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

Fei Teng,1,2,3 Fang Yang,3,5 Shi Huang,2,5 Cunpei Bo,2 Zhenjiang Zech Xu,4 Amnon Amir,4 Rob Knight,4 Junqi Ling,1,6 and Jian Xu1,2,*

1Department of Operative Dentistry and Endodontics, Guanghua School and Hospital of Stomatology, Sun Yat-sen University, Guangzhou, 510055, China
2Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, 266101, China
3Department of Stomatology, Qingdao Municipal Hospital, Qingdao, Shandong, 266101 China
4Departments of Pediatrics and Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093, USA
5Co-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 (Meijare 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 (Meijare 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 to both 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) (Anaïse, 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 ≤ 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
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

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>K1012</td>
<td>C2C</td>
<td>47</td>
<td>3^a</td>
<td>5^a</td>
<td>8^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1014</td>
<td></td>
<td>53</td>
<td>7^a</td>
<td>10^a</td>
<td>16^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1017</td>
<td></td>
<td>48</td>
<td>3^a</td>
<td>8^a</td>
<td>9^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1020</td>
<td></td>
<td>51</td>
<td>23^a</td>
<td>30^a</td>
<td>35^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1021</td>
<td></td>
<td>48</td>
<td>2^a</td>
<td>5^a</td>
<td>8^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1025</td>
<td></td>
<td>46</td>
<td>5^a</td>
<td>7^a</td>
<td>10^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1028</td>
<td></td>
<td>49</td>
<td>26^a</td>
<td>33^a</td>
<td>51^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1030</td>
<td></td>
<td>52</td>
<td>5^a</td>
<td>8^a</td>
<td>14^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1035</td>
<td></td>
<td>50</td>
<td>14^a</td>
<td>14^a</td>
<td>20^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1024</td>
<td></td>
<td>46</td>
<td>2^a</td>
<td>11^a</td>
<td>14^a</td>
<td>28^a</td>
<td></td>
</tr>
<tr>
<td>K1041</td>
<td></td>
<td>59</td>
<td>–</td>
<td>–</td>
<td>24^a</td>
<td>32^a</td>
<td></td>
</tr>
<tr>
<td>K1056</td>
<td></td>
<td>54</td>
<td>–</td>
<td>2^a</td>
<td>3^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1001</td>
<td>H2C</td>
<td>42</td>
<td>0^b</td>
<td>1^a</td>
<td>2^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1002</td>
<td></td>
<td>52</td>
<td>0^b</td>
<td>3^a</td>
<td>4^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1004</td>
<td></td>
<td>50</td>
<td>0^b</td>
<td>3^a</td>
<td>3^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1006</td>
<td></td>
<td>49</td>
<td>0^b</td>
<td>8^a</td>
<td>8^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1023</td>
<td></td>
<td>49</td>
<td>0^b</td>
<td>3^a</td>
<td>5^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1003</td>
<td></td>
<td>50</td>
<td>0^b</td>
<td>2^a</td>
<td>3^a</td>
<td>4^a</td>
<td></td>
</tr>
<tr>
<td>K1005</td>
<td></td>
<td>46</td>
<td>0^b</td>
<td>5^a</td>
<td>5^a</td>
<td>6^a</td>
<td></td>
</tr>
<tr>
<td>K1026</td>
<td></td>
<td>47</td>
<td>0^b</td>
<td>1^a</td>
<td>1^a</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1008</td>
<td></td>
<td>48</td>
<td>0^c</td>
<td>0^c</td>
<td>2^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1009</td>
<td></td>
<td>51</td>
<td>0^c</td>
<td>0^c</td>
<td>2^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1011</td>
<td></td>
<td>50</td>
<td>0^c</td>
<td>0^c</td>
<td>3^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1016</td>
<td></td>
<td>51</td>
<td>0^c</td>
<td>0^c</td>
<td>2^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1013</td>
<td></td>
<td>46</td>
<td>0^c</td>
<td>0^c</td>
<td>4^a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1007</td>
<td></td>
<td>45</td>
<td>0^c</td>
<td>0^c</td>
<td>4^a</td>
<td>6^a</td>
<td></td>
</tr>
<tr>
<td>K1040</td>
<td></td>
<td>60</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1042</td>
<td></td>
<td>58</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1043</td>
<td></td>
<td>58</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1054</td>
<td></td>
<td>57</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1059</td>
<td></td>
<td>56</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1060</td>
<td></td>
<td>58</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1062</td>
<td></td>
<td>56</td>
<td>–</td>
<td>–</td>
<td>0^b</td>
<td>2^a</td>
<td></td>
</tr>
<tr>
<td>K1015</td>
<td>H2H</td>
<td>44</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1018</td>
<td></td>
<td>47</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1019</td>
<td></td>
<td>43</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1022</td>
<td></td>
<td>42</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1027</td>
<td></td>
<td>48</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1029</td>
<td></td>
<td>47</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1031</td>
<td></td>
<td>45</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1032</td>
<td></td>
<td>42</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1033</td>
<td></td>
<td>48</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1034</td>
<td></td>
<td>49</td>
<td>0^c</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1044</td>
<td></td>
<td>59</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1045</td>
<td></td>
<td>55</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

aCaries.  
bRelative Health.  
cConfident Health.

Table 1. Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>K1049</td>
<td></td>
<td>60</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td>0^c</td>
<td></td>
</tr>
<tr>
<td>K1052</td>
<td></td>
<td>62</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1053</td>
<td></td>
<td>63</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1061</td>
<td></td>
<td>58</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1063</td>
<td></td>
<td>55</td>
<td>–</td>
<td>–</td>
<td>0^c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Third, compared to children in the H2H group, the C2C group due to differences between the two time points of T1 and T3. The relative microbiota maturity level actually change with age (Friedman test, p > 0.05); as for plaque, microbiota maturity of saliva microbiota did not substantially change with age (Friedman 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, 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, ρ = 0.54 in plaque and ρ = 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.
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 samples.

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).
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
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). Disciminatory 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 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.
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; Stahninger 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

![Figure 5. Stratification of RelativeH Children Based on Oral Microbiota Profile](image-url)

(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.
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., z and 
 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 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.

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 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.

### 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 throughout 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 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.chem.2015.08.005.

AUTHOR CONTRIBUTIONS


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; 2012017110060 from Research Fund for Doctoral Programs in Higher Education; and SKLCO20150F07 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


