or treatment strategies based on microbial agents. These data are also consistent with complex ecological interactions and metabolic crosstalk among multiple strains underlying dental caries (Gross et al., 2012; Jagathrakshakan et al., 2015; Marsh, 2003). The microbial ecology of dental caries may be highly individualized, rather than defined by health and disease ecotypes or single-microbe etiology across populations.

The role that taxa modulated by genetic and environmental forces play in disease prevalence should be further assessed based on personalized functional surveys of the microbiome (metagenomics, metatranscriptomics, and metabolomics) and the host genome (genome-wide association studies). These approaches could provide improved resolution on how potentially heritable and environmentally derived microbial features interact in the oral ecosystem, and with intrinsic and extrinsic forces, to modify disease progression on specific individuals. Likewise, due to the significant age-driven variation in heritability estimates among oral taxa, analyses expanded across longitudinal scales with the same individuals are needed. This strategy is key to trace how health and disease-associated oral biomarkers change along a person's lifetime in the context of genotype and environment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study Subjects
  - Caries Assessment
- METHOD DETAILS
  - Sample Collection and Processing
  - Molecular Analyses
  - 16S rRNA Sequence Analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Microbial Community Analyses
  - Dendrograms Bray-Curtis Distances
  - Heritability Calculations
  - Intraclass Correlation Coefficients
  - Caries Indicators
  - Nonlinear Relationships between OTU Abundances and Age
  - Network View of OTU Co-occurrence
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2017.08.013.

AUTHOR CONTRIBUTIONS

Conceptualization: A.G., J.L.E., C.L.D., and K.E.N.; Methodology: A.G., J.L.E., R.K., D.M.H., M.T., J.I., C.L.D., and K.E.N.; Software: J.L.E., D.M.H., and J.I.; Formal Analysis: A.G., J.L.E., R.K., D.M.H., M.T., C.K., and J.I.; Investigation: P.L., R.S., M.B., T.H., J.M.C., S.K.H., and M.B.J.; Resources: P.L., R.S., M.B., T.H., J.M.C., and K.E.N.; Data Curation: J.L.E., D.M.H., R.K., and J.I.; Writing – Original Draft: A.G., Writing – Review & Editing: A.G., J.L.E., D.M.H., P.L., R.S., M.B., M.T., T.H., J.M.C., S.K.H., C.L.D., and K.E.N.; Visualization: A.G. and J.L.E.; Supervision: K.E.N.; Project Administration: K.E.N.; Funding Acquisition: K.E.N.

ACKNOWLEDGMENTS

We wish to thank all twins and their families; Grant Townsend and Nicky Kilpatrick for their dental expertise; and Tina Vaiano, Jane Loke, Anna Czajko, Blessy Manil, Chrissie Robinson, Mihiri Silva, and Supriya Raj for their expertise and assistance with collection of data and samples. We also want to thank Dr. Anna Edlund, Dr. Nicholas Schork, and Thomas Clarke at the JCVI. The research in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under Award Number R01DE019665 to K.E.N. The Peri/Postnatal Epigenetic Twins Study (PETS) was supported by grants from the Australian National Health and Medical Research Council (grant numbers 437015 and 607358 to J.M.C. and R.S.), the Bonnie Babes Foundation (grant number BBF20704 to J.M.C.), the Financial Markets Foundation for Children (grant number 032-2007 to J.M.C.), and the Victorian Government's Operational Infrastructure Support Program. The Craniofacial Biology Research Group (CBRG) was supported by grants from the Australian National Health and Medical Research Council (grant numbers 349448 and 1006294 to T.H.) and the Financial Markets Foundation for Children (grant number 223-2009 to T.H.).

Received: April 24, 2017
Revised: June 30, 2017
Accepted: August 18, 2017
Published: September 13, 2017

REFERENCES

Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., Bradshaw, C.J.A., Townsend, G., Sołtysiak, A., Alt, K.W., et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat. Genet. *45*, 450–455.e1.

Azad, M.B., Konya, T., Maughan, H., Guttman, D.S., Field, C.J., Chari, R.S., Sears, M.R., Becker, A.B., Scott, J.A., and Kozyrskyj, A.L.; CHILD Study Investigators (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. CMAJ *185*, 385–394.

Beck, J.D., Slade, G., and Offenbacher, S. (2000). Oral disease, cardiovascular disease and systemic inflammation. Periodontol. 2000 *23*, 110–120.

Benjamin, R.M. (2010). Oral health: the silent epidemic. Public Health Rep. *125*, 158–159.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. USA *107*, 18933–18938.

Blaser, M.J. (2014). The microbiome revolution. J. Clin. Invest. *124*, 4162–4165.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., and Clark, A.G. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Bodmer, W.F. (1961). Biometrics *17*, 329–330.

Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S.P., et al. (2016). The effect of host genetics on the gut microbiome. Nat. Genet. *48*, 1407–1412.

Bretz, W.A., Corby, P.M.A., Hart, T.C., Costa, S., Coelho, M.Q., Weyant, R.J., Robinson, M., and Schork, N.J. (2005). Dental caries and microbial acid production in twins. Caries Res. *39*, 168–172.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4516–4522.

Cephas, K.D., Kim, J., Mathai, R.A., Barry, K.A., Dowd, S.E., Meline, B.S., and Swanson, K.S. (2011). Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. PLoS One *6*, e23503.

Corby, P.M.A., Bretz, W.A., Hart, T.C., Filho, M.M., Oliveira, B., and Vanyukov, M. (2005). Mutans streptococci in preschool twins. Arch. Oral Biol. *50*, 347–351.

Corby, P.M., Bretz, W.A., Hart, T.C., Schork, N.J., Wessel, J., Lyons-Weiler, J., and Paster, B.J. (2007). Heritability of oral microbial species in caries-active and caries-free twins. Twin Res. Hum. Genet. *10*, 821–828.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. PLoS One *10*, e0140301.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. USA *107*, 11971–11975.

Eaves, L.J., Last, K.A., Young, P.A., and Martin, N.G. (1978). Model-fitting approaches to the analysis of human behaviour. Heredity (Edinb) *41*, 249–320.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods *10*, 996–998.

Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. *8*, e1002687.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe *19*, 731–743.

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

Hagberg, A.A., Schult, D.A., and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference (SciPy2008), G. Varoquaux, T. Vaught, and J. Millman, eds., pp. 11–15.

Hansen, E.E., Lozupone, C.A., Rey, F.E., Wu, M., Guruge, J.L., Narra, A., Goodfellow, J., Zaneveld, J.R., McDonald, D.T., Goodrich, J.A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4599–4606.

Holgerson, P.L., Vestman, N.R., Claesson, R., Ohman, C., Domellöf, M., Tanner, A.C.R., Hernell, O., and Johansson, I. (2013). Oral microbial profile discriminates breast-fed from formula-fed infants. J. Pediatr. Gastroenterol. Nutr. *56*, 127–136.

Holmstrup, P. (1999). Non-plaque-induced gingival lesions. Ann. Periodontol. *4*, 20–31.

Holst, K.K., and Scheike, T. (2015). mets: Analysis of Multivariate Event Times. https://cran.r-project.org/web/packages/mets/index.html.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. *33*, 1635–1638.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Ismail, A.I., Sohn, W., Tellez, M., Amaya, A., Sen, A., Hasson, H., and Pitts, N.B. (2007). The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. Community Dent. Oral Epidemiol. *35*, 170–178.

Jagathrakshakan, S.N., Sethumadhava, R.J., Mehta, D.T., and Ramanathan, A. (2015). 16S rRNA gene-based metagenomic analysis identifies a novel bacterial co-prevalence pattern in dental caries. Eur. J. Dent. *9*, 127–132.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. *79*, 5112–5120.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2/33*, 18–22. http://www.bios.unc.edu/~dzeng/BIOS740/randomforest.pdf.

Lif Holgerson, P., Harnevik, L., Hernell, O., Tanner, A.C.R., and Johansson, I. (2011). Mode of birth delivery affects oral microbiota in infants. J. Dent. Res. *90*, 1183–1188.

Lif Holgerson, P., Öhman, C., Rönnlund, A., and Johansson, I. (2015). Maturation of oral microbiota in children with or without dental caries. PLoS One *10*, e0128534.

Loesche, W.J. (2011). Microbiology of dental decay and periodontal disease. In Medical Microbiology, S. Baron, ed. (University of Texas Medical Branch at Galveston).

Loke, Y.J., Novakovic, B., Ollikainen, M., Wallace, E.M., Umstad, M.P., Permezel, M., Morley, R., Ponsonby, A.-L., Gordon, L., Galati, J.C., et al. (2013). The Peri/postnatal Epigenetic Twins Study (PETS). Twin Res. Hum. Genet. *16*, 13–20.

Marsh, P.D. (2003). Are dental diseases examples of ecological catastrophes? Microbiology *149*, 279–294.

Morou-Bermudez, E., Rodriguez, S., Bello, A.S., and Dominguez-Bello, M.G. (2015). Urease and dental plaque microbial profiles in children. PLoS One *10*, e0139315.

Oksanen, J., Kindt, R., Legendre, P., and O'Hara, R.B. (2006). Vegan: Community Ecology Package ver. 1.8.–3. Physics Chem. Earth: B: Hydrol (Oceans Atmosphere).

Papapostolou, A., Kroffke, B., Tatakis, D.N., Nagaraja, H.N., and Kumar, P.S. (2011). Contribution of host genotype to the composition of health-associated supragingival and subgingival microbiomes. J. Clin. Periodontol. *38*, 517–524.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.

R Development Core Team. (2013). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Rôças, I.N., Alves, F.R.F., Rachid, C.T.C.C., Lima, K.C., Assunção, I.V., Gomes, P.N., and Siqueira, J.F., Jr. (2016). Microbiome of deep dentinal caries lesions in teeth with symptomatic irreversible pulpitis. PLoS One *11*, e0154653.

Rubinstein, M.R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y.W. (2013). Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe *14*, 195–206.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D.R., Kultima, J.R., Martin, J., et al. (2013). Genomic variation landscape of the human gut microbiome. Nature *493*, 45–50.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Schmidt, B.L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P.M., Queiroz, E.L.S., Nightingale, K., Kerr, A.R., DeLacure, M.D., Veeramachaneni, R., et al. (2014). Changes in abundance of oral microbiota associated with oral cancer. PLoS One *9*, e98741.

Schult, D.A., and Swart, P. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conferences (SciPy 2008), pp. 11–16.

Seymour, G.J., Ford, P.J., Cullinan, M.P., Leishman, S., and Yamazaki, K. (2007). Relationship between periodontal infections and systemic disease. Clin. Microbiol. Infect. *13* (*Suppl 4*), 3–10.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. *9*, 279–290.

Stahringer, S.S., Clemente, J.C., Corley, R.P., Hewitt, J., Knights, D., Walters, W.A., Knight, R., and Krauter, K.S. (2012). Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood. Genome Res. *22*, 2146–2152.

Takahashi, N., and Nyvad, B. (2011). The role of bacteria in the caries process: ecological perspectives. J. Dent. Res. *90*, 294–303.

Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. Annu. Rev. Microbiol. *65*, 411–429.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Dental Plaque Swabs	University of Adelaide Craniofacial Biology, Murdoch Children's Research Institute	N/A
Critical Commercial Assays		
QIAGEN PCR purification kit	QIAGEN Inc.	Cat# 28106
Illumina MiSEQ reagent kit	Illumina Inc.	Cat# MS-102-2003
Deposited Data		
Raw and analyzed data	This paper	16S rRNA sequence data (NCBI) BioProject: PRJNA383868, Sequence Read Archive, SRA: SRR5467515–SRR5467785, SRR5467788–SRR5468062
Oligonucleotides		
16S rRNA-Forward Primer 515-534F: GTGCCAGCMGCCGCGGTAA 16S rRNA-Reverse Primer 806-787 GGACTACHVGGGTWTCTAAT	Caporaso et al., 2011.	N/A
Software and Algorithms		
UPARSE	Edgar, 2013	http://www.drive5.com/uparse/
mothur	Schloss et al., 2009	https://www.mothur.org/
mets: Analysis of Multivariate Events	N/A	https://cran.r-project.org/web/packages/mets/index.html
vegan: Community Ecology Package	Oksanen et al., 2006	https://cran.r-project.org/web/packages/vegan/index.html
Cytoscape	Shannon et al., 2003	http://www.cytoscape.org/
ETE3: A Python framework to work with trees	Huerta-Cepas et al., 2016	http://etetoolkit.org/
randomForest: Breiman and Cutler's Random Forests for Classification and Regression	Liaw and Wiener, 2002	https://cran.r-project.org/web/packages/randomForest/index.html
R 3.4.1	R Core Team	https://www.r-project.org/
Continuum Analytics Scientific Python 3.5.2	N/A	https://anaconda.org/anaconda/python/files?version=3.5.2?version=3.5.2
NetworkX	Hagberg et al., 2008	https://networkx.readthedocs.io/en/stable/
SparCC	Friedman and Alm, 2012	https://bitbucket.org/yonatanf/sparcc

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Karen Nelson (kenelson@jcvi.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study Subjects

Dental plaque samples were collected from participants of the University of Adelaide Craniofacial Biology Research Group (CBRG), and the Murdoch Children's Research Institute (MCRI)'s Peri/Postnatal Epigenetic Twins Study (PETS) (Loke et al., 2013) in 2015 and 2016. The PETS (n = 193) and CBRG (n = 292) cohorts were composed of male (48.3%) and female (51.4%) twins, 5 to 11 years old (Table S1). The samples were collected after ensuring that each participant had refrained from brushing their teeth the evening before and the morning of the clinical visit. Patients who had had antibiotic treatment within six months prior to sample collection were excluded. Work with PETS was approved by the Royal Children's hospital Human Research Ethic Committee, reference number 33174 and work with the CBRG cohort was approved by The University of Adelaide Human Research Ethics Committee, project No. H-2013-097. Research at J. Craig Venter Institute was approved by the JCVI Institutional Review Board, protocol number 2013-182.

Caries Assessment

The entire dentition of each participant was assessed using the International Caries Detection and Assessment System (ICDAS II) (Ismail et al., 2007). The ICDASII is used to assess and define dental caries at the initial and early enamel lesion stages through to dentinal and finally stages of the disease. It enables reliable, informative, and comparative data that is widely used for population health and clinical studies. Examiners were experienced clinicians who had undergone rigorous calibration and were routinely recalibrated across measurement sites to minimize error. Caries experience in each participant was initially reduced to a whole-mouth score and three classifications were utilized: no evidence of current or previous caries; experience; evidence of current caries affecting the enamel layer only on one or more tooth surfaces; evidence of previous or current caries experience that has progressed through the enamel layer to involve the dentin on one or more tooth surfaces (including restorations or tooth extractions due to caries). For the purpose of this analysis, we classified diseased phenotypes in twins as presence of caries in enamel or dentin. Metadata available for this cohort, collected through a series of validated questionnaires, included data on dietary habits (including 'sugar' intake), oral health habits (such as tooth brushing frequency and toothpaste amount and type), and antibiotic exposure post-natally (Table S1).

METHOD DETAILS

Sample Collection and Processing

Dental plaque samples were collected by thoroughly swabbing along the gingival margins and buccal surfaces of the dentition. The swab was then expressed into a labeled microcentrifuge tube containing 500 μL of RNAprotect bacteria (QIAGEN, Inc.). The samples were immediately stored at −80°C and transported on dry ice to JCVI.

Molecular Analyses

DNA from plaque samples was extracted using lysozyme digest followed by bead beating, phenol/chloroform isoamyl alcohol extraction and ethanol precipitation. Bacterial community composition was targeted using the V4 variable region of the 16S rRNA gene. This was amplified using adaptor and barcode ligated V4 primers (Kozich et al., 2013). PCR amplicons were generated from approximately 100ng of extracted DNA using Platinum Taq polymerase (ThermoFisher Inc) and the following cycling conditions; 94°C followed by 35 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 30 s, followed by a final extension cycle of 72°C for 5 min. Amplicons were purified using the Qiaquick PCR purification kit (QIAGEN Inc.) following manufacturer's specifications. Purified amplicons were quantified using SybrGold, normalized and pooled in preparation for Illumina MiSeq sequencing. 16S libraries were sequenced using the dual index V2 chemistry kit 2x250bp on an Illumina MiSeq according to manufacturer's specifications.

16S rRNA Sequence Analyses

Operational taxonomic units (OTUs) were generated de novo from raw Illumina sequence reads using an *in-house* analyses pipeline relying on the UPARSE (Edgar, 2013) and mothur (Schloss et al., 2009) open-source bioinformatics tools. Briefly, paired-end reads were trimmed of adaptor sequences, barcodes, and primers prior to assembly, followed by discarding low quality reads and singletons. The pipeline empirically determines quality and expected error rates for each position on the assembled reads. In this case, we used the conservative 'HI_EE' value of 0.11. After a de-replication step and abundance determination, sequences were filtered for chimeras and clustered into OTUs. UPARSE has a built-in filter for chimera detection and removal, UCHIME2 (http://www.biorxiv.org/content/early/2016/09/09/074252), which uses the highly curated SILVA database (Quast et al., 2013). To predict taxonomy, we used the Wang classifier, and bootstrapped using 100 iterations. We set mothur to report full taxonomies only for sequences where 80 or more of the 100 iterations were identical (cutoff = 80). Taxonomies were assigned to the OTUs with mothur using version SSU Ref NR 99 123 of the SILVA 16S ribosomal RNA database as the reference. Tables with OTUs and the corresponding taxonomy assignments were generated and used in subsequent analyses. All molecular analyses took place in 2015 and 2016.

QUANTIFICATION AND STATISTICAL ANALYSIS

Microbial Community Analyses

All microbiome community analyses and tests were conducted using R statistical software (R Development Core Team, 2013). To avoid the inclusion of possible sequencing artifacts, OTUs found in fewer than 5 individuals or fewer than 5 times in the dataset were not included in subsequent analyses. Principal coordinates analyses were performed on Bray-Curtis distance matrices after transforming raw OTU data counts into relative abundances using the *ape* and *vegan* R packages (Oksanen et al., 2006; Paradis et al., 2004). The latter was also used to calculate alpha diversity. Random forest was performed using the randomForest R package (Liaw and Wiener, 2002).

Dendrograms Bray-Curtis Distances

Pairwise Bray-Curtis distance matrices were calculated separately for dentin and enamel microbiomes from relative abundance data. Ward linkage hierarchical clustering dendrograms were constructed from distance matrices using *Continuum Analytics Scientific*

Python (v3.5.2) distribution. The dendrogram tree structure was converted to Newick format and ported to *ETE3 (v3.0.0)*(Huerta-Cepas et al., 2016) for circular tree representation. Distances between twin nodes along the dendrogram were calculated using 2 measures: (1) topological distance and (2) branch distance.

Heritability Calculations

Heritability calculations were performed in the full cohort. Previous to all heritability calculations, OTU (raw) count tables were filtered and transformed as follows: (1) Only OTUs present in > 50% individuals were taken into account leaving 91 total OTUs; and (2) raw counts were then normalized by sum, log transformed and scaled (mean-centered and divided by the square root of the standard deviation of each OTU). The ACE model for heritability was calculated using the *mets* package in R (Holst and Scheike, 2015). The model implemented controlled for age and sex in the calculation of A, C and E parameters. Falconer's formula was calculated based on ICCs in the following way: $2^{*}(ICC_{mz}-ICC_{dz})$.

Intraclass Correlation Coefficients

ICC calculations were executed by: (1) separating twins into monozygotic, dizygotic, and unrelated subsets; (2) splitting twins-pairs into vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ where the *i*-th index contains a patient from the *i*-th family; (3) calculating the Pearson's correlation between $\boldsymbol{u}$ and $\boldsymbol{v}$; (4) shuffling the twin membership between vectors; and (5) permuting steps (3-4) 10,000 times for each zygosity subset and taking the mean ICC.

Caries Indicators

Wilcoxon Rank Sum tests as implemented in the *stats* R package were used to determine whether transformed raw counts of a given OTU was associated with absence or presence of caries, age, and sugar consumption patterns. A false discovery rate adjustment (FDR) using the Benjamin-Hochberg method was applied to all *P*-values.

Nonlinear Relationships between OTU Abundances and Age

Nonlinear regression models to estimate the effect of age on OTU abundances were run using the nonlinear least square (*nls*) function in R, controlling for zygosity and sex and in the transformed OTU data, as described above.

Network View of OTU Co-occurrence

The 91 OTUs used in the heritability analysis were used for generating co-occurrence networks. Pairwise correlation matrices were calculated from the abundance of these 91 taxa using *SparCC* (Friedman and Alm, 2012) with 20 iterations. Node connections between taxa were determined by evaluating the scale-free topology index of various hard-thresholds ranging from 0.1 - 0.99 *SparCC* correlation using *NetworkX* (v1.11) (Schult and Swart, 2008). We determined a threshold of 0.5 to contain the best tradeoff between scale-free topology = 0.876, node representation = 42, and edge connections = 127. The *NetworkX* graph was ported to *Cytoscape* (v3.4.0) for data visualization and aesthetics (Shannon et al., 2003).

DATA AND SOFTWARE AVAILABILITY

16S rRNA sequence data have been deposited in the National Center for Biotechnology Information (NCBI) database on BioProject accession number: PRJNA383868, and under Sequence Read Archive (SRA) accession IDs SRR5467515–SRR5467785 and SRR5467788–SRR5468062.