Microb Ecol (2018) 75:543–554
DOI 10.1007/s00248-017-1056-1

HUMAN MICROBIOME

Taxonomic and Functional Analyses of the Supragingival Microbiome from Caries-Affected and Caries-Free Hosts

Jinzhi He[1] · Qichao Tu[2,3] · Yichen Ge[1] · Yujia Qin[3] · Bomiao Cui[1] · Xiaoyu Hu[1] · Yuxia Wang[1] · Ye Deng[4] · Kun Wang[1] · Joy D. Van Nostrand[3] · Jiyao Li[1] · Jizhong Zhou[3,5,6] · Yan Li[1] · Xuedong Zhou[1]

Received: 20 February 2017 / Accepted: 22 August 2017 / Published online: 20 September 2017

Abstract Caries is one of the most prevalent and costly infectious diseases affecting humans of all ages. It is initiated by cariogenic supragingival dental plaques forming on saliva-coated tooth surfaces, yet the etiology remains elusive. To determine which microbial populations may predispose a patient to caries, we report here an in-depth and comprehensive view of the microbial community associated with supragingival dental plaque collected from the healthy teeth of caries patients and healthy adults. We found that microbial communities from caries patients had a higher evenness and inter-individual variations but simpler ecological networks compared to healthy controls despite the overall taxonomic structure being similar. Genera including *Selenomonas*, *Treponema*, *Atopobium*, and *Bergeriella* were distributed differently between the caries and healthy groups with disturbed co-occurrence patterns. In addition, caries and healthy subjects carried different *Treponema*, *Atopobium*, and *Prevotella* species. Moreover, distinct populations of 13 function genes involved in organic acid synthesis, glycan biosynthesis, complex carbohydrate degradation, amino acid synthesis and metabolism, purine and pyrimidine metabolism, isoprenoid biosynthesis, lipid metabolism, and co-factor biosynthesis were present in each of the healthy and caries groups. Our results suggested that the fundamental differences in dental plaque ecology partially explained the patients' susceptibility to caries, and could be used for caries risk prediction in the future.

Keywords Caries · Supragingival dental plaque · 16S rRNA sequencing · Human microbial functional gene microarray · Taxonomic composition · Functional diversity

Jinzhi He and Qichao Tu contributed equally to this study.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00248-017-1056-1) contains supplementary material, which is available to authorized users.

✉ Jizhong Zhou
jzhou@ou.edu

✉ Yan Li
feifeiliyan@163.com

✉ Xuedong Zhou
zhouxd@scu.edu.cn

[1] State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China

[2] Department of Marine Sciences, Ocean College, Zhejiang University, Hangzhou, Zhejiang, China

[3] Institute for Environmental Genomics, Department of Microbiology and Plant Biology, and School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK 73019, USA

[4] Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China

[5] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China

[6] Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Introduction

Dental caries, or tooth decay, is one of the most prevalent and costly infectious diseases worldwide and contributes significantly to the burden of tooth pain and the associated marked decrease in quality of life [1, 2]. Past or present coronal caries affect 93.8% of the US adult population [3] and 88.1% of adults in China [4]. Thus, studies providing information on prevention, prognosis, and early diagnosis of dental caries are of particular clinical significance.

Microbes colonizing tooth surfaces (collectively known as supragingival dental plaque or supragingival dental biofilm) are responsible for tooth decay. By metabolizing carbohydrates, microbes produce both organic acids and exopolysaccharides (EPS). With the help of the EPS-rich matrix, organic acids maintain a low-pH micro-environment within the biofilm [5], which not only causes demineralization of tooth hard tissue but also imposes a stress on the biofilm, driving the microbial community to a cariogenic state [6]. Since Clark's first isolation of *Streptococcus mutans* from caries lesions in the 1920s, the sugar-fermenting, acidogenic and aciduric species has been regarded as the etiological agent of tooth decay, and a large number of studies have focused on this microorganism [7, 8]. Recently, the introduction of molecular approaches has revealed that the human oral cavity (mouth) is colonized by hundreds of bacterial species and that those species regarded as causative agents of caries are also frequently detected in healthy individuals [9]. This has given rise to the ecological hypothesis that caries are a poly-microbial infectious disease caused by potentially pathogenic microbial communities rather than specific pathogens. Under this scenario, an increased interest in using "holistic" approaches to define the complex microbial communities that contribute to caries using advanced metagenomic technologies has emerged. Several studies have compared the taxonomic composition of the oral microbiome from caries-affected and caries-free individuals, and the data clearly showed that taxa, including the genera *Veillonella*, *Actinomyces*, *Granulicatella*, *Leptotrichia*, *Thiomonas*, and *Prevotella*, in plaque or saliva were significantly associated with tooth decay [10–13]. These data strongly support the poly-microbial infection nature of caries but fail to provide more information regarding the microbial community changes other than differently distributed taxa.

Moreover, it should be noted that obtaining taxonomic differences is far from understanding disease mechanisms, as inter-individual microbial composition is highly variable and the combinations of possible consortia are numerous [10, 14]. Although individual human microbiomes are radically distinct in terms of taxonomic composition, they are surprisingly global in regard to functional capabilities [15, 16]. Thus, focusing on the functions associated with the microbial communities regardless of the microorganisms involved could provide another clue for better understanding and controlling caries formation [9].

From an applied viewpoint, we believe that the combination of taxonomic and functional profiles would be more informative than either one alone in understanding the risk factors for dental caries, and could provide potential diagnostic value. In the present study, we collected supragingival dental plaque samples from the healthy tooth surfaces of both caries patients and caries-free adults to determine the microbial taxonomic and functional characteristics that may be increasing the hosts' predisposition to caries. We analyzed these specimens using 16S rRNA gene amplicon sequencing and a human microbial functional gene array (i.e., HuMiChip 1.0) that has been successfully used in a previous periodontal microbial investigation by our group [17]. Our data showed that the healthy tooth surfaces of caries patients were colonized by microbial communities with higher evenness and inter-individual variations, disturbed inter-microbial co-occurrence patterns, and lower ecological network complexity. In addition, there were distinct differences in functional genes involved in organic acid synthesis, glycan biosynthesis, complex carbohydrate degradation, amino acid synthesis and metabolism, purine and pyrimidine metabolism, isoprenoid biosynthesis, lipid metabolism, and co-factor biosynthesis between the healthy and caries groups. These data suggested that teeth within caries subjects may be "primed" for caries due to fundamental differences in the ecology of the supragingival dental plaque microflora.

Results

Clinical Investigation

In total, 37 subjects of both genders were recruited from Chengdu after a strict clinical survey at the West China Hospital of Stomatology. The subjects were adults with permanent dentition and (I) were medically healthy, (II) had no antibiotic use within the past 3 months, (III) had normal oral mucosa, (IV) had no less than 28 teeth, and (V) were judged to be free of periodontal diseases and had > 90% of sites with periodontal tissues showing no clinical signs of inflammation, such as redness, swelling, or bleeding on probing. The clinical features of subjects are shown in Tables 1 and S1. No significant difference was observed in age or sex between the study and control groups.

Healthy Teeth of Caries Patients Were Colonized by a Microbial Community with Higher Evenness and Inter-individual Variation

After preprocessing and filtering, the average amplicon length was 252 bp. High-quality reads were assigned to 1340

Table 1 Clinical parameters of the study population

Variations	($\overline{X}$ ± SD)		*P* value
	Healthy (*n* = 12)	Caries (*n* = 25)	
Age (years)	36 ± 9	37 ± 9	0.69[a]
Sex			0.17[b]
Female	58%	68%	
Male	42%	32%	
Caries			
DMFT	0	1~11	

DMFT Decayed, missing, filled teeth

[a] Mann-Whitney *U* test

[b] Fisher's exact test

operational taxonomic units (OTUs) at 3% dissimilarity. The rarefaction curve revealed that the sequence depth was not sufficient to recover the diversity of this community (Fig. S1).

The community structures of all 37 samples were analyzed by detrended correspondence analysis (DCA). As shown in Fig. S2, caries patients and healthy subjects were not well separated. The dissimilarity tests also showed no significant differences between these two groups (PERMANOVA: $R = 0.045$, $P = 0.097$; ANOSIM: $R = 0.006$, $P = 0.429$; MRPP: $\delta = 0.572$, $P = 0.094$). Similar tendencies were also found for richness (Chao 1) and alpha diversity (Shannon) comparisons, with no statistically significant differences detected between the groups ($P > 0.05$, Fig. 1). However, the supragingival dental plaque of caries patients demonstrated a higher evenness, as revealed by the Buzas index ($P = 0.04$, Fig. 1), and the degree of variation among healthy individuals was significantly lower than among the caries patients based on the Bray-Curtis dissimilarity and the weighted-UniFrac distance comparisons ($P < 0.05$, Fig. 1 and Fig. S3).

Microbes from the Intact Teeth of Caries Patients Formed Simpler Ecological Network

We constructed co-occurrence ecological networks at the OTU level to better understand how supragingival communities assemble and to determine whether the oral microbial community network topology had distinct inter-groups. Several interesting findings could be detected here. First, the constructed caries network contained 132 nodes (OTUs), 231 links, and 16 modules (a module in a network is a sub-network that has more internal edges than external edges [18]; in this network, 12 modules with ≥ 3 nodes), with an average connectivity of 3.5, an average geodesic distance of 1.156, and a modularity of 0.803 (Table 2, Fig. 2a), whereas the healthy group network had 159 nodes, 531 links, and 6 modules (5 with ≥ 3 nodes), with an average connectivity of 6.679, an average geodesic distance of 4.157, and a modularity of 0.691 (Table 2, Fig. 2b). Thus, the caries network contained fewer nodes and links and was simpler than the healthy network based on the average connectivity and modularity (Table 2). Sub-networks constructed by extracting the first neighbors of the ten nodes with the highest connectivity were less complex for the caries patients even though more nodes were selected to construct that sub-network (Fig. 2c, d).

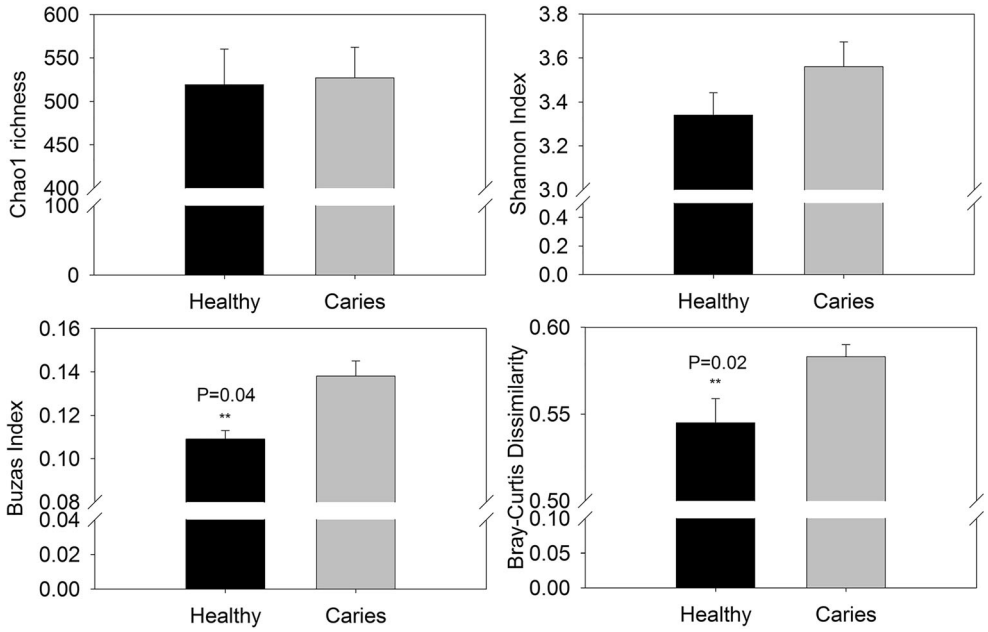

Fig. 1 Comparison of richness (Chao1), α-diversity (Shannon), evenness (Buzas), and β-diversity (Bray-Curtis distance) between caries and control groups. The supragingival microbial community of caries patients displayed higher evenness and inter-individual variation compared to healthy individuals (two asterisks indicate $P < 0.05$)

Table 2 Topological properties of co-occurrence networks of microbial communities and their corresponding random networks

Community	Empirical networks								Random networks[d]		
	No. of OTUs[a]	Network size[b]	Total links	*tbcolw45ptr*^{2} of scale free[c]	tbcolw45ptModules (≥ 3 nodes)	Avg. connectivity	tbcolw45ptAvg. path (GD)	tbcolw50ptAvg. clustering coefficient	Modularity	tbcolw40ptAvg. path (GD)	tbcolw50ptAvg. clustering coefficient	Modularity
Healthy	162	159	531	0.538	5/6	6.679	4.157[e]	0.446[e]	0.691[e]	2.913	0.067	0.329
Caries	164	132	231	0.766	12/16	3.500	1.156[e]	0.384[e]	0.803[e]	3.648	0.031	0.516

GD geodesic distance

[a] The number of OTUs that were originally used for network construction using the RMT-based approach

[b] The number of OTUs (i.e., nodes) in a network

[c] The correlation coefficient (r) of the linear relationship in log $[P(k) \sim \gamma \log(k)]$, where $P(k)$ is the fraction of connectivity k and γ is a constant

[d] The random networks were generated by rewiring all of the links of a MEN with the identical numbers of nodes and links to the corresponding empirical MEN

[e] These parameters of empirical networks are significantly different from the random networks ($P < 0.001$)

Second, for modules with ≥ 3 nodes in the network, there were fewer inter-modular connections linking different modules together in the caries network than in the healthy network, despite detecting more modules in the caries network. Third, the negative relationship (red line) between OTUs was greatly reduced in the caries network compared to the healthy network (Fig. 2a, b). Fourth, the caries sub-network was dominated by OTUs from *Bacteroidales* (belongs to the phylum Bacteroidetes), whereas the healthy sub-network was dominated by OTUs from *Actinomycetales* (belongs to the phylum Actinobacteria) (Fig. 2c, d).

Differently Distributed Taxa Also Show Disturbed Inter-microbial Co-occurrence in Caries Patients

In total, 12 known bacterial phyla and 70 genera were identified from the clinical samples. At the phylum level, both caries and healthy individuals were dominated by *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, *Fusobacteria*, and *Spirochaetes*. Among these phyla, the relative abundance levels of *Bacteroidetes* and *Spirochaetes* were significantly higher in the caries patients than in the healthy individuals (Fig. 3a and Table S2, $P < 0.05$). At the genus level, we found only four genera with different abundances between caries and healthy individuals. *Selenomonas*, *Treponema*, and *Atopobium* showed higher relative abundance in the caries group (caries-associated genera), whereas the genus *Bergeriella* had a higher relative abundance in healthy subjects (healthy-associated taxa, Fig. 3b and Table S3, $P < 0.05$). We further explored the distribution of OTUs belonging to these four genera using a detection frequency-based analysis, and found that caries patients and normal individuals carried different arrays of *Prevotella*, *Treponema*, and *Leptotrichia* species (Table 3). A total of seven *Prevotella* species, with a detection frequency higher than 50%, were found only in the caries group, while another two OTUs (OTU_475 and OTU_344) assigned to *Prevotella* sp. were exclusively detected in the healthy group. *Treponema* species were only detected in caries patients. In addition, three OTUs belonging to *Leptotrichia* (OTU_265, OTU_255, and OTU_321) were only detected in the healthy group while another three OTUs (OTU_280, OTU_273, OTU_340) were only found in caries patients.

We also investigated whether there were differences in the microbial relationships of these four differentially distributed genera and the genus *Prevotella* (a recently recognized caries-associated taxa) between the caries patients and healthy controls. To determine this, we employed a co-occurrence and co-exclusion analysis based on the Pearson correlation (Fig. 4 and Table S4). In healthy individuals, nearly half of the inter-genus co-occurrence relationships of these four genera were negative (Fig. 4a). However, in the caries patient group, the percentage of negative co-occurrence was ~ 20% (Fig. 4b).

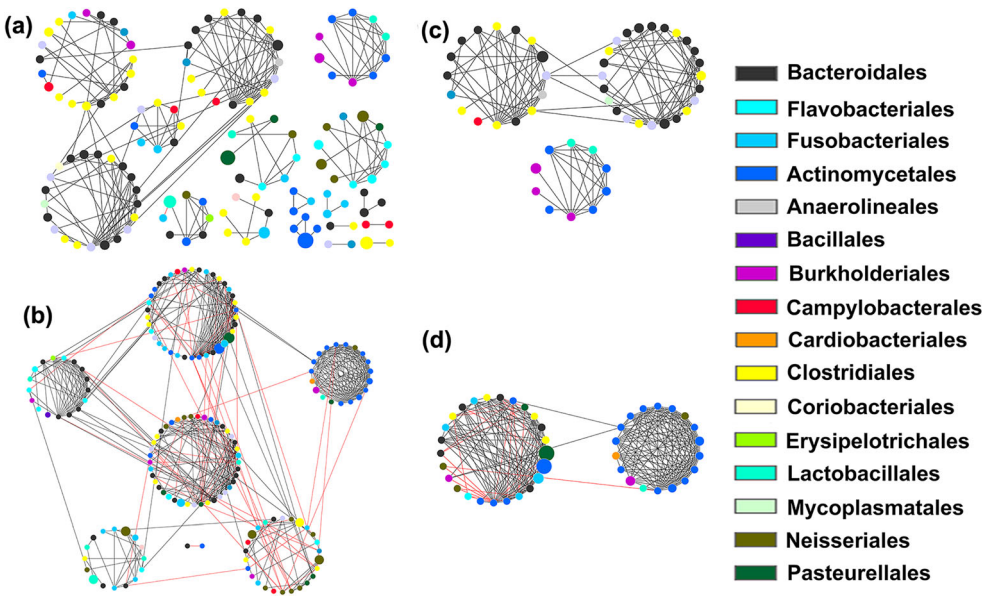

Fig. 2 Network inferences of the complex microbial relationships in caries and healthy subjects. Each node represents an OTU, and each edge represents a significant pairwise association. Red lines indicate negative relationships; black lines represent positive relationships. **a** Ecological networks of the caries group, **b** ecological networks of the healthy group, **c**, **d** sub-networks of caries and healthy groups, respectively, constructed by extracting the first neighbors of nodes with the highest connectivity

Functional Genes Provided Another Important Clue in the Caries Pathogenesis Investigation

To determine if the functional composition of the oral microbiota was different between caries and healthy groups, we analyzed genomic DNA using the functional gene array HuMiChip 1.0. In total, 12,345 probes, distributed among 20 categories, were identified. Among these, 1773 probes were detected only in the caries samples and 179 were unique to the healthy group. For each sample, the number of detected probes ranged from 4025 to 10,015. The overall functional gene diversity was similar, as no significant differences were found for the detected probe number and the Shannon index between groups ($P > 0.05$) (Table 4). In addition, caries patients and healthy subjects shared similar supragingival microbial functional structures, as revealed by DCA analysis (Fig. S4) and dissimilarity tests (PERMANOVA: $R = 0.023$, $P = 0.527$; ANOSIM: $R = -0.06$, $P = 0.832$; MRPP: $\delta = 0.27$, $P = 0.487$). However, 13 genes showed significantly different distributions between groups (Fig. 5). These genes were involved in organic acid synthesis, glycan biosynthesis, complex carbohydrate degradation, amino acid synthesis and metabolism, purine and pyrimidine metabolism, isoprenoid biosynthesis, lipid metabolism, and co-factor biosynthesis. Specifically, genes encoding folylpolyglutamate synthase, UDP-*N*-acetylglucosamine acyltransferase, 1-hydroxy-2-

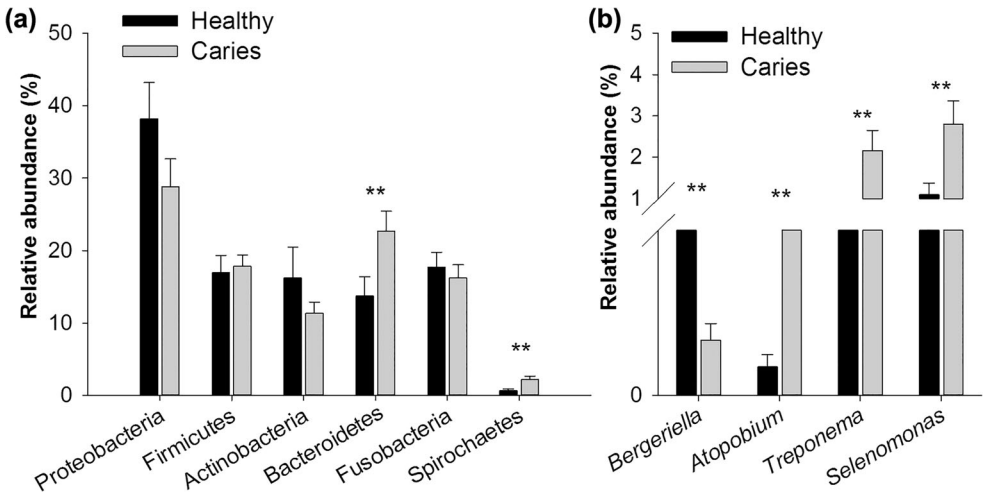

Fig. 3 Taxa distribution between caries-affected and healthy subjects at the phylum (**a**) and genus (**b**) levels. At the phylum level, the relative abundance of *Bacteroidetes* and *Spirochaetes* were significantly higher in the caries patients than in the healthy individuals. Four genera were distributed differently between the healthy and caries groups, of which *Selenomonas*, *Treponema*, and *Atopobium* showed higher relative abundance levels in the caries group, whereas the genus *Bergeriella* had a lower relative abundance. Two asterisks indicate $P < 0.05$

Table 3 Detection frequency of *Prevotella*, *Treponema*, and *Leptotrichia* species carried by caries patients and normal individuals

OTU_ID	Genus	Frequency		HOMD	
		Health	Caries	Species	Identity
OTU_475	*Prevotella*	8/12	0	*Prevotella* sp. | HOT_472	98.4%
OTU_344		11/12	0	*Prevotella* sp. | HOT_472	97.2%
OTU_2170		0	17/27	*Prevotella tannerae* | HOT_466	98.8%
OTU_335		0	17/27	*Prevotella nigrescens* | HOT_693	99.2%
OTU_334		0	16/27	*Prevotella* sp. | HOT_293	99.6%
OTU_686		0	14/27	*Prevotella* sp. | HOT_301	100%
OTU_447		0	14/27	*Prevotella marshii* | HOT_665	99.2%
OTU_297		0	14/27	*Prevotella multiformis* | HOT_685	99.2%
OTU_306		0	15/27	*Prevotella baroniae* | HOT_553	99.2%
OTU_531	*Treponema*	0	13/27	*Treponema* sp. | HOT_258	97.6%
OTU_391		0	14/27	*Treponema* sp. | HOT_270	99.6%
OTU_562		0	15/27	*Treponema* sp. | HOT_257	98.8%
OTU_265	*Leptotrichia*	7/12	0	*Leptotrichia* sp. | HOT_212	98.0%
OTU_255		8/12	0	*Leptotrichia wadei* | HOT_222	97.6%
OTU_321		8/12	0	*Leptotrichia buccalis* | HOT_563	98.0%
OTU_280		0	14/27	*Leptotrichia* sp. | HOT_215	98.8%
OTU_273		0	21/27	*Leptotrichia* sp. | HOT_219	99.2%
OTU_340		0	15/27	*Leptotrichia* sp. | HOT_498	99.6%

methyl-2-(*E*)-butenyl 4-diphosphate reductase 4Fe-4S protein, 1-phosphofructokinase, thymidine phosphorylase, purine-nucleoside phosphorylase, L-lactate dehydrogenase (LDH), proline dehydrogenase, and aspartate racemase displayed higher relative abundance levels in the caries group, whereas the structural genes encoding ornithine carbamoyltransferase, conjugated bile salt hydrolase, acetate kinase (Ack), and cellulase were less abundant in this group (Fig. 5, $P < 0.05$).

Discussion

The oral microbiota composition is a critical variable in the formation of caries. In this study, we collected supragingival dental plaque from healthy tooth surfaces of both caries patients and healthy controls to explore microbial factors involved in caries susceptibility. We found taxonomic and functional differences between caries and healthy controls, and our data suggested that the teeth of caries subjects may be "primed" for caries due to fundamental differences in the ecology of supragingival dental plaque microflora.

We compared the overall taxonomic composition between caries and healthy patients first, and found they were similar. Another question of concern was if there were any community topological differences between caries patients and healthy controls. To determine if differences existed, an ecological network analysis was performed. Ecological network analysis is a systems-level method to identify species interactions within an ecosystem that cannot be directly observed [21]. Comparative analysis of bacterial networks in this study indicated that the microbial communities colonized on the healthy tooth surfaces of caries patients and caries-free people were not the same. First, caries patients had simpler networks within the supragingival microbiome and the species in cariogenic dental plaque were more sparsely connected with each other, as evidenced by fewer nodes and links, as well as a lower average connectivity and modularity. Second, the average inter-module connectivity of the caries network was less than that of the healthy group, and was even diminished between some modules. Since the modules in the ecological networks could be regarded as putative microbial ecological niches or sub-communities [22], this suggests decreased relationships between different oral bacterial community "niches" or sub-communities in the caries patients. Third, negative relationships were greatly reduced in the caries network, indicating decreased antagonism among the dental plaque residents. Based on the "ecological plaque hypothesis," caries result from a community shift of the resident microbiota driven by environmental changes in oral conditions [23]. Our data suggested that the "shift" could be observed at the community topology level without significant microbial composition changes.

No significant differences were observed in the microbial richness and α-diversity between the disease and healthy groups, which is highly consistent with the previous observation obtained by sequencing plaques from children [24] and the saliva of young college students [10]. In contrast, the β-diversity within the caries group was markedly higher than in

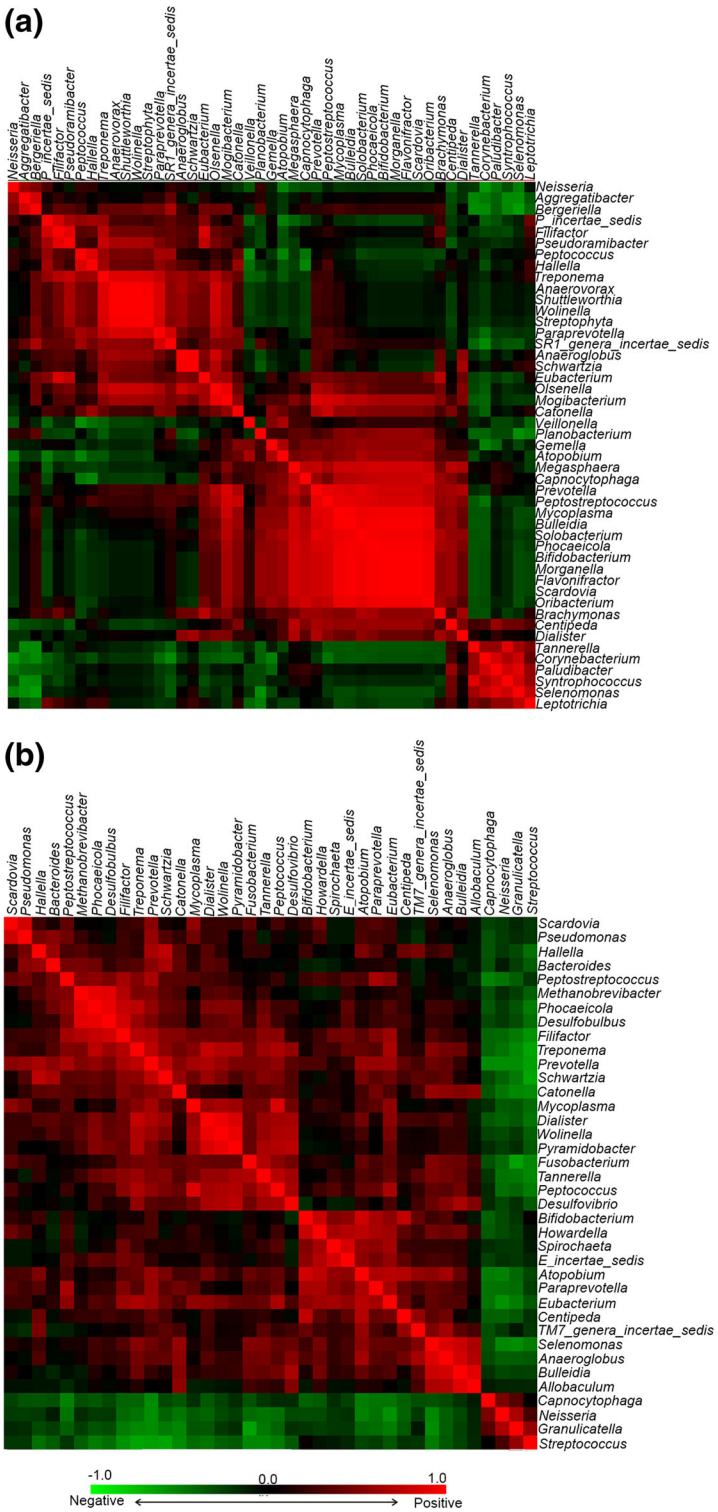

Fig. 4 Co-occurrence relationships of differently distributed genera. Only genera that had significant correlations with *Selenomonas*, *Treponema*, *Atopobium*, *Bergeriella*, and *Prevotella* were selected and plotted. The microbial co-occurrence was explored by calculating the Pearson correlation, clustering the correlations using Cluster 3.0 [19], and visualizing the results using Java TreeView version 1.8.0-31 [20]. Nearly half of the inter-genus co-occurrence relationships were negative in healthy individuals (**a**), whereas the majority (~ 80%) of inter-genus co-occurrence relationships was positive in patients with dental caries (**b**)

the healthy individuals, indicating that the cariogenic microbiome was significantly more variable, whereas the microbiome of healthy individuals was relatively conserved, which is consistent with the findings of Yang et al. [10]. Interestingly, when using unweighted-UniFrac distance to compared β-diversity, no statistical difference was observed, suggesting the variability of the cariogenic microbiome was mainly caused by distinct relative abundances of detected OTUs.

Second, we pinpointed that caries patients and non-caries controls carried different levels and/or arrays of taxa. Among these genera, *Selenomonas*, *Treponema*, and *Atopobium* showed higher relative abundances in the caries group, whereas the genus *Bergeriella* had a lower relative abundance in this

Table 4 Functional diversity of the healthy and caries-active dental plaque samples

	α-Diversity		β-Diversity	
	No. of detected probes	Shannon	Bray-Cutis	Sørensen
Caries	7508.83 ± 1173.72	8.90 ± 0.19	0.27 ± 0.07	0.21 ± 0.06
Healthy	7507.40 ± 1335.21	8.90 ± 0.16	0.26 ± 0.08	0.22 ± 0.05

group. Apart from these four genera, the genus *Prevotella* had a higher relative abundance in the supragingival dental plaque of the caries patients although the difference was not statistically significant. *Prevotella*, an anaerobic, gram-negative bacterium belonging to the phylum Bacteroidetes, has long been regarded as a potential periodontal pathogen, and its role in caries has been recently recognized. There was higher relative abundance of *Prevotella* spp. and different *Prevotella* populations in the oral cavities of caries patients and subjects of different ages (e.g., children, young college students, and adults) and dentition stages (e.g., primary dentition, mixed dentition, and permanent dentition) [10, 11, 25] and in different oral sites (e.g., saliva and supragingival dental plaque) [10, 11]. In this study, although the enrolled subjects were older and a different primer set was used, we detected similar trends for *Prevotella* species, as well as species from the genera *Treponema* and *Leptotrichia*. In addition, Teng et al. [11] found that the early childhood caries (ECC) predictive performance of the model derived from eight *Prevotella* species was comparable to that from the whole oral microbiota. Whether the different distributions of *Prevotella*, *Treponema*, and *Leptotrichia* species could be used for caries risk prediction in adults requires further study. In parallel, we found that not only the relative abundance but also the inter-genus relationships of *Selenomonas*, *Treponema*, *Atopobium*, *Bergeriella*, and *Prevotella* were significantly different between caries patients and healthy individuals. For example, *Prevotella* and *Atopobium* were positively correlated with *Capnocytophaga* (of which relative abundance levels in healthy and caries subjects were similar) in health, whereas negative associations were observed in the caries patients (Table S4). The lower proportion of negative inter-genus co-occurrence relationships among *Selenomonas*, *Treponema*, *Atopobium*, *Bergeriella*, and *Prevotella* in caries patients (~20%) compared to healthy controls (~50%) was highly consistent with the ecological network analysis and further confirmed that mutual antagonism was not as active among the dental plaque residents of caries patients.

For a more comprehensive understanding of the cariogenic microbial community, we were ultimately interested in not just the community composition ("who's there and with whom") and structure ("who's where") but also the community function ("who's doing what, where and when"). To determine this, we used the HuMiChip 1.0 microarray, which was constructed for analyzing human microbiomes and contains 36,802 probes targeting approximately 50,000 CDSs from 139 key functional genes in various metabolic pathways, such as the metabolism of amino acids, carbohydrates, energy, lipids, glycan, cofactors, vitamins, and nucleotides [26]. We found that the overall functional gene diversity and structure were similar between the groups. Although phylogenic β-

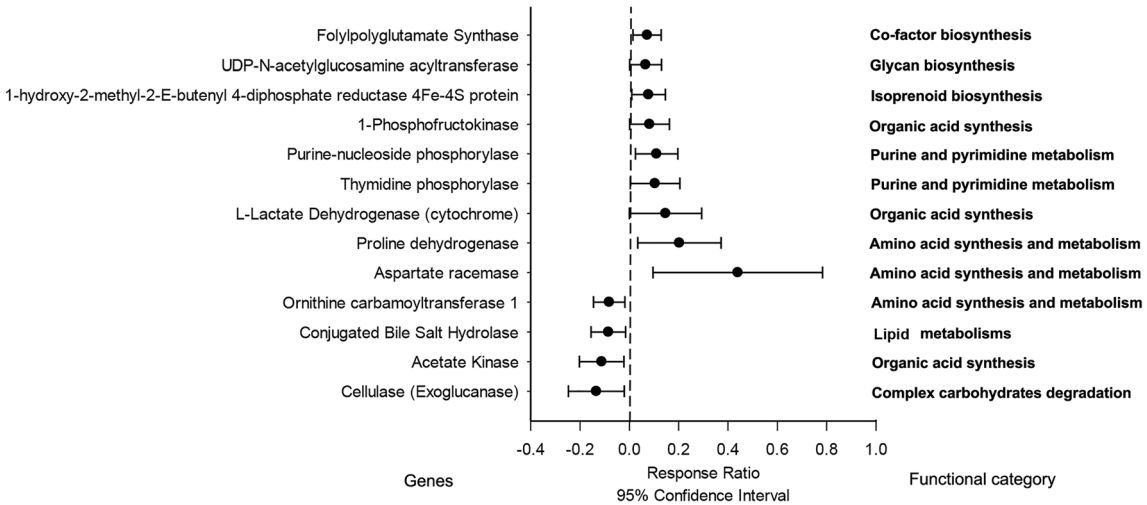

Fig. 5 Functional genes with different relative abundance levels between groups. The response ratio (caries-affected vs. healthy individuals) was calculated using the method of Luo et al. [18] at a 95% confidence interval. Error bars plotted at the right of the dashed line indicate higher relative abundance levels in the caries group, whereas error bars plotted at the left of the dashed line indicate lower relative abundance levels in the caries group. The functional categories of the genes are also listed

diversity was higher in the dental plaque of caries patients, no significant difference was observed. One possible reason might be that human microbiomes share global functional capabilities [15, 16]. However, we still found 13 gene families with distinct distributions between the groups, indicating the different functions of the biofilms.

Carbohydrates have long been regarded as the "arch-criminal" of dental caries because bacteria can produce both organic acid and exopolysaccharides (EPS) by metabolizing carbohydrates. With the help of an EPS-rich matrix, the organic acid can maintain a low pH micro-environment within the biofilm [5], which not only causes the demineralization of a tooth's hard tissue but also imposes a stress on the biofilm, driving the microbial community to a cariogenic state with acidic and aciduric bacterial dominance [6]. Thus, the lower pH resulting from carbohydrate metabolism is one of the most important virulence factors of the cariogenic biofilm. The pH decrease is due primarily to the organic acid, especially lactic acid [27, 28]. In the present study, three genes in the category of organic acid synthesis, encoding LDH, Ack, and 1-phosphofructokinase, showed different distributions between the healthy and caries groups. The enzyme 1-phosphofructokinase catalyzes the committed step of glycolysis, the metabolic pathway that converts glucose into pyruvate, and the pyruvate generated by glycolysis is acted on by LDH to produce lactate (the dominant acid in active dentin lesions [29] under anaerobic conditions or by pyruvate formate lyase/pyruvate dehydrogenase, phosphotransacetylase, and Ack sequentially to synthesize acetate under aerobic conditions [30]. We found that the genes encoding 1-phosphofructokinase and LDH displayed higher relative abundance levels in the caries group, whereas the structural gene encoding Ack was less abundant in the caries group ($P < 0.05$), suggesting that the microbiome on the intact tooth surfaces of the caries patients had a higher lactate production capacity, which could in part explain why these people are more sensitive to caries. With regard to amino acid synthesis and metabolism, the relative abundance level of genes encoding ornithine carbamoyltransferase was lower in caries group. Microbial catabolism of amino acids has long been known to contribute to plaque acid neutralization and thus partially accounts for caries inhibition [31, 32]. For example, arginine metabolism is considered to be the main alkali resource in the oral cavity. Caries-free individuals had significantly higher concentrations of arginine in their saliva than those with caries [31], and positive correlations between enzyme activity related to arginine metabolism and the absence of caries have been clinically demonstrated in adults [33] and children [34]. In the oral cavity, arginine is catalyzed sequentially primarily via arginine deiminase, ornithine carbamoyltransferase, and carbamate kinase, with the end products being putrescine, CO_2, ATP, and ammonia, respectively [35]. In the present study, we found a lower level of genes encoding ornithine carbamoyltransferase, which is consistent with studies linking compounds with neutralization functions to caries [32]. Combined with the organic acid synthesis analysis, the functional gene analysis suggested that caries-sensitive hosts harbored microbial communities with disturbed acid and alkali production.

Other functional genes also showed significantly different relative abundance levels between the groups. Aspartate racemase shows racemization activity on both aspartate and alanine, and the end products, D-forms of amino acids, are important components in the peptidoglycan layer of bacterial cell walls [36]. Proline dehydrogenase catalyzes the oxidation of L-proline to delta-pyrroline-5-carboxylate, leading to electron release; this enzyme has multiple physiological roles, including energy production, reactive oxygen species (ROS) production for cell cycle regulation, and adaptive responses to environmental stresses [37]. UDP-*N*-acetylglucosamine acyltransferase (belonging to the glycan biosynthesis category) catalyzes the first reaction of lipid A biosynthesis [38, 39]; this enzyme is required for cell division, growth under hypoosmotic conditions, and general viability [40, 41], and loss of its function leads to a substantial increase in antibiotic susceptibility owing to the loss of outer membrane impermeability [42, 43]. The enzyme 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase (ispH) is involved in isoprenoid biosynthesis. This enzyme has either a [4Fe-4S] or a [3Fe-4S] cluster that is critical for its catalytic activity [44], and *isp*H mutants are sensitive to penicillin and prone to cell lysis [45–47]. Cellulase (involved in complex carbohydrate degradation) is produced chiefly by fungi, bacteria, and protozoans that catalyze cellulolysis and the decomposition of cellulose and certain related polysaccharides [48]. Two genes in the category of purine and pyrimidine metabolism, including thymidine phosphorylase and purine-nucleoside phosphorylase, exhibited distinct distribution patterns between caries patients and healthy controls. The thymidine phosphorylase catalyzes thymidine to thymine. Substrates of purine-nucleoside phosphorylase are purine nucleoside and phosphate, and its products are purine and alpha-D-ribose 1-phosphate. Genes listed were novel in caries research, and larger cohorts will be necessary to validate their roles in caries and caries risk prediction.

In conclusion, the present study showed that the healthy tooth surfaces of caries patients were colonized by a microbial community with higher evenness and inter-individual variations, lower ecological network complexity, and disturbed inter-microbial interactions. Moreover, several taxa and functional genes were distinctly distributed between these groups. The differences listed above might contribute to the occurrence of caries, and taxa and functional genes with different distribution patterns might be useful for investigation of caries pathogenesis and risk prediction.

Methods

Subjects and Sampling

Written informed consent was obtained from volunteers in this study, and all procedures were approved by the ethics committee of the West China School of Stomatology, Sichuan University (WCHSIRB-ST-2011-005). Subjects with caries (n = 25) and sex- and age-matched healthy controls (n = 12; decayed, missing, filled teeth = 0) were recruited from the West China Hospital of Stomatology, Sichuan University. Demographic information was obtained and oral examination was performed on each subject, and mixed supragingival dental plaque was collected by swiping the intact buccal surfaces of two selected dentition quadrants (one from each of the maxillary and mandibular dentitions) with a dental explorer. Infected or filled teeth were excluded from sampling. Plaques were placed in 700 μL of Tris-EDTA buffer (10 mM Tris-HCl [pH 7.5] and 1 mM EDTA [pH 8.0]), carried to the laboratory within 4 h, and stored at − 80 °C before further processing. The methods were carried out in accordance with the approved guidelines.

DNA Extraction

Genomic DNA was extracted using a QIAamp ™ DNA micro Kit (QIAamp Sciences, Maryland, USA) with an added lysozyme (3 mg/mL, 1.5 h) treatment step to lyse gram-positive bacteria. The DNA quality was evaluated with a NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA), and the final concentration was quantified via a Pico-Green kit (Invitrogen, Carlsbad, CA, USA). Only DNA samples with A260/A280 > 1.7 and A260/A230 > 1.8 were used for downstream experiments. The extracts were stored at − 20 °C until use.

16S rRNA Amplicon Sequencing

The barcoded 16S rRNA amplicon sequencing was performed on an Illumina MiSeq at the Institute for Environmental Genomics, University of Oklahoma. Primers used in the present study were F515 (5′-GTGCCAGCMGCCGCGG-3′) and R806 (3′-TAATCTWTGGGVHCATCAG-5′). Both of the primers have over 90% coverage for bacteria and archaea in RDP, and are of an appropriate length for sequencing. A unique 12-mer tag for each DNA sample was added to the 5′-end of both primers to pool multiple samples for one run. Each 50 μL PCR amplification mix contained 0.2 μL Accuprime Taq DNA polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA), 5 μL Accuprime buffer II, 0.2 μM of each primer and 10 ng of genomic DNA. Cycling conditions were 94 °C for 1 min followed by 25 cycles of 94 °C for 20 s, 50 °C for 25 s, and 68 °C for 45 s, and a final elongation step at 68 °C for 10 min. PCR products were visualized on 3% agarose gels and then gel purified and quantified with a Pico-Green kit (Invitrogen, Carlsbad, CA, USA). The amplicons were then pooled in equamolar concentrations, assessed using Agilent BioAnalyzer 2100 (Invitrogen, Carlsbad, CA, USA), and sequenced on the MiSeq. Paired-end reads were merged into longer reads by FLASH [49]. After merging, sequences were trimmed using Btrim based on their quality score [50] with a trimming window size of 5 and an average cutoff of 20. Different sizes of trimming window and cutoffs were used to test trimming strategies. Sequences less than 200 bp or containing ambiguous residues were discarded. Chimeric sequences were discarded based on prediction by Uchime [51] using the reference database mode. The number of raw sequences was 481,653; after trimming, there were 413,556 sequences. The final sequences were used for operational taxonomic unit (OTU) clustering with the Uclust program at a 97% similarity level [52]. Final OTUs were generated based on the clustering results, and taxonomic annotations were assigned to the representative sequences of each OTU by RDP's 16S Classifier [53]. The pre-processed data was further analyzed as follows: (1) Rarefaction curves were constructed using the mothur software [54]. (2) Detrended correspondence analysis (DCA) was used to compare the phylogenetic structures of caries and healthy controls. Moreover, three non-parametric analyses for multivariate data, including analysis of similarities (ANOSIM) [55], non-parametric multivariate analysis of variance (PERMANOVA) using distance matrices [56], and multiresponse permutation procedure (MRPP), were used to examine the community differences between the two groups. (3) The richness (Chao 1), α-diversity (Shannon), evenness (Buzas), and β-diversity (Bray-Curtis and UniFrac) indices were calculated by the Bio-Community toolkit in BioPerl [57]. (4) Microbial-microbial co-occurrence networks were constructed by the MENA pipeline [58]. A Pearson correlation cutoff of 0.7 was determined using the random matrix theory approach by observing a transition point of the nearest-neighbor spacing distribution of eigenvalues from a Gaussian to a Poisson distribution, two universally extreme distributions. In such networks, OTUs were represented by network nodes while correlations were transformed into links between them. The co-occurrence networks were then visualized using Cytoscape 3.2.0 with a force-directed algorithm [59], and network topological parameters were computed using NetworkAnalyzer [60]. (5) The relative abundances of bacterial taxa at the phylum and genus levels were calculated. Student's t test was used to test the significance of differences of the relative abundances of bacterial taxa between healthy individuals and patients with dental caries at the phylum and genus levels. (6) Assigning OTUs of interest, including those belonging to *Prevotella*, *Treponema*, and *Leptotrichia*, to the species level was carried out by searching against the HOMD database at a stringent cutoff (> 98% identity). Representative

OTU sequences were blasted against the HOMD database using the HOMD online BLAST tool. (7) Correlations among different taxonomic groups were inferred by Pearson correlation coefficients, clustered by Cluster 3.0 [19] and visualized via Java TreeView version 1.8.0-31 [20]. Only taxonomic groups significantly ($P \leq 0.05$) correlated with microbial genera of interest (i.e., *Selenomonas*, *Treponema*, *Atopobium*, *Bergeriella*, and *Prevotella*) were plotted and analyzed.

Functional Gene Array Analysis

The HuMiChip 1.0 was used for functional gene profiling of the oral supragingival microbiome. This functional gene array contains 36,802 probes targeting 139 gene families [26], and additional control probes for quality control, normalization, and evaluation. Details of standard protocols for DNA labeling, hybridization, scanning, and preprocessing have been described previously [17]. Raw data was uploaded to the laboratory microarray data manager pipeline (http://ieg.ou.edu/microarray/). Pre-processed data were further analyzed as follows: (1) Diversity indices including detected probes, Shannon index, Bray-Cutis, and Sørensen were calculated using the Bio-Community toolkit in BioPerl [57]. (2) Community functional structure comparison was achieved by DCA analysis and dissimilarity tests. (3) The normalized signal intensity for each functional gene was the average of the total signal intensity from all the replicates and all data are presented as mean ± SD. Unpaired Student's *t* test was used to compare the significance of relative abundance differences of each gene category or family between healthy and diseased groups. Response ratios (caries vs. health) were calculated using the method of Luo et al. [18] at a 95% confidence interval.

Data Accession

The 16S rRNA sequencing data was added to the SRA database under accession number SRP078943, and the microarray data can be publicly accessed at http://ieg.ou.edu/microarray/.

Acknowledgements This study was supported by the National Natural Science Foundation of China (grant nos. 81430011, 81600874, 30901689, and 81172579), the International Science and Technology Cooperation Program of China (grant no. 2014DFE30180), the Sichuan Provincial Department of Science and Technology Project (grant no. 2013SZ0039), the key projects of Sichuan Provincial Health and Family planning Commission (grant No.:16ZD021), and the Fund for Young Scholars of Sichuan University (grant no.: 2016SCU11041).

We thank the members of the Zhou Lab at the University of Oklahoma for helpful discussions and technical assistance.

References

1. Kassebaum NJ, Bernabé E, Dahiya M, et al. (2015) Global burden of untreated caries: a systematic review and metaregression. J Dent Res 94(5):650–658
2. Petersen PE (2004) Challenges to improvement of oral health in the 21st century—the approach of the WHO Global Oral Health Programme. Int Dent J 54(S6):329–343
3. Shivakumar KM, Vidya SK, Chandu GN (2009) Dental caries vaccine. Indian J Dent Res 20(1):99–106
4. Qi X (2008) The third national sampling epidemiological survey on oral health. People's Medical Publishing House, Beijing
5. Xiao J, Klein MI, Falsetta ML, et al. (2012) The exopolysaccharide matrix modulates the interaction between 3D architecture and virulence of a mixed-species oral biofilm. PLoS Pathog 8:e1002623
6. Leme AFP, Koo H, Bellato CM, et al. (2006) The role of sucrose in cariogenic dental biofilm formation—new insight. J Dent Res 85(10):878–887
7. Plonka KA, Pukallus ML, Barnett AG, et al. (2012) Mutans streptococci and lactobacilli colonization in predentate children from the neonatal period to deven months of sge. Caries Res 46(3):213–220
8. Shanmugam K, Masthan K, Balachander N, et al. (2013) Dental caries vaccine—a possible option? J. Clin Diagn Res 7(6):1250–1253
9. Simón-Soro A, Mira A (2015) Solving the etiology of dental caries. Trends Microbiol 23(2):76–82
10. Yang F, Zeng X, Ning K, et al. (2011) Saliva microbiomes distinguish caries-active from healthy human populations. ISME J 6(1):1–10
11. Teng F, Yang F, Huang S, et al. (2015) Prediction of rarly childhood caries via spatial-temporal variations of oral microbiota. Cell Host Microbe 18(3):296–306
12. Johansson I, Witkowska E, Kaveh B, et al. (2015) The microbiome in populations with a low and high prevalence of caries. J Dent Res 95:80–86
13. Jiang W, Ling Z, Lin X, et al. (2014) Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood. Microb Ecol 67(4):962–969
14. Xu X, He J, Xue J, et al. (2015) Oral cavity contains distinct niches with dynamic microbial communities. Environ Microbiol 17(3):699–710
15. Vaishampayan PA, Kuehl JV, Froula JL, et al. (2010) Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. Genome Biol Evol 6:53–66
16. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. (2009) A core gut microbiome in obese and lean twins. Nature 457(7228):480–484
17. Li Y, He J, He Z, et al. (2014) Phylogenetic and functional gene structure shifts of the oral microbiomes in periodontitis patients. ISME J 8(9):1879–1891
18. Luo F, Yang Y, Chen CF, et al. (2007) Modular organization of protein interaction networks. Bioinformatics 23(2):207–214
19. de Hoon MJL, Imoto S, Nolan J, et al. (2004) Open source clustering software. Bioinformatics 20(9):1453–1454
20. Saeed AI, Bhagabati NK, Braisted JC, et al, *[9] TM4 microarray software suite*, in *Methods in enzymology*2006, Academic Press, 134–193
21. Fath BD, Scharler UM, Ulanowicz RE, et al. (2007) Ecological network analysis: network construction. Ecol Model 208(1):49–55
22. Zhou J, Deng Y, Luo F, et al (2010) Functional molecular ecological networks. mBio 1(4), doi: https://doi.org/10.1128/mBio.00169-10
23. Marsh PD (1994) Microbial ecology of dental plaque and its significance in health and disease. Adv Dent Res 8(2):263–271
24. Ling Z, Kong J, Jia P, et al. (2010) Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing. Microb Ecol 60(3):677–690

25. Gomar-Vercher S, Cabrera-Rubio R, Mira A, et al. (2014) Relationship of children's salivary microbiota with their caries status: a pyrosequencing study. Clin Oral Invest 18(9):2087–2094
26. Tu Q, He Z, Li Y, et al. (2014) Development of HuMiChip for functional profiling of human microbiomes. PLoS One 9:e90546
27. Higham SM, Edgar WM (1989) Human dental plaque pH, and the organic acid and free amino acid profiles in plaque fluid, after sucrose rinsing. Arch Oral Biol 34(5):329–334
28. Marsh PD, Martin M (2009) Oral microbiology, 5ed edn. Churchill Livingstone Elsevier, London
29. Hojo S, Komatsu M, Okuda R, et al. (1994) Acid profiles and pH of carious dentin in active and arrested lesions. J Dent Res 73(12): 1853–1857
30. Kim JN, Ahn SJ, Burne RA (2015) Genetics and physiology of acetate metabolism by the Pta-Ack pathway of *Streptococcus mutans*. Appl Environ Microb 81(15):5015–5025
31. Van Wuyckhuyse BC, Perinpanayagam HER, Bevacqua D, et al. (1995) Association of free arginine and lysine concentrations in human parotid saliva with caries experience. J Dent Res 74(2): 686–690
32. He J, Hwang G, Liu Y, et al. (2016) L-arginine modifies the exopolysaccharides matrix and thwarts *Streptococcus mutans* outgrowth within mixed-species oral biofilms. J Bacteriol. https://doi.org/10.1128/jb.00021-16
33. Nascimento MM, Gordan VV, Garvan CW, et al. (2009) Correlations of oral bacterial arginine and urea catabolism with caries experience. Oral Microbiol Immunol 24(2):89–95
34. Nascimento MM, Liu Y, Kalra R, et al. (2013) Oral arginine metabolism may decrease the risk for dental caries in children. J Dent Res 92(7):604–608
35. Nascimento M, Burne R (2014) Caries prevention by arginine metabolism in oral biofilms: translating science into clinical success. Curr Oral Health Rep 1(1):79–85
36. Lamont HC, Staudenbauer WL, Strominger JL (1972) Partial purification and characterization of an aspartate racemase from *Streptococcus faecalis*. J Biol Chem 247(16):5103–5106
37. Servet C, Ghelis T, Richard L, et al. (2012) Proline dehydrogenase: a key enzyme in controlling cellular homeostasis. Front Biosci (Landmark Ed) 1(17):607–620
38. Anderson MS, Bulawa CE, Raetz CR (1985) The biosynthesis of gram-negative endotoxin formation of lipid A precursors from UDP-GlcNAc in extracts of *Escherichia coli*. J Biol Chem 260(29):15536–15541
39. Crowell DN, Anderson MS, Raetz CR (1986) Molecular cloning of the genes for lipid A disaccharide synthase and UDP-N-acetylglucosamine acyltransferase in *Escherichia coli*. J Bacteriol 168(1):152–159
40. Fujishima H, Nishimura A, Wachi M, et al. (2002) kdsA mutations affect FtsZ-ring formation in *Escherichia coli* K-12. Microbiology 148(1):103–112
41. Vuorio R, Vaara M (1995) Comparison of the phenotypes of the lpxA and lpxD mutants of *Escherichia coli*. FEMS Microbiol Lett 134(2–3):227–232
42. Vuorio R, Vaara M (1992) The lipid A biosynthesis mutation lpxA2 of *Escherichia coli* results in drastic antibiotic supersusceptibility. Antimicrob Agents Chemother 36(4):826–829
43. Vaara M, Nurminen M (1999) Outer membrane permeability barrier in *Escherichia coli* mutants that are defective in the late acyltransferases of lipid A biosynthesis. Antimicrob Agents Chemother 43(6):1459–1462
44. Wolff M, Seemann M, Tse Sum Bui B, et al. (2003) Isoprenoid biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB/IspH) from *Escherichia coli* is a [4Fe-4S] protein. FEBS Lett 541(1–3): 115–120
45. McAteer S, Coulson A, McLennan N, et al. (2001) The lytB gene of *Escherichia coli* is essential and specifies a product needed for isoprenoid biosynthesis. J Bacteriol 183(24):7403–7407
46. Harkness RE, Kusser W, Qi BJ, et al. (1992) Genetic mapping of the lytA and lytB loci of *Escherichia coli*, which are involved in penicillin tolerance and control of the stringent response. Can J Microbiol 38(9):975–978
47. Gustafson CE, Kaul S, Ishiguro EE (1993) Identification of the *Escherichia coli* lytB gene, which is involved in penicillin tolerance and control of the stringent response. J Bacteriol 175(4):1203–1205
48. Sangrila S, Kanti MT (2013) Cellulase production by bacteria: a review. Br Microbiol Res J 3(3):235–258
49. Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21): 2957–2963
50. Kong Y (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98(2):152–153
51. Edgar RC, Haas BJ, Clemente JC, et al. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16): 2194–2200
52. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460–2461
53. Wang Q, Garrity GM, Tiedje JM, et al. (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. App Environ Microbiol 73(16):5261–5267
54. Schloss PD, Westcott SL, Ryabin T, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537–7541
55. Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 18(1):117–143
56. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Aust Ecol 26(1):32–46
57. Angly FE, Fields CJ, Tyson GW (2014) The bio-community Perl toolkit for microbial ecology. Bioinformatics 30:1926–1927
58. Deng Y, Jiang YH, Yang Y, et al. (2012) Molecular ecological network analyses. BMC Bioinformatics 13(1):1–20
59. Smoot ME, Ono K, Ruscheinski J, et al. (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–432
60. Assenov Y, Ramírez F, Schelhorn SE, et al. (2008) Computing topological parameters of biological networks. Bioinformatics 24(2):282–284