OPEN

ORIGINAL ARTICLE

On the ecosystemic network of saliva in healthy young adults

Egija Zaura[1,2,9], Bernd W Brandt[1,2,9], Andrei Prodan[1,3], Maarten Joost Teixeira de Mattos[4],
Sultan Imangaliyev[1,2,5], Jolanda Kool[1,5], Mark J Buijs[1,2], Ferry LPW Jagers[1,5],
Nienke L Hennequin-Hoenderdos[6], Dagmar E Slot[6], Elena A Nicu[6], Maxim D Lagerweij[7],
Marleen M Janus[1,2], Marcela M Fernandez-Gutierrez[1,8], Evgeni Levin[1,5], Bastiaan P Krom[1,2],
Henk S Brand[1,3], Enno CI Veerman[3], Michiel Kleerebezem[1,8], Bruno G Loos[6],
GA van der Weijden[6], Wim Crielaard[1,2] and Bart JF Keijser[1,2,5]

[1]Top Institute Food and Nutrition, Wageningen, The Netherlands; [2]Department of Preventive Dentistry, Academic Centre for Dentistry Amsterdam, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [3]Department of Oral Biochemistry, Academic Centre for Dentistry Amsterdam, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [4]Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands; [5]Research Group Microbiology and Systems Biology, TNO Earth, Life and Social Sciences, Zeist, The Netherlands; [6]Department of Periodontology, Academic Centre for Dentistry Amsterdam, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [7]Department of Cariology, Endodontology, Pedodontology, Academic Centre for Dentistry Amsterdam, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands and [8]Department of Animal Sciences, Host-Microbe Interactomics Group, Wageningen University, Wageningen, The Netherlands

A dysbiotic state is believed to be a key factor in the onset of oral disease. Although oral diseases have been studied for decades, our understanding of oral health, the boundaries of a healthy oral ecosystem and ecological shift toward dysbiosis is still limited. Here, we present the ecobiological heterogeneity of the salivary ecosystem and relations between the salivary microbiome, salivary metabolome and host-related biochemical salivary parameters in 268 healthy adults after overnight fasting. Gender-specific differences in the microbiome and metabolome were observed and were associated with salivary pH and dietary protein intake. Our analysis grouped the individuals into five microbiome and four metabolome-based clusters that significantly related to biochemical parameters of saliva. Low salivary pH and high lysozyme activity were associated with high proportions of streptococcal phylotypes and increased membrane-lipid degradation products. Samples with high salivary pH displayed increased chitinase activity, higher abundance of *Veillonella* and *Prevotella* species and higher levels of amino acid fermentation products, suggesting proteolytic adaptation. An over-specialization toward either a proteolytic or a saccharolytic ecotype may indicate a shift toward a dysbiotic state. Their prognostic value and the degree to which these ecotypes are related to increased disease risk remains to be determined.
The ISME Journal (2017) **11**, 1218–1231; doi:10.1038/ismej.2016.199; published online 10 January 2017

Introduction

The human oral cavity is a complex ecosystem, where hundreds of bacterial taxa and other microorganisms interact with each other (Dewhirst *et al.*, 2010; Wade, 2013) and with their environment, which is provided by the host. The oral environment is a dynamic system, influenced by intrinsic biological host parameters, behavior and external factors such as the circadian rhythm that affects salivary flow and composition (Dawes, 1972). The most studied behavioral factors include diet, especially high consumption of carbohydrates, which has a profound effect on the oral ecosystem and dental decay (Jensen, 1999; Paes Leme *et al.*, 2006), and smoking, which has been shown to disturb host-microbe interplay and result in increased proinflammatory responses of the host (Joshi *et al.*, 2014). In addition, systemic conditions (Preshaw *et al.*, 2012) and use of medications (Yuan and Woo, 2015) are all linked to changes in the oral environment.

Despite the open structure of the oral ecosystem, the oral microbiota composition is relatively stable over time (Zhou *et al.*, 2013). Nevertheless, its activity undergoes large variations during the day with 'feast' and 'famine' periods (Lingstrom *et al.*, 1993; Zaura and ten Cate, 2004). Carbohydrate intake by the host (the 'feast') results in temporary metabolic overflow, during which bacterial fermentation results in acid formation (Gerardu *et al.*, 2007). During the resting ('famine') period, salivary flow clears the acids and contributes to pH recovery (Humphrey and Williamson, 2001). The resting period at night functions as a natural recovery and growth phase for the oral microbial communities, when saliva is their main nutrient source (Marsh *et al.*, 2016). The difference between compositional stability and functional variability at a daily scale suggests that the oral ecosystem has powerful mechanisms to maintain its stability. Saliva with its pH buffering capacity and antimicrobial properties may have a key role in the maintenance of the ecological balance (Humphrey and Williamson, 2001).

It is believed that a prolonged state of imbalance may eventually lead to irreversible changes in the oral ecosystem, contributing to the transition of a healthy oral ecosystem toward dysbiosis and disease (Marsh *et al.*, 2011). Although this notion is well accepted, we have little understanding on the processes that underlie this transition and to date we are unable to predict individual predisposition toward a dysbiotic or disease-prone state.

Here, we aimed at describing the boundaries of and the heterogeneity within the healthy oral ecosystem and at deciphering the relations between salivary bacterial composition (microbiome), salivary metabolites (metabolome) and the host-related biochemical salivary parameters and diet. We controlled for circadian variability and influences from the host environment by assessing these variables after overnight fasting in non-smoking, systemically and orally healthy young adults.

Materials and methods

This cross-sectional observational study was carried out on 268 individuals (18–32 years) who fasted overnight and refrained from all oral hygiene procedures for 24 h. Detailed materials and methods are available as Supplementary Methods. The study protocol was reviewed and approved by the Medical Ethics Committee and registered at the Dutch Trial Register (NTR3649). The inclusion and exclusion criteria of this study have been described previously (Oliveira *et al.*, 2015; Prodan *et al.*, 2015). Systemic health status was confirmed by routine clinical chemical analysis of peripheral blood.

Saliva collection and processing
Unstimulated saliva was collected as described previously (Prodan *et al.*, 2015) in two sessions of 5 min with a 5 min interval between the two samples. Sample collection occurred between 0900 and 1000 hours, before any food or drink intake that day, by drooling into a sterile ice-cooled vial. The first sample was immediately aliquoted into two vials and stored at −80 °C for microbiological and metabolome analyses, respectively, whereas the second sample was processed for biochemical analyses (Prodan *et al.*, 2015). Assessments included salivary flow rate measurements, pH and buffered pH measurements, determination of the protein content, salivary mucins (MUC5B and MUC7), albumin, lactoferrin, secretory-IgA, cystatin S and enzymatic activities of amylase, chitinase, lysozyme and protease (Supplementary Methods).

Dietary assessment
A Food Frequency Questionnaire (FFQ) was developed to assess habitual dietary intake in the previous month (Feunekes *et al.*, 1993; Siebelink *et al.*, 2011). The FFQ was based on consumption data of 20–40 year old participants of the Dutch National Food Consumption Survey of 2007–2010 (van Rossum *et al.*, 2011) and the Dutch food composition database (RIVM, 2010). The FFQ included questions on the frequency and amount of intake of 130 food items. Selection of food items and calculation of the weighted average nutrient composition of the food items were performed using the validated Dutch FFQ-tool (Molag, 2010).

16S rRNA gene amplicon sequencing and data processing
Microbial DNA was extracted and processed for amplicon sequencing of the V4 hypervariable region of the 16S rRNA gene on the Illumina MiSeq platform (Supplementary Methods). The sequences were processed with mothur v.1.31.2 (Schloss *et al.*, 2009). The data are available at NCBI BioProject database under accession number PRJNA348169. The data analyses are described in Supplementary Methods.

Metabolome assessment
The saliva samples were extracted, prepared and processed at Metabolon (Durham, NC, USA) as described previously (Evans *et al.*, 2009, Supplementary Methods). The normalized metabolome data set provided by Metabolon was range-scaled between 0 and 10. Metabolites with a single value were omitted from statistical analysis, resulting in a data set with 493 metabolites.

Statistical analyses

Spectral clustering (SC) was performed using the Neighborhood Co-regularized SC algorithm (Tsivtsivadze *et al.*, 2013; Imangaliyev *et al.*, 2015; http://www.learning-machines.com), based on the SC method (Luxburg, 2007, Supplementary Methods).

Principal component analysis (PCA), permutational analysis of variance (PERMANOVA), Shannon diversity index and Chao-1 estimate of species richness were calculated using PAST software v. 3.04 (Hammer *et al.*, 2001).

Inter-cluster differences in host parameters and in microbiome alpha diversity were assessed using the Kruskal–Wallis and Mann–Whitney test and the differences in gender by the Pearson χ^2 test in SPSS version 21. Benjamini and Hochberg False Discovery Rate correction (Benjamini and Hochberg, 1995) of the *P*-values for multiple comparisons was performed in R.

The LDA effective size (LEfSe) biomarker discovery tool (Segata *et al.*, 2011) was used with the 'one against all' strategy for multiclass analysis and logarithmic LDA score threshold of 2, $P < 0.05$.

Assessment of significant patterns of microbial co-occurrence or mutual exclusion was performed using CoNet (Faust *et al.*, 2012) and visualized in Cytoscape v. 3.3.0.

High-dimensional feature selection by significance analysis of microarrays analysis (Tusher *et al.*, 2001), using the TIGR MultiExperiment Viewer (111. tm4.org/mev.html), was used to compare scaled-metabolite data among metabolome sample clusters and to assess the differences in host-related clusters and microbiome clusters by metabolites.

Elastic Net regression (Zou and Hastie, 2005) with stability selection (Meinshausen and Bühlmann, 2010) was used to select features (OTUs or metabolites) that were related to the host-related biochemical salivary parameters.

The relation between the microbiome (OTUs) and the metabolome data set (all 493 metabolites) was assessed using Spearman Correlation with Bonferroni-corrected *P*-values. The concordance between the two data sets was assessed using Procrustes Analysis (Gower, 1975) and the Mantel test (Mantel, 1967) implemented in QIIME (Caporaso *et al.*, 2010) v.1.8.0, both with 10 000 random permutations.

Results

In this study, we quantified the heterogeneity of and the mutual interrelationships between the bacterial composition of unstimulated saliva (microbiome), salivary metabolites (metabolome), and host-related parameters in a systemically and orally healthy young adult population. The biochemical salivary parameters (Prodan *et al.*, 2015) and clinical parameters (Oliveira *et al.*, 2015) have been reported previously. Here, we first assessed the interindividual variance in the salivary microbiome and metabolome profiles and then related these data sets to each other and to different host parameters.

In total, 268 individuals (150 males, 118 females, mean age 22.6 years, s.d. 2.6, range 18–32), at good systemic health and no apparent dental caries or periodontal disease were included in the study (Supplementary Table S1). The FFQ was completed by 255 individuals, of whom 16 (6.3%) reported a vegetarian lifestyle.

Heterogeneity in salivary microbiome and metabolome profiles

We first assessed the heterogeneity within the salivary microbiome and metabolome data sets. For this we applied Neighborhood Co-regularized Spectral Clustering (NCSC)—a robust and in microbial ecological studies validated unsupervised machine-learning approach (Biesbroek *et al.*, 2014; Borgdorff *et al.*, 2014; Imangaliyev *et al.*, 2015). NCSC is a partitioning technique that constructs a weighted graph from the input data set where each node represents a pattern and each weighted edge takes into account the similarity between patterns. Unlike classical partitioning clustering algorithms, spectral clustering is able to produce nonlinear separating hypersurfaces among data, since it constructs an adjacency structure from the input data set (Filippone *et al.*, 2008).

Within the salivary microbiome data set (Supplementary Table S2), three main microbiome sample co-occurrence clusters (MIC1: $N = 191$; MIC2: $N = 59$, MIC3: $N = 17$) (Figure 1a, left panel) were identified. The MIC1 cluster was further subdivided into three sub-clusters (MIC1.1: $N = 48$; MIC1.2: $N = 63$; MIC1.3: $N = 80$). The samples that clustered into MIC1.2 showed the lowest alpha diversity: the species richness (number of OTUs/sample), the Shannon Diversity and the Chao-1 index (Figure 1a) were significantly lower in MIC1.2 samples compared with the samples from any other cluster ($P < 0.001$). The Shannon Diversity of the MIC1.2 samples was 3.530 (s.d. = 0.27), whereas for MIC1.1 it was 3.693 (s.d. 0.27), MIC1.3—3.776 (s.d. 0.24), MIC2—3.775 (s.d. 0.31) and MIC3—3.859 (s.d. 0.23).

Application of the spectral co-occurrence analysis to the metabolome data (Supplementary Table S3) resulted in four metabolome clusters, where the majority of samples (74%) showed high co-occurrence of metabolites (MET1, $N = 198$), followed by three smaller clusters with 44 (MET2), 19 (MET3) and 5 samples (MET4), respectively (Figure 1a, right panel). Samples in cluster MET4 had the highest average number of metabolites (417 metabolites/sample, s.d. 6.8), followed by MET3 (409, s.d. 11) and MET2 (386, s.d. 14), whereas samples in the cluster MET1 had the lowest number of metabolites per sample (315, s.d. 41) compared with the other

clusters ($P<0.0001$). The PCA and PERMANOVA confirmed the spectral clustering results of both data sets (Figure 1b).

Next, we determined which microbial OTUs and genera contributed to the observed differences among the microbiome sample clusters
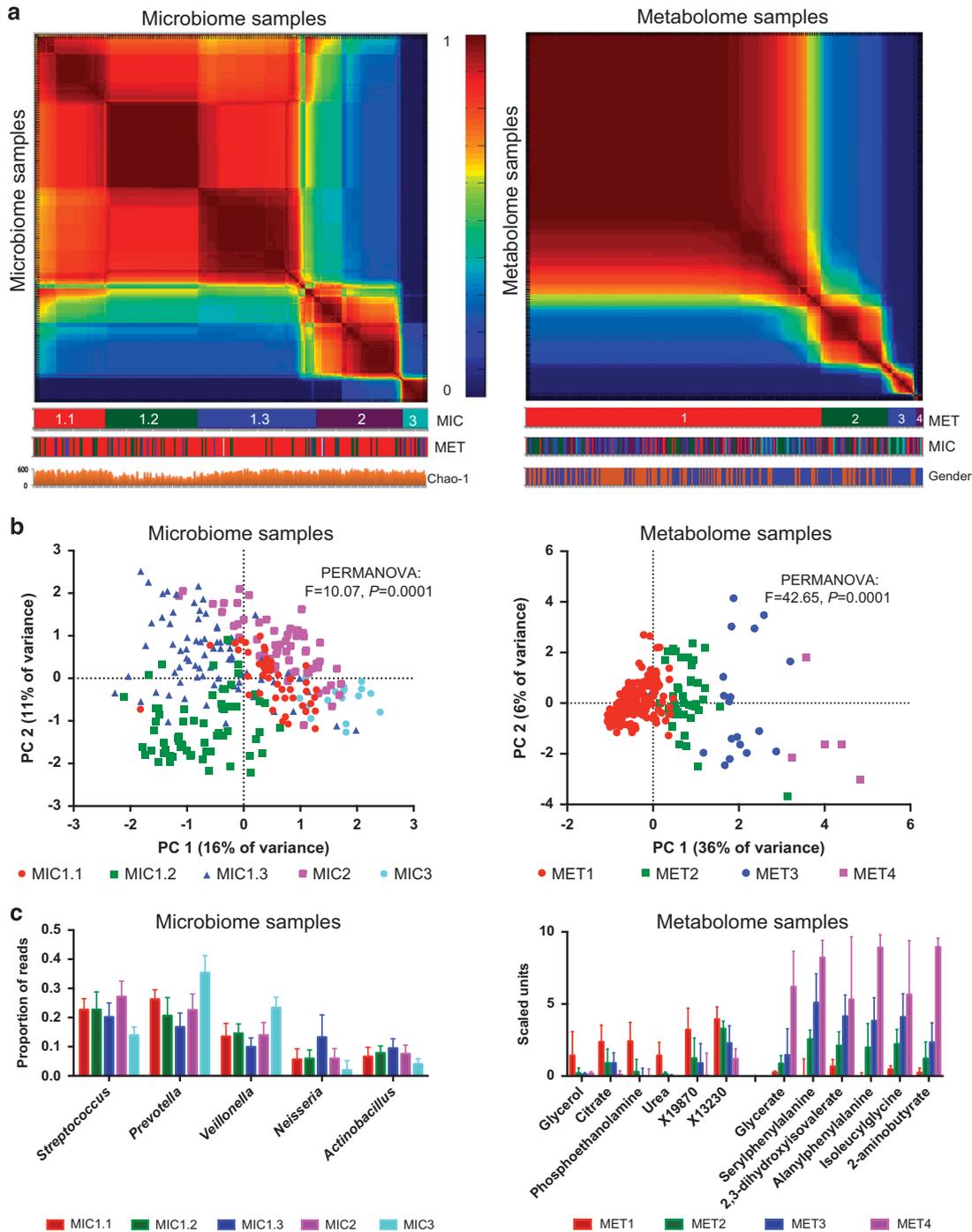
Figure 1 Heterogeneity of salivary microbiome (left panels) and metabolome (right panels) samples in healthy young adults: (**a**) spectral clustering co-occurrence plots of the microbiome or metabolome samples ordered along the axes according to the co-occurrence matrix: the more similar the sample profiles, the closer they are together on the axis. Co-occurrence values range from 0 (samples never cluster) to 1 (samples always cluster together) after multiple k-means clustering assignments. The clustering labels are shown below the graphs, according to salivary microbiome (MIC) and metabolome (MET) data sets. Chao-1—estimated species richness per individual microbiome sample. There were significantly more males (gender: blue) than females (gender: orange) in MET_2 and MET_3 clusters and MET_4 contained only males. (**b**) PCA plots based either on microbiome or metabolome samples. (**c**) Significantly positively and significantly negatively associated microbial genera and metabolites between samples belonging to different clusters. Only five most abundant genera are shown. Of the 217 negatively associated metabolites only the 6 with the highest negative fold change are shown.

(Supplementary Table S4, Figure 1c). The most divergent cluster, MIC3 ($N = 17$), had the highest proportion of *Veillonella atypica/Veillonella dispar* (OTU1) and *Prevotella* (OTU35, OTU94) among the five sample clusters. Cluster MIC2 was discriminated by the highest proportion of *Streptococcus mitis* group (OTU4), *Streptococcus gordonii/Streptococcus* sp. HOT056 (OTU86) and *Rothia mucilaginosa* (OTU8); cluster MIC1.3—by the highest proportion of *Neisseria flavescens/ Neisseria subflava* (OTU23) and *Haemophilus parainfluenzae* (OTU2); cluster MIC1.2—by *Streptococcus salivarius/Streptococcus vestibularis* (OTU14), *Streptococcus australis/Streptococcus parasanguinis* (OTU15) and *Granulicatella adiacens* (OTU41) and cluster MIC1.1—by *Prevotella* sp. HOT313 (OTU31) and *Paraprevotella/Alloprevotella* sp. HOT308 (OTU234).

In the metabolome data set, we identified 222 metabolites that significantly discriminated the large sample cluster (MET1) from the three other clusters. Only 6 out of the 222 metabolites—unidentified metabolites X19870 and X13230, phosphoethanolamine, glycerol, citrate and urea—were significantly higher in MET1, whereas the remaining metabolites showed a lower abundance as compared with the other clusters (Supplementary Table S5, Figure 1c).

Thereafter we assessed the structure of the mutual correlations among the most prevalent taxa within each microbiome cluster (Figure 2). The bacterial co-occurrence network of cluster MIC3 had the highest complexity due to the highest number of co-occurring taxa (66) and the highest average number of neighbors (2.88), followed by MIC2 (50 taxa and 2.84 neighbors), MIC1.3 (43 taxa and 2.65 neighbors), MIC1.2 (36 taxa and 1.83 neighbors) and MIC1.1 (31 taxa and 1.87 neighbors). The most frequent co-occurrences were found between the taxa belonging to the same genus. The highest number of neighbors belonging to different genera was found in MIC1.3, where *Prevotella salivae* co-occurred with *S. salivarius/S. vestibularis*, *Leptotrichia*, *Selenomonas*, *Megasphaera micronuciformis*, *V. atypica/V. dispar* and *Atopobium parvulum*. Depending on the cluster, different prevotellae species, *M. micronuciformis*, *Fusobacteria* or *V. atypica/V. dispar* displayed the most co-occurrences with the other taxa. Streptococci, although highly abundant in most of the sample clusters, did not form complex networks with other genera.

Mutual interrelationships between the microbiome, metabolome and host-related parameters

Relation between salivary microbiome and metabolome. Microbiome and metabolome data sets had similar structure when their heterogeneity was visualized by spectral clustering (Figure 1a). However, the samples within the respective clusters of one data set did not show any relation with the clusters of the other data set (MIC- and MET-code colored bars in Figure 1a).

We first tested the correlation of the two data sets using the Mantel test on Bray–Curtis distance matrices of the metabolome and the microbiome data. The matrices correlated significantly ($P = 0.0001$), though with a low correlation coefficient ($r = 0.199$). In addition, the PCA plots of both data sets were superimposed using Procrustes analysis, where the majority of the metabolome samples clustered at the center of the plot (Supplementary Figure S1). Although Procrustes output was significant ($P < 0.0001$), the sum of the squared deviations was very high ($M^2 = 0.905$), indicating low similarity between the principal components of the two data sets.

Although there was a low similarity between the metabolome and microbiome profile from each individual, we did find significant relationships between individual OTUs and metabolites. Of the top 400 OTUs, 167 correlated significantly with one or more (maximum 232) metabolites and of the 493 metabolites, 282 correlated with at least one (maximum 96) OTU (Bonferroni-corrected $P < 0.05$; $r < -0.2$ or $r > 0.2$). A clear dichotomy in the relation of these OTUs with metabolites was observed: the taxa had either strong positive or strong negative associations with metabolites (Supplementary Figure S2). The most significant correlations ($r < -0.5$ or $r > 0.5$) centered around two OTUs: one of the largest OTUs (13% of all reads)—OTU4 (*S. mitis* group)—and a relatively small OTU (0.8% of all reads)—OTU66 (*M. micronuciformis*) (Figure 3). OTU4 associated negatively with 228 metabolites (35 with $r < -0.5$) and positively with only four metabolites (urea: $r = 0.45$, phosphoethanolamine: $r = 0.39$, glycerol: $r = 0.33$ and glycerol-3-phosphate: $r = 0.32$). In contrast, OTU66 correlated negatively only with three metabolites (urea: $r = -0.34$, phosphoethanolamine: $r = -0.33$ and choline phosphate: $r = -0.32$) and positively with 224 metabolites (13 with $r > 0.5$). Among the metabolites the strongest network was observed for 3-phenylpropionate (hydrocinnamate), which had four negative and 29 positive correlations with OTUs. Of these, seven strongly correlated ($r > 0.5$) with OTUs classified as *Prevotella* (2 OTUs), *Alloprevotella, Fusobacterium periodonticum, Leptotrichia, Peptostreptococcus stomatis* and *Eubacterium sulci* (Figure 3).

Returning to the observed heterogeneity in the metabolome data set, we could relate the metabolome clusters with specific OTUs: Of the 912 OTUs, 207 were significantly discriminatory among the MET clusters (Supplementary Table S6). Of these, 30 OTUs discriminated one of the metabolome clusters from the others at an LDA score 3 or above. The largest cluster (MET1, $N = 198$ samples) had a significantly higher abundance of OTU4 (*S. mitis* group), OTU2 (*H. parainfluenzae*) and OTU27 (*Porphyromonas* sp. HOT278, HOT279), whereas other clusters were dominated by OTUs classified as veillonellae and prevotellae.
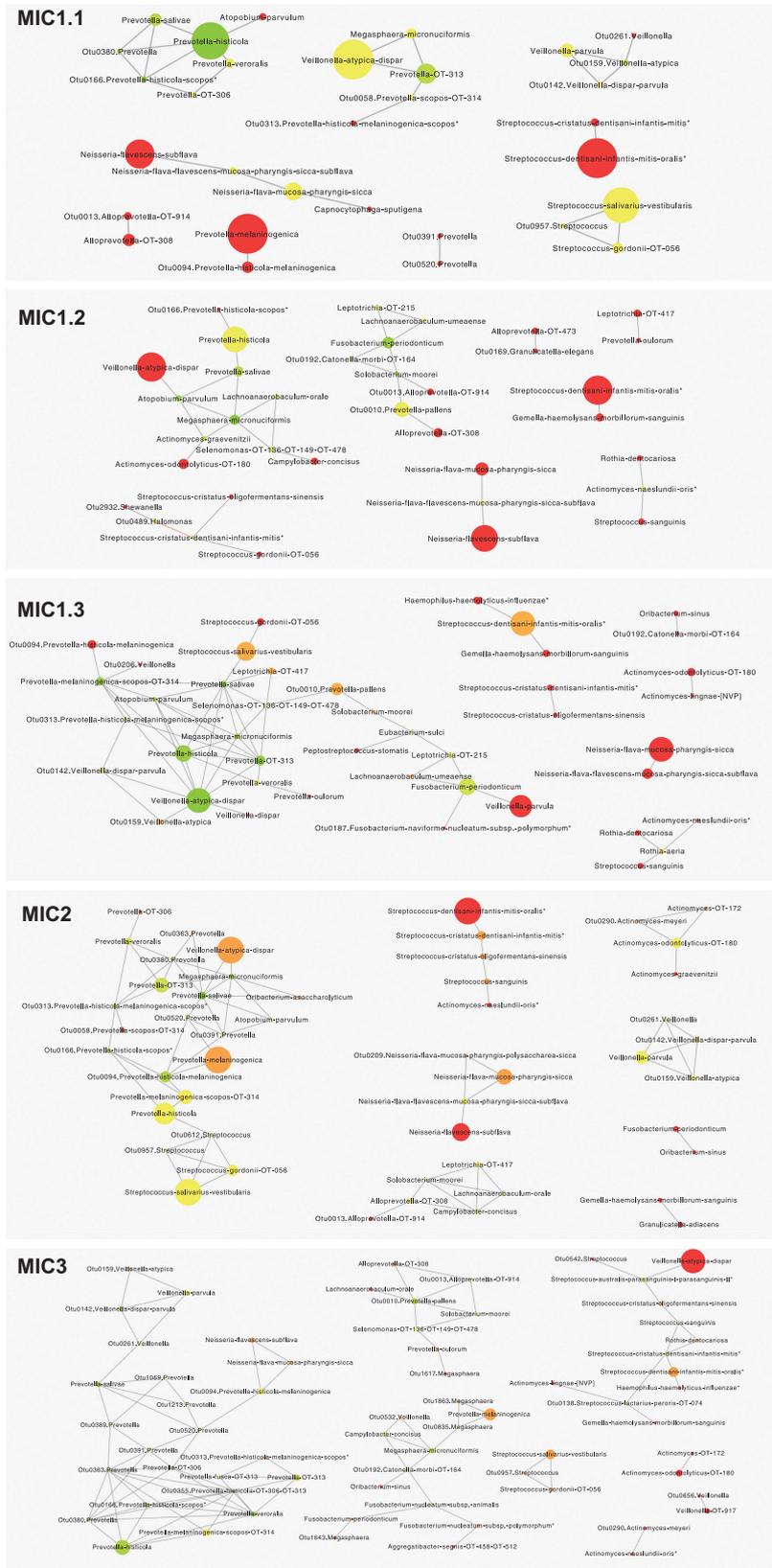
Figure 2 Salivary microbiome co-occurrence analysis results per individual microbiome cluster. The size of the nodes is related to the relative abundance of the taxa; the color of the node indicates the connectivity to the other nodes (red—low number of neighbors, green—high number of neighbors). Analysis was performed using CoNet v.1.0b6 in Cytoscape. Taxonomic names at species level were obtained using the representative sequences of the OTUs and the HOMD database. Taxonomic names marked with * have been truncated for legibility.

Figure 3 The network of the most significant Spearman correlations ($r < -0.5$ or $r > 0.5$) between the top 400 most abundant OTUs and the 493 metabolites. OTUs are shown with blue circles, the diameter of which is proportional to the abundance, metabolites—with yellow circles. Positive correlations are indicated with green lines, whereas negative correlations—with red lines. Taxonomic names marked with * have been truncated for legibility.

Similarly, the observed heterogeneity in the microbiome data set related with specific metabolites: Of the 493 metabolites, 64 were significantly discriminatory (significance analysis of microarrays analysis; median FDR 0%, delta 0.0033) among the microbiome clusters (Supplementary Figure S3), the majority of which were most abundant in the smallest microbiome cluster (MIC3, $N = 17$) compared with the other samples. The largest difference between MIC3 and the other samples was observed in higher abundance of the dipeptides: histidylhistidine, leucylserine, leucyltyrosine and phenylacetate, and two unidentified metabolites—X18037 and X18165.

Host-related parameters and saliva. Neither the age nor the clinical variables (gingival inflammation, amount of plaque, caries experience) related to the observed heterogeneity—the clusters—of salivary microbiome and metabolome data sets (results not shown).

Besides the microbiome and the metabolome, fourteen host-related biochemical parameters relevant to oral health were analyzed from the same saliva sample set (Prodan *et al.*, 2015). These included salivary flow rate, pH and various salivary enzymes. First we assessed if these parameters associated with the clusters in both, microbiome and metabolome data sets. Then we applied Elastic Net regression analysis, which is a sample cluster-independent approach that allows prediction of

variables based on most stable features (OTUs or metabolites in our case) in multivariate data sets.

Heterogeneity of the microbiome and metabolome data sets related to some of the biochemical parameters of saliva: the samples in microbiome cluster MIC3 had higher salivary chitinase and albumin and lower lysozyme activity than samples in the other MIC clusters, whereas clusters MIC1.2 and MIC2 had the lowest salivary pH of all clusters (Figure 4a). Regarding the metabolome data set, the biochemical salivary parameters changed gradually through the MET clusters: pH, buffered pH, chitinase activity, albumin and MUC5B increased, whereas lysozyme activity decreased from MET1 through MET4 (Figure 4b).

Elastic Net regression analysis revealed that salivary pH, buffered pH (Figure 5a) and lysozyme activity (Figure 5b) could be predicted with good accuracy based on microbiome composition. Among the 10 most stable OTUs that predicted salivary buffered pH, eight OTUs (classified as *Streptococcus*, *Actinomyces*, *Granulicatella adiacens*, *Rothia dentocariosa* and *Prevotella oris*) increased in abundance with decreasing pH (most abundant OTUs shown in Figure 5c), whereas only two OTUs (classified as *Capnocytophaga granulosa* and *Neisseria*) were more abundant at higher pH. Salivary lysozyme activity was positively associated with mainly Gram-positive taxa such as streptococci and actinomyces, and negatively associated with Gram-

Figure 4 Five of the 14 measured host-related biochemical salivary parameters where significant differences were observed among the samples belonging to the different (**a**) microbiome and (**b**) metabolome clusters. The lines connect significantly different clusters ($P < 0.05$, FDR corrected for multiple comparisons).

Figure 5 Results of Elastic Net regression on the salivary microbiome data set in predicting (**a**) buffered salivary pH and (**b**) salivary lysozyme activity, where distribution of the top four most stable and abundant OTUs that predicted (**c**) buffered pH and (**d**) lysozyme activity are shown. For Elastic Net regression, the mean of the buffered pH or lysozyme activity was subtracted of the measured value as to center the data to a mean of zero. Buffered pH was divided into the following quartiles: Q1: pH 4.1–5.9, Q2: pH 5.9–6.3, Q3: pH 6.3–6.6, Q4: pH 6.6–7.3. Lysozyme activity was divided into Q1: 28–464, Q2: 464–1514, Q3: 1514–2491, Q4: 2491–5635.

negatives such as selenomonae and neisseriae (Figure 5d).

Based on the metabolites in saliva, 8 of the 14 host parameters—salivary albumin concentration ($r^2 = 0.25$, NRMSE = 9.5), total protein ($r^2 = 0.296$, NRMSE = 12), secretory-IgA ($r^2 = 0.45$, NRMSE = 7.4), pH ($r^2 = 0.28$, NRMSE = 12.8), buffered pH ($r^2 = 0.37$, NRMSE = 13.9) and activity of lysozyme ($r^2 = 0.35$, NRMSE = 20), amylase ($r^2 = 0.28$, NRMSE = 13.9) and proteases ($r^2 = 0.23$, NRMSE = 10.58)—could be predicted using Elastic Net regression. For instance, albumin concentration was positively associated with pro-inflammatory metabolites (arachidonate, dihomo-linoleate and 1-stearoylglycerophosphoinositol) and negatively with anti-inflammatory palmitoyl ethanolamide, whereas low pH, low buffered pH and high lysozyme activity were associated with the typical membrane lipid degradation products phosphoethanolamine, ethanolamine and glycerol.

Gender and saliva. We have shown that biochemical parameters of saliva exhibited gender-related differences: salivary pH, buffering capacity, protein content, MUC5B, secretory-IgA, and chitinase activity were all lower in females compared with males, whereas MUC7 and lysozyme activity were higher (Prodan *et al.*, 2015). Here, we assessed if salivary microbiome and metabolome can be stratified by gender.

LEfSe identified 65 OTUs that differentiated between males and females (Supplementary Table S7), of which 44 OTUs (including 19 streptococcal OTUs) were significantly more abundant in females, whereas male salivary microbiomes had a higher abundance of *Veillonella*, *Prevotella* and *Megasphaera*. There was no difference in Shannon Diversity between males (mean 3.695, s.d. 0.27) and females (3.699, s.d. 0.30). The salivary microbiome clusters did not show any relation with gender (results not shown).

However, the heterogeneity observed within the metabolome data did relate to gender of the subjects ($P = 0.009$): there were significantly more males than females in the smaller metabolome sample clusters: MET1 contained 66.9% of all males and 83.8% of all

females, MET2—20.3% and 12%, MET3—9.5% and 4.3% of all males and females, respectively, whereas MET4 contained only males (Figure 1a, right panel, gender). In addition, 205 metabolites were differentially abundant by gender (Supplementary Table S3). Of these, only three metabolites, namely glycerol, caffeine and 3-hydroxyindolin-2-one, were overrepresented in females. The top five most abundant metabolites in males compared with females were glycerate, tryptophylserine, unidentified metabolite X19863, proline and alanylphenylalanine.

Males had significantly more metabolites in their saliva (mean number of metabolites 349, s.d. 48) than females (321, s.d. 52) ($P = 0.000013$).

Diet and saliva. Based on the information obtained from the FFQ, males had a significantly higher average daily dietary intake of energy and nutrients than females (Supplementary Table S1). No significant relation between self-reported dietary lifestyle habits (vegetarians versus the omnivores) or individual food components and microbiome profiles or microbiome clusters was found (results not shown). Similarly, none of the metabolites discriminated between the dietary lifestyle habits of the individuals (significance analysis of microarrays analysis, results not shown). However, high total dietary protein correlated positively with isovalerylcarnitine ($r = 0.221$, $P < 0.0001$) and negatively with glycerol ($r = -0.221$, $P < 0.0001$). Interestingly, dietary protein also correlated with buffered pH of saliva ($r = 0.266$, $P = 0.00003$).

Discussion

Our results provide a quantitative presentation of the heterogeneity of the microbiome and metabolome within a healthy oral ecosystem as inferred by an analysis of salivary samples from a homogeneous study population—orally and systemically healthy young adults. In addition, host-related salivary parameters have been included to draw a complete picture of the dynamic interactions between microorganisms, their intermediate and end metabolites and the oral environment. We emphasize that the analysis is limited to the salivary microenvironment and does not explicitly include other oral niches. Our microbiome data-driven analysis stratified the salivary microbiome of healthy volunteers into five distinct clusters. The samples assigned to these clusters differed not only in microbial community composition and diversity, but also in salivary pH, buffered pH, albumin levels and activity of lysozyme and chitinase, indicating the important role of the host. Despite significant differences in the microbiome, no direct relation with the metabolome profiles was observed. Indeed, the majority of the individuals had highly similar salivary metabolic profiles. Two other relevant observations were made: (1) microbiome and metabolome of saliva was

gender-related and (2) dietary protein intake was associated with salivary pH.

Several host-related biochemical salivary parameters have already been reported to relate with gender (Prodan et al., 2015). Males, generally more susceptible to periodontal disease (Haytac et al., 2013), presented with significantly higher salivary pH than females, who are more prone to dental decay (Lukacs, 2011). Low salivary pH in females has been attributed to physiological factors, such as the influence of sex hormones on salivary gland gene expression and a smaller salivary gland size (Prodan et al., 2015). The higher abundance of streptococci found in female saliva could be an indirect result of an ecological pressure due to this physiologically lower pH (Takahashi and Nyvad, 2011). A previous metabolome study using NMR spectroscopy to compare 20 males with 20 females also found that nearly all metabolites that were significantly different by gender were higher in concentration in males (Takeda et al., 2009). Since the total salivary protein concentration was higher in males (Prodan et al., 2015), this could reflect in higher amino acid and peptide abundance in male metabolomes, selecting for the more proteolytic microbiome observed here.

With respect to diet, we found a positive relation between total dietary protein intake and salivary pH and a higher protein intake in males compared with females. This indicates a potential role for diet in modifying salivary pH, one of the strongest environmental factors in the oral cavity. However, dietary lifestyle habits did not influence the salivary microbiome or the metabolome after overnight fasting. A recent study specifically designed to address the influence of dietary lifestyle concluded that salivary metabolome, but not the microbiome, was affected (De Filippis et al., 2014). Since details on controlling for timing of food consumption were not reported, saliva might have been collected considerably shorter after the last food intake compared with the current study with overnight fasting, confirmed by low peripheral blood glucose values during sample collection. Another potential explanation for the lack of relation with metabolome and dietary lifestyle habits could be a limited number of non-omnivores participating in our study.

Our findings on the heterogeneity of salivary microbiome confirm previous reports on large groups of healthy subjects, where three (De Filippis et al., 2014) or four (Ding and Schloss, 2014) salivary community types were distinguished. Three community types in the study on 161 Italians were driven by the most abundant and most prevalent genera into *Neisseria-Fusobacterium*, *Prevotella* or *Streptococcus-Gemella* dominated communities (De Filippis et al., 2014). The HMP salivary data set resulted into four communities based on complex configurations of numerous taxa (Ding and Schloss, 2014), although detailed composition of each of the communities was not reported. Interestingly, only two community types were detected in the largest

study on salivary microbiome to date (Takeshita *et al.*, 2016), where one of the two types was significantly associated with oral diseases. This implies that differences between healthy and dysbiotic microbial communities, such as during dental decay or periodontal disease (Marsh, 2003), are more pronounced than the natural heterogeneity observed at health.

Unlike the heterogeneity in salivary microbiome, the majority of the metabolome samples had a highly similar profile. This suggests that salivary metabolism after overnight fasting in different individuals carrying different microbiota is driven by similar processes and fits well to previously described functional redundancy of human microbial communities (The Human Microbiome Consortium, 2012). However, about one fourth of the individuals had clearly different (for example, more proteolytic activity-driven) and more diverse metabolome profiles. These profiles related with several biochemical parameters of saliva, such as a higher pH and a higher albumin concentration, and with microbial communities dominated by anaerobes and Gram-negatives, the taxa associated with mature oral microbial communities and gingival inflammation (Rosier *et al.*, 2014). Pro-inflammatory metabolites in saliva correlated with salivary albumin concentration, suggesting presence of early inflammatory state (Henskens *et al.*, 1993).

The clear dichotomy in the correlations between salivary microbiota and metabolites was highly unexpected. It was observed in positive correlations between species of *Prevotella, Megasphaera, Fusobacterium, Eubacterium, Veillonella* and metabolites

of protein and amino acid fermentation. On the other side, the saccharolytic taxa such as streptococci showed positive correlations with glycerol-3-phosphate, choline phosphate and phosphor-ethanolamine, all related to phospholipid metabolism. The phospholipids are membrane components of both eukaryotic and prokaryotic cells. Lipids may have an important role in the oral ecosystem. Parotid gland saliva contains neutral lipids, glycolipids and phospholipids (Larsson *et al.*, 1996) and their concentration correlates with the caries experience of the individuals (Tomita *et al.*, 2008). Although lipids could serve as possible nutrient source under fasting conditions, free fatty acids released after lingual lipase activity may also have antimicrobial properties (Hamosh and Burns, 1977), providing an additional host mechanism for shaping the oral microbiota. Therefore the association between saccharolytic taxa and metabolites of phospholipid degradation found in this study is intriguing and prompts further research.

The most noteworthy finding of this study was that the dichotomy in microbiota-metabolite associations was also reflected in the composition of microbiome clusters and their association with host-related biochemical parameters such as salivary pH, activity of lysozyme and chitinase, and salivary albumin concentration. Based on these findings we propose that the five salivary microbiota clusters identified in these healthy individuals may represent different ecological states or 'ecotypes' of the oral ecosystem, with different ecological properties and different levels of specialization. The samples grouped into clusters that were functionally adapted to proteolysis
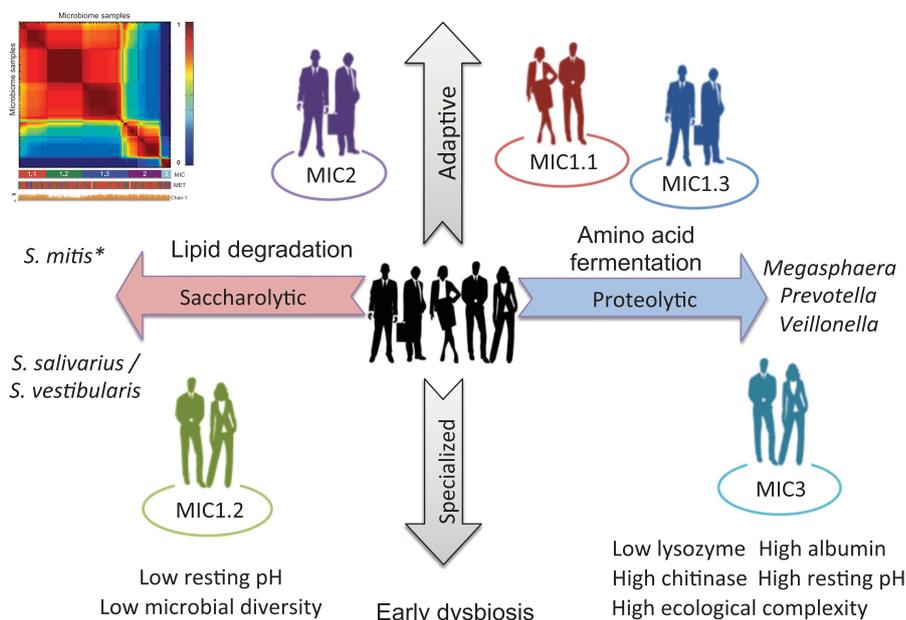
Figure 6 Proposed ecological states or ecotypes of the oral ecosystem and the positioning of the microbiome clusters according to these states. The dichotomy in bacteria-metabolite associations and the relation with salivary parameters is depicted in saccharolytic (left side) or proteolytic (right side) adaptations of the ecosystem. Based on the observed associations, the microbiome-based sample clusters MIC1.2 and MIC3 are positioned toward the specialized state of the system, whereas clusters MIC2, MIC1.1 and MIC1.3 are positioned within the adaptive state of the system. The more specialized the ecosystem becomes, the more it may shift toward dysbiosis.

and amino acid fermentation (MIC1.1, MIC1.3, MIC3), as well as clusters functionally adapted for saccharolysis (MIC1.2, MIC2) (Figure 6).

Amino acid fermentation is known to result in elevated pH due to bacterial deaminase and decarboxylase activity (Takahashi, 2015). Samples in cluster MIC3 had not only a higher pH, but also displayed signs of advanced ecological specialization toward protein and amino acid metabolism with the highest abundance of dipeptides among all clusters, with the most extensive microbiota community network linkage, low lysozyme and high chitinase activity, and high levels of salivary albumin. Elevated levels of salivary dipeptides have been associated with periodontal disease (Barnes et al., 2011), whereas high salivary albumin concentration—with gingival inflammation (Henskens et al., 1993). Collectively, even though our study population did not have clinical signs of oral disease, our findings indicate that individuals in cluster MIC3 may have entered an early inflammatory state with potential dysbiotic shift toward disease.

On the other side of the spectrum, we found ecological states that were more adapted to saccharolytic functions, represented by the *S. mitis* dominated cluster MIC2 and *S. salivarius/S. vestibularis* dominated cluster MIC1.2. Even in a resting state and during prolonged absence of external carbohydrate sources after an overnight fasting, the salivary pH in these individuals was lower compared with the clusters adapted to proteolysis. The saccharolytic cluster MIC1.2 showed a marked reduction in species diversity with the lowest number of ecological network connections, suggesting that streptococci might rather exploit the available nutrients than contribute to the community. Low pH pressure inhibits acid sensitive and selects aciduric oral species (Pham et al., 2009) that leads to decreased microbial diversity, as found during clinical stages of dental caries (Gomar-Vercher et al., 2014). Therefore we propose that cluster MIC1.2 represents an early phase of acidogenic adaptation, specialized in fast sugar metabolism and prone to a more cariogenic state (Takahashi, 2015). It is clear that longitudinal studies are required to establish the stability of these five ecosystem states, as well as their relation with the maintenance of oral health.

In conclusion, unstimulated saliva after overnight fasting in a healthy oral ecosystem is microbially heterogeneous. This heterogeneity is not related to salivary metabolites, but to biochemical host parameters of saliva. Host-related parameters, including gender, appear to affect the oral ecosystem in multiple ways. The clear dichotomy in the bacteria-metabolite associations and a relation with specific biochemical parameters suggests the presence of both, adaptive and highly specialized ecotypes of a healthy oral ecosystem. An over-specialization either toward proteolytic or saccharolytic ecotype may indicate a potential risk toward a dysbiotic state and consequently, disease. The degree with which the microbiome clusters of healthy individuals are related to a risk for the development of oral diseases remains to be assessed in clinical intervention studies.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Barnes VM, Ciancio SG, Shibly O, Xu T, Devizio W, Trivedi HM et al. (2011). Metabolomics reveals elevated macromolecular degradation in periodontal disease. *J Dent Res* **90**: 1293–1297.

Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical610 and powerful apporach to multiple testing. *J R Stat Soc* **57**: 289–300.

Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijser BJF et al. (2014). Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* **190**: 1283–1292.

Borgdorff H, Tsivtsivadze E, Verhelst R, Marzorati M, Jurriaans S, Ndayisaba GF et al. (2014). Lactobacillus-dominated cervicovaginal microbiota associated with reduced HIV/sti prevalence and genital HIV viral load in African women. *ISME J* **8**: 1781–1793.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Dawes C. (1972). Circadian rhythms in human salivary flow rate and composition. *J Physiol* **220**: 529–545.

De Filippis F, Vannini L, La Storia A, Laghi L, Piombino P, Stellato G et al. (2014). The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and vegan individuals. *PLoS One* **9**: e112373.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu WH et al. (2010). The human oral microbiome. *J Bacteriol* **192**: 5002–5017.

1230

Ding T, Schloss PD. (2014). Dynamics and associations of microbial community types across the human body. *Nature* **509**: 357–360.

Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* **81**: 6656–6667.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al*. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8**: e1002606.

Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. (1993). Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. *Am J Clin Nutr* **58**: 489–496.

Filippone M, Camastra F, Masulli F, Rovetta S. (2008). A survey of kernel and spectral methods for clustering. *Pattern Recognit* **41**: 176–190.

Gerardu VAM, Buijs MJ, van Loveren C, ten Cate JM. (2007). Plaque formation and lactic acid production after the use of amine fluoride/stannous fluoride mouthrinse. *Eur J Oral Sci* **115**: 148–152.

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

Gower JC. (1975). Generalized procrustes analysis. *Psychometrika* **40**: 33–51.

Hammer Ø, Harper DAT, Ryan PD. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* **4**: 1–9.

Hamosh M, Burns WA. (1977). Lipolytic activity of human lingual glands (Ebner). *Lab Invest* **37**: 603–608.

Haytac MC, Ozcelik O, Mariotti A. (2013). Periodontal disease in men. *Periodontol 2000* **61**: 252–265.

Henskens YMC, van der Velden U, Veerman ECI, Amerongen AVN. (1993). Protein, albumin and cystatin concentrations in saliva of healthy subjects and of patients with gingivitis periodonitis. *J Periodont Res* **28**: 43–48.

Humphrey SP, Williamson RT. (2001). A review of saliva: normal composition, flow, and function. *J Prosthet Dent* **85**: 162–169.

Imangaliyev S, Keijser B, Crielaard W, Tsivtsivadze E. (2015). Personalized microbial network inference via co-regularized spectral clustering. *Methods* **83**: 28–35.

Jensen ME. (1999). Diet and dental caries. *Dent Clin North Am* **43**: 615–633.

Joshi V, Matthews C, Aspiras M, de Jager M, Ward M, Kumar P. (2014). Smoking decreases structural and functional resilience in the subgingival ecosystem. *J Clin Periodontol* **41**: 1037–1047.

Larsson B, Olivecrona G, Ericson T. (1996). Lipids in human saliva. *Arch Oral Biol* **41**: 105–110.

Lingström P, Imfeld T, Birkhed D. (1993). Comparison of three different methods for measurement of plaque-pH in humans after consumption of soft bread and potato chips. *J Dent Res* **72**: 865–870.

Lukacs J. (2011). Sex differences in dental caries experience: clinical evidence, complex etiology. *Clin Oral Investig* **15**: 649–656.

Luxburg U. (2007). A tutorial on spectral clustering. *Stat Comput* **17**: 395–416.

Mantel N. (1967). The detection of disease clustering and a generalized regression apporach. *Cancer Res* **27**: 209–220.

Marsh PD. (2003). Are dental diseases examples of ecological catastrophes? *Microbiology* **149**: 279–294.

Marsh PD, Do T, Beighton D, Devine DA. (2016). Influence of saliva on the oral microbiota. *Periodontol 2000* **70**: 80–92.

Marsh PD, Moter A, Devine DA. (2011). Dental plaque biofilms: communities, conflict and control. *Periodontol 2000* **55**: 16–35.

Meinshausen N, Bühlmann P. (2010). Stability selection. *J R Stat Soc: Ser B Stat Methodol* **72**: 417–473.

Molag ML (2010). *Towards Transparent Development of Food Frequency Questionnaires: Scientific Basis of the Dutch FFQ-toolTM: A Computer System to Generate, Apply and Process FFQs*. Wageningen: Wageningen University.

Oliveira SC, Slot DE, Celeste RK, Abegg C, Keijser BJ, Van der Weijden FA. (2015). Correlations between two different methods to score bleeding and the relationship with plaque in systemically healthy young adults. *J Clin Periodontol* **42**: 908–913.

Paes Leme AF, Koo H, Bellato CM, Bedi G, Cury JA. (2006). The role of sucrose in cariogenic dental biofilm formation–new insight. *J Dent Res* **85**: 878–887.

Pham LC, van Spanning RJM, Röling WFM, Prosperi AC, Terefework Z, ten Cate JM *et al*. (2009). Effects of probiotic *Lactobacillus salivarius* W24 on the compositional stability of oral microbial communities. *Arch Oral Biol* **54**: 132–137.

Preshaw PM, Alba AL, Herrera D, Jepsen S, Konstantinidis A, Makrilakis K *et al*. (2012). Periodontitis and diabetes: a two-way relationship. *Diabetologia* **55**: 21–31.

Prodan A, Brand HS, Ligtenberg AJM, Imangaliyev S, Tsivtsivadze E, van der Weijden F *et al*. (2015). Interindividual variation, correlations, and sex-related differences in the salivary biochemistry of young healthy adults. *Eur J Oral Sci* **123**: 149–157.

RIVM (2010). Dutch food composition dataset (NEVO), vn 2010/2.0. National Institute for Public Health and the Environment, Ministry of 711 Health, Welfare and Sport (RIVM): Bilthoven.

Rosier BT, de Jager M, Zaura E, Krom BP. (2014). Historical and contemporary hypotheses on the development of oral diseases: are we there yet? *Front Cell Infect Microbiol* **4**: 92.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al*. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett W *et al*. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Siebelink E, Geelen A, de Vries JH. (2011). Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *Br J Nutr* **106**: 274–281.

Takahashi N. (2015). Oral microbiome metabolism: from 'who are they?' to 'what are they doing?'. *J Dent Res* **94**: 1628–1637.

Takahashi N, Nyvad B. (2011). The role of bacteria in the caries process. *J Dent Res* **90**: 294–303.

Takeda I, Stretch C, Barnaby P, Bhatnager K, Rankin K, Fu H *et al*. (2009). Understanding the human salivary metabolome. *NMR Biomed* **22**: 577–584.

Takeshita T, Kageyama S, Furuta M, Tsuboi H, Takeuchi K, Shibata Y *et al.* (2016). Bacterial diversity in saliva and oral health-related conditions: the hisayama study. *Sci Rep* **6**: 22164.

The Human Microbiome Consortium (2012). Structure, function and diversity of the healthy human micro-biome. *Nature* **486**: 207–214.

Tomita Y, Miyake N, Yamanaka S. (2008). Lipids in human parotid saliva with regard to caries experience. *J Oleo Sci* **57**: 115–121.

Zhou ZH, Schwenker F. *Partially Supervised Learning.* Berlin: Springer, 2013. pp 80–90.

Tusher VG, Tibshirani R, Chu G. (2001). Significance analysis of microarrays applied to the ionizing radia-tion response. *PNAS* **98**: 5116–5121.

van Rossum CTM, Fransen HP, Verkaik-Kloosterman J, Buurma-Rethans EJM, Ocke MC. (2011). *Dutch national food consumption survey 2007–2010.* RIVM: Bilthoven, The Netherlands. RIVM Report 350050006/2011.

Wade WG. (2013). Characterisation of the human oral microbiome. *J Oral Biosci* **55**: 143–148.

Yuan A, Woo S-B. (2015). Adverse drug events in the oral cavity. *Oral Surg Oral Med Oral Pathol Oral Radiol* **119**: 35–47.

Zaura E, ten Cate JM. (2004). Dental plaque as a biofilm: a pilot study of the effects of nutrients on plaque pH and dentin demineralization. *Caries Res* **38**: 9–15.

Zhou Y, Gao H, Mihindukulasuriya KA, La Rosa PS, Wylie KM, Vishnivetskaya T *et al.* (2013). Biogeography of the ecosystems of the healthy human body. *Genome Biol* **14**: R1.

Zou H, Hastie T (2005). Regularization and variable selection via the Elastic Net. *J R Stat Soc B* **67**: 301–320.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)