



# Oral Microbiome Alterations Associated with Early Childhood Caries Highlight the Importance of Carbohydrate Metabolic Activities

Yuan Wang,<sup>a</sup> Sa Wang,<sup>a</sup> Chunyan Wu,<sup>b,c</sup> Xi Chen,<sup>a</sup> Zhuhui Duan,<sup>a</sup> Qian Xu,<sup>c</sup> Wen Jiang,<sup>a</sup> Lei Xu,<sup>a</sup> Tingting Wang,<sup>c</sup> Lingkai Su,<sup>a</sup> Ying Wang,<sup>a</sup> Yadong Chen,<sup>a</sup> Jie Zhang,<sup>a</sup> Yun Huang,<sup>c</sup> Suman Tong,<sup>a</sup> Cheng Zhou,<sup>d</sup> Shuli Deng,<sup>a</sup> Nan Qin<sup>b,c,d</sup>

<sup>a</sup>Department of Conservative Dentistry and Periodontics, Affiliated Hospital of Stomatology, Zhejiang University School of Medicine, Hangzhou, China

<sup>b</sup>Shanghai Tenth People's Hospital Affiliated to Tongji University, Shanghai, China

<sup>c</sup>Realbio Genomics Institute, Shanghai, China

<sup>d</sup>State Key Laboratory for Diagnosis and Treatment of Infectious Disease, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

**ABSTRACT** Globally, dental caries is the most prevalent chronic oral disease and affects roughly half of all children. The aim of this report was to use metagenomic analyses to investigate the relationship between the oral microbiome and caries in preschool children. A total of 25 preschoolers, aged 3 to 5 years old with severe early childhood caries (ECC), and 19 age-matched, caries-free children as controls were recruited. Saliva samples were collected from the participants and were subjected to metagenomic analyses, whereby the oral microbial communities were investigated. The metagenomic analyses revealed substantial microbiota differences between the two groups, indicating apparent shifts of the oral microbiome present in the ECC group. At the species level, the ECC-enriched microbes included *Prevotella amnii*, *Shuttleworthia satelles*, *Olsenella uli*, and *Anaeroglobus geminatus*. Interestingly, *Actinomyces odontolyticus* and *Actinomyces graevenitzi* exhibited apparent differences at the strain level but not the species level between the ECC and control groups. Functional examination showed that the ECC group displayed extensive alterations in metabolic genes/pathways/modules, including enriched functions in sugar metabolism. Finally, an SVM (support vector machine) classifier comprising seven species was developed and generated a moderately good performance in predicting caries onset (area under the receiver operating characteristic curve [AUC] = 78.33%). Together, these findings indicate that caries is associated with considerable changes in the oral microbiome, some of which can potentially be exploited as therapeutic targets or diagnostic markers. (This study has been registered at ClinicalTrials.gov under registration no. NCT02341352.)

**IMPORTANCE** Dental caries is a highly prevalent oral disease that can lead to severe dental damage and may greatly compromise the quality of life of the affected individuals. Previous studies, including those based on 16S rRNA gene, have revealed that the oral microbiota plays a prominent role in development of the disease. But the approach of those studies was limited in analyzing several key microbiome traits, including species- or strain-level composition and functional profile. Here, we performed metagenomic analyses for a cohort of preschool children with or without caries. Our results showed that caries was associated with extensive microbiota differences at various taxonomic and functional levels. Some caries-associated species had not been previously reported, some of which may have significant clinical implications. A microbiome gene catalogue from children with caries was constructed for the first time. The results demonstrated that caries is associated with alterations of the oral microbiome, including changes in microbial composition and metabolic functional profile.

**Citation** Wang Y, Wang S, Wu C, Chen X, Duan Z, Xu Q, Jiang W, Xu L, Wang T, Su L, Wang Y, Chen Y, Zhang J, Huang Y, Tong S, Zhou C, Deng S, Qin N. 2019. Oral microbiome alterations associated with early childhood caries highlight the importance of carbohydrate metabolic activities. *mSystems* 4:e00450-19. <https://doi.org/10.1128/mSystems.00450-19>.

**Editor** Danilo Ercolini, University of Naples Federico II

**Copyright** © 2019 Wang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Shuli Deng, [dengshuli@zju.edu.cn](mailto:dengshuli@zju.edu.cn), or Nan Qin, [qinnan001@126.com](mailto:qinnan001@126.com).

Yuan Wang, Sa Wang, and Chunyan Wu contributed equally to this work.

This article is dedicated to Hui Chen, who guided us to accomplish this project.

**Received** 23 July 2019

**Accepted** 16 October 2019

**Published** 5 November 2019

**KEYWORDS** early childhood caries, metagenomics, oral microbiome, functional profile, preschool children

Dental caries, also known as tooth decay, is one of the most common oral infections in children (1). It is a destructive process that causes decalcification of tooth enamel and subsequently leads to continued breakdown of enamel and dentin (2). If left untreated, some pathogens or pathobionts in the oral microbiota can penetrate the enamel and dentin to reach the pulp, which leads to pulpitis and periapical periodontitis. In the absence of immediate and effective infection control, these local infections may expand and progress to culminate in more serious conditions, such as cellulitis (3, 4), osteomyelitis (5), bacteremia, and bacterial endocarditis (6).

The Global Burden of Disease (GBD) study reported that caries affected more than 10% of the world's population in 2015 and that the incidence of deciduous caries increased by 5.6% between 2005 and 2015 (7). According to an oral epidemiological investigation in China in 2018, there is a marked increase in the prevalence of childhood caries, up from 5.8% a decade ago (8). Therefore, study of the pathogenesis of childhood caries is of great significance in prevention, screening, and early intervention for vulnerable or affected children.

In previous studies, we performed 16S rRNA gene amplicon sequencing to examine the bacterial microbiota of dental plaques to study the microbial traits in severe cases of early childhood caries (ECC) (9), which revealed dynamic changes of oral microbiota at different stages of caries progression (10, 11). Nevertheless, this approach cannot provide some key information about oral microbiota, such as species-level and strain-level resolution and metabolic profile, which are likely important for caries pathogenesis (12).

Fang Yang et al. employed a microbial functional gene microarray to reconstruct the functional profiles of human saliva microbiota for healthy and caries-active adults; the results showed that saliva microbiota carried disease-associated functional signatures, which could be potentially exploited as diagnostic markers (13). However, the functional features of gene microarrays are dependent on preselected probe sets, thus limiting their scope in functional dissection of microbial communities.

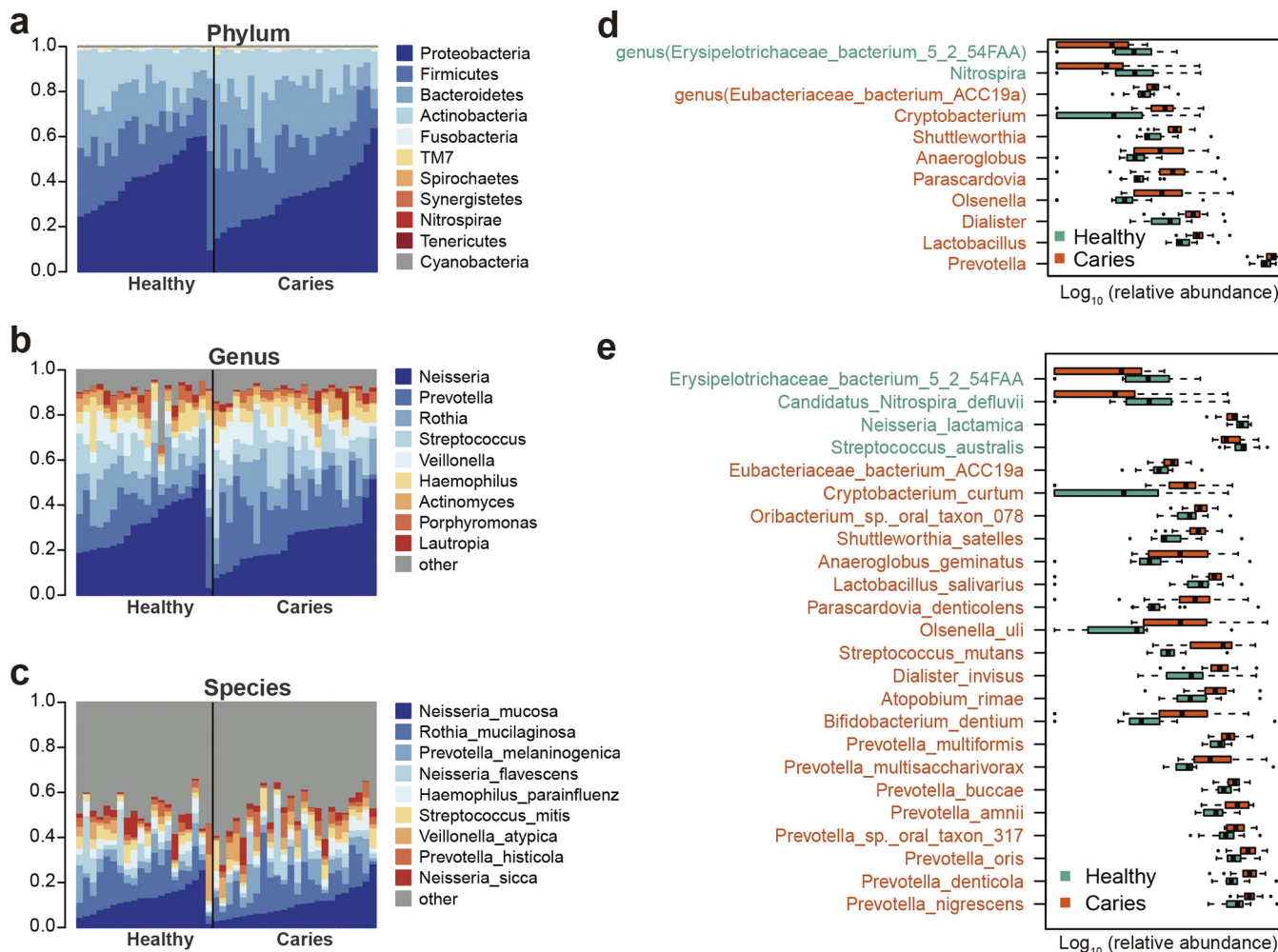
In this study, we analyzed the oral microbiome in preschoolers, whereby a gene catalogue was constructed for children with ECC. Our results not only corroborated previous findings that the microbiome has a great relevance in the occurrence of dental caries but also revealed new microbial species and functional groups associated with the disease.

## RESULTS

**Sample collection, sequencing, and quality control.** Saliva samples were collected from 25 preschool children with severe early childhood caries (ECC) (decayed, missing, and filled tooth surfaces [dmfs]  $\geq 8$ ) and 19 healthy control subjects (dmfs = 0) living in Lin'an, Zhejiang Province (see Table S1 in the supplemental material). There were no differences in age, gender, or body mass index (BMI) between the caries group and the healthy group. A total of 195-GB of raw data was generated from the Illumina HiSeq 2000 platform. After filtering out low-quality data and host contamination, an average of 3.08 GB (1.51 to 7.07 GB) of clean data were generated for each sample (Table S2).

To examine the association between oral microbiota and ECC development, we classified the 19 healthy children into two subgroups based on the changes in caries state during the 6 months after the initial sampling: (i) the "stay healthy" (H2H) subgroup, in which the 15 subjects maintained a healthy state, and (ii) the "caries-onset" (H2C) subgroup, in which the 4 subjects underwent the transition from a healthy to a caries-active state.

**Shifts of the oral microbiomes in preschoolers with caries.** After filtering out  $27.8\% \pm 16.7\%$  sequences/reads as host gene sequences, approximately  $49.8\% \pm 3.8\%$



**FIG 1** Relative abundances of phylotypes in healthy and ECC (caries) groups. (a to c) Relative abundances of phyla, genera, and species, respectively, are shown in a bar plot. (d and e) Relative abundances of genera (d) and species (e) with significantly different abundances (FDR < 0.1) are shown in a box plot.

of the reads from each sample were mapped to 7,312 reference bacterial genomes from GenBank and the HMP (Human Microbiome Project).

To investigate the diversity of salivary microbiome richness, the Shannon-Weiner index and Simpson index were calculated for species and genes, which showed that microbial diversity and richness were similar between the ECC and control groups ( $P > 0.05$ ) (Fig. S1a). To assess microbial structure alterations in the ECC group, we employed nonparametric analyses and principal-coordinate analysis (PCoA). Three nonparametric methods were applied, namely, the multiresponse permutation procedure (MRPP), analysis of similarity (ANOSIM), and permutational multivariate analysis of variance (Adonis). Apparent differences were detected by both the MRPP and Adonis at the phylum, class, order, family, and genus levels ( $P < 0.05$ ) or by the ANOSIM at the phylum, family, and genus levels ( $P < 0.05$ ). PCoA based on the Bray-Curtis distance of species abundance showed that the ECC and healthy groups displayed apparent microbiome structural differences (Fig. S1c).

At the phylum level, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* were the most abundant taxa in both groups (Fig. 1a). The healthy group displayed a higher abundance of *Nitrospirae* than did the ECC group (Wilcoxon test, false discovery rate [FDR] < 0.1) (Table S3).

At the genus level, *Neisseria*, *Prevotella*, *Rothia*, *Streptococcus*, *Veillonella*, and *Haemophilus* were among the major phylotypes in both groups (Fig. 1b). Subsequent

analysis of relative abundance revealed that nine genera, including *Prevotella*, were more abundant in the ECC group than in the healthy group, whereas *Nitrospira* and the genus *Erysipelotrichaceae bacterium 5\_2\_54FAA* were enriched in the healthy group (Fig. 1d; Table S3).

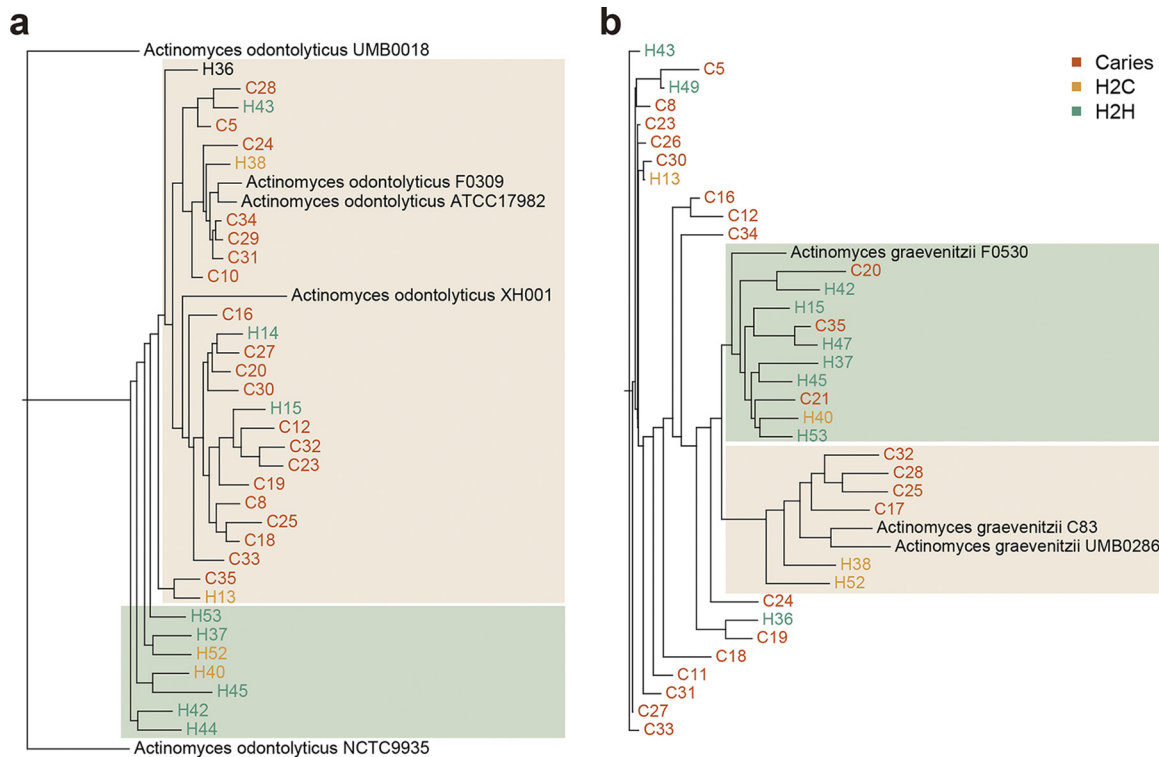
At the species level, *Neisseria mucosa*, *Rothia mucilaginosa*, and *Prevotella melanogenica* accounted for large shares of the total microbial abundance in both the children with caries and caries-free children, suggesting that these microbes belonged to the stable oral microflora (Fig. 1c). The caries-free subjects exhibited an increased relative abundance of *Neisseria lactamica* or *Streptococcus australis* (Fig. 1e; Table S3). Conversely, 20 species were found to be enriched in the severe ECC group (Fig. 1e; Table S3). These species included *Streptococcus mutans* (14–17) and multiple *Prevotella* spp. (12), which have been reported to be associated with dental caries, as well as *Prevotella amnii*, *Shuttleworthia satelles*, *Olsenella uli*, and *Anaeroglobus geminatus*, whose connections to the disease have not been reported.

To further delineate features of the ECC-associated saliva microbiome, we identified 26,264 differentially abundant genes (Wilcoxon rank sum test, FDR < 0.07) and clustered them into metagenomic species (MGS) on the basis of their correlated abundance variation across samples. We grouped the differentially abundant genes into 18 MGS, with 12 MGS enriched in the ECC group and 6 MGS enriched in the healthy controls (Fig. S2). Of the 12 MGS enriched in the ECC group, four were *Prevotella* species. On the other hand, two *Neisseria* species were more abundant in healthy subjects (Fig. S2). Importantly, these MGS-based results were in agreement with those derived from taxonomic analysis (Fig. 1d and e).

**Strain-level variations of the caries and healthy subjects.** It is being increasingly recognized that microbial species consist of distinct strains and that strain-level variations are a crucial factor for determining the functions of microbial communities. To examine the strain-level variants between the caries and healthy groups, we applied StrainPhlAn, an assembly-free strain-level phylogenetic method that identifies single nucleotide variants (SNVs) in species-specific marker genes (18). Using the SNV-based analysis, we built the phylogenetic trees of the common species from the samples with sufficient coverage and available reference genomes. We found considerable strain-level heterogeneity between the caries and healthy groups in two species, i.e., *Actinomyces odontolyticus* and *Actinomyces graevenitzii*, albeit neither of which displayed a difference in relative abundance at the species level. For *A. odontolyticus* (Fig. 2a), the dominant strains in caries individuals were phylogenetically closer to *A. odontolyticus* F0309, *A. odontolyticus* ATCC 17982, and *A. odontolyticus* XH001 than those in the control group. In addition, the dominant *A. graevenitzii* strains in the ECC and H2C subjects were closely related to *A. graevenitzii* C83 and *A. graevenitzii* UMB0286 strains, as were those in the healthy subjects to *A. graevenitzii* F0530 (Fig. 2b).

**Cooccurrence networks of saliva microbiota under healthy and ECC conditions.** To analyze the patterns of interbacterial interactions in oral microbial communities under healthy and ECC conditions, we constructed the cooccurrence networks for the two groups, respectively. We inferred the metacommunity cooccurrence networks based on Spearman correlation relationships and *P* values for correlations adjusted with the FDR (Benjamini and Hochberg). This generated a metacommunity cooccurrence network of the ECC group comprising 282 edges, reflective of the interbacterial associations, among 150 species/strains (Fig. 3a), as well as a network of the healthy group containing 374 edges among 164 species/strains (Fig. 3b). Whereas the healthy and ECC groups shared a considerable proportion of the edges (17.65% for the healthy network and 24.40% for the ECC network), most edges were condition specific (82.35% for the healthy network and 76.60% for the ECC network). In other words, approximately 20% of the interbacterial associations were shared by the two groups.

We calculated topological features for each node in the networks with the igraph package. This feature set included betweenness centrality (the number of shortest paths going through a node), closeness centrality (the number of steps required to access all other nodes from a given node), and degree (the number of adjacent edges).



**FIG 2** Strain-level phylogenetic trees of *A. odontolyticus* (a) and *A. graevenitzii* (b) of ECC (caries) and healthy (H2H and H2C) group samples. Available reference genomes were included in the phylogenetic trees.

Comparing these features between the two networks, we found that the closeness of nodes in the healthy group network was significantly higher than that in the ECC group network ( $P < 4E-11$ , Wilcoxon rank sum test), whereas degree and betweenness were not significantly different between the two groups.

Notably, there were one and two main clusters with  $>10$  nodes in the networks of ECC and control groups, respectively. *Bacteroides* spp. and *Prevotella* spp. were dominant in the two main clusters of the control group, as were *Streptococcus* spp. and *Prevotella* spp. in the main cluster of the ECC network.

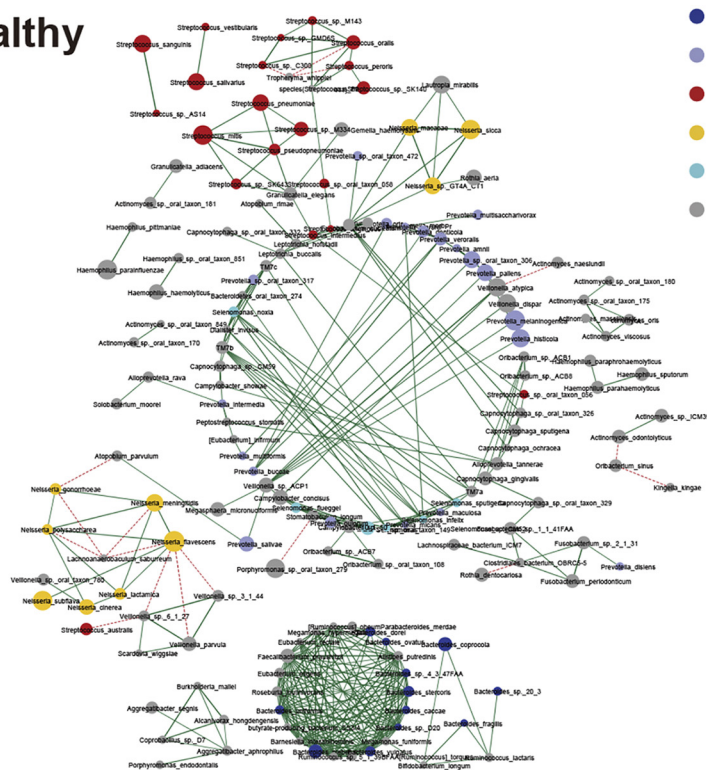
**Functional profiles of caries and healthy subjects.** To construct the gene catalogue, we combined the genes predicted from the assembled contigs and genes from the Human Oral Microbiome Database (HOMD). After filtering redundant genes, we generated a nonredundant oral microbial gene catalogue containing 2,200,443 genes. The healthy and ECC groups shared 107,151 genes, representing 71.05% and 81.58% of their core genes, respectively (Fig. S1b).

To investigate the functional role of the oral microbiome in ECC, we annotated the oral gene catalogues using KEGG (Kyoto Encyclopedia of Genes and Genomes database) and eggNOG (evolutionary genealogy of genes: Nonsupervised Orthologous Groups database). Correspondingly, three types of functional profiles were generated and compared between the caries and control groups: (i) gene profile, (ii) KEGG orthology profile, and (iii) eggNOG profile.

Analysis of the gene profile revealed a skewed pattern such that 1,200 genes were enriched in the ECC group, as opposed to only 62 genes in the healthy group (Wilcoxon rank sum test,  $FDR < 0.07$ ) (Table S4). Using a 0.85 identity threshold, these genes were mapped to GenBank via BLAT. Annotation of the differentially abundant genes to the KEGG and eggNOG databases revealed that extensive differences were present between the two groups in a variety of functions/pathways, including a relatively increased level of carbohydrate metabolism and decreased levels of translation, energy metabolism, coenzyme/cofactor/vitamin metabolism, and signal transduction in the

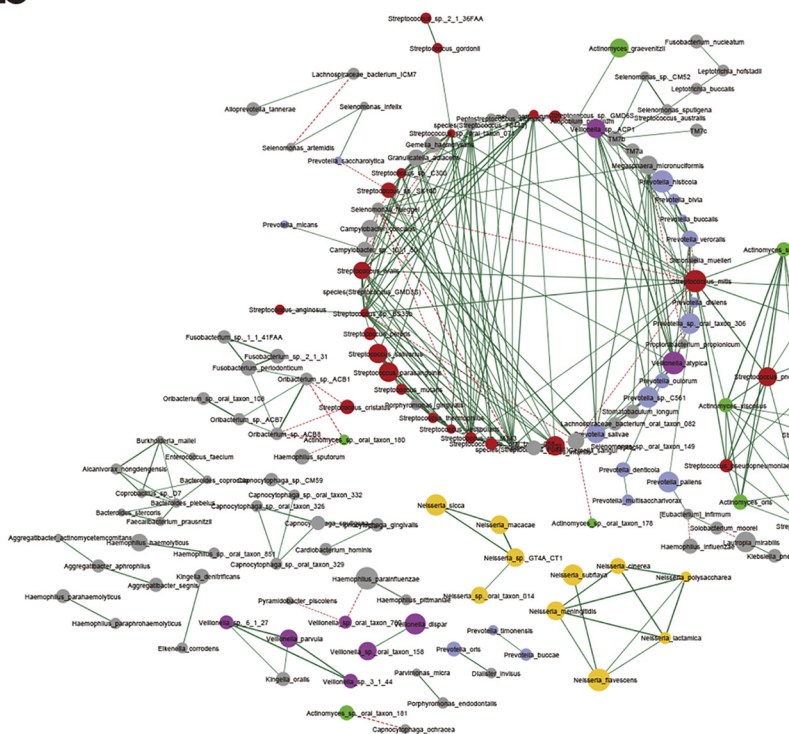


# a Healthy



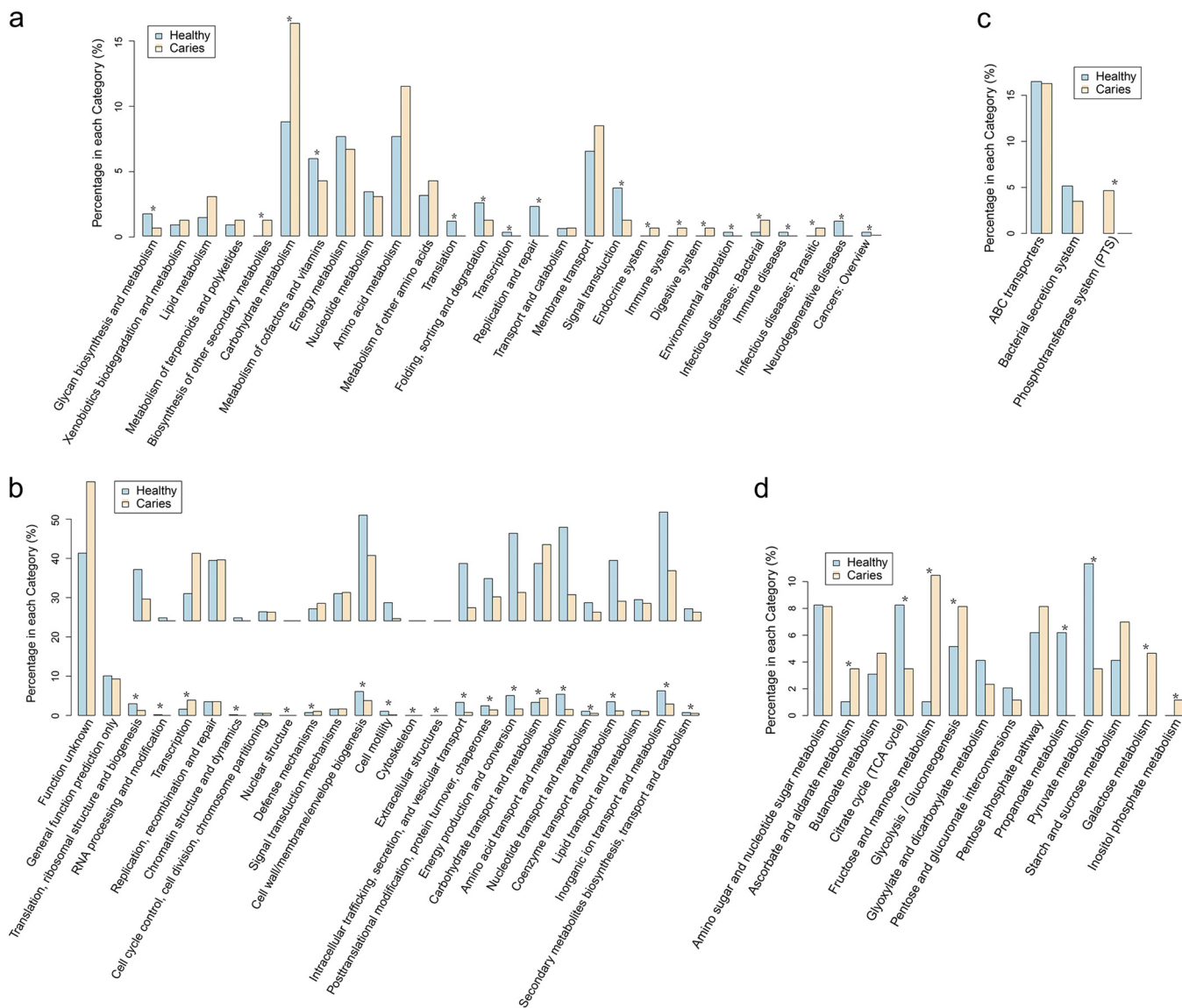
- Bacteroides
- Prevotella
- Streptococcus
- Neisseria
- Selenomonas
- Others

# b ECC



- Streptococcus
- Prevotella
- Actinomyces
- Veillonella
- Neisseria
- Others

**FIG 3** Networks in oral microbial communities under ECC and healthy conditions are shown, with each microbial species and cooccurrence relationship indicated by a node and an edge, respectively. A connection (line between dots) indicates a strong (Spearman's  $\rho > 0.6$ ) and significant (FDR  $< 0.05$ ) correlation. The size of each node is proportional to the relative abundance. Lines between nodes indicate positive correlations (green) or negative correlations (red). The top five abundant genera are indicated in color.



**FIG 4** Functional distribution of KEGG orthologous genes and eggNOG orthologous genes in healthy and ECC (caries) children. (a) Comparison between the KEGG orthologous genes enriched for healthy and ECC children for each KEGG functional category at the second functional level. (b) Comparison between the eggNOG orthologous genes enriched for healthy and ECC children for 24 eggNOG orthologue group functional categories. (c and d) Comparison between KEGG orthologous genes for healthy and ECC children for each KEGG functional category at the third functional level: membrane transport (c) and carbohydrate metabolism (d). Asterisks indicate hypergeometric distribution test results with phyper.R (\*, FDR < 0.05).

caries group (Fig. 4a and b). Differences were also detected at the third-level components of KEGG between the two groups. In the membrane transport pathway, the phosphotransferase system was enriched in the ECC children (Fig. 4c). In the carbohydrate metabolism pathway, citrate cycle (tricarboxylic acid [TCA] cycle) was enriched in healthy children, while glycolysis/gluconeogenesis was enriched in the ECC group (Fig. 4d). Notably, the glucosyltransferase (GTF) gene (12\_gene\_id\_1342) (Table S4) showed an increased relative abundance in ECC subjects, as did modules of the AI-2 (autoinducer-2) transport system, phosphotransferase (PTS) system, glucitol/sorbitol-specific II component, and nucleotide sugar biosynthesis (Table S5).

**Host factors associated with some microbial taxa and pathways.** Permutational multivariate analysis of variance (PERMANOVA) was performed to analyze the association between clinical factors and interpersonal distance (Bray-Curtis) of microbial composition (Table S6). For the 44 children, 3 factors were associated with interpersonal distance of microbial composition ( $P < 0.05$ ). Education background, height, and caries

status were the major sources of variance in the microbial species composition. In contrast, BMI, gender, income level, and dietary habit made comparatively minor or nonsignificant contributions to oral microbiome composition. We performed multivariate linear association analyses between the host clinical phenotypes and 205 representative species (>0.01% of total microbial reads and present in at least 10 individuals) or 315 MetaCyc pathways. When corrections were made for age and gender, we identified 28 associations with an FDR of <0.1 between 3 factors and 24 species (Table S7), as well as 52 associations between 5 factors and 176 MetaCyc pathways (Table S8). In our study, toothache in the past year was correlated not only with microbial composition but also with MetaCyc pathways. The relative abundances of *Prevotella amnii*, *Prevotella buccae*, and *Streptococcus mutans* were positively correlated with toothache in the past year and caries status. Other associations with phenotypical variables included a negative correlation between *Neisseria lactamica* and toothache in the past year and caries status. MetaCyc pathways, including biotin biosynthesis II, purine nucleobase degradation, and guanosine nucleotide degradation, were positively correlated with toothache in the past year and the frequency of dietary intake of biscuits, cakes, and bread. In comparison, the prevalence of L-lysine biosynthesis was inversely associated with decayed, missing, or filled tooth (dmft), dmfs, and toothache in the past year.

**Disease classification based on oral microbiota profiles.** By use of the mRMR algorithm, 7 species markers were chosen to construct the SVM (support vector machine) classifier, which exhibited the best performance (Fig. 5a). Of them, 5 species (i.e., *Streptococcus mutans*, *Prevotella amnii*, *Eubacteriaceae bacterium* ACC19a, *Shuttleworthia satelles*, and *Dialister invisus*) were enriched in the caries group, as were “*Candidatus Nitrospira defluvii*” and *Erysipelotrichaceae bacterium* in the control group (Fig. 5b). This classifier manifested an area under the receiver operating characteristic curve (AUC) of 98.53% and a 95% confidence interval (CI) of 95.81% to 100% (Fig. 5c). Notably, the relative levels of the 7 marker species in the H2C subgroup all exhibited a tendency of approaching that of the caries group (Fig. 5d). We used the classifier to predict the future new ECC onset of these 19 healthy controls and showed that the classifier was able to predict the onset of caries with a moderately good performance (AUC = 78.33%) (Fig. 5e).

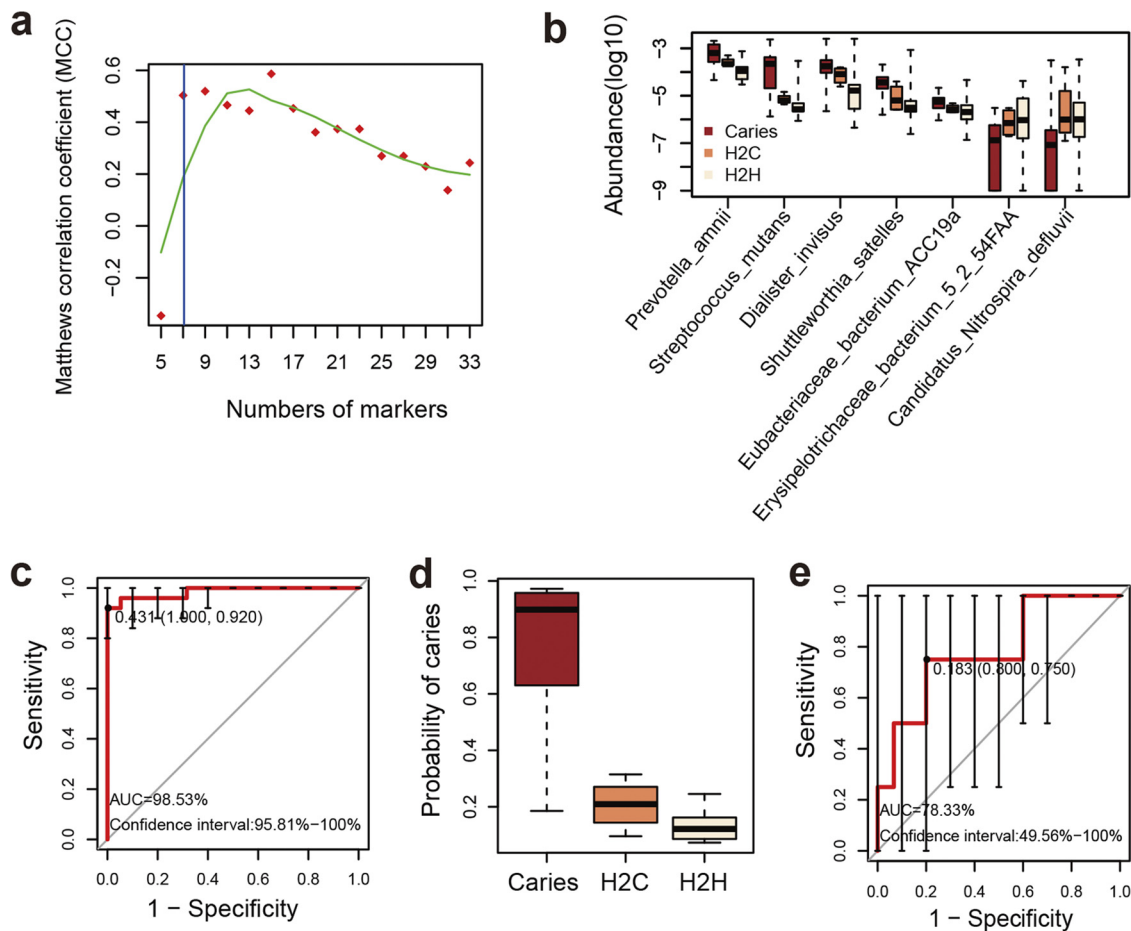
## DISCUSSION

Dental caries is a major oral health problem worldwide, affecting a great proportion of adults and children. The oral microbiome plays a crucial role in human health (19–22) and can profoundly affect the development of many diseases, including caries (20, 23). Microbial indicators of caries have been proposed as a method to predict future caries onset (24). There are approximately 700 prokaryotic species reportedly present in the human oral cavity, some of which may damage teeth under certain conditions (25). It has been reported that dental caries is directly caused by acid production on the enamel surface and that some microbes (e.g., *Streptococcus mutans*) play a significant role in this process. Dental caries is a polymicrobial disease that is not determined by one particular bacterium. Instead, it results from complex communal activities involving at least tens of bacterial species (26, 27). Thus far, however, a consensus has not been reached regarding cariogenic microbes and functional elements.

To study the effects of the oral microbiome on early childhood caries (ECC), we established an oral saliva gene catalogue (severe-ECC catalogue) from 44 children. The numbers of catalogue-specific genes were 22,824, 39,274, and 62,542 in the ECC group, healthy group, and HOMD-derived gene set, respectively. Our results revealed some differences between the oral microbiomes of the ECC and healthy groups. To facilitate subsequent analyses, we built an oral reference gene set by integrating our oral saliva genes with the gene set from the HOMD.

Our data revealed that the ECC and healthy groups exhibited considerable differences in taxonomic composition and functional profiles. For example, *Prevotella amnii*, *Shuttleworthia satelles*, *Olsenella uli*, and *Anaeroglobus geminatus*, whose connections





**FIG 5** Classifier used to distinguish ECC children from healthy controls. (a) The mRMR method was used to identify the ECC-associated markers. Sequential subsets were generated at five-species intervals. For each subset, the error rate was estimated using a leave-one-out cross-validation of a linear discrimination classifier. Using only the seven marker species as predictors, the SVM model exhibited predictive performance that was already comparable to the performance of the model derived from the optimum (highest value of the Matthews correlation coefficient) subset. (b) The relative abundances of seven marker species among the H2H group, the H2C group, and the caries group are shown in a box plot. (c) Receiver operating characteristic (ROC) curves for the ECC group and healthy controls; 95% confidence intervals (CIs) are indicated by error bars. (d) The probability of caries determined by the classifier among the H2H group, H2C group, and caries group is shown in a box plot. (e) ROC curves for the H2H group and H2C group.

with caries have not been reported, were enriched in the ECC group (Fig. 1e). The enrichment of *Prevotella amnii* in the ECC group is not surprising, as it is a species of the *Prevotella* genus, most of which have been reported to be related to caries and potentially have proteolytic/amino acid-degrading activities (12). *Shuttleworthia satelles*, *Olsenella uli*, and *Anaeroglobus geminatus* were reported to be present in the oral cavity and associated with periodontal disease (28, 29). Therefore, it appears that the presence of these microbes indicates an oral environment favoring caries onset. *N. lactamica* is considered a commensal microbe in the nasopharynx (30) and has been found to be the most abundant nasopharyngeal species in preschool children under the age of 5 years (31) (Fig. 1e). The higher levels of *N. lactamica* and *Streptococcus australis* in healthy subjects than in caries counterparts suggest that the two microbes correlate with dental health. Our results therefore not only corroborated previous findings on the relationship between oral microbiota and caries (14, 15, 17) but also identified new potential biomarkers of ECC.

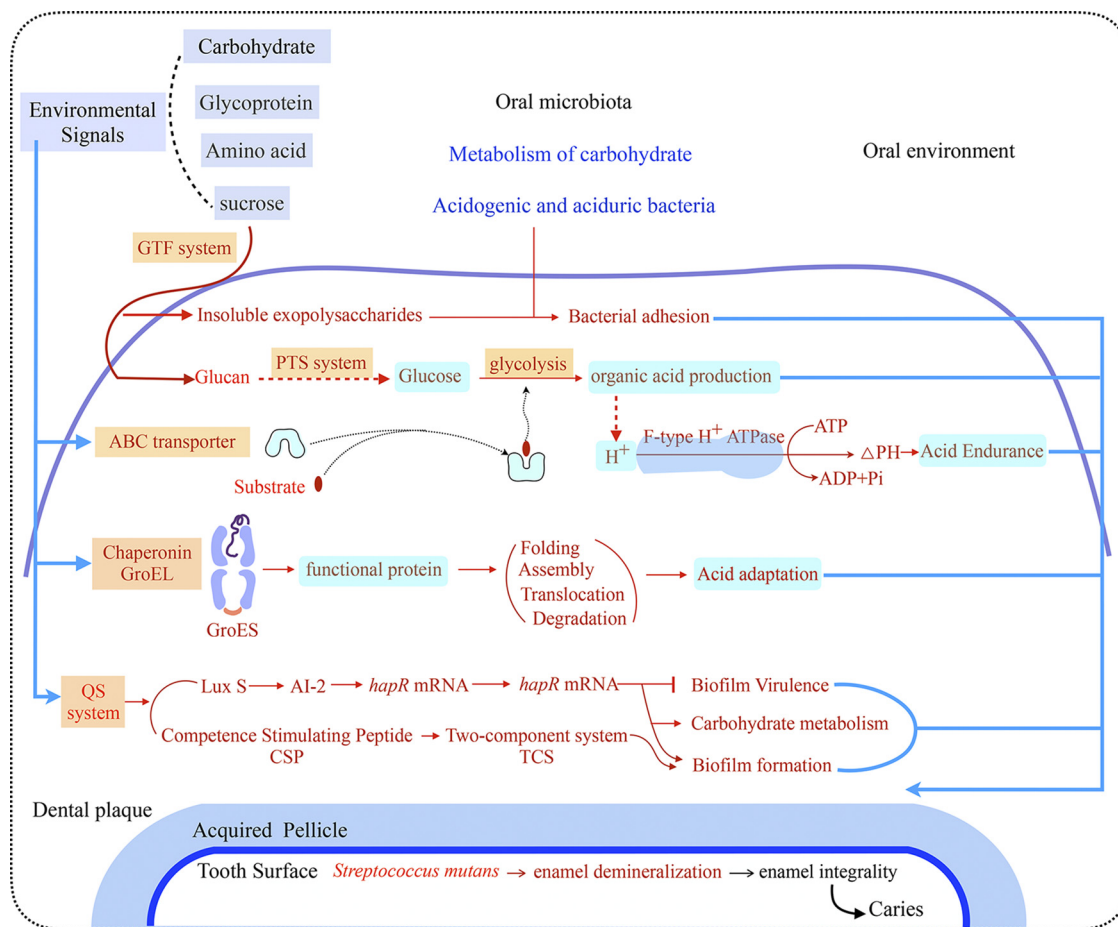
Interestingly, StrainPhlAn analysis detected clear strain-level, but not species-level, differences in *Actinomyces odontolyticus* and *Actinomyces graevenitzii* between the caries and healthy participants (Fig. 2). Besides the taxonomic alterations, analysis of the cooccurrence network also indicated distinct patterns of interbacterial interactions in

the saliva communities under healthy and ECC conditions (Fig. 3). For example, the main genera in the cooccurrence network were *Bacteroides* spp. and *Prevotella* spp. under the healthy condition, whereas *Streptococcus* spp. and *Prevotella* spp. were the main genera under the ECC condition.

Our analyses also revealed a prominent divergence between the caries and healthy groups at various functional levels. Functions related to carbohydrate metabolism were enriched in the ECC group, as revealed by the results of the gene profile, the KEGG orthology profile (Fig. 4a), and the eggNOG profile (Fig. 4b). This mirrored previous findings that enhanced carbohydrate activity in oral microbiota was a contributing factor of caries pathogenesis (32). In the membrane transport pathway, the phosphotransferase system was enriched in the ECC children (Fig. 4c), which indicated that transmembrane transport and phosphorylation were more active in this group. In the carbohydrate metabolism pathway, the citrate cycle (TCA cycle) was enriched in the healthy children, while glycolysis/gluconeogenesis was enriched in the ECC group. This result suggested that anaerobic metabolism of sugar was more active in the oral cavity of children with caries. At the module level, multiple metabolic components were enriched in the ECC group, including the AI-2 transport system, the PTS system, glucitol/sorbitol-specific II component, and nucleotide sugar biosynthesis (Table S5). Quorum sensing is an important mechanism underlying biofilm formation during the development of dental caries (32), which reportedly involves signal molecule autoinducer-2 (AI-2) in the interbacterial interaction (33, 34). While AI-2 is a quorum signaling component in some microbes, in other species the protein also harbors different activities (35) that may be implicated in caries development. The ABC transporter and two-component signal transduction system (TCS) regulates the expression of genes according to local environmental changes, which in turn influences bacterial competence, survival, and virulence. Moreover, the dramatic reduction of cell motility modules (Fig. 4b) in the ECC group could also be attributed to the increased biofilm formation in the ECC group, as inhibition of bacterial motility promotes biofilm formation (36).

Despite these findings, more research is needed to elucidate the precise mechanisms of oral microbiota in caries pathogenesis. Recent studies have shown that iron deficiency in young children is a risk factor for ECC (37), and this finding was confirmed in animal experiments (38). Our data revealed that iron complex transport was an enriched function in caries-free children, which was in agreement with previous results (Table S5). Other transport modules more enriched in the healthy group than in the caries group included sulfate transport, putrescine transport, and microcin C transport pathways (Table S5). It has been suggested that decreased activity of transporter proteins may lead to the accumulation of metabolic compounds, such as sugars and acids, and contribute to dental caries (37–39).

Our findings enabled us to propose a model to explain the roles of some microbes in caries pathogenesis (Fig. 6). It is known that dental plaque biofilms produce acids from carbohydrates that contribute to caries onset (40). Development of dental caries is a gradual process in which the first stage is characterized by oral biofilm formation and bacterial adhesion (41). Cariogenic bacteria, such as *Streptococcus mutans*, produce a GTF that synthesizes extracellular polysaccharides (EPSs) (42). The EPSs, especially water-insoluble glucans, play a critical role in dental plaque formation and biofilm stability, as these molecules allow cariogenic bacteria to adhere to enamel surfaces (43). The biofilm phenotype is regulated by its density-dependent quorum sensing (QS) system, which consists primarily of the competence stimulating peptide (CSP) and two-component signal transduction system (TCS) (44). In addition to biofilm formation, the CSP-mediated QS system also affects its acidogenicity and aciduricity (44). The PTS system is responsible for recognition, transmembrane transport, and phosphorylation of water-soluble glucans (45). When sugar is frequently consumed, glycolysis and acidification often ensue. F-type H<sup>+</sup> ATPase and chaperonin GroEL may enhance the acidogenicity and acidurance of the cariogenic bacteria (46, 47). In the acidogenic stage, the acidogenic and aciduric bacteria rapidly propagate, whereby the deminer-



**FIG 6** Taxonomic and functional characterization of oral microbiota in child caries. A schematic diagram shows the main functions of the oral microbes that are associated with caries. Red text denotes enriched functions in children with caries.

alization/remineralization balance is tilted toward net mineral loss and leads to dental caries. In the aciduric stage, more aciduric bacteria, such as *Streptococcus mutans* and *Lactobacillus* spp., become dominant and aggravate the symptoms. As such, environmental acidification is a main contributor to caries development.

In summary, our results present differences in the oral microbial communities of healthy preschoolers and those with caries at various taxonomic and functional levels. As demonstrated by other microbiota association studies, understanding such complex and delicate relationships is crucial for the prevention and treatment of these diseases. Nevertheless, to better understand the microbial contribution in caries development, metatranscriptome analyses are needed and may provide additional evidence in elucidating the roles of taxonomic and functional variables in the oral microbiota.

Our conclusions from this study are as follows. (i) A microbiome gene catalogue from children with caries was constructed for the first time. (ii) Preschool children with dental caries and their healthy counterparts exhibited differences in oral microbiomes and functional profiles. (iii) The results demonstrate that multiple *Prevotella* spp. and *Veillonella* spp. are associated with dental caries and that the potential functional differences between children with caries and caries-free children are mainly distributed on carbohydrate metabolism functions/pathways. (iv) A panel of seven species was developed to predict the onset of caries.

**MATERIALS AND METHODS**

**Study subjects.** Twenty-five children with severe ECC (decayed, missing, or filled tooth surfaces [dmfs] ≥ 8) and 19 caries-free (dmfs = 0) preschoolers, aged 45 to 73 months, were recruited in the

study. Their diagnoses were made by a dentist at the Affiliated Hospital of Stomatology, Zhejiang University School of Medicine, according to the International Caries Detection and Assessment System II (ICDAS-II) (48). Written informed consent was obtained from the parents or other guardians of all participants prior to enrollment. The study was approved by the ethical committee of the Affiliated Hospital of Stomatology, Zhejiang University School of Medicine. We obtained consent to publish from the participant (or legal parent or other guardian for children) to report individual patient data, including images, videos, voice recordings, etc.

Exclusion criteria were as follows: (i) children with <18 teeth, (ii) children who received antibiotics or fluoride treatment in the prior 3 months, and (iii) children who suffered active bacterial or viral infections in other parts of the body (49).

**Saliva sampling and isolation of bacterial genomic DNA.** All subjects were asked not to eat or drink 2 h before sampling, which was performed in the morning. To minimize stimulation of salivation, saliva needed to be kept in the mouth for 3 min. Subjects were then instructed to drool into sterile cryogenic vials for 3 min. Each saliva sample was pipetted into a sterile 1.5-ml Eppendorf tube, which was snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Bacterial genomic DNA was extracted using the QIAamp DNA mini kit (Qiagen, Hilden, Germany) as previously described (10). To reduce contamination by human DNA, every  $4\ \mu\text{g}$  DNA was incubated with  $160\ \mu\text{l}$  MBD-Fc-bound beads from a NEBNext microbiome DNA enrichment kit (New England Biolabs, Inc., Ipswich, MA, USA). The enriched microbial DNA samples were purified by ethanol precipitation. DNA concentration and sizes were determined using NanoDrop and agarose gel electrophoresis. The resulting DNA samples were stored at  $-20^{\circ}\text{C}$  until further processing.

**Illumina sequencing.** The metagenomic DNA libraries were constructed according to the Illumina TruSeq DNA sample prep v2 guide. The library insert sizes were checked using a DNA LabChip 1000 kit on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All libraries were sequenced on a HiSeq 2000 instrument with the PE100 mode (Illumina, San Diego, CA, USA).

**Quality control of reads.** The following steps were used for quality control: (i) remove reads with more than 5 ambiguous bases and 50 low-quality bases (low quality indicated by a Phred quality score of less than 2), (ii) trim low-quality base tails of reads (low quality indicated by a Phred quality score of less than 2), and (iii) remove reads that were mapped to the human genome (HG19) by SOAPaligner 2.1 (50) using default parameters. The percentage of human reads accounted for, on average, 28% of the total sequencing data after the human DNA removal step. The microbial yield was apparently higher than the 20% to 30% proportion of microbial reads reported previously (19, 51).

**Genome assembly, gene prediction, and gene catalogue construction.** We carried out a two-round assembly strategy to improve the read utility ratios. For the first round, SOAPdenovo (52) (version 2.04) was used to assemble reads *de novo* for each sample, with parameters “-d 1 -M 3 -F” at k-mers ranging from 39 to 63, before the contigs with the longest  $N_{50}$  value were selected. For the second round, unused reads were selected by aligning clean reads with SOAPaligner 2.1 (50), prior to being repeatedly assembled with the same parameters at k-mer 59. MetaGeneMark (53) (prokaryotic GeneMark.hmm version 3.5) was used to predict open reading frames (ORFs) in contigs. The program predicted 3,086,934 ORFs with a length cutoff of 100 bp, and the total length of the ORFs was 1,548,170,042 bp. In light of the possibility that some low-abundance microbes were not detected in the limited sequencing data, we combined the previously public gene set from HOMD ([http://www.homd.org/ftp/all\\_oral\\_genomes/20160329/](http://www.homd.org/ftp/all_oral_genomes/20160329/)) (54) to build a nonredundant oral gene catalogue for further analyses. This nonredundant gene catalogue was established by cd-hit-v4.6.1 (55) with parameters “-c 0.95 -aS 0.9 -r 0.” Redundant ORFs sharing 95% identity or greater and 90% coverage or greater were removed, resulting in a nonredundant gene catalogue composed of 2,200,443 genes.

**Profiling of microbial taxa and genes.** Organism and gene abundance were calculated according to previous studies (20, 56, 57). Briefly, clean reads were aligned against reference genomes and by SOAPalign2.21 with the parameters “-r 2 -m 100 -x 1000.” Matched paired-end reads were chosen for further abundance calculation and then assigned to two types: (i) multiple reads that aligned to more than one species and (ii) unique reads that matched only one species. For species  $S$ , abundance  $Ab(S)$  could be divided into unique abundance  $Ab(U)$  and multiple abundance  $Ab(M)$ . We then calculated  $Ab(S)$  as follows:

$$\begin{aligned} Ab(S) &= Ab(U) + Ab(M) \\ Ab(U) &= U/l \\ Ab(M) &= \left( \sum_{i=1}^M Co \times \{M_i\} \right) / l \\ Co &= \frac{Ab(U)}{\sum_{i=1}^N Ab(U)} \end{aligned}$$

For each species ( $S$ ),  $U$  and  $M$  are the number of unique and multiple reads, respectively, and  $l$  is the average genome length of species  $S$ . For each multiple read in  $\{M\}$ , there is a species-specific coefficient  $Co$ , and the  $N$  is the number of aligned species of this read.

Likewise, this method was also used to calculate gene abundance.

**MGS identification.** To cluster genes into metagenomic species (MGS), we applied the method described by Le Chatelier et al. (56), Qin et al. (20), and Nielsen et al. (57). First, gene markers with differential abundances were identified using `wilcox.test` in R (FDR < 0.07, Wilcoxon rank sum test corrected by the Benjamini and Hochberg method). Next, we clustered the marker genes using a



Spearman's correlation coefficient ( $\rho$ ) of  $>0.9$  according to their abundances across all the individuals. After removal of clusters with fewer than 25 genes, a second hierarchical clustering was performed with the Spearman's correlation coefficient between the mean abundance of genes in each cluster and a new threshold of 0.8. The final gene clusters were called MGS.

**Strain-level analysis.** StrainPhlAn was used for strain-level profiling. For each sample, clean reads were first mapped against the MetaPhlAn2 markers by Bowtie2 (58) to generate the consensus sequence, which represented the most abundant strain for each species in a sample. Similarly, the consensus sequences of public reference genomes of strains for each species were obtained by aligning the markers to these genomes. Finally, the extracted consensus sequences of references and samples were multiply aligned by MUSCLE (59), and the phylogenetic trees were built by RAxML (60) (parameters:  $-m$  GTRCAT and  $-p$  1234).

**Gene function analysis.** Protein sequences were aligned to the KEGG gene database (KEGG release 71 July 2014) (61) and eggNOG v4.0 (62) by BLAT with parameters " $-prot -out = blast8 -minidentity = 30 -minScore = 60$ ." The best hit was selected for each gene based on score and identity. The abundances of eggNOG and KEGG orthologs were calculated as the sum of the abundances of all genes annotated to that ortholog.

Samples were functionally profiled using HUMAnN2 (<http://huttenhower.sph.harvard.edu/humann2>) (63). HUMAnN2 used the MetaCyc pathway database (<https://metacyc.org/download.shtml>) and MinPath to identify a parsimonious set of pathways which explain observed reactions in the community.

**Cooccurrence network.** To construct the metacommunity cooccurrence network, we first removed species with relative abundances of less than 0.01%. The Spearman correlation coefficients between species were computed using R, and all the  $P$  values were adjusted for multiple testing using the Benjamini and Hochberg false discovery rate (FDR)-controlling procedure. The cooccurrence networks were generated based on correlation coefficients ( $>0.6$ ) and FDR ( $<0.05$ ) for correlation and visualized by Cytoscape 3.0.2. Network properties were calculated with the igraph package.

**Association analysis between microbes and clinical variables.** To identify significant associations between oral microbial and phenotypic variables (see Table S1 in the supplemental material), we applied a statistical program of Multivariate Association with Linear Model (MaAsLin; <https://huttenhower.sph.harvard.edu/maaslin>) (64). In this study, age, gender, and BMI were included as potential confounders in each model. To test the association for each species, we first filtered low-abundance species and confined our analysis to 205 species that had relative abundances of  $>0.01\%$  and were present in more than 10 individuals. These 205 species accounted for, on average, 99.3% of microbial reads. The percentage of each species was arcsine-square-root transformed by taking the arcsine of the square root of the proportional value of each species. For MetaCyc pathways, the same filtering criteria were used, and a total of 315 pathways were further associated with different factors using MaAsLin. In each analysis, the false discovery rate was controlled at a  $q$  value of 0.1 using the Benjamini and Hochberg method (p.adjust package in R).

**Classifier construction.** We used an SVM (support vector machine) (R 3.1.3; the e1071 R package) to build the classifier for ECC. The differentially abundant species ( $P < 0.01$ ) were chosen as features. To filter out redundant features, the mRMR algorithm (65) (the sideChannelAttack R package) and the leave-one-out cross-validation LDA (linear discriminant analysis) (the paleoMAS R package) were applied. The feature set which has the highest Matthews correlation coefficient (MCC) was chosen to build the SVM classifier. The receiver operating characteristic (ROC) figures were drawn by using the pROC R package.

**Statistical analyses.** To detect significant differences in relative abundance of metagenomics features, the nonparametric Wilcoxon test (wilcox.test package in R) was performed. The FDR was calculated using the Benjamini and Hochberg method (p.adjust package in R).

**Data availability.** The Illumina raw read data have been deposited at the National Center for Biotechnology Information (NCBI) under accession number SRP103050.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00450-19>.

**FIG S1**, PDF file, 0.8 MB.

**FIG S2**, PDF file, 0.7 MB.

**TABLE S1**, XLSX file, 0.02 MB.

**TABLE S2**, XLSX file, 0.01 MB.

**TABLE S3**, XLSX file, 0.04 MB.

**TABLE S4**, XLSX file, 0.8 MB.

**TABLE S5**, XLSX file, 0.01 MB.

**TABLE S6**, XLSX file, 0.01 MB.

**TABLE S7**, XLSX file, 0.01 MB.

**TABLE S8**, XLSX file, 0.01 MB.

## ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (81371142, 31970111, 31670118), by 2011 China State Key Clinical department grants,

by the Natural Science Foundation of Zhejiang Province, China (grant no. LY17H140004, LGF18H140004, and LQ19H140002), and by the General Project of Health and Family Planning Commission of Zhejiang Province (2016KYA118).

We declare that we have no competing financial interests.

## REFERENCES

- Bradshaw DJ, Lynch RJ. 2013. Diet and the microbial aetiology of dental caries: new paradigms. *Int Dent J* 63(Suppl 2):64–72. <https://doi.org/10.1111/idj.12082>.
- Andlaw RJ. 1960. The relationship between acid production and enamel decalcification in salivary fermentations of carbohydrate foodstuffs. *J Dent Res* 39:1200–1209. <https://doi.org/10.1177/00220345600390061401>.
- Arunkumar KV. 2016. Orbital infection threatening blindness due to carious primary molars: an interesting case report. *J Maxillofac Oral Surg* 15:72–75. <https://doi.org/10.1007/s12663-015-0801-6>.
- Moschos MM, Brouzas D, Mezitis M, Zachariadis N. 2005. Visual loss due to a carious tooth. *Lancet* 366:1504. [https://doi.org/10.1016/S0140-6736\(05\)67602-7](https://doi.org/10.1016/S0140-6736(05)67602-7).
- Romagna A, Troeltzsch M, Birkenmaier C, Schwartz C, Suchorska B, Zausinger S. 2018. Oral cavity infection: an underestimated source of pyogenic spondylodiscitis? *J Neurol Surg A Cent Eur Neurosurg* 79: 218–223. <https://doi.org/10.1055/s-0037-1608823>.
- Aoyagi S, Oda T, Wada K, Nakamura E, Kosuga T, Yasunaga H. 2018. Infective endocarditis associated with atopic dermatitis. *Int Heart J* 59:420–423. <https://doi.org/10.1536/ihj.17-078>.
- Anonymous. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388:1545–1602. [https://doi.org/10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6).
- Wang X. 2018. The fourth national oral health epidemiological survey report. People's Medical Publishing House, Beijing, China.
- Jiang W, Zhang J, Chen H. 2013. Pyrosequencing analysis of oral microbiota in children with severe early childhood dental caries. *Curr Microbiol* 67:537–542. <https://doi.org/10.1007/s00284-013-0393-7>.
- Jiang W, Ling Z, Lin X, Chen Y, Zhang J, Yu J, Xiang C, Chen H. 2014. Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood. *Microb Ecol* 67:962–969. <https://doi.org/10.1007/s00248-014-0372-y>.
- Xu L, Chen X, Wang Y, Jiang W, Wang S, Ling Z, Chen H. 2018. Dynamic alterations in salivary microbiota related to dental caries and age in preschool children with deciduous dentition: a 2-year follow-up study. *Front Physiol* 9:342. <https://doi.org/10.3389/fphys.2018.00342>.
- Yang F, Zeng X, Ning K, Liu KL, Lo CC, Wang W, Chen J, Wang D, Huang R, Chang X, Chain PS, Xie G, Ling J, Xu J. 2012. Saliva microbiomes distinguish caries-active from healthy human populations. *ISME J* 6:1–10. <https://doi.org/10.1038/ismej.2011.71>.
- Yang F, Ning K, Chang X, Yuan X, Tu Q, Yuan T, Deng Y, Hemme CL, Van Nostrand J, Cui X, He Z, Chen Z, Guo D, Yu J, Zhang Y, Zhou J, Xu J. 2014. Saliva microbiota carry caries-specific functional gene signatures. *PLoS One* 9:e76458. <https://doi.org/10.1371/journal.pone.0076458>.
- Caufield PW, Li Y, Dasanayake A. 2005. Dental caries: an infectious and transmissible disease. *Compend Contin Educ Dent* 26:10–16.
- Marchant S, Brailsford SR, Twomey AC, Roberts GJ, Beighton D. 2001. The predominant microflora of nursing caries lesions. *Caries Res* 35:397–406. <https://doi.org/10.1159/000047482>.
- Tanzer JM, Livingston J, Thompson AM. 2001. The microbiology of primary dental caries in humans. *J Dent Educ* 65:1028–1037.
- Loesche WJ. 1986. Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* 50:353–380.
- Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* 27:626–638. <https://doi.org/10.1101/gr.216242.116>.
- Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486:215–221. <https://doi.org/10.1038/nature11209>.
- Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, Zhou J, Ni S, Liu L, Pons N, Batto JM, Kennedy SP, Leonard P, Yuan C, Ding W, Chen Y, Hu X, Zheng B, Qian G, Xu W, Ehrlich SD, Zheng S, Li L. 2014. Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513:59–64. <https://doi.org/10.1038/nature13568>.
- Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
- Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F. 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498:99–103. <https://doi.org/10.1038/nature12198>.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JL. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444:1027–1031. <https://doi.org/10.1038/nature05414>.
- Teng F, Yang F, Huang S, Bo C, Xu ZZ, Amir A, Knight R, Ling J, Xu J. 2015. Prediction of early childhood caries via spatial-temporal variations of oral microbiota. *Cell Host Microbe* 18:296–306. <https://doi.org/10.1016/j.chom.2015.08.005>.
- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* 43:5721–5732. <https://doi.org/10.1128/JCM.43.11.5721-5732.2005>.
- Xu H, Hao W, Zhou Q, Wang W, Xia Z, Liu C, Chen X, Qin M, Chen F. 2014. Plaque bacterial microbiome diversity in children younger than 30 months with or without caries prior to eruption of second primary molars. *PLoS One* 9:e89269. <https://doi.org/10.1371/journal.pone.0089269>.
- Head DA, Marsh PD, Devine DA. 2014. Non-lethal control of the cariogenic potential of an agent-based model for dental plaque. *PLoS One* 9:e105012. <https://doi.org/10.1371/journal.pone.0105012>.
- Downes J, Munson MA, Radford DR, Spratt DA, Wade WG. 2002. *Shuttleworthia satelles* gen. nov., sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol* 52:1469–1475. <https://doi.org/10.1099/00207713-52-5-1469>.
- Camelo-Castillo A, Novoa L, Balsa-Castro C, Blanco J, Mira A, Tomás I. 2015. Relationship between periodontitis-associated subgingival microbiota and clinical inflammation by 16S pyrosequencing. *J Clin Periodontol* 42:1074. <https://doi.org/10.1111/jcpe.12470>.
- Hollis DG, Wiggins GL, Weaver RE. 1969. *Neisseria lactamica* sp. n., a lactose-fermenting species resembling *Neisseria meningitidis*. *Appl Microbiol* 17:71–77.
- Cartwright KA, Stuart JM, Jones DM, Noah ND. 1987. The Stonehouse survey: nasopharyngeal carriage of meningococci and *Neisseria lactamica*. *Epidemiol Infect* 99:591–601. <https://doi.org/10.1017/s0950268800066449>.
- Peterson SN, Snesrud E, Schork NJ, Bretz WA. 2011. Dental caries pathogenicity: a genomic and metagenomic perspective. *Int Dent J* 61(Suppl 1):11–22. <https://doi.org/10.1111/j.1875-595X.2011.00025.x>.
- McNab R, Ford SK, El-Sabaeny A, Barbieri B, Cook GS, Lamont RJ. 2003. LuxS-based signaling in *Streptococcus gordonii*: autoinducer 2 controls carbohydrate metabolism and biofilm formation with *Porphyromonas gingivalis*. *J Bacteriol* 185:274–284. <https://doi.org/10.1128/jb.185.1.274-284.2003>.
- Huang Z, Meric G, Liu Z, Ma R, Tang Z, Lejeune P. 2009. luxS-based quorum-sensing signaling affects biofilm formation in *Streptococcus mutans*. *J Mol Microbiol Biotechnol* 17:12–19. <https://doi.org/10.1159/000159193>.
- Rezzonico F, Duffy B. 2008. Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for luxS in most bacteria. *BMC Microbiol* 8:154. <https://doi.org/10.1186/1471-2180-8-154>.
- Guttenplan SB, Kearns DB. 2013. Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev* 37:849–871. <https://doi.org/10.1111/1574-6976.12018>.
- Iranna Koppal P, Sakri MR, Akkareddy B, Hinduja DM, Gangolli RA, Patil BC. 2013. Iron deficiency in young children: a risk marker for early childhood caries. *Int J Clin Pediatr Dent* 6:1–6. <https://doi.org/10.5005/jp-journals-10005-1176>.
- Eshghi A, Kowsari-Isfahan R, Rezaiefar M, Razavi M, Zeighami S. 2012. Effect of iron containing supplements on rats' dental caries progression. *J Dent (Tehran)* 9:14–19.

39. Kleemola-Kujala E, Räsänen L. 1979. Dietary pattern of Finnish children with low high caries experience. *Community Dent Oral Epidemiol* 7:199–205. <https://doi.org/10.1111/j.1600-0528.1979.tb01216.x>.
40. Tanner AC, Kent RL, Jr, Holgerson PL, Hughes CV, Loo CY, Kanasi E, Chalmers NI, Johansson I. 2011. Microbiota of severe early childhood caries before and after therapy. *J Dent Res* 90:1298–1305. <https://doi.org/10.1177/0022034511421201>.
41. Zhao W, Li W, Lin J, Chen Z, Yu D. 2014. Effect of sucrose concentration on sucrose-dependent adhesion and glucosyltransferase expression of *S. mutans* in children with severe early-childhood caries (S-ECC). *Nutrients* 6:3572–3586. <https://doi.org/10.3390/nu6093572>.
42. Koo H, Xiao J, Klein MI, Jeon JG. 2010. Exopolysaccharides produced by *Streptococcus mutans* glucosyltransferases modulate the establishment of microcolonies within multispecies biofilms. *J Bacteriol* 192:3024–3032. <https://doi.org/10.1128/JB.01649-09>.
43. Chen L, Ren Z, Zhou X, Zeng J, Zou J, Li Y. 2016. Inhibition of *Streptococcus mutans* biofilm formation, extracellular polysaccharide production, and virulence by an oxazole derivative. *Appl Microbiol Biotechnol* 100:857–867. <https://doi.org/10.1007/s00253-015-7092-1>.
44. Senadheera D, Cvitkovich DG. 2008. Quorum sensing and biofilm formation by *Streptococcus mutans*. *Adv Exp Med Biol* 631:178–188. [https://doi.org/10.1007/978-0-387-78885-2\\_12](https://doi.org/10.1007/978-0-387-78885-2_12).
45. Deutscher J, Francke C, Postma PW. 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* 70:939–1031. <https://doi.org/10.1128/MMBR.00024-06>.
46. Quivey RG, Jr, Kuhnert WL, Hahn K. 2000. Adaptation of oral streptococci to low pH. *Adv Microb Physiol* 42:239–274. [https://doi.org/10.1016/S0065-2911\(00\)42004-7](https://doi.org/10.1016/S0065-2911(00)42004-7).
47. Matsumi Y, Fujita K, Takashima Y, Yanagida K, Morikawa Y, Matsumoto-Nakano M. 2015. Contribution of glucan-binding protein A to firm and stable biofilm formation by *Streptococcus mutans*. *Mol Oral Microbiol* 30:217–226. <https://doi.org/10.1111/omi.12085>.
48. Ismail AI, Sohn W, Tellez M, Amaya A, Sen A, Hasson H, Pitts NB. 2007. The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dent Oral Epidemiol* 35:170–178. <https://doi.org/10.1111/j.1600-0528.2007.00347.x>.
49. Ling Z, Kong J, Jia P, Wei C, Wang Y, Pan Z, Huang W, Li L, Chen H, Xiang C. 2010. Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. *Microb Ecol* 60:677–690. <https://doi.org/10.1007/s00248-010-9712-8>.
50. Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714. <https://doi.org/10.1093/bioinformatics/btn025>.
51. Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simón-Soro A, Pignatelli M, Mira A. 2012. The oral metagenome in health and disease. *ISME J* 6:46–56. <https://doi.org/10.1038/ismej.2011.85>.
52. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>.
53. Noguchi H, Park J, Takagi T. 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 34:5623–5630. <https://doi.org/10.1093/nar/gkl723>.
54. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* (Oxford) 2010:baq013. <https://doi.org/10.1093/database/baq013>.
55. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
56. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jorgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clement K, Dore J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker J-D, Raes J, Hansen T, Kaas RS, Wang J, Ehrlich SD, Pedersen O, MetaHIT Consortium. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500:541–546. <https://doi.org/10.1038/nature12506>.
57. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto J-M, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezeur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, Renault P, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32:822–828. <https://doi.org/10.1038/nbt.2939>.
58. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
59. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
60. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
61. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–80. <https://doi.org/10.1093/nar/gkh063>.
62. Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36:D250–4. <https://doi.org/10.1093/nar/gkm796>.
63. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henriksen B, White O, Kelley ST, Methe B, Schloss PD, Gevers D, Mitreva M, Huttenhower C. 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 8:e1002358. <https://doi.org/10.1371/journal.pcbi.1002358>.
64. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13:R79. <https://doi.org/10.1186/gb-2012-13-9-r79>.
65. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. 2013. mRMR: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* 29:2365–2368. <https://doi.org/10.1093/bioinformatics/btt383>.