Caries-associated Salivary Microbiota of Children at Mixed Dentition from Different Geographic Locations

Shanshan Li
School of Stomatology, Qingdao University
Shi Huang
Qingdao Institute of Bioenergy and Bioprocess Technology Chinese Academy of Sciences
Ying Zhang
School of Stomatology, Qingdao University
Lijuan Zhang
Women & Children's Health Care Hospital of Linyi
Fan Li
School of Stomatology, Qingdao University
Kaixuan Tan
Qingdao Municipal Hospital
Jie Lu
Qingdao Municipal Hospital
Yi Guo
Tongji University
Zhenggang Chen
Qingdao Municipal Hospital
Qingyuan Guo
Qingdao Municipal Hospital
Yongping Tang
Qingdao Municipal Hospital
Fei Teng (✉ tengfei@qibebt.ac.cn)
Qingdao Institute of Bioenergy and Bioprocess Technology Chinese Academy of Sciences
Fang Yang (✉ yangf82@sina.com)
Qingdao Municipal Hospital

Research Article

Keywords: caries, geography, saliva microbiota, mixed dentition, diagnosis models

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-290144/v1

Caries-associated Salivary Microbiota of Children at Mixed Dentition from Different Geographic Locations

Shanshan Li[1*], Shi Huang[2,6*], Ying Zhang[1], Lijuan Zhang[4], Fan Li[1], Kaixuan Tan[3], Jie Lu[3], Yi Guo[5], Zhenggang Chen[3], Qingyuan Guo[3], Yongping Tang[3], Fei Teng[2,#], Fang Yang[1,3,#]

[1]School of Stomatology, Qingdao University, Qingdao, 266003, China
[2]Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology Chinese Academy of Sciences, Qingdao, Shandong, 266101, China
[3]Stomatology Center, Qingdao Municipal Hospital, Qingdao, Shandong, 266071, China
[4]Stomatology Department, Women & Children's Health Care Hospital of Linyi, Linyi 276002, China
[5]Department of Computer Science and Technology, the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, 200092, China
[6]Department of Pediatrics and Center for Microbiome Innovation at Jacobs School of Engineering, University of California, San Diego, USA

*These authors contributed equally to this work.
[#]Correspondence:
Corresponding Author
Fang Yang: yangf82@sina.com and Fei Teng: tengfei@qibebt.ac.cn

Abstract

The microbial composition of dental caries may depend on age, diet, and geography, yet the effect of geography on these microbiomes is largely underexplored. Here, we profiled and compared saliva microbiota from 130 individuals aged 6 to 8 years old, representing both healthy children (H) and children with severe caries (C) from two geographical regions of China: Qingdao Guangzhou. First, the saliva microbiota exhibited profound differences in diversity and composition between the C and H groups. The caries microbiota featured a lower alpha diversity and more variable community structure than the healthy microbiota. Furthermore, the relative abundance of several genera (e.g., *Lactobacillus*, *Gemella* and *Cryptobacterium*) was significantly higher in the C group than in the H group. Next, geography dominated over disease status in shaping salivary microbiota, and a wide array of salivary bacteria was highly predictive of the individuals' city of origin. Finally, we built a universal diagnostic model based on 14 bacterial species, which can diagnose caries with 87% and 85% accuracy within each city and 83% accuracy across cities. These findings demonstrated that despite the large effect size of geography, a universal model based on salivary microbiota has the potential to diagnose caries across human populations.

Key words: caries, geography, saliva microbiota, mixed dentition, diagnosis models

Introduction

Dental caries is one of the most prevalent chronic infectious diseases, affecting approximately half of children worldwide [1,2]. Once started, the damage to teeth is irreversible [3]. Severe caries, an aggressive form of dental caries, can lead to acute pain, sepsis, and potential tooth loss and even interfere with children's quality of life, nutrition, and school participation [4]. Therefore, preventive measures against caries, as well as improved tools for prognosis early diagnosis, are of particular clinical significance.

Human oral microbiome dysbiosis is increasingly implicated in various local and systemic human diseases, such as dental caries [5], gingivitis [6], and obesity [7]. The oral microbial composition depends on many factors, including age, diet, and geography. Accumulating evidence supports that changes in oral microbiota continue throughout human life [8-11], especially among three dentitions (i.e., deciduous/primary, mixed, and permanent dentition) [12,13]. Wim et al. found that *Prevotella* increased from deciduous, mixed, to permanent dentitions in healthy individuals, and there was a higher proportion of Proteobacteria in deciduous dentition than in mixed and permanent dentition [13]. Another study showed that *Lactobacillus spp*. and *Propionibacterium FMA5* were enriched in primary teeth from caries samples, while *Atopobium genomospecies C1* was enriched in permanent teeth [12]. The mixed dentition stage is a crucial transitional period during which deciduous teeth exfoliate successively and new permanent teeth erupt [14]. It is not only the main growth and development period of children's maxillofacial and dental arches but also subject to tremendous changes in host hormones and the immune system [15], which may promote maturation of oral microbiota [16]. Notably, most of the previous microbial studies were focused on early childhood or adult caries [5,8,17,18], and there are rare reports on the association of the oral microbiome with health and caries in mixed dentition [14,19,20].

Regarding geographical factors, former studies reported that adult populations from different continental regions or even countries had microbial variations in saliva [21,22], and supragingival microbiota differed among ethnic groups (i.e., African American, Burmese, Caucasian, and Hispanic) in children from the same geographic location (i.e., Burma) [23].

Early microbiota development has a significant impact on oral health and diseases of adulthood [24]. Understanding the oral microbiota differences in children in different geographic locations will shed light on the factors that might drive oral health disparities. However, the influence of geographic factors, such as city-scale differences, on the oral microbiome of healthy and diseased children is largely underexplored.

In this study, we address three general questions: (i) During the mixed dentition period, do oral communities assemble differently at different host states (i.e., healthy and caries)? (ii) How is bacterial diversity partitioned across biogeography, host states and biological gender? (iii) Should the geographic factor be taken into account when building classifiers to distinguish children with caries from healthy controls? Here, we conducted a comparison of the saliva microbiome from severe caries and healthy child cohorts between 6 and 8 years old from two cities in China (Qingdao and Guangzhou) by 16S rRNA gene sequencing (Figure 1). Ecological modeling techniques were further employed to dissect the role of saliva microbiota in caries and geography and probe the predictive value of the microbiome for diagnosing caries by identifying both biogeography- and disease-associated taxa.

Results

Dental Caries Altered Saliva Microbiota in the Mixed Dentition

To investigate whether and how caries affects oral microbiota in the mixed dentition stage, we first compared beta diversity within and between disease status (i.e., health and severe caries) and gender based on the Jensen-Shannon distances. We found that disease status exhibited a remarkable effect on shaping salivary microbiota (p<0.01, F=3.20) rather than gender (p>0.05; Figure 2A). Furthermore, the C group exhibited significant variability, while the H group was relatively conserved in microbial community structure (p<0.05; Figure 2B). Next, we assessed the impact of the disease status on the alpha diversity represented by Shannon, Simpson, and Pielou's evenness indices. The results showed that the alpha diversity was significantly lower in the C group than in the H group (all p<0.01; Figure 2C). Finally, we quantitatively profiled the bacterial taxa from the phylum to species level to characterize the mixed-dentition microbial composition (Figure S1A) and then tested whether there were any caries-enriched and caries-depleted taxa. All sequences were distributed in 13 bacterial phyla that included six predominant phyla (accounting for > 99% of the microbial diversity; Figure S1A), namely, *Firmicutes* (78.0%), *Actinobacteria* (11.9%), *Bacteroidetes* (5.0%), *TM7* (2.0%), *Proteobacteria* (1.6%) and *Fusobacteria* (1.4%). At the genus level, a total of 124 genera were identified, among which the most frequently detected genera (the four most abundant genera that each represented at least 5% in the average relative abundance) were *Streptococcus* (51.4%), *Gemella* (11.2%), *Actinomyces* (8.7%) and *Granulicatella* (5.8%; Figure S1A). Moreover, no 'caries-specific' taxon (present in one status but absent in the other) was detected between the two groups. At the genus level, *Lactobacillus*, *Gemella*, *Cryptobacterium* and *Mitsuokella* were found to have significantly higher relative abundances in the C group, while *Leptotrichia*, *Porphyromonas*, *Peptococcus*, *TM7*, and *Tannerella* were higher in the H group (all p<0.05, Figure S1B). At the species level, *Actinomyces IP073*, *Lactobacillus gasseri*, *Prevoella denticola*, *Propionbavterium FMA5*, *Streptococcus anginosus*, *Streptococcus mutans*, *Streptococcus sobrinus* and *Actinomyces gerencseriae* were found to have significantly higher relative abundances in the C group, while *Porphyromonas catoniae*, *Porphyromonas CW034*, *Propionibacterium propionicum*, *Tannerlla oral taxon 808*, *TM7 oral taxon 352* and *uncultured Lachnospiraceae oral taxon 100* were higher in the H group (all p<0.05, Figure S1C).

Geography Affected the Saliva Microbiota More than Caries Status

To elucidate the impact of geography on the oral microbiota, we included a second group of 34 age-matched children (17 with severe caries and 17 healthy subjects) from the southern city in China (Guangzhou group), approximately 1900 kilometers southwest of the northern city (Qingdao group; Materials and Methods). The analyses over the two cities showed that geography exhibited a higher effect on defining microbiota composition (p=0.001) than did caries status (p<0.05), and the two factors jointly explained up to 54% of the variation in microbiota, suggesting that they were the major factors shaping oral microbiota (Figure 3A). Furthermore, the Qingdao microbiota communities were more similar to each other than the Guangzhou microbiota: Guangzhou city samples showed higher within-group variability in Jensen-Shannon metrics (Figure 3B; p<0.05), and Qingdao city samples were significantly more diverse in Shannon indices than that in Guangzhou (Figure 3C; p<0.05).

To identify geography-specific markers contributing to predicting city origins, we first built classification models via the random forest (RF) machine learning algorithm using healthy samples as the training set. The city origin was predicted from healthy samples with 78.88% accuracy (area under the concentration curve [AUC]: 97.30%; CI: 93.80%-100.00%, Figure 4A). The probability of Guangzhou city was significantly higher in Guangzhou city samples than in Qingdao city samples from the H group (Wilcoxon test, p<0.05, Figure 4B). Next, the RF model ranked the contribution of each predictor based on the variable importance, where we can identify the most discriminatory bacteria between two cities. Performance improvement was minimal when the top eight most discriminatory species were included (Figure 4C). Eight geography-specific marker species underlying the power of the healthy model were identified, namely, *Veillonella atypica/dispar/parvula*, *Granulicatella elegans*, *Corynebacterium durum*, *Rothia aeria*, *Bergeyella 602D02*, *Granulicatella adiacen*, *Peptostreptococcus stomatis* and *Streptococcus parasanguinis oralis* (Figure 4D). Among them, the relative abundance of the former five species was higher in Qingdao city samples than in the Guangzhou city samples (Wilcoxon test, adjusted p<0.05), while that of the latter three taxa

significantly increased in Qingdao city samples (Wilcoxon test, adjusted $p<0.05$, Figure S2A). Moreover, these taxa were shared in caries samples, representing 12.88% and 13.13% abundance for healthy and caries samples, respectively. Finally, application of the eight-marker-based model on the caries samples resulted in 92.31% accuracy (AUC: 95.00%; Figure S2B), and the probability of Guangzhou city was significantly higher in the Guangzhou samples than in the Qingdao samples from the C group (Wilcoxon test, $p<0.05$; Figure S2C). Thus, geography-specific differences in the salivary microbiome were consistent, irrespective of health status.

A universal disease diagnosis model for all samples across geographic locations

Consistent with the results for the Qingdao city samples, a reduction in alpha diversity was associated with caries ($p<0.05$; Shannon index; Figure S3A) in all samples from both cities, and the beta diversity was distinct between caries and healthy microbiota ($p<0.05$, F=1.00; Figure S3B). These results suggested the feasibility of caries diagnosis based on oral microbiota in different geographic locations.

There were three strategies to construct and optimize the caries diagnosis model. First, to test the effect of taxonomic level on the discriminatory power of the RF model, the models were constructed based on taxa at the phylum, genus, and species levels to discriminate between healthy and caries samples using two city datasets. We found that the use of species-level taxa maximized (AUC: 88.56%, CI: 83.56%−94.61%) compared with that of the others at the phylum (AUC: 64.11%, CI: 54.54%−73.67%) and genus (AUC:77.61%, CI: 69.48%−85.74%) levels (Figure S4). Second, to test whether differences in oral microbiota in caries were consistent by city, we built RF models in each city (i.e., Qingdao and Guangzhou) and achieved diagnosis accuracies of 84.38% and 76.47%. Furthermore, training a diagnosis model in one dataset and applying it to another led to lower yet still decent and meaningful performance (Figure S5). Specifically, application of the Qingdao model (i.e., the Qingdao cohort as training data) on the Guangzhou dataset led to a reduction in the AUC from 91.10% to 83.00%, and similarly, application of the Guangzhou model (i.e., the Guangzhou cohort as training data) on the Qingdao dataset led to a reduction in the AUC from 85.81% to 80.00% (Figure S5). Third, we built RF models using all caries and healthy samples from the two geographic locations. Unexpectedly, excluding eight geography-specific signatures from the species profile rarely affected the classification performance, with AUCs from 88.56% to 88.99% (Figure 5A). Moreover, intriguingly, these most discriminatory taxa associated with caries state did not show correlation with geography in the healthy samples (Figure 5C) and vice versa (Figure S6) in either the geographic or caries diagnosis model. Underlying the power of the model using the species-level profile that ruled out these geographic signatures, fourteen bacterial species markers were identified based on both the rank order of important scores (Figure 5B and 5C) and the Wilcoxon test results (adjusted $p<0.05$; Figure S1C). Among them, eight taxa (i.e., *Streptococcus mutans*, *Actinomyces gerencseriae*, *Propionibacterium FMA5*, *Actinomyces IP073*, *Streptococcus anginosus*, *Lactobacillus gasseri*, *Streptococcus sobrinus*, and *Prevotella denticola*) were caries-enriched, while the other six taxa (i.e., *Tannerella oral taxon 808*, *Propionibacterium propionicum*, *Uncultured Lachnospiraceae oral taxon 100*, *Porphyromonas CW034*, *Porphyromonas catoniae* and *TM7 oral taxon 352*) were caries-depleted (adjusted $p<0.05$; Figure S1C). Consequently, we constructed the final caries diagnosis model based on the fourteen species selected, which led to an increase in predictive performance in Qingdao city (AUC: 91.02%; CI: 85.27%-96.76%; Figure 6A), Guangzhou city (AUC: 86.16%; CI: 71.57%-100.00%; Figure 6A) and across two cities (AUC: 92.17%; CI: 87.45%-96.88%; Figure 6B). Notably, *Streptococcus mutans* (*S. mutans*) with the top importance score in the model (Wilcoxon test, adjusted $p<0.05$, Figure 5C and Figure S1C) has previously been documented to play a critical role in caries pathogenesis. Using only *S. mutans* as a predictor, the simplified random forest model led to a lower yet decent performance (AUC=81.62%, CI: 74.40%-88.84%, Figure S7). However, *S. mutans* was not detected in any of the samples (the occurrence rate in the caries sample=78.5%, the occurrence rate in the healthy sample=30.8), as well as the others (Figure S8), suggesting that dental caries is not associated with a single taxon but in fact with a complex community.

Discussion

It has been well documented that in dental caries, environmental perturbation alters the balance of the oral microbiota and eventually leads to a predominance of cariogenic bacteria, resulting in sustained demineralization of tooth hard tissue [25]. Evidence has recently emerged that the oral microbiome may depend on age, oral dentition, diet and geography [8,13,26-28]. Effectively reducing dental caries burden requires a better understanding of its determinants. To address this issue, we profiled and compared saliva microbiota from 130 individuals aged 6 to 8 years old, representing both severe caries and healthy control children from two geographical regions of China: a northern city (Qingdao) and a southern one (Guangzhou).

First, we characterized the dysbiotic saliva microbiome in caries in human populations in terms of alpha diversity, beta diversity and bacterial composition. Similar to our observations, previous studies of various age stages of individuals have shown that caries status favored reduced microbial diversity [8,29,30]. Such a reduction in alpha diversity is likely caused by increased carbohydrate consumption and fermentation, leading to acid production and secretion. The low-pH environment probably selects acidogenic and aciduric taxa, which could thrive under the condition [31]. The beta-diversity analysis showed that saliva microbial communities significantly differed between diseased and healthy children. Moreover, caries children also have higher Jensen-Shannon distances than healthy children. This is likely because caries

microbiomes have higher intro-group variation and more personalized microbiomes than healthy microbiomes, which are more similar to each other [5]. Moreover, our data substantiate existing evidence that organisms other than *Streptococcus mutans* and *Lactobacilli* play a role in the development and progression of dental caries. At the genus level, the caries microbiome harbored a higher abundance of *Lactobacillus, Gemella* and *Cryptobacterium* than healthy controls, which is in line with previous studies [32-34]. At the species level, the increase in *non-mutans streptococci* (i.e., *S. anginosus* and *S. sobrinus*) and *Actinomyces_gerencseriae* in the C group was not surprising. They were recognized as acidogenic and aciduric bacteria, which have been reported to produce weaker acid resulting in caries initiation and thrive during caries progression in low pH conditions (e.g., pH=5.0; [19,35,36]. Similarly, according to our and other studies [5], *Prevotella denticola* was significantly enriched in caries and was identified as the main predictor of caries, which potentially have proteolytic/amino acid-degrading activities. *Propionibacterium FMA5* was implicated in dental caries from young permanent teeth [18]and root caries from elderly individuals [37]. In addition, *S. mutans* was identified in relatively low abundance, and the detection rate was relatively low (AUC=81.62%). Consistently, previous studies found that despite a significant enrichment of *S. mutans* with caries development, several bacteria were far more abundant in the carious lesions [38]. Our findings illustrated that dental caries in the mixed dentition resulted from widespread shifts in the oral microbial community instead of any particular taxa from healthy to diseased status, supporting the "ecological plaque hypothesis" [35,39].

Next, our data included children from two cities of China: Qingdao (N group) and Guangzhou (S group), between which the distance was approximately 1900 kilometers. We found that alpha and beta diversity in background oral microbiomes are radically distinct across geographic locations. Thus, geography accounted for the highest variance in the salivary bacterial profiles compared to other confounding factors, such as caries state or host gender. In previous studies, saliva microbial profiles can vary greatly across large-scale geographic locations (e.g., the continental region [21] or country [22]) or by ethnicity within one nation [40]. However, few studies to date have systematically investigated the oral microbiome from the mixed dentition of Chinese subjects residing in different cities. This makes it challenging to directly compare caries microbiomes across studies and test the generalizability of microbiome-based diagnosis models across geography. Geography is a considerable yet complex factor influencing the development of microbiome-based diagnostic models. First, diet can contribute greatly to geographic differences across populations. The diet in Qingdao city typically encompasses a wider variety of carbohydrates than that in Guangzhou city, and fatty foodstuffs may supply a more complex array of substrates and allow more diverse bacterial species to thrive in the oral cavity. [22]. Additionally, these unique food nutrients have an indispensable effect on the microbial ecology of dental caries [41]. Second, the population composition in cities may largely determine intraindividual microbiome variation. With the rapid economic development in Guangzhou, an increasing number of citizens from other parts of China have migrated to well-developed southern cities to seek job opportunities, which has resulted in higher population-level diversity in Guangzhou city. In contrast, those in Qingdao city reflected a more homogenous group from a relatively restricted area. This might be a plausible explanation why we observed a higher interindividual microbiome diversity in Guangzhou's population. Although the mechanism for the city-dependent microbiome remains obscure, host genetics, climate, dietary patterns, built environments [42] and other epidemiological factors should be further considered in developing a microbiome-based diagnostic model of ECC [43].

Finally, we built a universal classification to diagnose caries using oral bacterial species by appropriately detrending the geographic effect in microbiome data. Despite the considerable differences between the two cohorts, a caries diagnosis model built from a single city can still be applied across the two cities with decent accuracy. Moreover, although geographic factors showed a larger effect size in defining oral microbiome data than caries state, city-specific markers had little impact on the prediction performance of caries classification models. Intriguingly, disease-specific biomarkers showed no correlation with geography. These results suggested the feasibility of universal caries diagnosis independent of geographic distances among populations. As a result, the caries diagnosis model consisting of the top 14 bacterial oral species can reliably diagnose caries with 83.08% accuracy (AUC=92.17%) across cities. Our previous and other studies have verified the diagnostic and predictive efficacy of oral microbiota using random forest classification models in deciduous and permanent dentition [6,8,44,45]. Together, these results suggested that caries diagnosis models were biogeography-independent using saliva microbial profiles.

There were several limitations in the current study, and different factors might affect the results. For example, (*i*) the sample size of Guangzhou city here is relatively small, which should be increased to allow a better statistical comparison of the microbial diversity with Qingdao (n=96) samples. (*ii*) Cross-sectional data to examine the link between disease status and geography with dental caries are relatively limited. To further test the causal relationship, a longitudinal design should be conducted. (*iii*) Whether universal diagnosis models are generally applicable in other cities is not yet clear, and more environments and geographic regions must be observed. Future efforts tackling these questions are key to more precise dental caries therapies.

Conclusions

To our knowledge, this is the first study to use current molecular techniques to the differences between the bacterial composition of the saliva microbiota in mixed dentitions of severe caries and healthy children living in different geographic locations: either Qingdao or Guangzhou of China. Using machine learning approaches, we also revealed that although geography has the most remarkable effect size on salivary microbiota (the saliva microbiome can predict the originated city with near 100% accuracy), a universal model based on fourteen bacterial species can diagnose caries with 83.08% accuracy across cities (area under the concentration-time curve [AUC], 92.17%). Our study underscores the possibility of employing saliva microbiota for a universal diagnosis method, which can be probed for other dentition stages of oral caries and for caries in other geographic locations.

Materials and Methods

Study Design and Sample Collection

This study was reviewed and approved by the Ethical Committee of Qingdao University (Qingdao, China),and followed the Declaration of Helsinki. Written informed consent was obtained from the legal parents or other guardians of all participants prior to enrollment. All experiments were performed following relevant guidelines and regulations. The children employed in this study were from two primary schools in the northern city (Qingdao, Shandong Province) and the southern city (Guangzhou, Guangdong Province) in mainland China. They were all unrelated students of both genders, aged between 6 and 8 years old and shared a relatively homogeneous primary school campus living environment. After oral clinical examination, 96 (Qingdao) and 34 (Guangzhou) children were chosen for saliva sample collection. Among these, the Qingdao-originated samples were from 48 severe caries (DMFT≧6) and 48 healthy (DMFT=0) subjects, while the Guangzhou-originated samples were from 17 severe caries (DMFT≧6) and 17 healthy (DMFT=0) children. All reported no antibiotic intake for at least the preceding 6 months and were asked to avoid eating or drinking for 1 h before oral sampling. Each sample was collected by expectorating approximately 3 ml saliva into sterile plastic 50 ml tubes. Then, we individually numbered and sealed the samples, placed them in a 4°C sample storage tank, and stored them in a freezer at -80°C for long-term storage.

DNA Extraction, PCR Amplification, and Sequencing of the Oral Microbiome

Microbial genomic DNA was isolated using lysozyme-containing enzymatic lysis buffer and zirconia-silica beads (BioSpec, Bartlesville, OK) and a DNeasy® Blood and Tissue Kit (Qiagen Valencia, CA). The V1-V3 hypervariable regions of the 16S rRNA gene were subjected to high-throughput sequencing at Beijing Auwigene Tech, Ltd. (Beijing, China) using the Illumina MiSeq PE300 sequencing platform (Illumina, Inc., CA, USA). PCR amplification of the V3-V4 region of the bacterial 16S rRNA gene was performed using universal primers 5'-TGGAGAGTTTGATCCTGGCTCAG-3' (forward) and 5'-TACCGCGGCTGCTGGCAC-3' (reverse) incorporating a sample barcode sequence. The PCR conditions were as follows: 2 min initial denaturation at 95°C; 25 cycles of denaturation at 94°C (30 s), annealing at 56°C (25 s), and elongation at 72°C (25 s); and final extension at 72°C for 5 min. The PCR products were separated by 1.2% agarose gel electrophoresis, and the approximately 500 bp fragments were purified using Agencourt AMPure XP (Beckman Coulter, Inc., CA, USA). Sequencing was performed using Roche 454 FLX Titanium chemistry.

Raw sequencing data were processed by Beijing Auwigene Tech, Ltd. (Beijing, China) using the pipeline tools MOTHUR [46] and QIIME [47], and pyrosequencing data were analyzed using customized R scripts. Noise reduction was carried out using MOTHUR. The sequences were binned into operational taxonomic units (OTUs) with 97% similarity. OTUs are groups of sequences that are clustered based on similarity, allowing taxonomic assignment.

Statistical Analysis

Overall, the saliva microbiota was compared in two dimensions: (*i*) between severe caries samples and healthy controls to discover the potential microbial factors associated with caries and (*ii*) between samples from Qingdao and Guangzhou cities to identify the effect size of geography on saliva microbiota. The Jensen-Shannon distance metric (JSD) was used to visualize the differential distribution of the between-microbiome difference between sampling groups (e.g., diseased states, city of origin). PERMANOVA analyses were further applied to determine the significance (p-value) and strength (F values) of a given confounding factor in explaining the variation in the oral microbiome. The pairwise p-values from Adonis were corrected for multiple comparisons. To compare the quantitative data in the alpha and beta diversity analysis and biomarker selection, the Kruskal-Wallis rank-sum test was used, and p-values were corrected via false discovery rate (FDR) for multiple pairwise comparisons.

Building the Diagnostic Models of Caries

Random forest (RF) was applied to identify features that are differentially abundant (i.e., present in different abundances) across sample groups and diagnosis models. The N top-ranking caries-discriminatory taxa and geography-discriminatory taxa that led to reasonably good fit were identified based on the 'rfcv' function in the random forest package

(https://cran.rproject.org/web/packages/ randomForest/ index.html). RF models were trained to identify disease status in the training set, which included samples from the healthy and severe caries groups using the taxonomy profiles. The results were evaluated with a 10-fold cross-validation approach, and model performance was evaluated by receiver operating characteristic (ROC) curves. Using the species profiles, the performance of the models based on microbiota was evaluated with a 10-fold cross-validation approach where the original samples were randomly partitioned into 10 groups with a similar distribution of healthy and caries samples. In each cross-validation iteration, nine groups of samples were used as training data and tested samples in the remaining group. The cross-validation process was then repeated 10 times, and per-sample prediction was reported as ones in the test fold. Based on the optimization step that selects the taxonomic level that maximizes model performance, the final RF models were based on the taxonomic profiles at the species level. ROC analysis was then used to evaluate the diagnostic performance of the RF models (https://cran.r-project.org/web/packages/pROC/ index.html). In the ROC plots, the x axis represents the true-positive rate (TPR, or sensitivity), and the y axis presents the false-positive rate (FPR, or specificity). The area under the ROC curve (AUC) was calculated to quantify the performance of the RF model.

Conflict of Interest

The authors declare that they have no competing interests.

Author contributions

S.L., S.H., F.Y., F.T., Z.C., Q.G.and Y.T. designed the study; L.Z., F.L.and Y.Z. performed clinical examination and sample collection; K.T. and J.L.performed sample processing and 16S rDNA pyrosequencing; S.H., F.T. contributed to statistical analysis methods; S.L., F.Y., Y. G. and F.T. performed bioinformatics analysis; S.L.,F.Y., F.T.wrote the paper. All authors have reviewed and approved the final version of the manuscript.

Acknowledgments

Our work was supported in part by grants 81670979 and 31600099 from the Natural Science Foundation of China. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Reference

1 Min Quan, D. *et al.* Dental caries status and its associated factors among 3-to 5-year-old children in China: a National Survey. **21**, 167-179 (2018).

2 Quan, J. K. *et al.* Permanent teeth caries status of 12-to 15-year-olds in China: findings from the 4th National Oral Health Survey. **21**, 181-193 (2018).

3 Selwitz, R. H., Ismail, A. I. & Pitts, N. B. Dental caries. *Lancet* **369**, 51-59, doi:10.1016/S0140-6736(07)60031-2 (2007).

4 Li, M. Y., Zhi, Q. H., Zhou, Y., Qiu, R. M. & Lin, H. C. Impact of early childhood caries on oral health-related quality of life of preschool children. *Eur J Paediatr Dent* **16**, 65-72 (2015).

5 Yang, F. *et al.* Saliva microbiomes distinguish caries-active from healthy human populations. *The ISME journal* **6**, 1-10, doi:10.1038/ismej.2011.71 (2012).

6 Huang, S. *et al.* Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *The ISME journal* **8**, 1768-1780, doi:10.1038/ismej.2014.32 (2014).

7 Yang, F. *et al.* Caries experience and its association with weight status among 8-year-old children in Qingdao, China. *Journal of International Society of Preventive & Community Dentistry* **5**, 52-58, doi:10.4103/2231-0762.151978 (2015).

8 Teng, F. *et al.* Prediction of Early Childhood Caries via Spatial-Temporal Variations of Oral Microbiota. *Cell host & microbe* **18**, 296-306, doi:10.1016/j.chom.2015.08.005 (2015).

9 Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with their dogs. *Elife* **2**, e00458, doi:10.7554/eLife.00458 (2013).

10 Ling, Z., Liu, X., Wang, Y., Li, L. & Xiang, C. Pyrosequencing analysis of the salivary microbiota of healthy Chinese children and adults. *Microb Ecol* **65**, 487-495, doi:10.1007/s00248-012-0123-x (2013).

11 Xu, X. *et al.* Oral cavity contains distinct niches with dynamic microbial communities. *Environmental microbiology* **17**, 699-710, doi:10.1111/1462-2920.12502 (2015).

12 Aas, J. A. *et al.* Bacteria of dental caries in primary and permanent teeth in children and young adults. *Journal of clinical microbiology* **46**, 1407-1417, doi:10.1128/jcm.01410-07 (2008).

13 Crielaard, W. *et al.* Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC medical genomics* **4**, 22, doi:10.1186/1755-8794-4-22 (2011).

14 Shi, W., Qin, M., Chen, F. & Xia, B. Supragingival Microbial Profiles of Permanent and Deciduous Teeth in Children with Mixed Dentition. *PLoS One* **11**, e0146938, doi:10.1371/journal.pone.0146938 (2016).

15 Bimstein, E. & Matsson, L. J. P. D. Growth and development considerations in the diagnosis of gingivitis and periodontitis in children. **21**, 186 (1999).

16 Feres, M., Teles, F., Teles, R., Figueiredo, L. C. & Faveri, M. The subgingival periodontal microbiota of the aging mouth. *Periodontology 2000* **72**, 30-53, doi:10.1111/prd.12136 (2016).

17 Ling, Z. *et al.* Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. *Microbial ecology* **60**, 677-690, doi:10.1007/s00248-010-9712-8 (2010).

18 Gross, E. L. *et al.* Bacterial 16S sequence analysis of severe caries in young permanent teeth. *Journal of clinical microbiology* **48**, 4121-4128, doi:10.1128/jcm.01232-10 (2010).

19 Xu, Y., Jia, Y. H., Chen, L., Huang, W. M. & Yang, D. Q. Metagenomic analysis of oral microbiome in young children aged 6-8 years living in a rural isolated Chinese province. *Oral diseases* **24**, 1115-1125, doi:10.1111/odi.12871 (2018).

20 Shi, W., Tian, J., Xu, H., Zhou, Q. & Qin, M. Distinctions and associations between the microbiota of saliva and supragingival plaque of permanent and deciduous teeth. *PLoS One* **13**, e0200337, doi:10.1371/journal.pone.0200337 (2018).

21 Nasidze, I., Li, J., Quinque, D., Tang, K. & Stoneking, M. Global diversity in the human salivary microbiome. *Genome research* **19**, 636-643, doi:10.1101/gr.084616.108 (2009).

22 Li, J. *et al.* Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. **14**, 316 (2014).

23 Premaraj, T. S. *et al.* Ethnic variation of oral microbiota in children. *Scientific reports* **10**, 14788, doi:10.1038/s41598-020-71422-y (2020).

24 Gomez, A. & Nelson, K. E. The Oral Microbiome of Children: Development, Disease, and Implications Beyond Oral Health. *Microbial ecology* **73**, 492-503, doi:10.1007/s00248-016-0854-1 (2017).

25 Kidd, E. & Fejerskov, O. Changing concepts in cariology: forty years on. *Dental update* **40**, 277-278, 280-272, 285-276, doi:10.12968/denu.2013.40.4.277 (2013).

26 Dashper, S. G. *et al.* Temporal development of the oral microbiome and prediction of early childhood caries. *Scientific reports* **9**, 19732, doi:10.1038/s41598-019-56233-0 (2019).

27 Hansen, T. H. *et al.* Impact of a vegan diet on the human salivary microbiota. *Scientific reports* **8**, 5847, doi:10.1038/s41598-018-24207-3 (2018).

28 Mark Welch, J. L., Ramírez-Puebla, S. T. & Borisy, G. G. Oral Microbiome Geography: Micron-Scale Habitat and Niche. *Cell host & microbe* **28**, 160-168, doi:10.1016/j.chom.2020.07.009 (2020).

29 Li, Y., Ge, Y., Saxena, D. & Caufield, P. W. Genetic profiling of the oral microbiota associated with severe early-childhood caries. *Journal of clinical microbiology* **45**, 81-87, doi:10.1128/jcm.01622-06 (2007).

30 de Jesus, V. C. *et al.* Sex-Based Diverse Plaque Microbiota in Children with Severe Caries. *Journal of dental research* **99**, 703-712, doi:10.1177/0022034520908595 (2020).

31 Marsh, P. D. Dental diseases--are these examples of ecological catastrophes? *International journal of dental hygiene* **4 Suppl 1**, 3-10; discussion 50-12, doi:10.1111/j.1601-5037.2006.00195.x (2006).

32 Ledder, R. G., Kampoo, K., Teanpaisan, R. & McBain, A. J. Oral Microbiota in Severe Early Childhood Caries in Thai Children and Their Families: A Pilot Study. *Frontiers in microbiology* **9**, 2420, doi:10.3389/fmicb.2018.02420 (2018).

33 Jiang, W. *et al.* Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood. *Microbial ecology* **67**, 962-969, doi:10.1007/s00248-014-0372-y (2014).

34 Ma, C. *et al.* Comparison of oral microbial profiles between children with severe early childhood caries and caries-free children using the human oral microbe identification microarray. *PLoS One* **10**, e0122075, doi:10.1371/journal.pone.0122075 (2015).

35 Takahashi, N. & Nyvad, B. The role of bacteria in the caries process: ecological perspectives. *Journal of dental research* **90**, 294-303, doi:10.1177/0022034510379602 (2011).

36 Svensäter, G., Borgström, M., Bowden, G. H. & Edwardsson, S. The acid-tolerant microbiota associated with plaque from initial caries and healthy tooth surfaces. *Caries Res* **37**, 395-403, doi:10.1159/000073390 (2003).

37 Preza, D. *et al.* Bacterial profiles of root caries in elderly patients. *Journal of clinical microbiology* **46**, 2015-2021, doi:10.1128/jcm.02411-07 (2008).

38 Simón-Soro, A., Belda-Ferre, P., Cabrera-Rubio, R., Alcaraz, L. D. & Mira, A. A tissue-dependent hypothesis of dental caries. *Caries Res* **47**, 591-600, doi:10.1159/000351663 (2013).

39 Marsh, P. D. Are dental diseases examples of ecological catastrophes? *Microbiology (Reading, England)* **149**, 279-294, doi:10.1099/mic.0.26082-0 (2003).

40 Mason, M. R., Nagaraja, H. N., Camerlengo, T., Joshi, V. & Kumar, P. S. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. *PloS one* **8**, e77287, doi:10.1371/journal.pone.0077287 (2013).

41 Navia, J. M. J. T. A. j. o. c. n. Carbohydrates and dental health. **59**, 719S-727S (1994).

42 Chase, J. *et al.* Geography and Location Are the Primary Drivers of Office Microbiome Composition. *mSystems* **1**, doi:10.1128/mSystems.00022-16 (2016).

43 Gupta, V. K., Paul, S. & Dutta, C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. **8**, doi:10.3389/fmicb.2017.01162 (2017).

44 Wang, Y. *et al.* Oral Microbiome Alterations Associated with Early Childhood Caries Highlight the Importance of Carbohydrate Metabolic Activities. *mSystems* **4**, doi:10.1128/mSystems.00450-19 (2019).

45 Hemadi, A. S., Huang, R., Zhou, Y. & Zou, J. Salivary proteins and microbiota as biomarkers for early childhood caries risk assessment. *International journal of oral science* **9**, e1, doi:10.1038/ijos.2017.35 (2017).

46 Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* **75**, 7537-7541, doi:10.1128/aem.01541-09 (2009).

47 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).

Figures

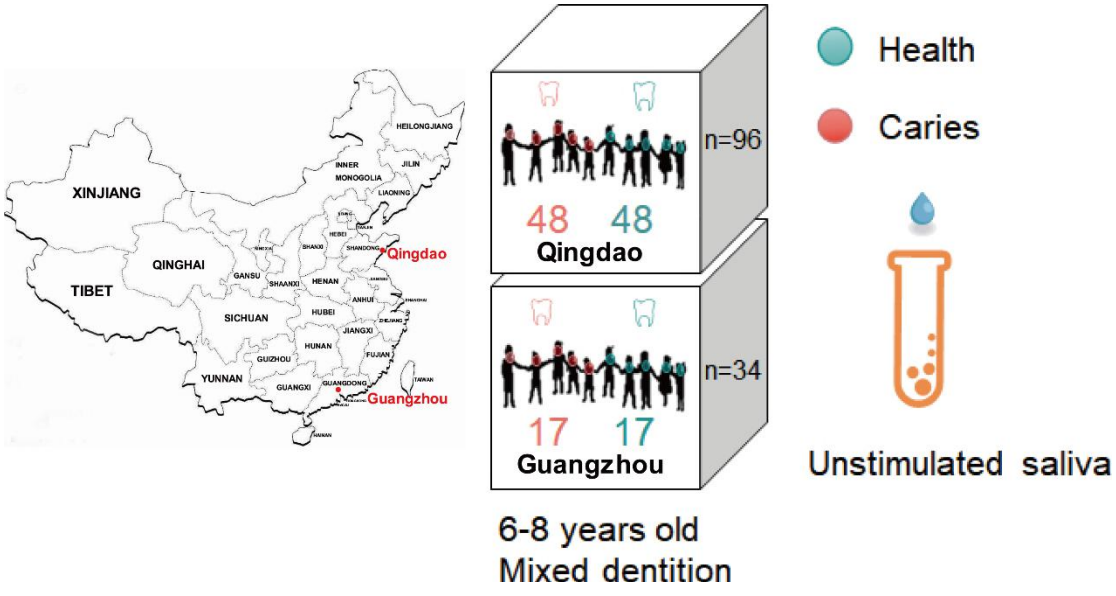


Figure 1. Experimental design that sampled saliva microbiome from caries-active children and healthy controls in the two Chinese cities of Qingdao and Guangzhou. Unstimulated saliva microbiota from 130 individuals (Qingdao, n=96; Guangzhou, n=34) were compared.

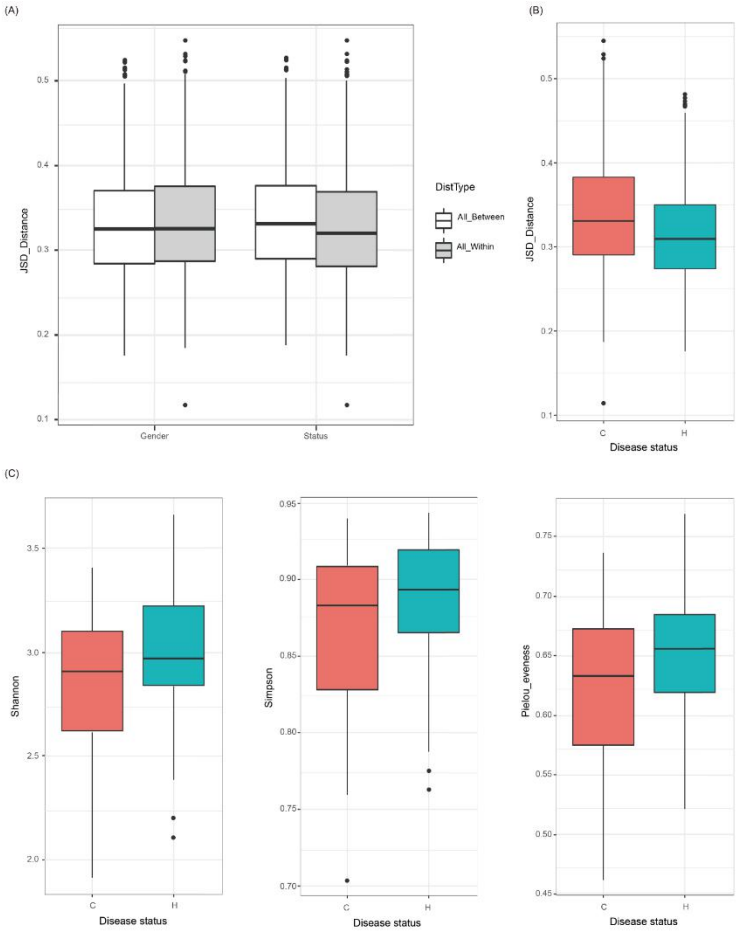


Figure 2. Oral microbial diversity comparisons between caries and healthy children in Qingdao cohort. (**A**) Salivary microbiota variation was compared within and between disease status (i.e., H or C), or gender based on the Jensen-Shannon distances. (**B**) Caries-free children have more conservative microbiota than do children with caries (***p*<0.01). (**C**) Alpha diversity comparisons between the C and H groups using on Shannon, Simpson, and Pielou's evenness index.

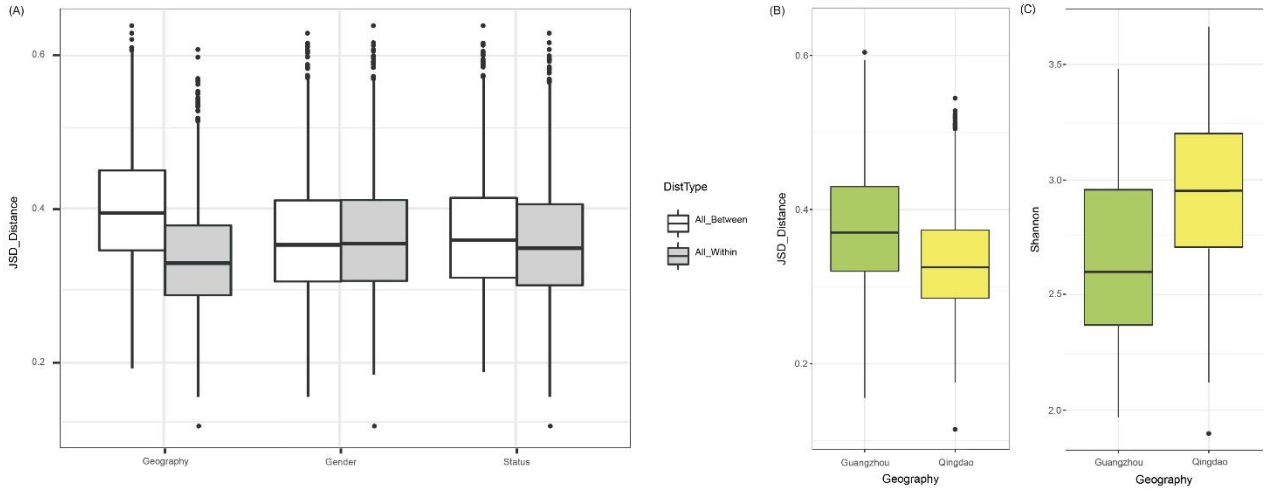


Figure 3. The remarkable impact of city of origin on oral microbiomes. (**A**) The effect size of geography, gender and host's disease status on saliva microbiota based on Jensen-Shannon distance. The city of origin exhibited the strongest effect on bacterial composition of the saliva microbiome, followed by host status and gender factor. (**B**) Beta-diversity difference between Qingdao and Guangzhou groups measured by JSD distances. (**C**) Alpha diversity difference between Qingdao and Guangzhou groups measured by Shannon index.

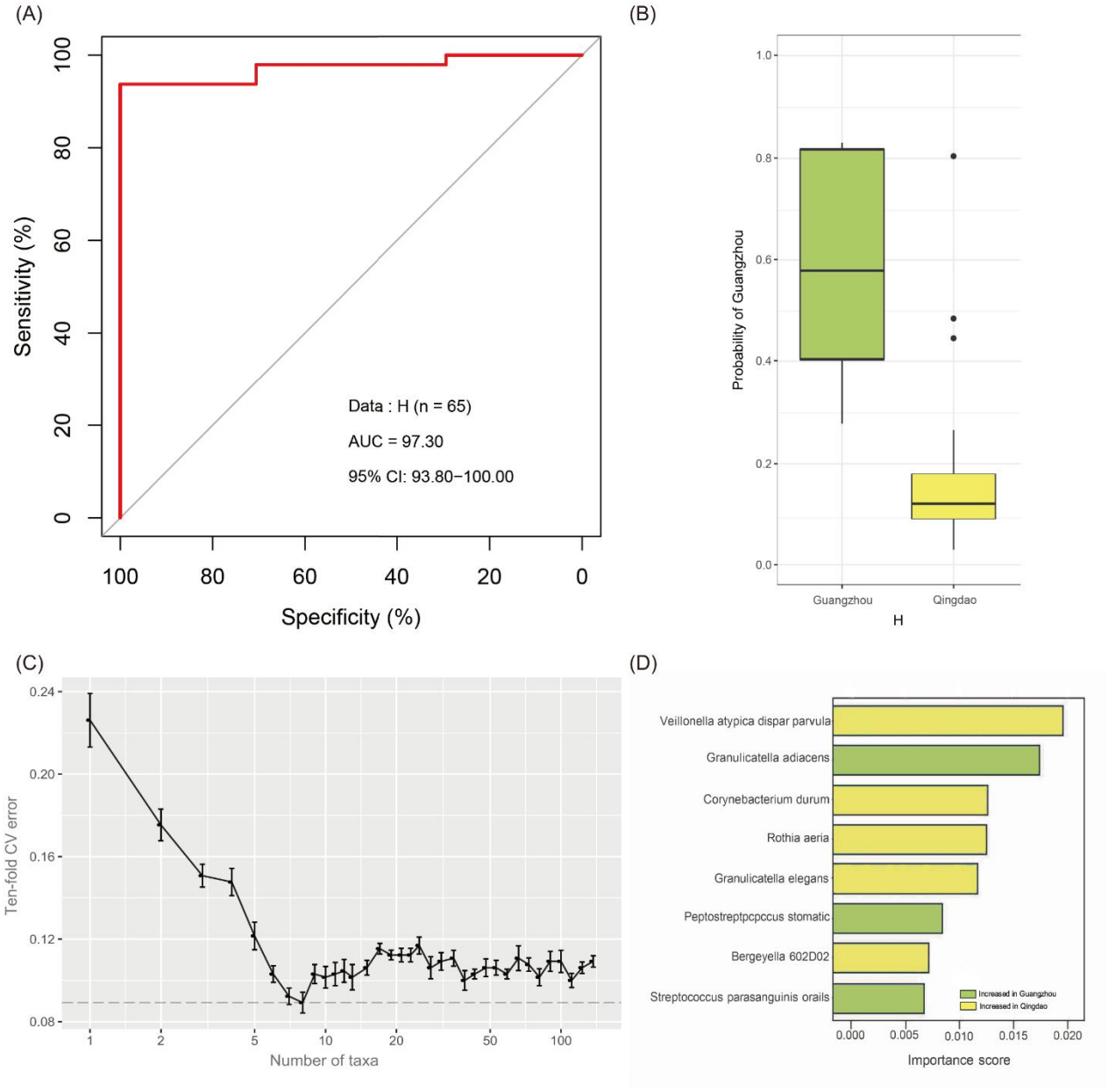
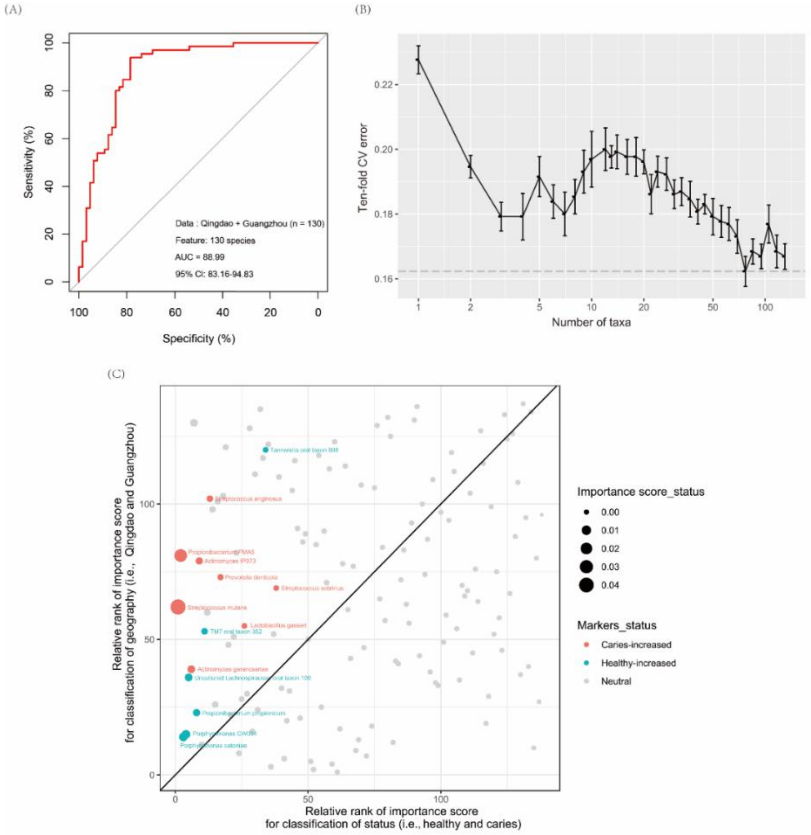
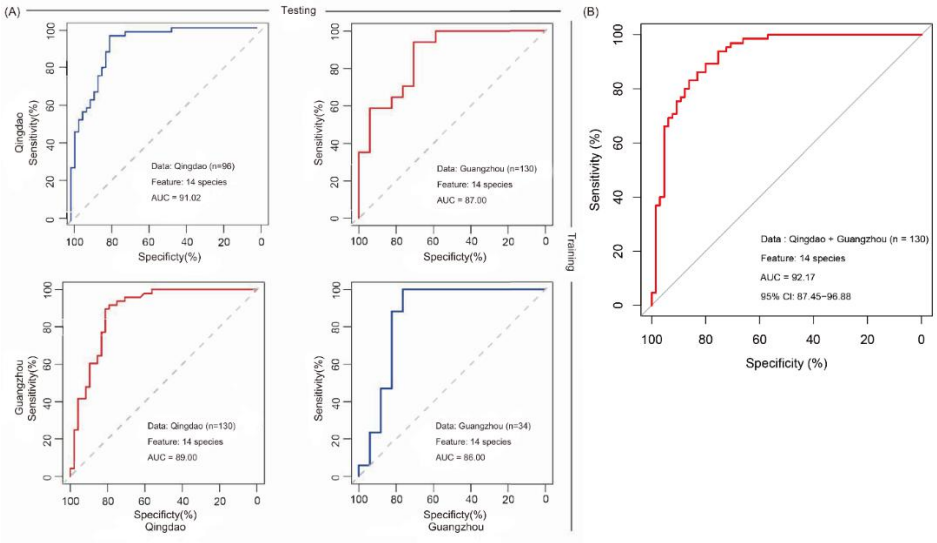


Figure 4. The strong geographical background of the healthy oral microbiota and key drivers. (**A**) Microbiome can classify the city of origin of healthy samples with a high accuracy. (**B**) Box plot indicates the prediction probability of Guangzhou city in healthy samples. (**C**) Relationship between the numbers of variables used in the reduced models and the corresponding predictive performance (the error bar denotes SD). (**D**) The importance score of eight the most discriminating species in the diagnosis model to predict city origin. The bar length at each row indicates relative contribution of the species to the RF model.



423

424 **Figure 5. Caries diagnostic models based on oral microbiome detrended for geography.** (**A**) Saliva microbiota can
425 predict caries status with a remarkably high accuracy (AUC=92.17%). (**B**) The relationship between the numbers of
426 variables used in the reduced Random Forest model and the corresponding predictive performance (the error bar denotes
427 SD). (**C**) The most caries-discriminatory taxa (N=14) do not correlate with geography. The scatterplot shows the relative
428 rank of microbial markers in both Random Forest models for classifying disease status and geographic locations. Any
429 dots on the reference line which slope=1 suggests a taxon is equally important to both disease states and geography.



430

431 **Figure 6. Cross-applications of caries diagnosis models based on microbiomes from Qingdao and Guangzhou**
432 **cohorts.** (**A**) The prediction performance of models in the Qingdao (AUC=91.02%, Guangzhou (AUC=86.16%%) and
433 model application from one city to another. A classification model trained in Qingdao data and tested in Guangzhou Data
434 resulting in a AUC=87.00; A classification model trained in Guangzhou data and tested in Qingdao Data resulting in a
435 AUC=89.00%. (**B**) The predictive performance using data from two cities (AUC=92.17%).

Figures

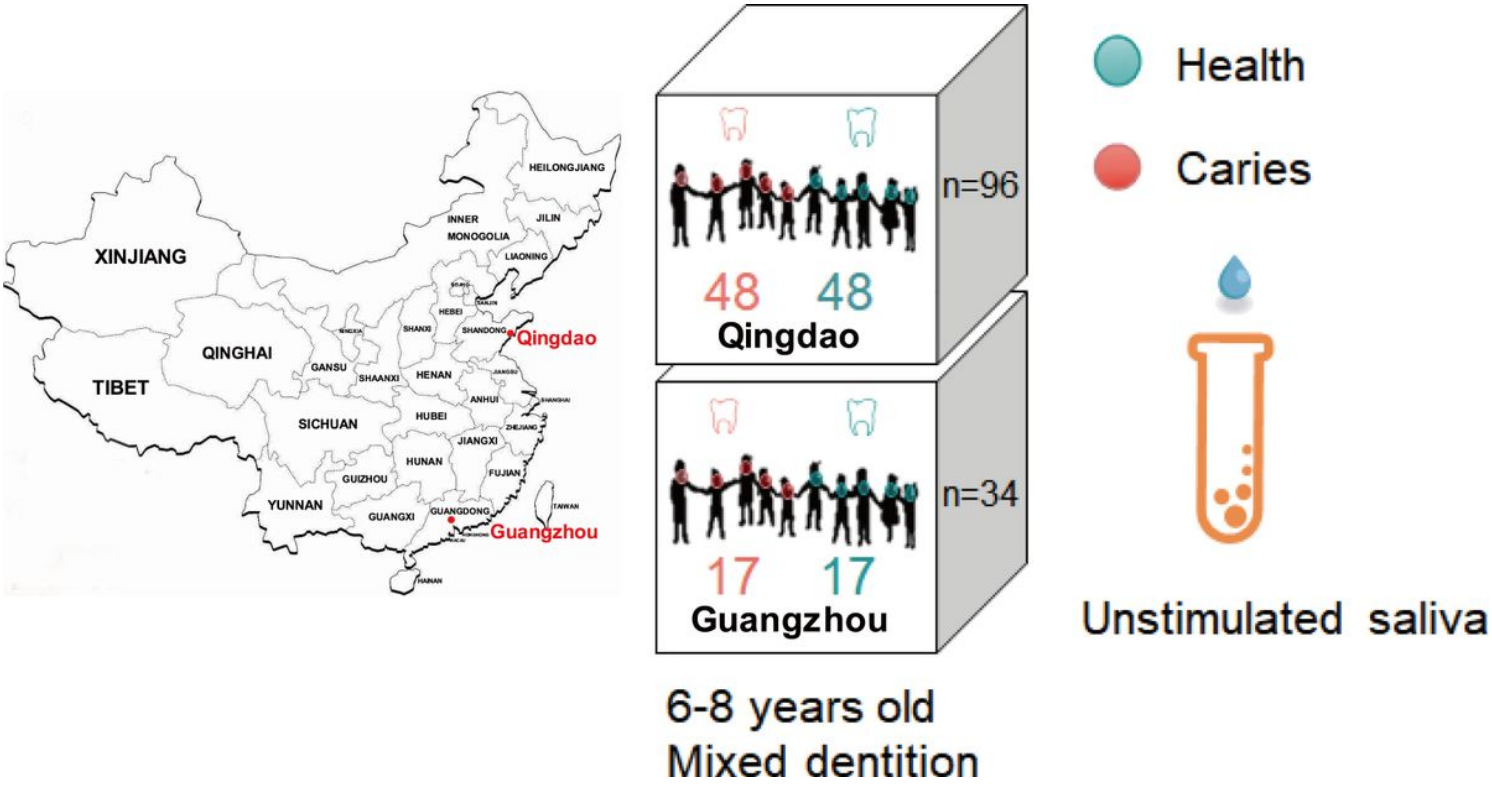


Figure 1

Experimental design that sampled saliva microbiome from caries-active children and healthy controls in the two Chinese cities of Qingdao and Guangzhou. Unstimulated saliva microbiota from 130 individuals (Qingdao, n=96; Guangzhou, n=34) were compared.

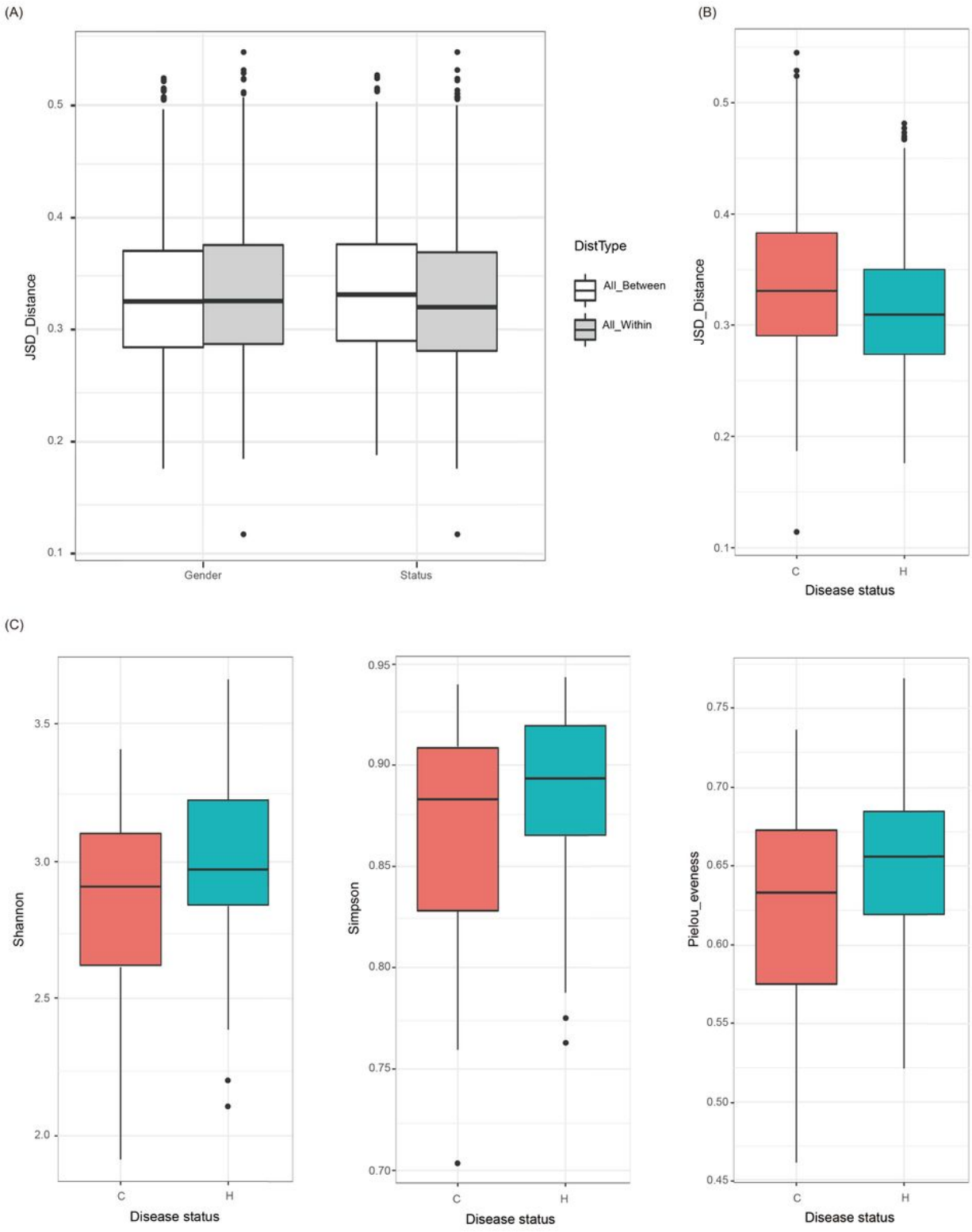


Figure 2

Oral microbial diversity comparisons between caries and healthy children in Qingdao cohort. (A) Salivary microbiota variation was compared within and between disease status (i.e., H or C), or gender based on the Jensen-Shannon distances. (B) Caries-free children have more conservative microbiota than do children with caries (**p<0.01). (C) Alpha diversity comparisons between the C and H groups using on Shannon, Simpson, and Pielou's evenness index.

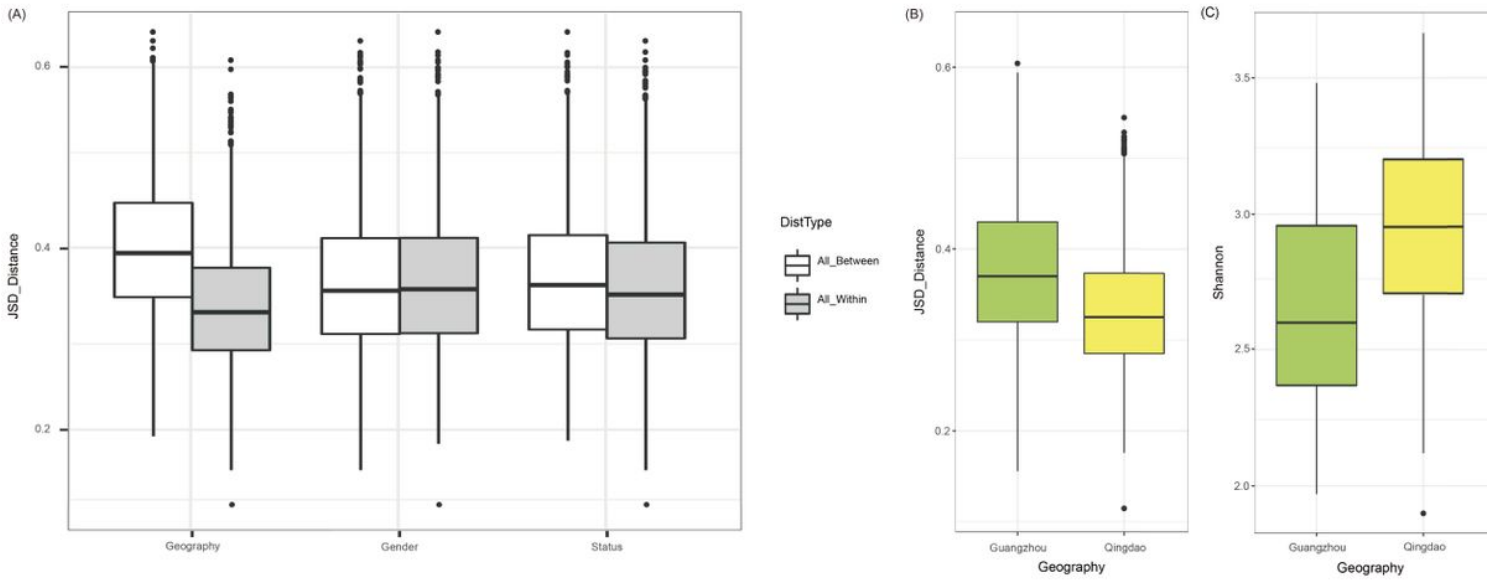


Figure 3

The remarkable impact of city of origin on oral microbiomes. (A) The effect size of geography, gender and host's disease status on saliva microbiota based on Jensen-Shannon distance. The city of origin exhibited the strongest effect on bacterial composition of the saliva microbiome, followed by host status and gender factor. (B) Beta-diversity difference between Qingdao and Guangzhou groups measured by JSD distances. (C) Alpha diversity difference between Qingdao and Guangzhou groups measured

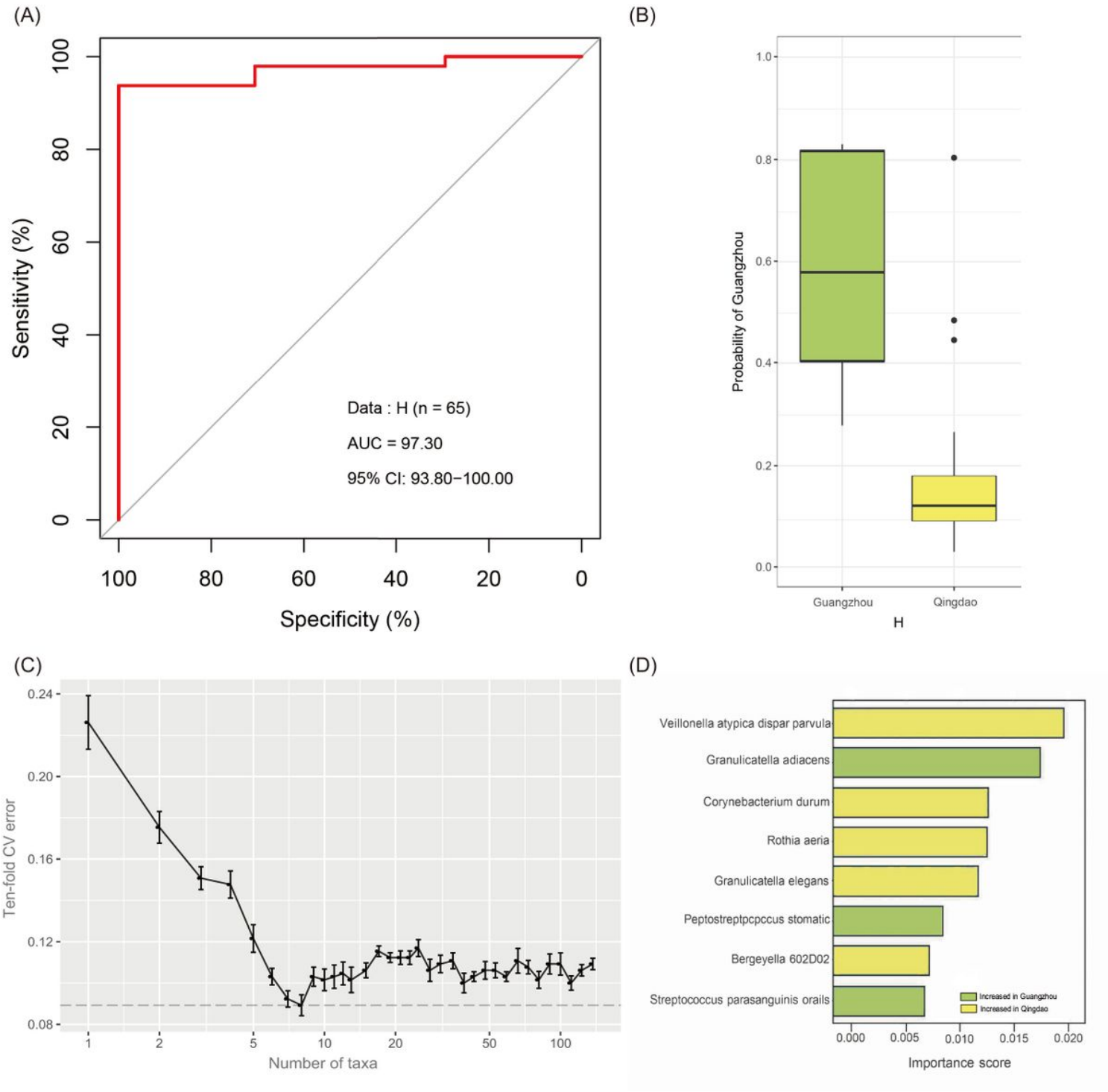


Figure 4

The strong geographical background of the healthy oral microbiota and key drivers. (A) Microbiome can classify the city of origin of healthy samples with a high accuracy. (B) Box plot indicates the prediction probability of Guangzhou city in healthy samples. (C) Relationship between the numbers of variables used in the reduced models and the corresponding predictive performance (the error bar denotes SD). (D) The importance score of eight the most discriminating species in the diagnosis model to predict city origin. The bar length at each row indicates relative contribution of the species to the RF model.

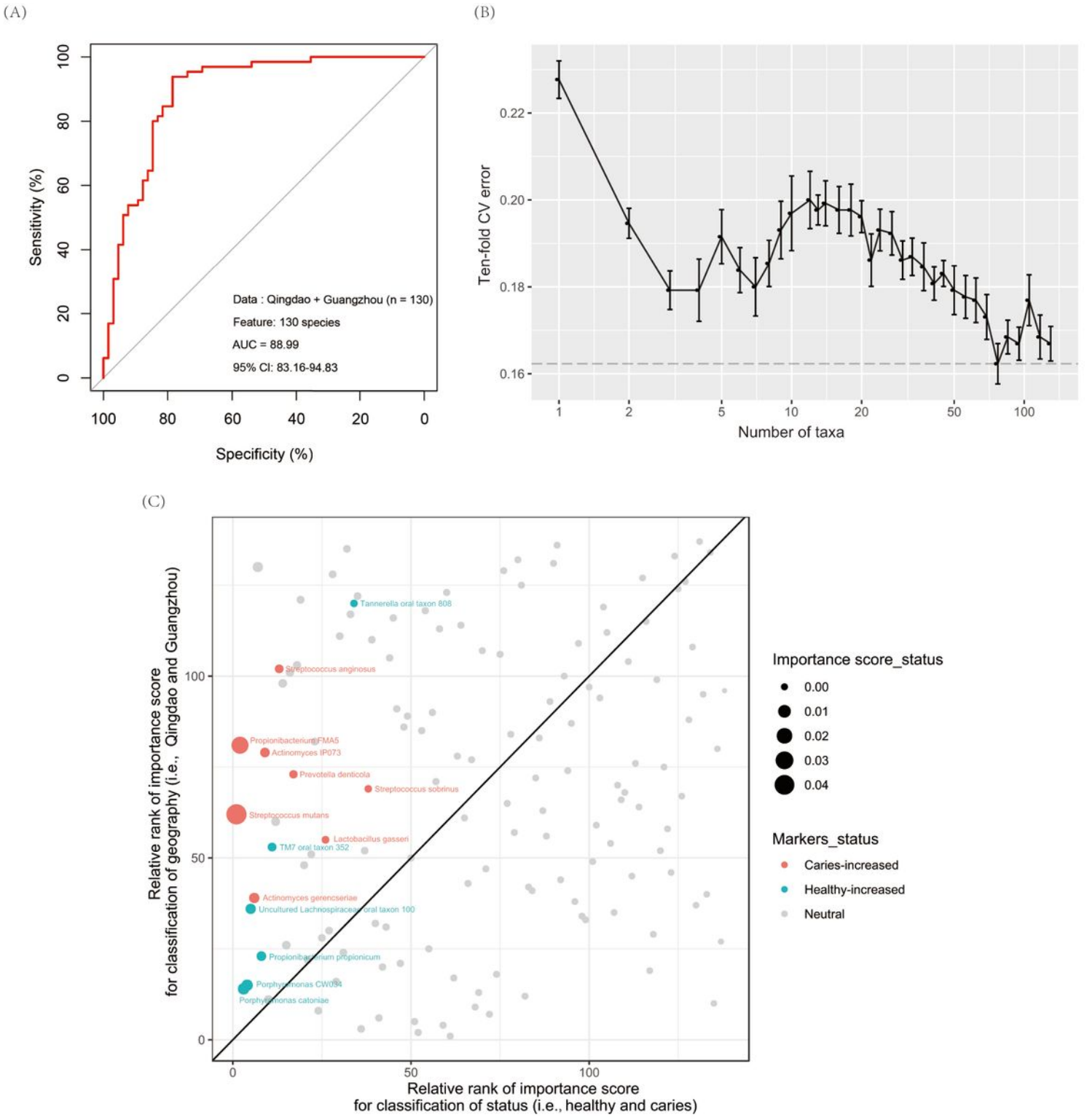


Figure 5

Caries diagnostic models based on oral microbiome detrended for geography. (A) Saliva microbiota can predict caries status with a remarkably high accuracy (AUC=92.17%). (B) The relationship between the numbers of variables used in the reduced Random Forest model and the corresponding predictive performance (the error bar denotes SD). (C) The most caries-discriminatory taxa (N=14) do not correlate with geography. The scatterplot shows the relative rank of microbial markers in both Random Forest

models for classifying disease status and geographic locations. Any dots on the reference line which slope=1 suggests a taxon is equally important to both disease states and geography.

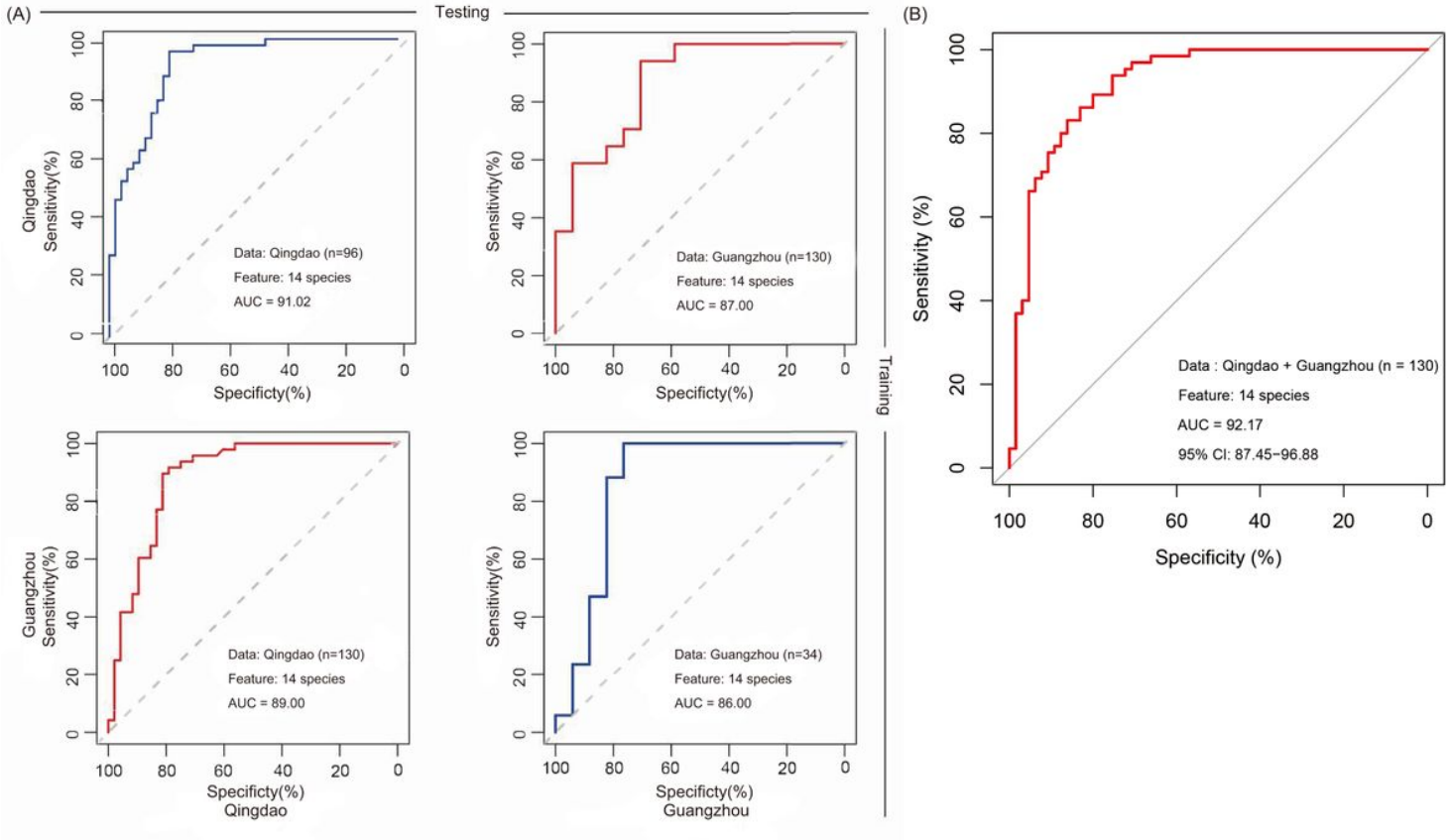


Figure 6

Cross-applications of caries diagnosis models based on microbiomes from Qingdao and Guangzhou cohorts. (A) The prediction performance of models in the Qingdao (AUC=91.02%, Guangzhou (AUC=86.16%%) and model application from one city to another. A classification model trained in Qingdao data and tested in Guangzhou Data resulting in a AUC=87.00; A classification model trained in Guangzhou data and tested in Qingdao Data resulting in a AUC=89.00%. (B) The predictive performance using data from two cities (AUC=92.17%).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigures.docx