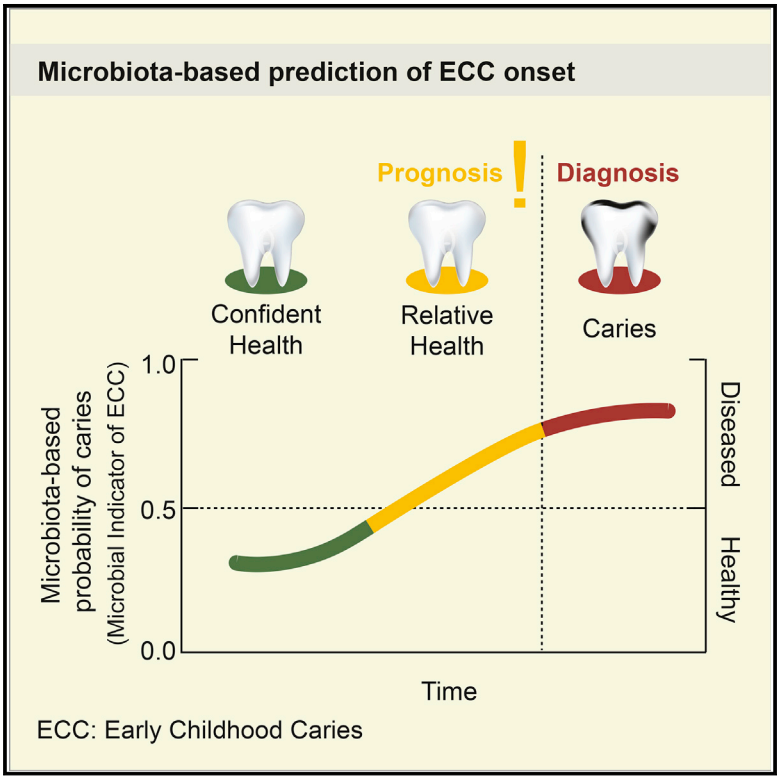
Article

Cell Host & Microbe

Prediction of Early Childhood Caries via Spatial-Temporal Variations of Oral Microbiota

Graphical Abstract



Highlights

- Oral microbiota in 50 four-year-old children were tracked for 2 years
- Age-dependent microbiota development is perturbed by early childhood caries (ECC) onset
- Shifts in microbiota precede manifestation of clinical symptoms of ECC
- Microbial Indicators of Caries, when de-trended for age, can predict ECC onset

Authors

Fei Teng, Fang Yang, Shi Huang, ..., Rob Knight, Junqi Ling, Jian Xu

Correspondence

xujian@qibebt.ac.cn (J.X.), lingjq@mail.sysu.edu.cn (J.L.)

In Brief

Teng et al. tracked plaque and saliva microbiota of 50 4-year-old children for 2 years. By distinguishing between aging- and disease-associated taxa and exploiting the distinct microbiota dynamics between disease onset and progression, a predictive model, Microbial Indicators of Caries, is proposed as a method to predict future caries onset.

Teng et al., 2015, Cell Host & Microbe *18*, 296–306
September 9, 2015
http://dx.doi.org/10.1016/j.chom.2015.08.005

CellPress

Cell Host & Microbe
Article

Prediction of Early Childhood Caries via Spatial-Temporal Variations of Oral Microbiota

Fei Teng,[1,2,5] Fang Yang,[3,5] Shi Huang,[2,5] Cunpei Bo,[2] Zhenjiang Zech Xu,[4] Amnon Amir,[4] Rob Knight,[4] Junqi Ling,[1,*] and Jian Xu[2,*]

[1]Department of Operative Dentistry and Endodontics, Guanghua School and Hospital of Stomatology, Sun Yat-sen University, Guangzhou, 510055, China

[2]Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, 266101, China

[3]Department of Stomatology, Qingdao Municipal Hospital, Qingdao, Shandong, 266101 China

[4]Departments of Pediatrics and Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093, USA

[5]Co-first author

*Correspondence: xujian@qibebt.ac.cn (J.X.), lingjq@mail.sysu.edu.cn (J.L.)

http://dx.doi.org/10.1016/j.chom.2015.08.005

SUMMARY

Microbiota-based prediction of chronic infections is promising yet not well established. Early childhood caries (ECC) is the most common infection in children. Here we simultaneously tracked microbiota development at plaque and saliva in 50 4-year-old preschoolers for 2 years; children either stayed healthy, transitioned into cariogenesis, or experienced caries exacerbation. Caries onset delayed microbiota development, which is otherwise correlated with aging in healthy children. Both plaque and saliva microbiota are more correlated with changes in ECC severity (dmfs) during onset than progression. By distinguishing between aging- and disease-associated taxa and exploiting the distinct microbiota dynamics between onset and progression, we developed a model, Microbial Indicators of Caries, to diagnose ECC from healthy samples with 70% accuracy and predict, with 81% accuracy, future ECC onsets for samples clinically perceived as healthy. Thus, caries onset in apparently healthy teeth can be predicted using microbiota, when appropriately de-trended for age.

INTRODUCTION

One key goal of human microbiome projects worldwide is to classify and predict host states based on human microbiota (Knights et al., 2011). The potential of microbiota-based classification of disease states has been tested in several studies such as plaque for gingivitis (Huang et al., 2014) and fecal samples for obesity (Le Chatelier et al., 2013), diabetes (Qin et al., 2012), and liver cirrhosis (Qin et al., 2014), etc. However, few studies have yet reported successful microbiota-based prediction of future disease outcome, especially for disease onset (Gevers et al., 2014). One explanation is that few experimental designs have simultaneously considered the spatial and temporal variation of microbiota during disease development. Along the temporal scale, microbiota change as the host ages, and their diversity and composition can change substantially due to the particular physiology, diet, and environmental exposure at a specific stage of host development (Song et al., 2013; Stahringer et al., 2012; Yatsunenko et al., 2012). For example, oral microbiota change during the normal aging process (Xu et al., 2015). On the other hand, at the spatial scale, microbiota from various physical niches (e.g., spatially distinct habitats such as the saliva and the plaque) can differ greatly in community structure (Costello et al., 2009; Huang et al., 2011; Huttenhower et al., 2012). Association of microbiota from multiple habitats with disease symptoms have been shown (Ling et al., 2010), although how the niches differed in their ability to model infections remained elusive.

Early childhood caries (ECC) is the most common oral disease in children. It affects approximately half of children worldwide and incurs enormous societal costs (Casamassimo et al., 2009). ECC leads to sustained demineralization of enamel and dentin, and the infection can spread from the affected tooth to the surrounding soft tissues, resulting in swelling and inflammation in highly progressed cases. Once started, the damage to teeth is irreversible, with child patients continuing to suffer from a higher risk for new lesions and even tooth loss over their entire lifespan (Chen et al., 2012; Leroy and Declerck, 2013; Selwitz et al., 2007). Therefore, preventive intervention of ECC is of particular clinical significance (Lancet, 2009; Selwitz et al., 2007). However, prediction of future ECC, particularly for new disease onset, has been difficult (Mejàre et al., 2014).

For the assessment of ECC risk, oral bacteria count (e.g., *Streptococcus mutans* count and salivary *Lactobacillus* count); chemical characteristics of saliva (e.g., buffering capacity and pH), baseline caries status (host oral condition at the first oral examination), as well as personal oral hygiene (e.g., visible plaque levels), behavior, diet (e.g., sugar intake) or socioeconomic level, have been employed as single-predictors or variables in multivariate models (Mejàre et al., 2014). However, limitations of these existing methods are apparent (Tellez et al., 2013). (i) Most of these risk factors are subjective, prone to human bias and error, and are not satisfactorily reproducible among

observers. For example, examination of baseline caries status typically depends on visual observation and the individual judgment of the examiner, making comparison across studies challenging. Personal information such as oral hygiene, frequency of food consumption, sugar consumption, and fluoride sources is usually collected voluntarily from patients and thus hard to accurately quantitate. (ii) Current chair-side checks such as teeth probing, saliva characteristics test, and/or X-ray examinations might not be accessible to many children and are burdensome both to children and parents. (iii) Collection of patient information via questionnaires and counting bacterial taxa can be time, cost, and labor intensive, whereas the cost of molecular analyses continues to decline. (iv) Past caries experience remains the sole consistent predictor of recurring ECC, but this metric cannot apply to new-onset ECC. For example, in one report from Australia, among children with at least one tooth surface with caries experience at baseline, the overall accuracy (sensitivity + specificity) was 1.28, yet merely 1.05 for children with non-caries experience at baseline (Ha et al., 2014). Additionally, the majority of susceptible children lack past caries experience (Hallett, 2013). (v) Existing methods exhibit a moderate to good performance for sorting elder individuals into caries risk groups, but they are of limited prediction utility in preschool children (Tellez et al., 2013). Although *Streptococcus mutans* has long been regarded as the specific pathogen underlying ECC, significant difference in microbial community structure between the caries and caries-free hosts was revealed, which supports the ecological plaque hypothesis (Gross et al., 2010; Jiang et al., 2013, 2014). Thus, the current notion is that in ECC, environmental perturbation shifts the balance of the oral microbiota and eventually leads to a predominance of cariogenic bacteria that cause sustained demineralization of tooth tissue (Gross et al., 2010, 2012; Russell, 2009). However, little is known about the within-host development or inter-host variation of the oral microbiota during the full course of ECC development, which can take years (Crielaard et al., 2011). In the spatial dimension, it is not clear whether and how microbiota from distinct oral niches are associated with ECC. For example, although plaques were usually sampled for microbiota in ECC (Jiang et al., 2011, 2013; Ling et al., 2010) and in adult caries (Belda-Ferre et al., 2012; Russell, 2009), salivary microbiota might also be implicated, as their sensitivity to adult caries was recently reported (Yang et al., 2014; Yang et al., 2012), and they are considerably easier to sample.

Therefore, employing ECC as a model, we designed a combined cross-sectional and longitudinal study to test whether and how the spatial and temporal variation of microbiota can be exploited to develop predictive models for chronic infections in humans. Temporal variation in the oral microbiota from the two distinct niches of supragingival plaque and saliva was tracked in a cohort of 50 children for up to 2 years (284 samples), representing the three most common circumstances of ECC development: the stay-healthy mode, the caries-onset mode, and the caries-progression mode. Ecological modeling techniques were employed to dissect the role of the microbiota in ECC progression, to compare it to that in gingivitis development, and to probe the predictive value of the microbiome for diagnosing and predicting ECC by identifying both aging- and disease-associated taxa.

RESULTS

A Dual-Niche Longitudinal View of Microbiota Development in ECC

Fifty children 51 (± 6) months of age were recruited and sampled for their plaque and saliva microbiota at four time points (T1, T2, T3, and T4). Intervals between the first three sampling time points (i.e., T1, T2, and T3) were 6 months, and between T3 and T4 was 1 year, together spanning up to 2 years (Table 1; Experimental Procedures). Thus, there were 284 samples in total, 142 from each of the two niches. The clinical state of caries was monitored with the dmfs (i.e., the number of decayed, missing, and filled teeth-surfaces in deciduous dentition) (Anaise, 1984). At a given time, microbiota with dmfs of zero were designated as "Healthy" ("H"); otherwise were as "Caries" ("C"), which consist of "low caries" ($1 \leq$ dmfs < 6) and "severe caries" (dmfs ≥ 6).

Based on the individual change of caries state over T1, T2, T3, and T4, the 50 children were classified into three groups (Table 1; Experimental Procedures). (i) The "stay healthy" (H2H) group, in which the 17 subjects (94 samples) maintained healthy state, with dmfs staying zero. (ii) The "caries-onset" (H2C) group; these 21 subjects (120 samples) underwent the transition from healthy to caries-active state. (iii) The "caries-progression" (C2C) group, where 12 subjects (70 samples) started with caries and evolved into an exacerbated disease state. Thus, our study design encompassed the most common circumstances of caries development in natural human populations.

Impact of Chronologic Age of Host on Oral Microbiota

The oral microbiota can exhibit profound alterations in diversity and composition over a relatively protracted time frame (Ling et al., 2013; Song et al., 2013; Xu et al., 2015), suggesting that developmental dynamics of oral microbiota can potentially respond to host development or maturation. For example, structure of oral microbiota can differ substantially among distinct age groups (i.e., infants without teeth, preschool children with deciduous teeth, children with mixed teeth, adolescents with permanent teeth, and adults with permanent teeth) (Crielaard et al., 2011; Song et al., 2013; Xu et al., 2015). However, the microbiota dynamics for the dentition phase (i.e., preschoolers of 4–6 years old) is not yet clear; moreover, whether healthy development of the oral microbiota is perturbed by ECC onset or progression is unknown.

To address the first question, impact of the various factors on the oral microbiota in our children cohort was assessed. In both plaque and saliva, in decreasing order, time (i.e., T1, T2, T3, and T4), disease status (i.e., H or C), individual host, and host grouping (i.e., H2H, H2C, and C2C) all exhibit influence on microbiota composition (Figures S1A and S1B). This suggests correlation of oral microbiota with the age of the child. In the H2H group, the link between temporal changes of the healthy microbiota with age was probed using the Jensen-Shannon metric (Figure 1). Within each niche, chronologic age of host has a strong effect on oral microbial diversity of healthy hosts (Figure 1; PERMANOVA, $p < 0.05$, $F > 3.5$). The effect of age is stronger than that of the host factor (Figure 1; PERMANOVA, $p > 0.05$ for plaque and $p < 0.05$, $F = 1.52$ for saliva). The effect of age factor was stronger in plaque than in saliva, suggesting that the plaque microbiota are more sensitive than that of saliva to host

Table 1. Study Design for Microbiota-Based Diagnosis and Prediction of Early Children Caries in Natural Population

Host ID	Group	Age at Sampling Start (Month)	dmfs Jun 6, 2011	Dec 19, 2011	Jun 15, 2012	Jun 19, 2013
K1012	C2C	47	3[a]	5[a]	8[a]	–
K1014		53	7[a]	10[a]	16[a]	–
K1017		48	3[a]	8[a]	9[a]	–
K1020		51	23[a]	30[a]	35[a]	–
K1021		48	2[a]	5[a]	8[a]	–
K1025		46	5[a]	7[a]	10[a]	–
K1028		49	26[a]	33[a]	51[a]	–
K1030		52	5[a]	8[a]	14[a]	–
K1035		50	14[a]	14[a]	20[a]	–
K1024		46	2[a]	11[a]	14[a]	28[a]
K1041		59	–	–	24[a]	32[a]
K1056		54	–	–	2[a]	3[a]
K1001	H2C	42	0[b]	1[a]	2[a]	–
K1002		52	0[b]	3[a]	4[a]	–
K1004		50	0[b]	3[a]	3[a]	–
K1006		49	0[b]	8[a]	8[a]	–
K1023		49	0[b]	3[a]	5[a]	–
K1003		50	0[b]	2[a]	3[a]	4[a]
K1005		46	0[b]	5[a]	5[a]	6[a]
K1026		47	0[b]	1[a]	1[a]	2[a]
K1008		48	0[c]	0[b]	2[a]	–
K1009		51	0[c]	0[b]	2[a]	–
K1011		50	0[c]	0[b]	3[a]	–
K1016		51	0[c]	0[b]	2[a]	–
K1013		46	0[c]	0[b]	4[a]	–
K1007		45	0[c]	0[b]	4[a]	6[a]
K1040		60	–	–	0[b]	2[a]
K1042		58	–	–	0[b]	2[a]
K1043		58	–	–	0[b]	2[a]
K1054		57	–	–	0[b]	2[a]
K1059		56	–	–	0[b]	2[a]
K1060		58	–	–	0[b]	2[a]
K1062		56	–	–	0[b]	2[a]
K1015	H2H	44	0[c]	0[c]	0	–
K1018		47	0[c]	0[c]	0	–
K1019		43	0[c]	0[c]	0	–
K1022		42	0[c]	0[c]	0	–
K1027		48	0[c]	0[c]	0	–
K1029		47	0[c]	0[c]	0	–
K1031		45	0[c]	0[c]	0	–
K1032		42	0[c]	0[c]	0[c]	0
K1033		48	0[c]	0[c]	0[c]	0
K1034		49	0[c]	0[c]	0[c]	0
K1044		59	–	–	0[c]	0
K1045		55	–	–	0[c]	0
K1049		60	–	–	0[c]	0
K1052		62	–	–	0[c]	0
K1053		63	–	–	0[c]	0
K1061		58	–	–	0[c]	0
K1063		55	–	–	0[c]	0

Each value represents the subject's dmfs at each sampling time point, which indicates the severity of ECC. Here we define those samples with dmfs higher than 0 as "caries" and samples with dmfs of 0 as "health." To derive a predictive model for future disease outcome, we further define those clinically-defined-as-healthy samples at the immediately preceding time point of caries onset as "Relative Health" (i.e., "RelativeH"). Consistently, samples whose hosts did not develop to caries in the next 6 months or 1 year were defined as "Confident Health" (i.e., "ConfidentH"). The other clinically defined "health" samples that are without any follow-up clinical state, i.e., those white-color cells filled with 0 were defined as "Not Determined" [ND]). Those time points where no valid samples were available were filled with "–." Both caries and ConfidentH samples were finally used for model training, while RelativeH samples were used for prediction.

[a]Caries.

[b]Relative Health.

[c]Confident Health.

aging (Figure 1, PERMANOVA, $p < 0.05$, $F = 5.29$ for plaque and $p < 0.05$, $F = 3.66$ for saliva). These results were consistent with those based on UniFrac metrics (Figures S1C and S1D), suggesting that both common and rare taxa contribute to the result and that it is robust to choice of metric.

To identify age-discriminatory taxa in healthy children, the relative abundance of all species in plaque and saliva within the H2H group was regressed against the chronologic age of each child at the time of sample collection using Random Forests machine learning algorithm (Breiman, 2001), which suggested that plaque and saliva microbiota age of unrelated healthy children can explain over 50% of the variance in microbiota. Inclusion of taxa beyond the top 10 taxa in plaque or the top 15 taxa in saliva, respectively, produced only minimal improvement in performance of the predictive model (Figures 2A, 2B, and S2). The most age-discriminatory taxa were thus operationally defined to be those bacterial taxonomic biomarkers able to differentiate the maturity of oral community composition in healthy children aged from 4 to 6. Remarkably, these 25 taxa from either plaque or saliva were mainly *Streptococcus*, *Neisseria*, *Fusobacterium*, *Capnocytophaga*, *Prevotella*, and *Porphyromonas*, which have been frequently observed as abundant members of the healthy oral microbiota in children (Crielaard et al., 2011; Ling et al., 2013; Xu et al., 2015). The 10-taxon plaque model and the 15-taxon salivary model were then employed to determine the relative microbiota maturity in plaque and saliva, respectively (Figures 2A, 2B and S2). For a particular sample, the relative microbiota maturity was calculated as the difference between the "microbiota age of a child" and the "microbiota age of healthy children of the same chronologic age" (Subramanian

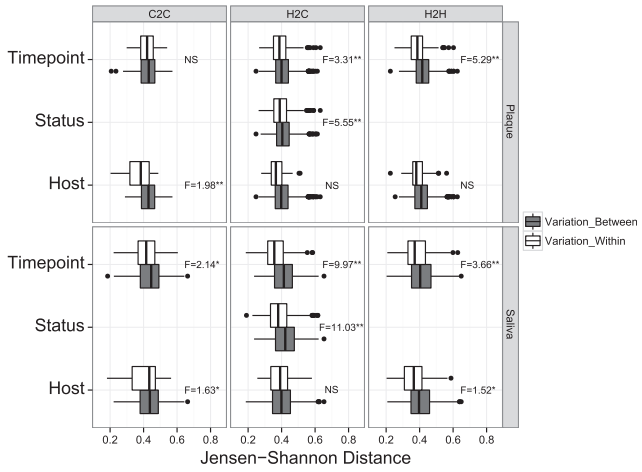


Figure 1. Impact of Chronologic Age of Child Host on Oral Microbiota Variation

Microbiota variation (from plaque or saliva) was compared within and between time points, disease status, or hosts in the three host groups (i.e., H2H, H2C, and C2C). Interpersonal variation in the bacterial component from both oral niches is not higher than variation between either the various time points or the different status. In the "H2H" and "C2C" panels, healthy oral microbiota significantly changed over time, while bacterial composition showed no significant changes during caries progression. In the "H2C" panel, despite the strong variation in bacterial composition between the time points, a high degree of disease-driven variability was also evident in both of the oral niches (i.e., plaque and saliva; *$p < 0.05$; **$p < 0.01$; NS, not significant).

et al., 2014), where the chronologic age was interpolated from 4 to 6 years of life using a spline fit (Figures 2C and 2D).

To probe the effect of caries on microbiota maturation, development of microbiota defined by both overall bacterial diversity and age-discriminatory taxa identified above were thus monitored over the same time frame from the C2C group and the H2C group. Here, the Random Forest model derived from healthy children was used to define the microbiota age of children in both the C2C and the H2C groups. The relative microbiota maturity of both plaque and saliva for children in the C2C and the H2C groups was further compared to those in the H2H group (Figures 2E and 2F).

Interestingly, in the C2C group, the microbiota maturity of the oral microbiota was highly distinct. First, in contrast to the H2H group, plaque diversity is not significantly affected by age (PERMANOVA $p > 0.05$; Figure 1), whereas only a small amount of variance in salivary diversity can be explained by age in the C2C group (PERMANOVA, $p < 0.05$, $F = 2.14$; Figure 1). The effect of age appears to be noisiest in the C2C group, which could be due to the smallest size of this group. Second, microbiota age calculated for the C2C group exhibited an apparent plateau in disease duration, especially when caries symptoms were severe (dmfs > 10; Figures 2C and 2D). Specifically, within-host relative microbiota maturity level of saliva microbiota did not substantially change with age (Friedman test, $p > 0.05$); as for plaque, the within-host relative microbiota maturity level actually decreased with age (Friedman test, $p < 0.05$), which is solely due to differences between the two time points of T1 and T3. Third, compared to children in the H2H group, the C2C group of children exhibited lower relative microbiota maturity (Wilcoxon rank-sum test, $p < 0.05$; Figures 2E and 2F). The exhibited significant microbiota immaturity in children with caries suggests that caries progression may retard or inhibit the otherwise normal oral microbial maturation process.

On the other hand, in the H2C group, for the overall microbial structure, the effect of ECC onset on bacterial diversity (PERMANOVA, $p < 0.05$, $F = 5.55$ in plaque and $p < 0.05$, $F = 11.03$ in saliva; Figure 1) was more important that of age (PERMANOVA, $F = 3.31$ in plaque and $F = 9.97$ in saliva; Figure 1), suggesting that the health-to-caries shift in the oral microbiota can be distinguished from the "normal" variation attributed to host aging. Thus, the age factor should be taken into account when modeling ECC development. Intriguingly, host variation did not affect the bacterial diversity in the H2C group (PERMANOVA, $p > 0.05$; Figure 1), suggesting that oral microbiota can serve as a proxy for tracking ECC onset of human populations in this age range. In addition, the microbiota age of diseased children in the H2C group was slightly lower than those in the H2H group (Wilcoxon rank-sum test, $p < 0.05$; Figures 2E and 2F), although the relative microbiota maturity of all children in the H2C group is equivalent to that in the H2H group (Wilcoxon rank-sum test, $p > 0.05$; Figures 2E and 2F). Therefore, the normal development of the oral microbiota in preschool children is perturbed in ECC.

Microbiota Change at ECC Onset Is Distinct from That during ECC Progression

We further investigated the temporal pattern that healthy hosts acquire caries-like microbiota over time. We here provided an integrated view of the ECC development based on the hosts separately from the H2C and C2C groups. We first quantified the microbial diversity within each subject at a given time point (α diversity) and the difference of oral microbiota between pairs of time points for each subject (β diversity, as represented by Jensen-Shannon metrics). Then the correlation between oral microbiota structure and clinical symptoms (as indicated by dmfs) was calculated in both the H2C and C2C groups.

Intriguingly, during ECC development, the most profound change in the oral microbiota took place at ECC onset rather than during ECC progression. In fact, during ECC progression, α diversity of neither plaque nor saliva microbiota correlated with clinical disease severity (Figure 3A, $p > 0.05$; Spearman correlation, which we term "plateau phase of microbiota change"). However, in both plaque and saliva, the value is strongly correlated with dmfs during ECC onset (Figure 3A, $p < 0.05$, rho = 0.54 in plaque and rho = 0.62 in saliva; Spearman correlation; termed "acute phase of microbiota change"). Likewise, analysis of β diversity of both plaque and saliva microbiota from ECC-progressing hosts also suggested this pattern of a relatively stable diseased microbiota during ECC progression. First, in the C2C group, Jensen-Shannon distance was not correlated with dmfs change and dmfs (Figures 3B and S3, $p > 0.05$, Spearman correlation), suggesting plaque-microbiota development became somehow retarded after the onset of ECC. Second, in the H2C group, although hosts exhibit a lower degree of dmfs changes (i.e., less change in clinical symptoms) than those in the C2C group, microbiota in either plaque or saliva were significantly more temporally correlated with dmfs changes than those in the C2C group (Figures 3B and S3). Taken together, the results revealed the microbiota change at ECC onset is more dramatic

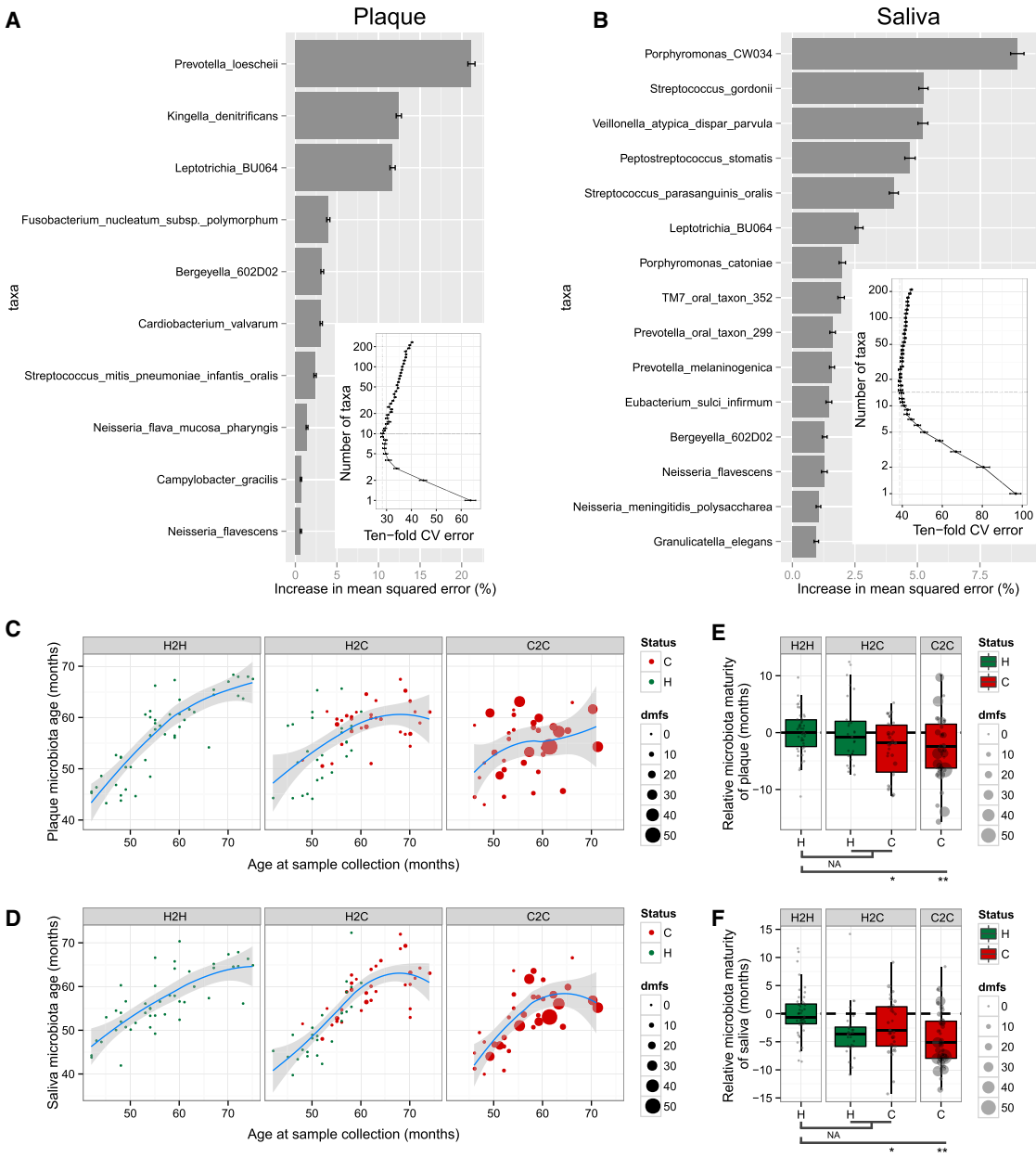


Figure 2. Bacterial Taxonomic Biomarkers for Defining Oral-Microbiota Maturation in Healthy Children

(A and B) The most age-discriminatory bacterial taxa from plaque (A) and saliva (B) were identified by applying Random Forests regression of their relative abundances in oral samples against chronologic age in 17 healthy children. Species are shown in their corresponding niche ranked in descending order of their importance to the accuracy of the model. Importance was determined based on the percentage increase in mean-squared error of microbiota age prediction when the relative abundance values of each taxon were randomly permuted (mean importance ± SD, n = 100 replicates). The inset of (A) and (B) shows 10-fold cross-validation error as a function of the number of input species-level taxa used to regress against the age of children in the training set, in order of variable importance. (C and D) The microbiota age predictions (green circles, each circle represents an individual oral sample) in the healthy children of the H2H group based on plaque (C) and saliva (D). The trained model was subsequently applied to two sets of unhealthy children in the H2C and C2C groups (red circles, represents an individual oral sample from a diseased host). The curve is a smoothed spline fit between microbiota age and chronologic age in both training and test sets (right two panels of [C] and [D]).

(E and F) The relative microbiota maturity of plaque (E) and saliva (F) were defined in children of the H2H, H2C, and C2C groups (*$p < 0.05$; **$p < 0.01$; NS, not significant).

than that during ECC progression, which thus can potentially be exploited for microbiota-based disease prediction.

Intriguingly, this temporal pattern of microbiota development during disease onset and progression was highly distinct from what we observed previously in gingivitis, which together with caries is the most prevalent plaque-dependent oral disease. The longitudinal dataset for gingivitis, generated from our group with a similar sequencing strategy, consists of 150 plaque

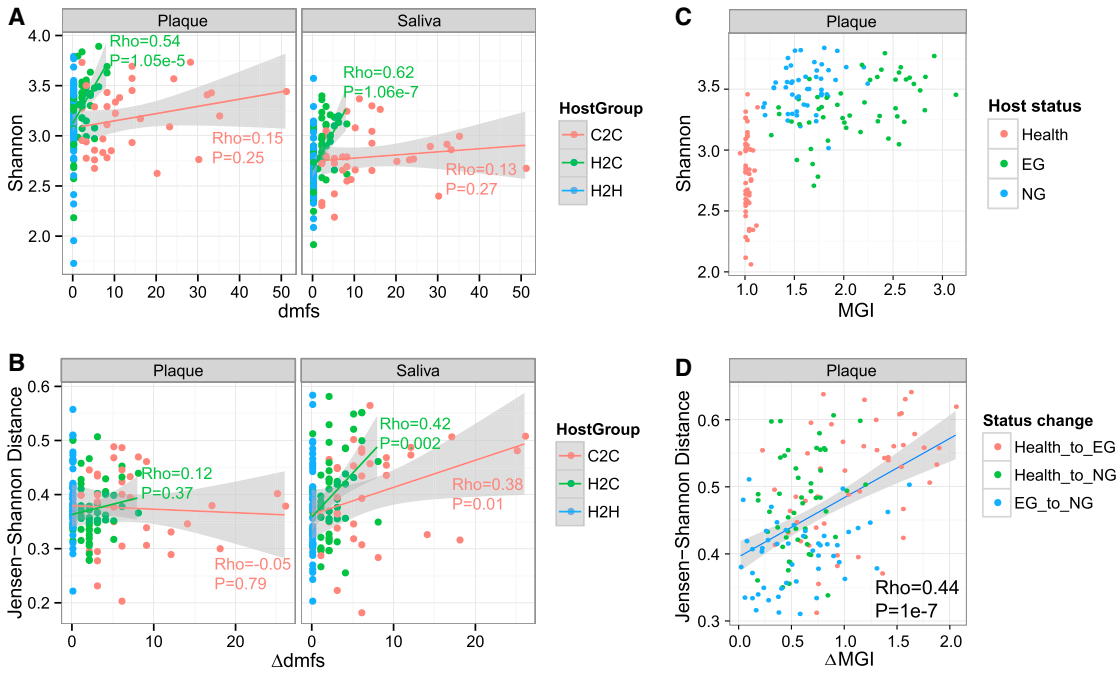


Figure 3. Correlation between Bacterial Diversity and Clinical Symptoms in ECC and in Gingivitis, Respectively

(A) In ECC, the α diversity in both plaque and saliva microbiota significantly changed with dmfs during caries onset (the H2C group), whereas it exhibited no correlation with dmfs in caries progression (C2C group).

(B) In ECC, plaque microbiota (β diversity) showed no correlation with caries progression (i.e., dmfs change within hosts) in the C2C group.

(C) In the gingivitis study where temporal changes of both clinical symptom and microbiota were tracked during gingivitis retrogression and progression, the Shannon index change substantially with MGI.

(D) In contrast to ECC, at all of the levels of gingivitis severity tested, plaque microbiota (β diversity) were significantly correlated with gingivitis progression and retrogression (i.e., the within-host MGI change).

microbiota from fifty adults that underwent controlled transitions from naturally occurring gingivitis to healthy gingiva and then to experimental gingivitis (Huang et al., 2014). In contrast to the temporal pattern that links symptom and plaque microbiota in ECC development, the degree of within-subject change in gingivitis symptom was highly correlated with the extent of plaque microbiota change during gingivitis development (Figures 3C and 3D, $p < 0.05$, Spearman correlation). Thus, the temporal pattern of microbiota variation during caries development is highly distinct from that during gingivitis development, although temporal changes in the oral microbiota can be associated with onset and progression of both diseases. Specifically, we observed a plateau of microbial development as hosts advanced to severe ECC, while no such plateau was observed during gingivitis progression, which instead featured gradual and continuous microbiota change at a largely consistent rate throughout the disease course (Figure 3D). The distinct temporal pattern of microbiota between ECC and gingivitis, both plaque-induced oral diseases, suggests two types of microbiota ecological succession in pathogenesis and raises the possibility of exploiting such temporal characteristics of microbiota for predictive modeling of disease courses.

Predictive Modeling of ECC Onset Using Plaque and Saliva Microbiota

The substantial temporal changes of oral microbiota at ECC onset raise the question of whether substantial alteration of oral microbiota occurs rapidly at a switching point to caries onset or if a gradual progressive dysbiosis eventually results in caries. We reasoned that if the oral microbiota gradually shift to a disease state, then "healthy" microbiota sampled prior to disease onset (i.e., "RelativeH"; Table 1) would reflect dysbiosis before the physical symptoms of caries. Interestingly, in the H2C group, the microbiota shift from ConfidentH to caries is greater than that from RelativeH to caries (Figure S4A). Furthermore, the oral microbiota of RelativeH resembles that of both ConfidentH and caries (Figure S4B). These results suggested that RelativeH microbiota is an intermediate state indicating a high risk of future disease outcomes. To test the hypothesis, we trained a random forest model in a training set of the "ConfidentH" and caries samples (208 samples in total from plaque and saliva; Table 1) using the profile of taxa at the six different phylogenetic levels (from phylum to species) and then derived a microbial indicator of caries (MiC) (Experimental Procedures). To probe how spatial variation of microbiota affects model performance, models were built from plaque microbiota alone, saliva microbiota alone, and both plaque and saliva microbiota (the composite model), respectively. Model performance was evaluated using a 10-fold cross-validation approach, and the predictive power was scored in a receiver-operating characteristic (ROC) analysis (Experimental Procedures; Supplemental Information). The discriminatory power was calculated as the ROC curve (AUC).

To construct an ECC diagnosis model, we first excluded taxa highly correlated with age in healthy children from the training

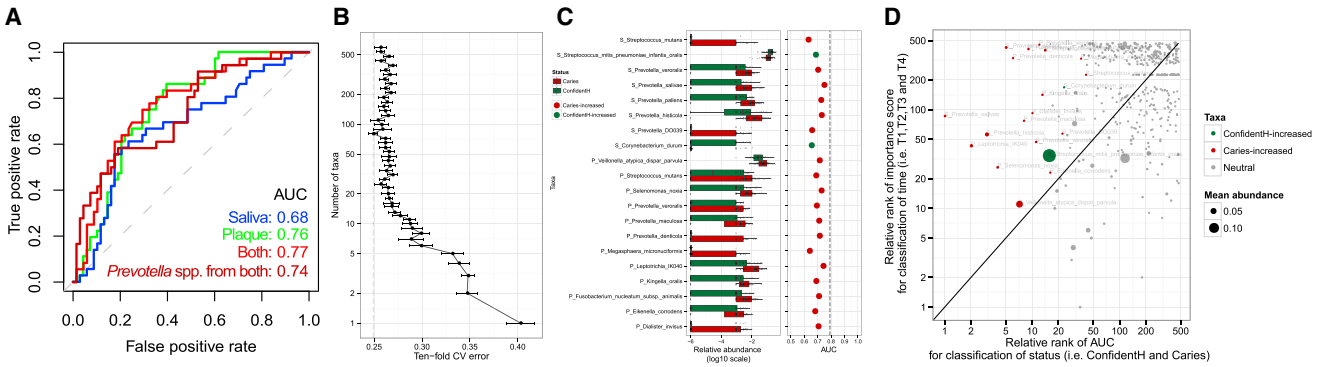


Figure 4. Disease Classification Based on Oral Microbiota Profiles

(A) Classification performance of Random Forest model using species profiles of plaque, saliva, and both, assessed by area under the ROC. From both plaque and saliva, we next collected 20 key drivers of the microbial dysbiosis that are associated with ECC status.

(B) Relationship between the numbers of variables included in MiC and the corresponding predictive performance (the error bar denotes SD).

(C) The 20 most discriminant species in predictive model were showed as boxplot. Each row indicates the log10-transformed abundance of each predictor (i.e., bacterial taxon) from either plaque or saliva microbiota. The utility of each taxon as a potential caries marker is assessed by the area under the ROC curve (AUC). Samples are colored by status: health (green; i.e., "ConfidentH") or caries (red). Boxes represent the interquartile range (IQR), and the lines inside represent the median. Whiskers denote the lowest and highest values within 1.5× IQR.

(D) The contributions of the 20 most discriminant species to microbiota-based classification of host status and time points.

datasets. Then to test whether model performance depends on the taxonomic level analyzed, AUC of the classifiers was estimated at each of the six levels from phylum to species (Experimental Procedures). Use of species-level taxa maximized AUC within each niche; moreover, at each of the levels (except Genus), the best performance was obtained when both niches were used (followed closely by use of plaque samples alone; Figure S4C). Thus, the dual-niche-based species-level model, which exhibited the best performance, was selected as the final MiC (AUC = 0.77; Figures 4A, S4D, and S4E). Discriminatory power of the model was then validated via saliva microbiota of an independent, 40-member, 46- to 50-month-old cohort (20 healthy children and 20 with severe ECC; Table S1). The AUC was 0.72 (Figure S4F), supporting the predictive value of our Random Forests classifier for ECC status.

In the Random Forest model using microbiota profiles from both plaque and saliva, improvement in predictive performance (estimated by 10-fold cross-validation) was minimal once the top 20 most discriminatory taxa associated with caries state were included, showing that these 20 had the most discriminatory power (12 from plaque and eight from saliva; Figure 4B). These 20 taxa include *Streptococcus mutans* and *Veillonella atypical/Veillonella dispar/Veillonella parvula*, which have previously been documented to play key roles in caries pathogenesis. Consequently, caries onset is not associated with a single taxon but in fact with a complex community. Remarkably, in both saliva and plaque, the caries-enriched microbes largely consisted of *Prevotella* spp. (Figure 4C), consistent with recent observations in caries-active children (Gomar-Vercher et al., 2014) and caries-active adults (Yang et al., 2012). Moreover, using only the eight marker *Prevotella* species as predictors, the simplified Random Forest model exhibited predictive performance (AUC = 0.74; Figure 4A) already comparable to the performance of the model derived from the whole microbiota (AUC = 0.77; Figure 4A). Furthermore, *Prevotella* members do not show correlation with age in the H2H group ($p > 0.05$, Spearman correlation; Figure 4D), suggesting that they can detect ECC onset and are not affected by age.

While patients are not diagnosed with caries until clinical symptoms are apparent, we can test whether the microbial dysbiosis takes place earlier by employing the model to predict whether a subject with RelativeH microbiota would develop to caries (i.e., estimate the risk of future disease outcome). The MiC model was applied to the 42 RelativeH samples (21 samples in plaque and 21 in saliva) (Table 1). In plaque, 17 (81% given the cutoff probability for classification is 50%) were correctly predicted as caries, congruent with clinical symptoms arising in the subsequent sampling time point, whereas only three samples were classified as ConfidentH state (one of the 42 failed to be classified as its probability of being either caries or ConfidentH was equal; i.e., at 0.50; Figure 5A). Intriguingly, plaque MiC identified more pre-disease samples as caries state than both saliva (11 samples were correctly predicted) and composite MiC (15 correctly predicted), suggesting higher accuracy of plaque than saliva for predicting ECC onset (Figure 5A). The abundance of species markers in *Prevotella* genus was significantly higher in the subgroup identified as caries than that classified as ConfidentH (Wilcoxon rank-sum test, $p = 0.002$; Figure 5B). In comparison, no significant difference in the abundance of *Streptococcus mutans* was observed between the two MiC-stratified subgroups (Figure 5C). The crucial role of these marker *Prevotella* spp. in classification and prediction of caries status was further supported by the predicted functional profiles of the microbiota (Supplemental Information; Figure S5A), as these *Prevotella* species contribute greatly to the activity of carbohydrate-derived acid production (i.e., one of the most prominent microbial functions associated with caries status in our dataset; Figure S5B). These findings, together with observations that the effect sizes of both caries and gingivitis are greater than host variation in oral microbiota (Supplemental Information; Figure S1E), underscore the advantages and potential of oral microbiota as disease biomarkers in human populations.

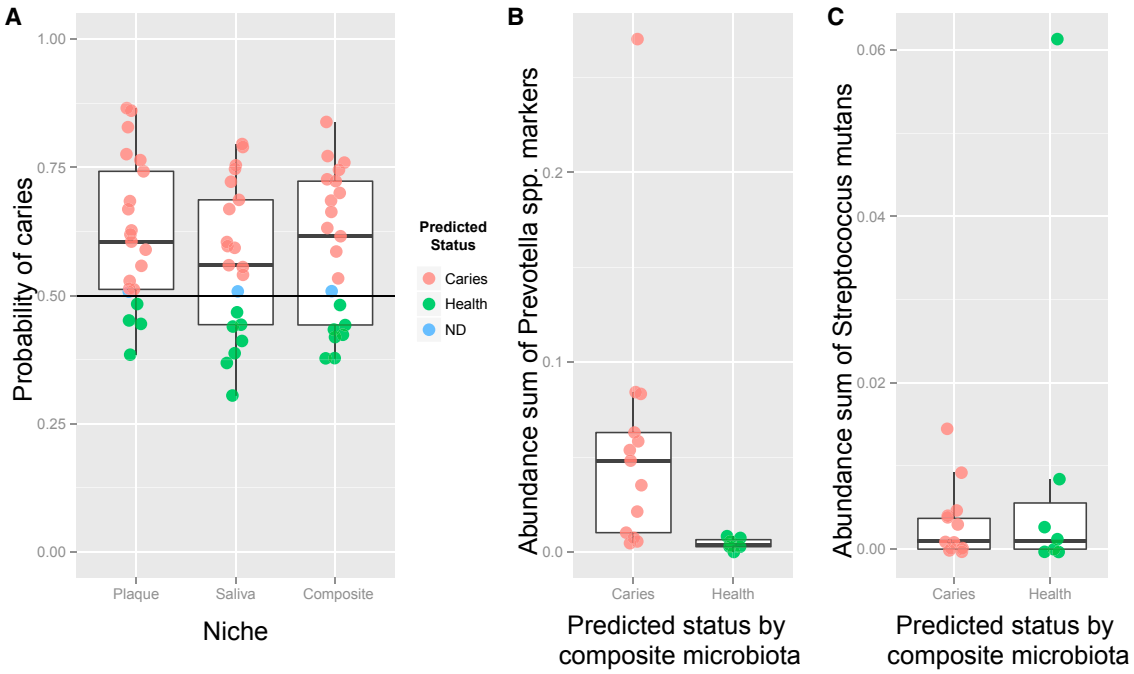


Figure 5. Stratification of RelativeH Children Based on Oral Microbiota Profile

(A) Use of the species-based MiC trained for discriminating ConfidentH and caries microbiota to stratify pre-disease samples.

(B) Those RelativeH children predicted to be caries had higher abundance of *Prevotella* spp. markers ($p = 0.002$, Wilcoxon rank-sum test).

(C) *Streptococcus mutans*, which was frequently considered as caries pathogen, showed no significant difference in abundance between predicted status ($p > 0.05$, Wilcoxon rank-sum test). Boxes denote the IQR between the first and third quartiles, and the line within denotes the median; whiskers denote the lowest and highest values within 1.5 times IQR from the first and third quartiles, respectively.

DISCUSSION

Method development for early diagnosis and risk assessment of future new ECC onsets are both significant and urgent (Lancet, 2009; Selwitz et al., 2007). Here we developed a method, called MiC, for predicting future new ECC onsets based on oral microbiota. As the collection of plaque or saliva is rapid and non-invasive, and can be readily performed at home, MiC can potentially serve as an objective, sensitive, and patient-friendly measure of ECC susceptibility and contribute to preventive intervention of ECC and cross-study evaluation of oral care products in children populations.

Evidence has recently emerged that changes in the human microbiota continue through human life (Song et al., 2013; Stahringer et al., 2012; Xu et al., 2015). The development of the oral microbiota involved profound alterations in diversity and composition that took place not only over the first three postnatal years of life (Song et al., 2013; Subramanian et al., 2014), but also in different ages (i.e., childhood, adolescence, adulthood, and old age) (Crielaard et al., 2011; Xu et al., 2015). However, to date, the temporal changes in the healthy oral microbiota are ill defined in preschoolers (nominally, 4–6 years in age), which is the most vulnerable stage in developing caries (Lancet, 2009). In comparing the temporal dynamics of the oral microbiota in healthy and ECC children, we showed that plaque and saliva microbiota age of unrelated healthy children can explain over 50% of the variance in microbiota. The age-dependent changes observed could be caused by biological changes in tissues around the teeth during exfoliation of teeth (Crielaard et al., 2011), contact with external microbes (Könönen, 2000), and/or development of immune systems (Costello et al., 2012), etc. Furthermore, the age-associated development of the healthy oral microbiota appears to be driven in part by changes in the relative abundance of taxa rather than the acquisition/loss of unique taxa with age, similar to the development of the skin microbiota at different ages (Song et al., 2013) but different from the gut microbiota at the age of 0–2 (Subramanian et al., 2014). These may suggest the importance of defining microbiota maturation at different age segments and in each body habitats. Microbiome maturation may serve as a microbial measure of childhood development as well as a microbial baseline to define susceptibility, onset, and progression of a range of diseases.

In ECC hosts, the general age-dependent development of oral microbiota was perturbed by disease, which was also observed in gut microbiota (e.g., fecal microbiota of Kwashiorkor-afflicted children failed to develop with increasing age) (Smith et al., 2013). Notably, although here health-associated taxa were used to define age of disease-associated microbiota maturation (Kostic et al., 2015; Subramanian et al., 2014), this strategy does not exclude the possibility that maturation of oral microbiota in those children predisposed to disease might actually entail an entirely different group of taxa. Given the pronounced microbiota alterations over time in healthy children, the top age discriminatory features were excluded from the microbial biomarkers of ECC onset (i.e., MiC), as their variation in the healthy children can be mistakenly perceived as that observed in children that

are entering ECC onset. Thus, defining temporal variations within the healthy populations is crucial before drawing robust conclusions based on observed microbiota dysbiosis associated with disease onset and development.

Intriguingly, the microbial diversity (i.e., α and β diversity) varied during the course of this chronic illness. In the onset phase (when host symptoms were not yet detectable), microbial diversity significantly increased with ECC severity (i.e., dmfs) and its change (i.e., Δdmfs); however, it tended to be stable during caries progression (when host symptom became detectable). Overabundance of certain taxa contributed to the diversity switch from healthy to caries, including *Streptococcus* spp., *Prevotella* spp., and *Veillonella* spp., all commensal bacteria that can turn pathogenic. Such changes in organismal signatures were correlated with a shift in abundance of certain metabolic pathways such as carbohydrate and amino acid metabolisms. Interestingly, in gingivitis, also a plaque-dependent disease, the temporal pattern of microbiota is distinct from that of ECC, with the former exhibiting gradual and continuous diversity change at a largely consistent rate throughout disease course. Therefore, several key lessons are apparent. (i) Definition of age-dependent variation in both healthy and diseased populations is important. (ii) Tracking temporal dynamics of the microbiota along full disease course that includes both prior to and after disease-onset phases may reveal phase-specific microbiota features that can be exploited for preventive intervention. (iii) Comparison of temporal microbiota development patterns across different plaque-based diseases may provide novel insights into disease etiology and intervention strategy.

Furthermore, spatial heterogeneity of oral microbiota in healthy humans has been documented (Huttenhower et al., 2012), but its link to microbiota-based disease modeling has been rarely tested (Gevers et al., 2014). Several features of this link were revealed here. First, the model incorporating plaque microbial data is superior to the model based only on salivary data. This is consistent with plaque providing more biomarkers in the MiC than saliva. Second, salivary microbiota is also of value for predictive models and has considerable practical advantages as a sampling site (due to non-invasiveness and patient compliance). Although the taxon of *Prevotella* spp. was shared by both niches, the ECC-discriminatory taxa in saliva were mostly distinct from those in plaque, and saliva offers more functional markers than plaque. This finding is in agreement with the generally accepted notion that microbiota attached to the teeth and soft issue surfaces are continuously shed into the saliva, making saliva a reservoir of the overall oral microbiota (Parisotto et al., 2010). Third, the composite MiC outperformed the single-niche-based MiC in accuracy: incorporating both the lesion niche (plaque) and the "cosmopolitan" niche (saliva) could provide the most "informative" microbiota for modeling oral infections. Therefore, for microbiota-based modeling of infections, (i) not every site is equal in contribution, (ii) more sites should bring performance improvement, and (iii) comparing microbiota dynamics of among distinct niches may provide novel insights into perturbation/intervention of disease onset.

On the other hand, the ability of whole-mouth sampling data to distinguish and predict host ECC states appears a surprise, as microbiota maturation and onset of caries disease can both be site specific. Within the same individual, difference in microbial composition can be present among distinct healthy tooth surfaces (Simón-Soro et al., 2013), between a carious lesion and non-carious lesion, and between dentinal carious lesions and enamel lesions (Simón-Soro et al., 2014). However, multiple reports have suggested caries affect the microbiota of not just the dentition sites but also the other apparently healthy teeth for a given individual (Gross et al., 2010, 2012; Jiang et al., 2013, 2014). In our cohort, the difference between an H2H individual and an H2C individual is typically of multiple carious lesions (i.e., 3.2 ± 1.8 lesions; mean ± SD); thus, it is possible that oral microbiota as a whole have changed significantly. Follow-up studies are warranted that sample microbiota based on individual dentition sites and individual teeth.

Here we report that the spatial and temporal variations of microbiota can be exploited for predictive modeling of plaque-associated diseases. Notably, we can predict the onset of caries in apparently healthy teeth using the microbial community composition, when samples are appropriately de-trended for age. In providing an objective strategy for predicting new ECC onsets and showing that shifts in microbiota can precede manifestation of clinical symptoms, our work suggests that monitoring "oral-microbiota age" in children populations, based on an age-stratified reference of oral microbiota in healthy children, might be warranted. Despite the challenges associated with tracking microbiota from a sufficient number of ECC hosts starting from their pre-disease stage, future surveys with higher sampling frequency and larger child cohorts are needed to refine the landscape of key microbiota events along ECC onset and development. Moreover, comparison of microbiota dynamics among different age segments of caries-susceptible hosts should allow probing whether MiC are possible only in children or are generally applicable to populations at different ages. Finally, our study underscores the possibility of employing microbiota for preventive disease intervention, which can be probed for other types of oral infections and for infections in other body sites.

EXPERIMENTAL PROCEDURES

Study Design and Sample Collection

The 50 preschool children were recruited from the same kindergarten using a protocol approved by ethics committee of the Stomatology Hospital, Sun Yat-sen University between June 2011 and June 2013. The participants were all unrelated individuals of both genders, aged around 4 years old at baseline (June 2011), and shared a relatively homogeneous living environment throughout the study. As part of our study design, these children were all from the so-called "All-Day Care Kindergarten" in China, where they live in the kindergarten all day long, 5 days a week, with planned, trackable, and regular menus for their daily meal consumption. Therefore, these children have undergone a relatively homogeneous and consistent candy/carbonated drink consumption, dental visits, diets, and oral health habit. Specifically, the children underwent consistent and similar levels of tooth brushing (i.e., twice daily in the days leading up the sample collection), and this habit was maintained along the whole duration of this study. Those who had antibiotics intake for the preceding at least 3 months or other oral or systemic diseases were excluded. The examinations and assessments of children's caries and oral sample collection were all conducted by professional dentists during the four visits to the kindergarten within 2 years. Finally, all 50 subjects were divided into three types of host groups (i.e., H2H, H2C, and C2C) according to longitudinal clinical status (Table 1). For validation of the diagnosis model, an additional independent cohort consisting of 40 children of 46 to 50 months old (20 healthy children and 20 children with severe ECC; Table S1) were recruited for analysis of saliva microbiota.

DNA Extraction and Sequencing of Supragingival Plaque and Saliva

Total DNA was extracted from dental plaque and saliva, respectively, from each host. Barcoded 16S rRNA amplicons (V1-V3 hypervariable region; Huang et al., 2014) of all samples were sequenced on using Roche 454 FLX Titanium chemistry. Pyrosequencing data were analyzed using scripts from MOTHUR (Schloss et al., 2009), QIIME (Caporaso et al., 2010), and customized R scripts. All raw sequences were deposited at Sequence Read Archive under Accession ID SRP040945 and SRP040947.

Predictive Modeling of ECC Onset

For calculation of oral microbiota age, the relative abundance profile of all species-level taxa in healthy samples (H2H group) was fit against its corresponding chronologic age (months) using default parameters in the "randomForest" package in R (3.1.1). The 25 top-ranking important age-discriminatory taxa led to reasonably good fit from either of the niches were identified based on "rfcv" function in the randomForest package. Random Forests models were then trained to identify disease status in the training set that included samples from the "ConfidentH" and the "Caries" groups using the taxonomy profiles, which were then assessed by area under the ROC. The models were termed "MiC" (microbial indicator of caries). To construct and optimize the MiC, we first tested how taxonomical levels, oral niches, and host aging influenced the AUC of MiC. Using the profiles of species, the performance of models based on microbiota from different oral niches was further evaluated with a 10-fold cross-validation approach. According to "rfcv" function in the randomForest package, the 20 top-ranking important taxa from either of the niches led to reasonably good classification of ECC status. The Random Forests models trained as above were applied to stratify the "RelativeH" samples.

Full description of methodology is provided in the Supplemental Information.

ACCESSION NUMBERS

The accession numbers for all raw sequences reported in this paper are National Center for Biotechnology Information (NCBI) Sequence Read Archive: SRP040945 and SRP040947.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, one table, one data file, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2015.08.005.

AUTHOR CONTRIBUTIONS

F.T., F.Y., J.L., and J.X. envisioned the study. F.T. and F.Y. designed experiments and collected clinical samples. F.T. and C.B. prepared and sequenced samples. F.T., S.H., F.Y., Z.Z.X., A.A., R.K., and J.X. analyzed data. J.X., F.T., S.H., F.Y., and R.K. wrote the paper.

ACKNOWLEDGMENTS

We thank Xiaowei Zeng and Jie Chen for technical support and Yuejiao Zhang for assistance with oral sample collection. This work was funded by Grants 2012AA02A707 and 2012BAI07B03 from the Ministry of Science and Technology of China; 31425002, 31300424, 31400089, and 81430011 from National Natural Science Foundation of China; 20120171110060 from Research Fund for Doctoral Programs in Higher Education; and SKLOD2015OF07 from State Key Laboratory of Oral Diseases. The authors declare no conflicts of interest.

Received: April 10, 2015
Revised: June 29, 2015
Accepted: August 14, 2015
Published: September 9, 2015

REFERENCES

Anaise, J.Z. (1984). Measurement of dental caries experience–modification of the DMFT index. Community Dent. Oral Epidemiol. *12*, 43–46.

Belda-Ferre, P., Alcaraz, L.D., Cabrera-Rubio, R., Romero, H., Simón-Soro, A., Pignatelli, M., and Mira, A. (2012). The oral metagenome in health and disease. ISME J. *6*, 46–56.

Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Casamassimo, P.S., Thikkurissy, S., Edelstein, B.L., and Maiorini, E. (2009). Beyond the dmft: the human and economic cost of early childhood caries. J. Am. Dent. Assoc. *140*, 650–657.

Chen, X., Clark, J.J., and Naorungroj, S. (2012). Length of tooth survival in older adults with complex medical, functional and dental backgrounds. J. Am. Dent. Assoc. *143*, 566–578.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science *326*, 1694–1697.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J.M., and Relman, D.A. (2012). The application of ecological theory toward an understanding of the human microbiome. Science *336*, 1255–1262.

Crielaard, W., Zaura, E., Schuller, A.A., Huse, S.M., Montijn, R.C., and Keijser, B.J. (2011). Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. BMC Med. Genomics *4*, 22.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe *15*, 382–392.

Gomar-Vercher, S., Cabrera-Rubio, R., Mira, A., Montiel-Company, J.M., and Almerich-Silla, J.M. (2014). Relationship of children's salivary microbiota with their caries status: a pyrosequencing study. Clin. Oral Investig. *18*, 2087–2094.

Gross, E.L., Leys, E.J., Gasparovich, S.R., Firestone, N.D., Schwartzbaum, J.A., Janies, D.A., Asnani, K., and Griffen, A.L. (2010). Bacterial 16S sequence analysis of severe caries in young permanent teeth. J. Clin. Microbiol. *48*, 4121–4128.

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.

Ha, D.H., Spencer, A.J., Slade, G.D., and Chartier, A.D. (2014). The accuracy of caries risk assessment in children attending South Australian School Dental Service: a longitudinal study. BMJ Open *4*, e004311.

Hallett, K.B. (2013). The application of caries risk assessment in minimum intervention dentistry. Aust. Dent. J. *58* (*Suppl 1*), 26–34.

Huang, S., Yang, F., Zeng, X., Chen, J., Li, R., Wen, T., Li, C., Wei, W., Liu, J., Chen, L., et al. (2011). Preliminary characterization of the oral microbiota of Chinese adults with and without gingivitis. BMC Oral Health *11*, 33.

Huang, S., Li, R., Zeng, X., He, T., Zhao, H., Chang, A., Bo, C., Chen, J., Yang, F., Knight, R., et al. (2014). Predictive modeling of gingivitis severity and susceptibility via oral microbiota. ISME J. *8*, 1768–1780.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al.; Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Jiang, W., Jiang, Y., Li, C., and Liang, J. (2011). Investigation of supragingival plaque microbiota in different caries status of Chinese preschool children by denaturing gradient gel electrophoresis. Microb. Ecol. *61*, 342–352.

Jiang, W., Zhang, J., and Chen, H. (2013). Pyrosequencing analysis of oral microbiota in children with severe early childhood dental caries. Curr. Microbiol. *67*, 537–542.

Jiang, W., Ling, Z., Lin, X., Chen, Y., Zhang, J., Yu, J., Xiang, C., and Chen, H. (2014). Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood. Microb. Ecol. *67*, 962–969.

Knights, D., Parfrey, L.W., Zaneveld, J., Lozupone, C., and Knight, R. (2011). Human-associated microbial signatures: examining their predictive value. Cell Host Microbe *10*, 292–296.

Könönen, E. (2000). Development of oral bacterial flora in young children. Ann. Med. *32*, 107–112.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al.; DIABIMMUNE Study Group (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260–273.

Lancet (2009). Oral health: prevention is key. Lancet *373*, http://dx.doi.org/10.1016/S0140-6736(08)61933-9.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., et al.; MetaHIT consortium (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Leroy, R., and Declerck, D. (2013). Impact of caries onset on number and distribution of new lesions in preschool children. Int. J. Paediatr. Dent. *23*, 39–47.

Ling, Z., Kong, J., Jia, P., Wei, C., Wang, Y., Pan, Z., Huang, W., Li, L., Chen, H., and Xiang, C. (2010). Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. Microb. Ecol. *60*, 677–690.

Ling, Z., Liu, X., Wang, Y., Li, L., and Xiang, C. (2013). Pyrosequencing analysis of the salivary microbiota of healthy Chinese children and adults. Microb. Ecol. *65*, 487–495.

Mejàre, I., Axelsson, S., Dahlén, G., Espelid, I., Norlund, A., Tranæus, S., and Twetman, S. (2014). Caries risk assessment. A systematic review. Acta Odontol. Scand. *72*, 81–91.

Parisotto, T.M., Steiner-Oliveira, C., Silva, C.M., Rodrigues, L.K., and Nobre-dos-Santos, M. (2010). Early childhood caries and mutans streptococci: a systematic review. Oral Health Prev. Dent. *8*, 59–70.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. Nature *513*, 59–64.

Russell, R.R. (2009). Changing concepts in caries microbiology. Am. J. Dent. *22*, 304–310.

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.

Selwitz, R.H., Ismail, A.I., and Pitts, N.B. (2007). Dental caries. Lancet *369*, 51–59.

Simón-Soro, A., Tomás, I., Cabrera-Rubio, R., Catalan, M.D., Nyvad, B., and Mira, A. (2013). Microbial geography of the oral cavity. J. Dent. Res. *92*, 616–621.

Simón-Soro, A., Guillen-Navarro, M., and Mira, A. (2014). Metatranscriptomics reveals overall active bacterial composition in caries lesions. J. Oral Microbiol. *6*, 25443.

Smith, M.I., Yatsunenko, T., Manary, M.J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A.L., Rich, S.S., Concannon, P., Mychaleckyj, J.C., et al. (2013). Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science *339*, 548–554.

Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D., Caporaso, J.G., Knights, D., Clemente, J.C., Nakielny, S., et al. (2013). Cohabiting family members share microbiota with one another and with their dogs. eLife *2*, e00458.

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.

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., et al. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature *510*, 417–421.

Tellez, M., Gomez, J., Pretty, I., Ellwood, R., and Ismail, A.I. (2013). Evidence on existing caries risk assessment systems: are they predictive of future caries? Community Dent. Oral Epidemiol. *41*, 67–78.

Xu, X., He, J., Xue, J., Wang, Y., Li, K., Zhang, K., Guo, Q., Liu, X., Zhou, Y., Cheng, L., et al. (2015). Oral cavity contains distinct niches with dynamic microbial communities. Environ. Microbiol. *17*, 699–710.

Yang, F., Zeng, X., Ning, K., Liu, K.L., Lo, C.C., Wang, W., Chen, J., Wang, D., Huang, R., Chang, X., et al. (2012). Saliva microbiomes distinguish caries-active from healthy human populations. ISME J. *6*, 1–10.

Yang, F., Ning, K., Chang, X., Yuan, X., Tu, Q., Yuan, T., Deng, Y., Hemme, C.L., Van Nostrand, J., Cui, X., et al. (2014). Saliva microbiota carry caries-specific functional gene signatures. PLoS ONE *9*, e76458.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.