



Compositional Data Analysis of Periodontal Disease Microbial Communities

Laura Sisk-Hackworth^{1†}, Adrian Ortiz-Velez^{1†}, Micheal B. Reed² and Scott T. Kelley^{1*}

¹ Department of Biology, San Diego State University, San Diego, CA, United States, ² Department of Nanoengineering, Joint School of Nanoscience and Nanoengineering, North Carolina Agricultural and Technical State University, Greensboro, NC, United States

OPEN ACCESS

Edited by:

Marcelo Freire,
J. Craig Venter Institute (La Jolla),
United States

Reviewed by:

Karim El Kholy,
Harvard School of Dental Medicine,
United States
Ahmed Moustafa,
American University in Cairo, Egypt

*Correspondence:

Scott T. Kelley
skelley@sdsu.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Microbial Immunology,
a section of the journal
Frontiers in Microbiology

Received: 15 October 2020

Accepted: 23 March 2021

Published: 17 May 2021

Citation:

Sisk-Hackworth L, Ortiz-Velez A,
Reed MB and Kelley ST (2021)
Compositional Data Analysis
of Periodontal Disease Microbial
Communities.
Front. Microbiol. 12:617949.
doi: 10.3389/fmicb.2021.617949

Periodontal disease (PD) is a chronic, progressive polymicrobial disease that induces a strong host immune response. Culture-independent methods, such as next-generation sequencing (NGS) of bacteria 16S amplicon and shotgun metagenomic libraries, have greatly expanded our understanding of PD biodiversity, identified novel PD microbial associations, and shown that PD biodiversity increases with pocket depth. NGS studies have also found PD communities to be highly host-specific in terms of both biodiversity and the response of microbial communities to periodontal treatment. As with most microbiome work, the majority of PD microbiome studies use standard data normalization procedures that do not account for the compositional nature of NGS microbiome data. Here, we apply recently developed compositional data analysis (CoDA) approaches and software tools to reanalyze multiomics (16S, metagenomics, and metabolomics) data generated from previously published periodontal disease studies. CoDA methods, such as centered log-ratio (clr) transformation, compensate for the compositional nature of these data, which can not only remove spurious correlations but also allows for the identification of novel associations between microbial features and disease conditions. We validated many of the studies' original findings, but also identified new features associated with periodontal disease, including the genera *Schwartzia* and *Aerococcus* and the cytokine C-reactive protein (CRP). Furthermore, our network analysis revealed a lower connectivity among taxa in deeper periodontal pockets, potentially indicative of a more "random" microbiome. Our findings illustrate the utility of CoDA techniques in multiomics compositional data analysis of the oral microbiome.

Keywords: periodontal disease, CLR, compositional data analysis, microbiome, oral microbiome, C-reactive protein

INTRODUCTION

Periodontal disease (PD) manifests as bacterial biofilms (plaque) that lead to gum inflammation, recession, and, in later stages, degradation of the bone and tooth loss. Despite the prevalence of the disease, which affects over 45% of United States adults, the precise role of the oral microbiome in the progression of PD remains elusive (Eke et al., 2012). Prior to the development of next-generation sequencing (NGS) technologies, a cluster of three species deemed the "red complex," consisting of *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*, was found to be associated with the PD clinical factors gum pocket depth and bleeding (Socransky et al., 1998).

While some individual species of the oral microbiome contributing to PD, such as the members of the red complex, have been studied extensively (Lamont and Jenkinson, 1998), the presence of these specific species is not enough to explain the occurrence of PD (Ximénez-Fyvie et al., 2000). NGS technologies have revealed greater diversity of the oral microbiome and a complex relationship between microbiome composition and periodontal disease states, including an association between increasing microbial diversity and pocket depth (Kroes et al., 1999; Paster et al., 2001; Faveri et al., 2008; Griffen et al., 2012). Analysis of periodontal disease metagenomes has also revealed a novel bacterium strongly associated with the red complex and periodontal disease (Torres et al., 2019). Furthermore, high inter-patient diversity of the oral microbiome complicates deciphering the relationship between periodontal treatments or changes in disease state on the associated microbiome (Kumar et al., 2006; Schwarzberg et al., 2014; Califf et al., 2017).

While NGS technologies illuminate a great deal of information about the oral microbiome, most microbiome analyses ignore the compositional structure of NGS microbiome data, which presents problems in statistical and biological interpretation. Microbiome data are compositional for two main reasons. First, sequencing only captures a proportion of the microbes in a sample, so the counts of taxa in each sample are relative rather than absolute. As the measurement of one taxon increases, the measurement of another taxon must decrease regardless of whether its absolute abundance is actually lower. Second, as the count total obtained in a run of NGS sequencing is capped by sequencing depth limitations, each sample size is different, rendering the counts of taxa between samples incomparable. Common normalization methods for microbiome data, such as rarefaction, transcripts per million, and library size normalizations, attempt to make samples with different library sizes comparable, but generate proportional data still constrained by its relative nature (Gloor et al., 2017). Metabolomics, another data type that is commonly used in conjunction with microbiome analysis, is also relative in nature and therefore compositional. Furthermore, the integration of multiomics data, or different “omics” datasets like proteomics, metabolomics, and metagenomics, from the same sample is challenging due to the different scales with which these data are measured.

Most statistical tests assume that the sample data exist in real space, where Euclidian geometry and distance formulas can be used to describe the distance between points. However, compositional data exist in a space known as the simplex where dimensions are arbitrary and values are subject to spurious correlations (Aitchison, 1982; Gloor et al., 2017). Compositional data analysis (CoDA) approaches have been developed to deal with these constraints of compositional data. One method gaining traction is the centered log-ratio (clr) transformation, which recasts relative count data with respect to the sample's geometric mean and creates scale-invariant data in Euclidian space where the use of multivariate statistical methods is valid (Gloor et al., 2017; Quinn et al., 2018). We recently showed that analyzing clr-transformed compositional datasets can reveal novel relationships, allow better discrimination between variables, and facilitate the integration of multiomics

16S, internal transcribed spacer (ITS), and metabolomic datasets (Sisk-Hackworth and Kelley, 2020).

In this work, we applied CoDA approaches, namely, clr transformation prior to standard methods such as non-metric multidimensional scaling (NMDS) ordination, Spearman's correlation, multiomics structure correlation, beta dispersion, random forest, and network analysis, as well as the log-ratio balance method used in the R package selbal (Rivera-Pinto et al., 2018), to 16S, metagenomic, cytokine, and metabolomic datasets from prior studies of patients with periodontal disease before and after treatments. By reanalyzing these data with a CoDA approach, we integrated these multiomics datasets to reveal patterns and correlations between the disease state, microbes, metabolites, and cytokines, in addition to the relationships between community structure and disease state not identified with standard normalization methods.

MATERIALS AND METHODS

Study Descriptions

This study incorporated data from two separate studies. The standard periodontal treatment (PT) study consisted of patients with periodontal disease and investigated the biofilms of periodontal pockets through 16S sequences, metagenomic sequences, and serum cytokine levels before and after standard periodontal treatments (Schwarzberg et al., 2014; Delange et al., 2018; Vijay Kumar et al., 2018). A total of 21 males and 38 females with an average age of 29 years were recruited from an American Indian/Alaska Native population in Southern California for the PT study. Eight patients had mild periodontitis (pocket depth less than 3 mm), 40 had moderate periodontitis (3–6 mm), and 11 had severe periodontitis (pocket depth over 6 mm). The second study measured pocket metabolites, 16S sequences, metagenomic sequences, and the serum cytokine levels of patients before and after treatments with 0.25% sodium hypochlorite (SHT) (Califf et al., 2017). For this study, 19 males and 15 females with an average age of 41 years were recruited among patients of the Ostrow School of Dentistry at the University of Southern California. In the SHT study, periodontal pocket depths ranged from 3 to 12 mm, while pocket depths in the PT ranged from 1.3 to 3.8 mm. The disease classes for the SHT study were separated into class “A” (pocket depth up to 6 mm), class “B” (pocket depth between 6 and 8 mm), and class “C” (pocket depth over 8 mm). Further details on the patient populations can be found in Schwarzberg et al. (2014) and Delange et al. (2018) for the PT study and in Califf et al. (2017) for the SHT study.

PT Study Data

The original PT data contained 76 samples of 247 16S operational taxonomic units (OTUs), 144 samples of six cytokine inflammatory markers, and 23 samples of 3,830 bacterial metagenomic OTUs. The 16S ribosomal RNA (rRNA) sequences and the mapping file from in this study are accessible at: <http://dx.doi.org/10.6084/m9.figshare.855613> and <http://dx.doi.org/10.6084/m9.figshare.855612>.

The serum cytokine data, raw reads from the 16S rRNA sequences, and metagenomic OTUs, classified by Kraken, were published previously (Delange et al., 2018; Vijay Kumar et al., 2018; Torres et al., 2019). Details on the study population, sampling, disease classification, and cytokine identification can be found in previously published papers (Schwarzberg et al., 2014; Delange et al., 2018; Vijay Kumar et al., 2018; Torres et al., 2019).

SHT Study Data

The SHT study contained 286 samples of 773 16S OTUs, 215 samples of 914 tandem mass spectrometry (MS/MS) features, and 24 samples of 3,770 bacterial metagenomic features. The 16S rRNA sequences used in this study were accessed through the European Nucleotide Archive under project PRJEB19122 (Califf et al., 2017). Metabolite data from tandem mass spectrometry were downloaded from the online MassIVE repository of the GNPS database under MassIVE ID number MSV000078894. Metagenomic sequence libraries, generated from 24 subgingival samples from the SHT study patients and classified *via* Kraken, were obtained from Dr. Pedro Torres (Torres et al., 2019).

16S Sequence Analysis

16S sequencing data were analyzed using QIIME 2020.2 (Bolyen et al., 2019). Sequences were clustered into 100% identity using v-search OTU clustering (Rognes et al., 2016). Taxonomy was assigned to sequences using the RDP Classifier (Wang et al., 2007) retrained on Greengenes 13_5 (McDonald et al., 2012) *via* QIIME 2.

Data Reduction and Transformation

Due to computational constraints, the numbers of features in the original sequencing and metabolomic datasets were reduced for *selbal* analysis (see below). The same reduced datasets were then used for the rest of the analyses. Genera of the 16S bacterial taxa present in greater than 10% of the samples, the 181 most abundant metagenomic taxa counts, and the 65 most abundant metabolites were selected for correlation analysis. For both PT and SHT studies, samples with a NA value for pocket depth in the mapping file were removed from all analyses. For the PT data, only samples with an overall response of improved or worsened were kept for all the analyses, determined by whether pocket depth decreased or increased, respectively (Schwarzberg et al., 2014). For the SHT data, only subgingival samples with disease class “A” or “C” were used on all analyses, as class “B” contained too few samples. Disease status was classified by maximum pocket depth (“A” = up to 6 mm, “B” = 6–8 mm, and “C” = over 8 mm) (Califf et al., 2017). For both the PT and SHT studies, OTUs were summed by genus for each sample in each of the 16S and metagenomic datasets. Zero replacement was performed with the pseudo-counts method from the R package zCompositions (Palarea-Albaladejo and Martín-Fernández, 2015) version 1.3.3. clr transformation was performed separately on all datasets (not on combined “multiomics” datasets). The clr transformation was computed for each sample j : each feature in that sample was divided by the geometric mean of all the feature counts

in the sample, then the natural log of that ratio was taken (Aitchison, 1982).

$$\text{clr}(X_j) = \left[\ln\left(\frac{X_{1j}}{g(X_j)}\right), \dots, \ln\left(\frac{X_{Dj}}{g(X_j)}\right) \right]$$

where X_j is the list of features in a sample, $g(X_j)$ is the geometric mean of the features in sample X_j , X_{1j} is the first feature in a sample, and X_{Dj} is the last feature in a sample of D values. To guide the reader, we have provided a diagram of the various datasets and analyses used in this study (Figure 1).

NMDS Ordination Plots

clr-transformed values were used to generate the NMDS ordination plots with the R package *vegan* version 2.5.6 using Euclidean distances (Oksanen et al., 2019). The NMDS plots were created in R using *ggplot2* version 3.2.1c (Wickham, 2016) and the samples were colored and shaped by periodontal treatment, overall response, and disease class. Permutational multivariate analysis of variance (PERMANOVA) was performed for each condition (periodontal treatment, overall response, and disease class) in every dataset (16S bacteria, cytokine, metabolites, and bacterial metagenomics) using the R package *vegan* with 9,999 permutations, with the p values corrected for multiple comparisons using the Benjamini–Hochberg method. The multivariate PERMANOVA test determines whether the centroid of a sample set is equal among the specified categories (e.g., periodontal treatment and disease class). The centroid was estimated using the between-sample Euclidean distances.

Beta Dispersion

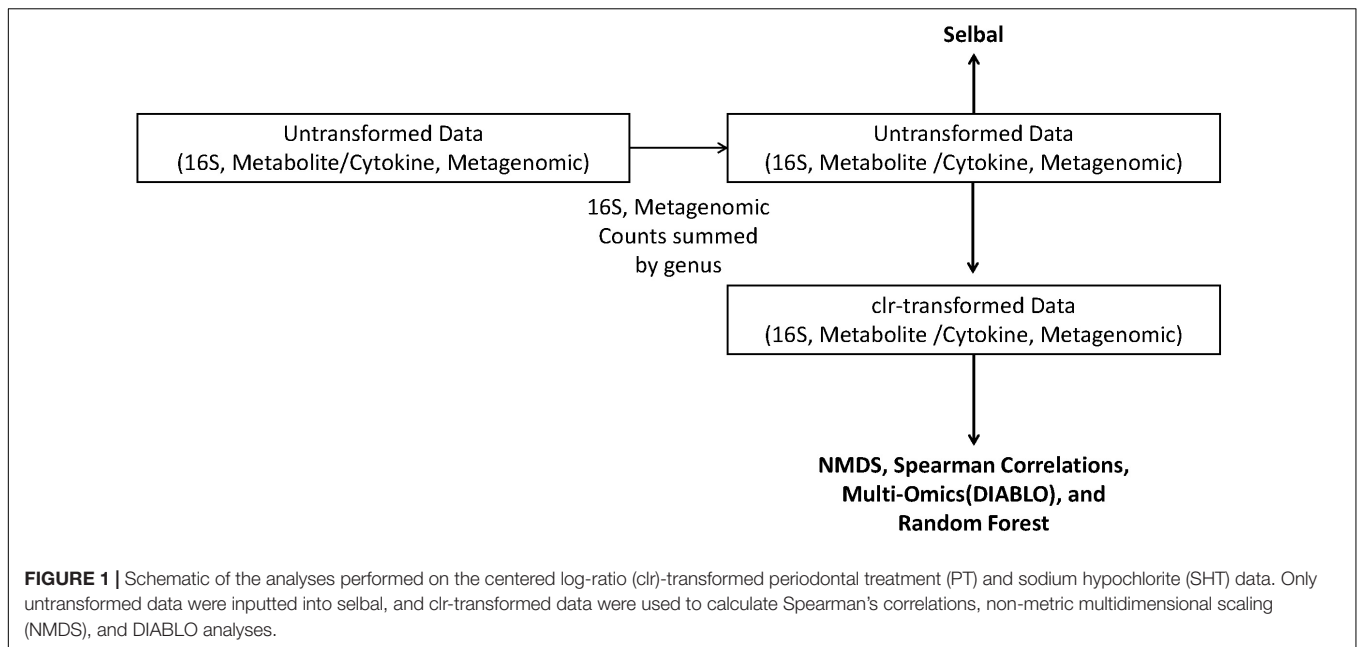
Beta dispersion, which measures the distance of each sample in a category from the centroid of that category, was estimated with the R package *vegan* using between-sample Euclidean distances. The beta dispersion test is a multivariate test used to determine whether the dispersion of samples is equivalent among categories. The p values were adjusted using the Benjamini–Hochberg method.

Spearman's Correlations

Spearman's correlations were computed using the R package *psych* v1.0.67 (Revelle, 2020). For each periodontal treatment, overall response, and disease class, we computed the correlations between genera from four different combined multiomics datasets: (1) 16S bacteria and cytokine (PT); (2) 16S bacteria, cytokine, and bacterial metagenomics (PT); (3) bacteria and metabolite (SHT); and (4) bacteria, metabolite, and bacterial metagenomics (SHT). The p values were adjusted with the Bonferroni correction.

Multiomics Integration

We integrated the same datasets as in the Spearman's correlations using the DIABLO framework, a method for multiomics classification and integration, in the *mixOmics* R package version 6.10.8c (Rohart et al., 2017). We assessed the correlation structure at the component level for each of the three conditions on



their respective dataset: periodontal treatment, overall response, and disease class.

Microbial Balances

We identified differentially abundant taxa, metabolites, and cytokines using the R package selbal version 0.1, a compositional data analysis method that detects microbial signatures between different sample types by identifying the smallest number of differentially abundant taxa that is predictive of sample condition. Raw measurements of cytokines, metabolites, 16S, and metagenomic taxa summed by genus were inputted to selbal, as it performs zero handling and transformation within the package. Although selbal was designed with microbial balances in mind, the method is valid for finding balances of other data types, such as metabolite and cytokine data. Furthermore, selbal only finds balances for dichotomous and continuous response variables, so we performed this analysis only for the variables periodontal treatment (dichotomous), disease class (dichotomous), and pocket depth (continuous).

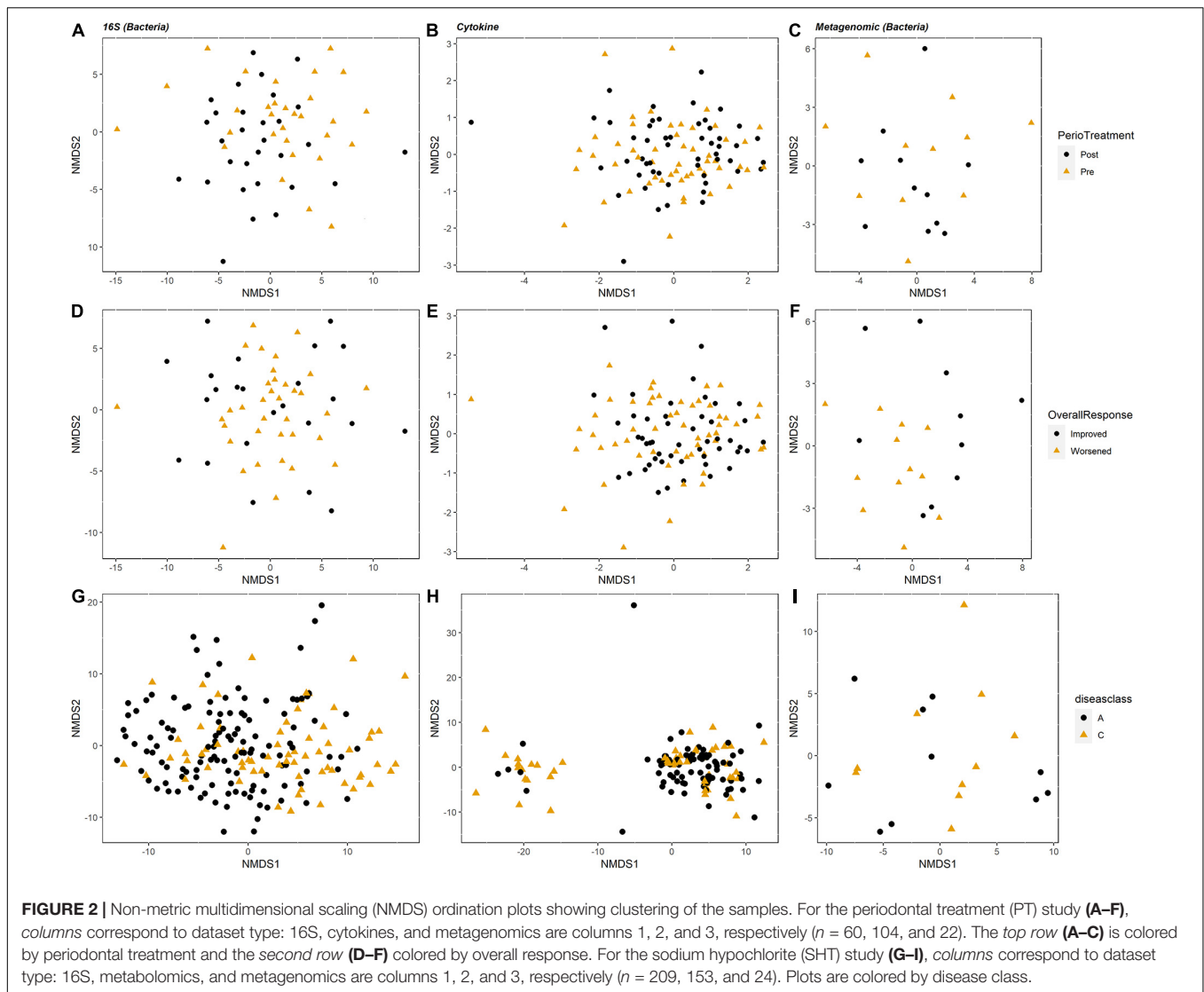
Random Forest

A random forest classifier was implemented in Python using the scikit-learn package (Pedregosa et al., 2012) to identify cytokines, metabolites, and 16S genera that discriminate between pocket depth (PT) and disease class (SHT). The metrics used to analyze the random forest classifier include accuracy, out-of-bag (OOB) score, mean accuracy, and area under the curve of the receiver of components (AUC-ROC). For the PT data, the pocket depth boundary used to distinguish high and low pocket depths was 2.6 mm. For the SHT study, only disease classes “A” and “C” were used as disease class “B” had few samples.

Network Analysis

Using the R package psych, we calculated Pearson's correlations for each of the following datasets: 16S OTU, combined 16S OTU–cytokine for the PT study, and 16S for the SHT study. Correlations with a magnitude of $|0.55|$ or greater were kept; all other values were changed to zero. Using the psych package, p values were calculated for each pairwise Pearson's correlation. The correlation matrix and the p value matrix were then filtered to contain only significant correlations (those with a Bonferroni-corrected p value below 0.05). The resulting adjacency matrix was transformed into an igraph object using a function from the SpiecEasi library (Kurtz et al., 2015). Using igraph v.1.2.5 package (Csardi and Nepusz, 2006), a network was constructed from the adjacency matrix using the OTUs as nodes and the Pearson's correlation values as edge weights. Networks were constructed with nodes scaled according to the eigen centrality.

For each network, we calculated the number of nodes, edges, as well as the diameter and transitivity. Nodes represent individual genera or cytokines and edges are lines representing relationships between genera or cytokines. The diameter of a network is the shortest distance between the furthest apart nodes in a network. Transitivity, ranging from 0 to 1, measures the average connectedness of a network, with higher values signifying that a high proportion of nodes are connected to surrounding nodes, which indicates the presence of tightly connected clusters of nodes. To identify taxa that occupy important structures of the network, the R package igraph was used to calculate the eigenvector centrality (eigen centrality) and betweenness centrality. Eigen centrality identifies which highly connected nodes are connected to other highly connected nodes; these highly connected nodes therefore form most of the architecture that orders the network. Betweenness centrality represents the frequency that a node is traversed



when the shortest paths in a network are calculated; high betweenness centrality indicates nodes that facilitate correlations between other nodes.

RESULTS

Beta Diversity

We used NMDS ordination to determine the clustering of samples by condition (periodontal treatment, overall response, disease class, and pocket depth) for microbes, cytokines, and metabolites for the PT and SHT study datasets. For the 16S, cytokine, and metagenomic datasets from the PT study, we did not observe clustering of samples by periodontal treatment or overall response (Figure 2) or pocket depth (Supplementary Figure 1). The most distinct separation was seen in the metabolites for disease class in the SHT study, where most of the samples in disease class “A” clustered together and the samples in class “C” split into two groups (Figure 2H); a

similar pattern was observed in the metabolites for pocket depth (Supplementary Figure 2). PERMANOVA indicated a difference between periodontal treatment groups for the 16S and cytokine datasets in the PT study (p –adj = 0.0234 for both comparisons; Table 1) and for the 16S and metabolite datasets in the SHT study (p –adj = 0.0012 and 0.006, respectively; Table 1). Analysis of beta dispersion showed no differences between the periodontal treatment or overall response groups in the PT study or by disease class in the SHT class (Supplementary Figure 3).

Spearman's Correlations

clr transformation reduces spurious correlations in compositional data, such as microbiome and metabolome data, and allows the application of statistical methods such as Spearman's correlation (Quinn and Erb, 2019). We applied Spearman's correlation to analyze the relationships between the multiomics datasets from both the PT and SHT studies. Most correlations between the combined multiomics datasets

TABLE 1 | PERMANOVA results (9,999 permutations) for centered log-ratio (clr)-transformed periodontal treatment (PT) 16S, cytokine, and metagenomic datasets and sodium hypochlorite (SHT) 16S, metabolite, and metagenomic datasets.

Dataset	Variable	R^2	p	$p - \text{adj}^a$
PT study				
16S	Periodontal treatment	0.0343	0.0053	0.0234
	Overall response	0.0129	0.7848	0.7848
Cytokine	Periodontal treatment	0.0343	0.0078	0.0234
	Overall response	0.0175	0.1152	0.1683
Metagenomic	Periodontal treatment	0.0687	0.1403	0.1683
	Overall response	0.0972	0.0293	0.0586
SHT study				
16S	Disease class	0.032	0.0004	0.001
Metabolite	Disease class	0.028	0.0040	0.006
Metagenomic	Disease class	0.044	0.5482	0.548

^aBenjamini–Hochberg corrected for multiple comparisons.

were within the same datasets (e.g., bacteria to bacteria), while few between-omics correlations (e.g., bacteria to cytokines) were detected. For the PT datasets, no significant ($p < 0.05$) bacteria–cytokine correlations were observed, except in the posttreatment samples that had worsened. In these samples, *Prevotella* was strongly correlated ($R^2 = 0.808$) with interleukin (IL)-1 (Supplementary Figure 4).

In the SHT study, we observed many significant correlations among bacteria and metabolites when the datasets for samples of disease classes “A” and “C” were combined; *Acinetobacter*, *Rubrivivax*, and *Treponema* were positively correlated with six metabolites, while *Desulfovibrio*, *Paludibacter*, *Peptococcus*, *TG5*, and *Treponema* were negatively correlated with six different metabolites (Supplementary Figure 5A). In samples that were only disease class “A,” *Olsenella* and *Atopobium* were positively correlated with two metabolites, while *Treponema* was negatively correlated with one metabolite (Supplementary Figure 5B). No bacterial–metabolite correlations were observed in samples that were only disease class “C.”

Multiomics Integration

Using DIABLO, we found that the correlation structure between the 16S and cytokine datasets in the PT study was better when the overall response (improved or worsened) variable was included than when the time (pre vs. post) variable was incorporated (Figures 3A,C). When metagenomic data were included in the multiomics correlation (excluding samples that did not get the metagenome sequenced), the correlation structure did not change dramatically, but there was greater discrimination of samples by overall response (Figures 3B,D). For the SHT study, the metabolite and 16S combined datasets strongly distinguished between disease class, but the correlation structure was low (Figure 3E). When metagenomic data were included in the multiomics correlation (excluding samples that did not get the metagenome sequenced), the correlation structure greatly increased, but this may be an artifact of the severely reduced sample size (Figure 3F).

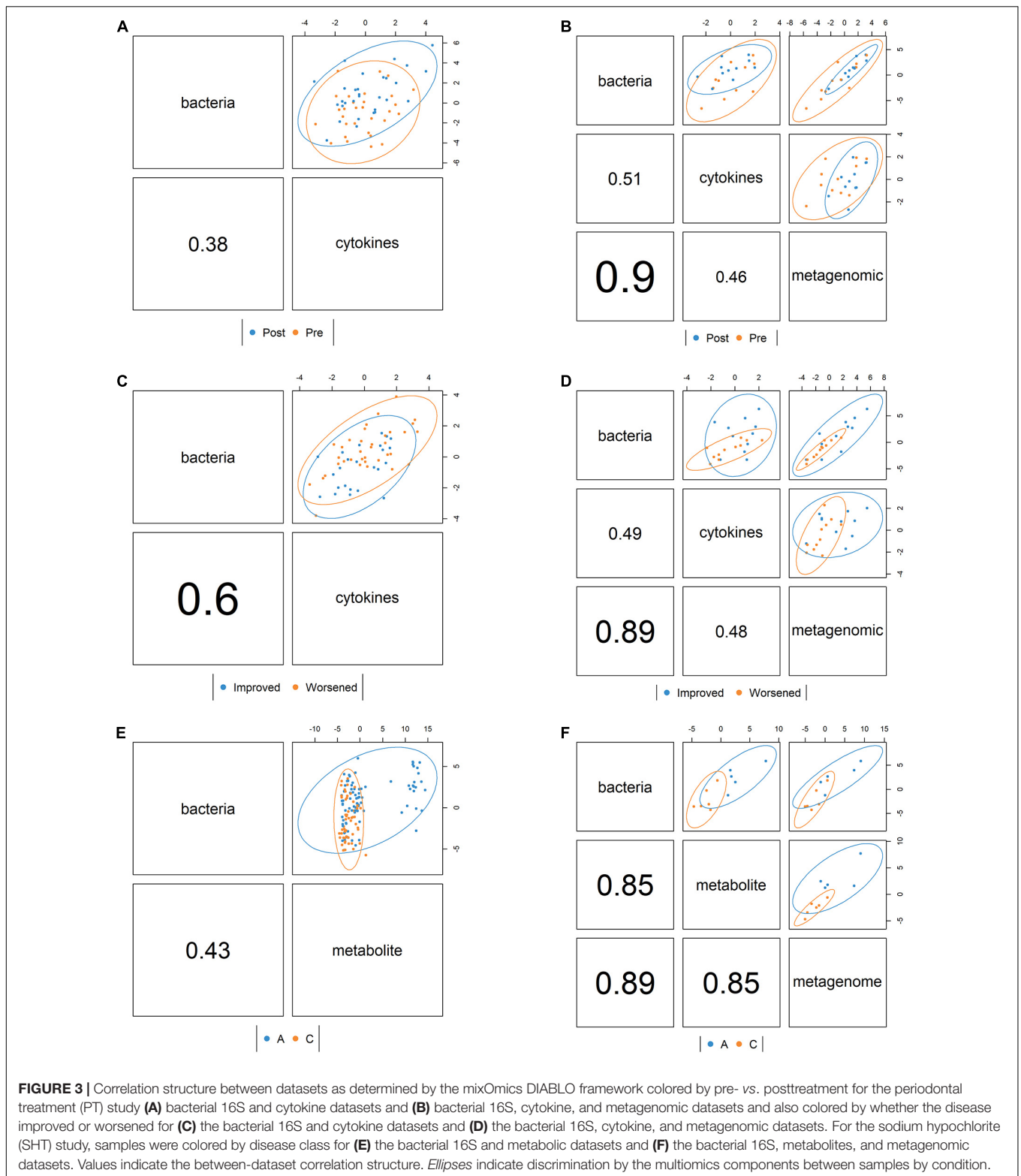
We generated Circos plots showing the correlations between the “omics” datasets using the DIABLO correlation structure. In the PT study, IL-6/IL-10 were strongly negatively correlated with *Treponema* and *Schwartzia* (Figure 4A). In the SHT study, *TG5* and *Treponema* were strongly correlated with many unidentified metabolites (Figure 4B).

Microbial Balances

Using selbal, we analyzed the differentially abundant genera, cytokines, and metabolites between different conditions. In the PT study, selbal identified many bacterial genera predictive of pretreatment samples, *Desulfovulbus*, *Bulleidia*, and *Hylemonella*, while *Abiotrophia* and *Haemophilus* were more predictive of posttreatment samples (Table 2 and Supplementary Figure 6A). For cytokines, selbal identified IL-6 as the most predictive of pretreatment samples and IL-6/IL-10 as the most predictive of posttreatment samples (Table 2 and Supplementary Figure 5B). From the metagenomic dataset, selbal identified that *Prevotella* predicted pretreatment and *Haemophilus* predicted posttreatment samples (Table 2 and Supplementary Figure 5C). In distinguishing overall response among the datasets, selbal identified *Desulfovulbus*, IL-6, and *Burkholderia* as more predictive in samples that improved and *Peptococcus*, C-reactive protein (CRP), and *Corynebacterium* as more predictive in samples that worsened (Table 2 and Supplementary Figures 6D–F). We then analyzed pre- and posttreatment balances for samples that improved and samples that worsened. In pretreatment samples that ended up improving, *Campylobacter*, *Bulleidia*, *Treponema*, IL-10, and *Actinomyces* were more relatively abundant, while in improved samples posttreatment *Peptococcus*, IL-11(*2), and *Rothia* were relatively abundant (Table 2 and Supplementary Figures 7A–C). In pretreatment samples that later worsened, *TG5*, CRP, and *Veillonella* were more relatively abundant, while in posttreatment samples that had worsened, *Fusobacterium*, IL-6/IL-10, and *Haemophilus* were more relatively abundant (Table 2 and Supplementary Figures 7D–F).

We also used selbal to explore which features’ relative abundance changed with the sum of all pocket depths. Within the improved samples, selbal identified *Selenomonas* and IL-10 as more associated with shallow pocket depths and *Filifactor*, IL-11(*2), and *Porphyromonas* as more associated with deeper pocket depths (Table 2 and Supplementary Figures 8A–C). Within the worsened samples, *Filifactor*, CRP, and *Veillonella* were associated with shallow pocket depths, while *Paludibacter*, IL-6/IL-10, and *Bacillus* were associated with deeper pocket depths (Table 2 and Supplementary Figures 8D–F).

We also used selbal to determine the metabolite and microbial balances in the SHT study 16S and metagenomic datasets for disease class and pocket depth. Both 16S and metagenomic microbial balances identified the genus *Tanerella* as more predictive for disease class “A” and *Fusobacterium* as more predictive of disease class “C” (Table 3 and Supplementary Figures 9A,C). Metabolite balances were clearly identified, but the specific metabolites in the balance remain unknown (Table 3 and Supplementary Figures 8B,E,H). In disease class “A” samples,



selbal identified *Desulfobulbus* and *Rothia* as more predictive of shallower pocket depths and *SHD-231* and *Fusobacterium* as more predictive of deeper pocket depth (Table 3 and Supplementary Figures 9D,F). For the samples in disease class

“C,” shallow pocket depth was more associated with *Lactobacillus* and *Parabacteroides*, while deeper pocket depth was more associated with *Desulfovibrio* and *Bacteroides* (Table 3 and Supplementary Figures 9G,I).

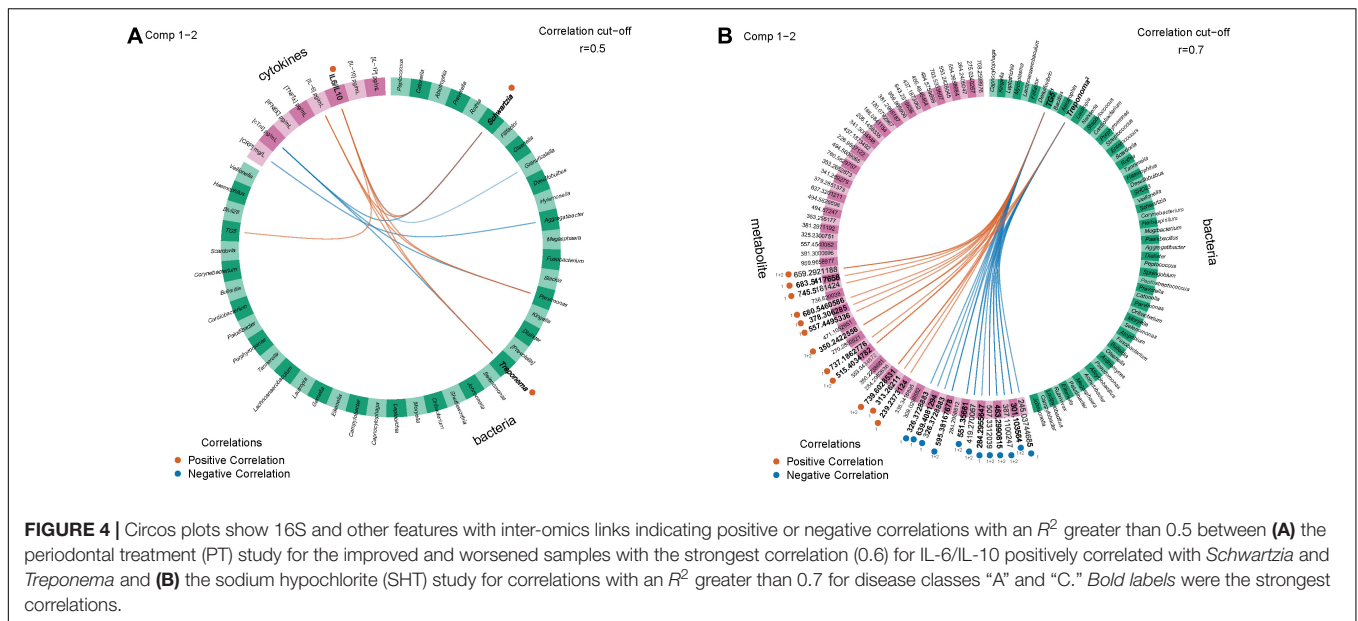


TABLE 2 | Summary of the microbial and metabolic balances generated from the periodontal treatment (PT) datasets.

Dataset	Variable	Denominator	Numerator	AUC-ROC
16S	Periodontal treatment	<i>Desulfohalobium</i> , <i>Bulleidia</i> , <i>Hylemonella</i>	<i>Abiotrophia</i> , <i>Haemophilus</i>	0.85
	Overall response	<i>Desulfohalobium</i>	<i>Peptococcus</i>	0.76
	Periodontal treatment for improved	<i>Campylobacter</i> , <i>Bulleidia</i> , <i>Treponema</i>	<i>Haemophilus</i>	0.83
	Periodontal treatment for worsened	<i>TG5</i>	<i>Fusobacterium</i>	0.76
	Pocket depth for improved	<i>Selenomonas</i>	<i>Filifactor</i>	0.73
	Pocket depth for worsened	<i>Filifactor</i>	<i>Paludibacter</i>	0.23
Cytokine	Periodontal treatment	IL-6	IL-6/IL-10	0.60
	Overall response	IL-6	CRP	0.66
	Periodontal treatment for improved	IL-10	IL-11(2)	0.66
	Periodontal treatment for worsened	CRP	IL-6/IL-10	0.58
	Pocket depth for improved	IL-10	IL-11(2)	0.42
	Pocket depth for worsened	CRP	IL-6/IL-10	0.06
Metagenomic	Periodontal treatment	<i>Prevotella</i>	<i>Haemophilus</i>	0.86
	Overall response	<i>Burkholderia</i>	<i>Corynebacterium</i>	0.90
	Periodontal treatment for improved	<i>Actinomyces</i>	<i>Rothia</i>	1.00
	Periodontal treatment for worsened	<i>Veillonella</i>	<i>Haemophilus</i>	1.00
	Pocket depth for improved	<i>Selenomonas</i>	<i>Porphyromonas</i>	0.90
	Pocket depth for worsened	<i>Veillonella</i>	<i>Bacillus</i>	0.74

AUC-ROC, area under the curve of the receiver of components.

Random Forest

The random forest machine learning classifier was trained to determine how accurately pocket depth class and disease class could be predicted from 16S bacteria OTUs and cytokines for the PT study or from 16S bacteria and metabolites for the SHT study. The most important cytokines were IL-10 and CRP (**Figure 5B**). The 16S features in the SHT study most predictive of disease class changed dramatically when metabolite data were included in the random forest analysis; the only feature recognized as highly important in both classifiers was *Abiotrophia* (**Figures 5A–C**). The addition of metabolite data to the 16S data increased the AUC-ROC scores compared with the 16S by itself in the SHT

study (**Table 4**). However, the inclusion of the cytokine features did not improve the AUC-ROC scores in the PT study (**Table 4**).

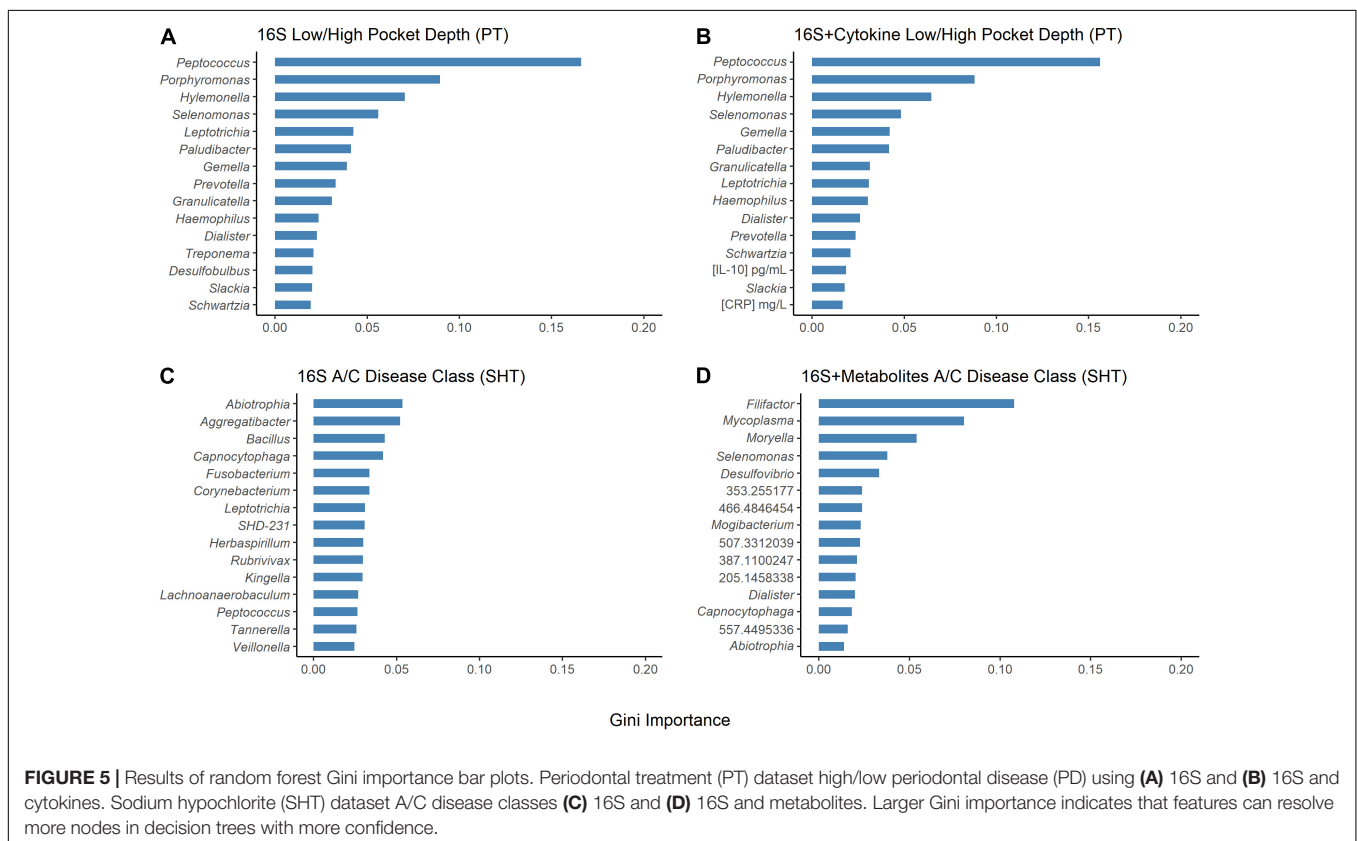
Network Analysis

In the samples with moderate disease conditions, the Pearson's correlation networks had approximately twice as many edges as those with gingivitis and a small diameter (**Table 5** and **Figures 6A,B**). Transitivity was not different between disease conditions, implying that both networks have similar levels of inter-nodal interactions. In the gingivitis network, seven taxa formed correlations, of which four had eigen centralities that were greater than 0.1: *Pelomonas*,

TABLE 3 | Summary of the microbial and metabolic balances generated from the sodium hypochlorite (SHT) datasets.

Dataset	Variable	Denominator	Numerator	AUC-ROC
16S	Disease class	<i>Capnocytophaga</i>	<i>Porphyromonas</i>	0.77
	Pocket depth for class A	<i>Desulfobulbus</i>	<i>SHD-231</i>	0.04
	Pocket depth for class C	<i>Lactobacillus</i>	<i>Desulfovibrio</i>	0.208
Metabolite	Disease class	301.1035645	637.32011211, 554.3658664	0.74
	Pocket depth for class A	419.270067	284.2958512	0.127
	Pocket depth for class C	245.0374468	494.5759899	0.264
Metagenomic	Disease class	<i>Tannerella</i>	<i>Fusobacterium</i>	0.88
	Pocket depth for class A	<i>Bacillus</i>	<i>Rothia, Fusobacterium</i>	0.892
	Pocket depth for class C	<i>Parabacteroides</i>	<i>Bacteroides</i>	0.848

AUC-ROC, area under the curve of the receiver of components.



Thermoanaerobacterium, *Aeribacillus*, and *Ralstonia* (Figure 6A and Supplementary Table 1). The degree distribution of the gingivitis network did not strictly follow the characteristic shape of a power law distribution, but it did reveal high amounts of “low connectivity” and low amounts of “high connectivity,” while the moderate network had a degree distribution that followed the power law trend much more closely (Supplementary Figures 10A,B).

Networks were also constructed for samples that had either shallow or deep pocket depths. Networks from the samples with shallow pockets had approximately three times as many edges as those with deep pockets (Table 5 and Figures 6C,D). The networks from the samples with shallower pockets had greater transitivity than those with deep pockets. No cytokines

were highly connected (high eigen centralities) within any of the networks we constructed (Supplementary Tables 1–4), suggesting that cytokines were not strongly correlated with many taxa. The degree distribution for these networks followed the typical shape of a power law distribution (Supplementary Figures 10C,D).

For the SHT study, networks were constructed with 16S data for the samples from class A and class C disease status. The network structure characteristics were similar between the two networks describing the microbiomes of patients diagnosed with different disease classes (Table 4 and Figures 6E,F). The taxa with the highest eigen centralities in disease class “A” were *Streptobacillus*, *Aerococcus*, and *Arthrobacter*, while *Aerococcus*, *Massilia*, and *Gemella* had the highest eigen centralities in

TABLE 4 | Results of random forest analysis with single and combined multiomics.

Comparisons	Accuracy	OOB score	Mean accuracy	AUC-ROC
PT study				
High vs. low (16S)	0.63	0.65	0.59	0.59
High vs. low (16S + cytokine)	0.63	0.57	0.59	0.59
SHT study				
A vs. C (16S)	0.72	0.69	0.73	0.6
A vs. C (16S + metabolite)	0.76	0.71	0.74	0.7

OOB score, out-of-bag score; AUC-ROC, area under the curve of the receiver of components.

TABLE 5 | Summary statistics of the networks generated from the periodontal treatment (PT) datasets comparing 16S and cytokines based on disease severity and pocket depth and the sodium hypochlorite (SHT) 16S datasets comparing oral microbiomes with different disease classes.

Data	Condition	Vertices	Edges	Diameter	Transitivity
16S	Gingivitis	48	32	10	0.49
	Moderate	48	69	6	0.45
16S + cytokine	Gingivitis	56	42	12	0.43
	Moderate	56	78	7	0.42
16S	Shallow	48	62	7	0.55
	Deep	48	22	6	0.20
16S + cytokine	Shallow	56	70	8	0.52
	Deep	56	26	6	0.18
16S	Class A	86	130	8	0.50
	Class C	86	124	8	0.46

class “C” (Supplementary Tables 6,7). The degree distribution for these networks followed the typical shape of a power law distribution (Supplementary Figures 10E,F).

DISCUSSION

Our reanalysis, and expanded analysis, of previously published data from periodontal disease studies using CoDA techniques improved our ability to detect patterns and correlations in these data and provided new insights into the relationships of organisms, cytokines, and metabolites to the disease process. Analysis of beta-diversity using the clr-transformed datasets detected distinct clustering not observed in the original studies. In the SHT study, unlike the original PCoA analysis that saw no separation of samples by disease class for any “omics” dataset (Califf et al., 2017), we identified clustering of clr-transformed data by both disease class and pocket depth in the NMDS ordination plots, with disease class C and deeper pocket depth in the left cluster and disease class A and shallower pocket depth in the right cluster (Figure 2H and Supplementary Figure 2B). In the PT study, we did not see clustering of the samples by 16S, cytokine, or metagenomic datasets for any of the classifications (periodontal treatment, overall response, or pocket depth) in the NMDS ordination (Figure 2), which concurs with

the findings from the original study (Schwarzberg et al., 2014). However, PERMANOVA showed differences between the pre- and posttreatment samples and disease class in the PT and SHT studies’ 16S data, findings not determined in the original studies due to the high level of individual variability (Tables 1–3).

The original PT study reported significant relationships between the combined abundance of *P. gingivalis*, *Fusobacterium nucleatum*, *T. forsythia*, and *T. denticola*, which are periodontal pathogens, and IL-1 β (Vijay Kumar et al., 2018). However, the only significant correlation we observed in the PT study dataset was between IL-1 and *Prevotella*. IL-1 has been associated with periodontal disease severity (Offenbacher et al., 2007), and *Prevotella* includes species in the orange complex associated with periodontal disease (Socransky et al., 1998). For the SHT study, we observed significant bacterial-metabolite correlations among the diseased samples, which the original study did not investigate (Supplementary Figures 5A,B). Among the samples in all disease classes, *Paludibacter*, a bacterial genus associated with plaque in healthy patients (Chen et al., 2018), was negatively correlated with two metabolites (Supplementary Figure 5A). However, selbal identified this genus as more predictive of worsened samples with high pocket depth (Table 3 and Supplementary Figure 8D), so this genus may not be “health-associated” as previously thought. Other genera negatively correlated with metabolites in the samples of all disease classes were *Desulfovibrio*, *Peptococcus*, and *TG5* (Supplementary Figure 4A). *Desulfovibrio* species may stimulate the immune response (Dzierżewicz et al., 2010) and have been observed in periodontal pockets (Loubinoux et al., 2002). *Peptococcus* and *TG5* have been found in the oral microbiomes of periodontitis patients (Kumar et al., 2005; Apatzidou et al., 2017). *Acinetobacter*, which has been previously associated with periodontitis samples (Souto et al., 2014), and *Rubrivivax* were positively correlated with metabolites (Supplementary Figure 5A). As *Rubrivivax* is not commonly associated with periodontal disease, this genus may be a new route of study. *Treponema*, a “red complex” species, was correlated with most metabolites in the samples from all disease classes (Supplementary Figure 5A) and was correlated with one metabolite among the samples that were only within disease class “A” (Supplementary Figure 4B). Also significantly correlated with metabolites in disease class “A” samples were *Olsenella* and *Atopobium* (Supplementary Figure 5B). In periodontitis patients, *Olsenella* species have been detected in abundance (Chávez de Paz et al., 2004) and *Atopobium* species have been associated with periodontal disease (Kumar et al., 2005; Apatzidou et al., 2017).

The DIABLO multiomics integration uncovered an additional negative relationship between IL-6/IL-10 and *Treponema* and *Schwartzia* in the PT study and many correlations between *Treponema* and *TG5* with many metabolites in the SHT study (Figures 3A,B). *Schwartzia*, *Treponema*, and *TG5* have been associated with the biofilms of periodontal disease patients (Socransky et al., 1998; Camelo-Castillo et al., 2015; Apatzidou et al., 2017). Additionally, *T. denticola* has been shown to degrade IL-1 β and IL-6 (Miyamoto et al., 2006), and infection

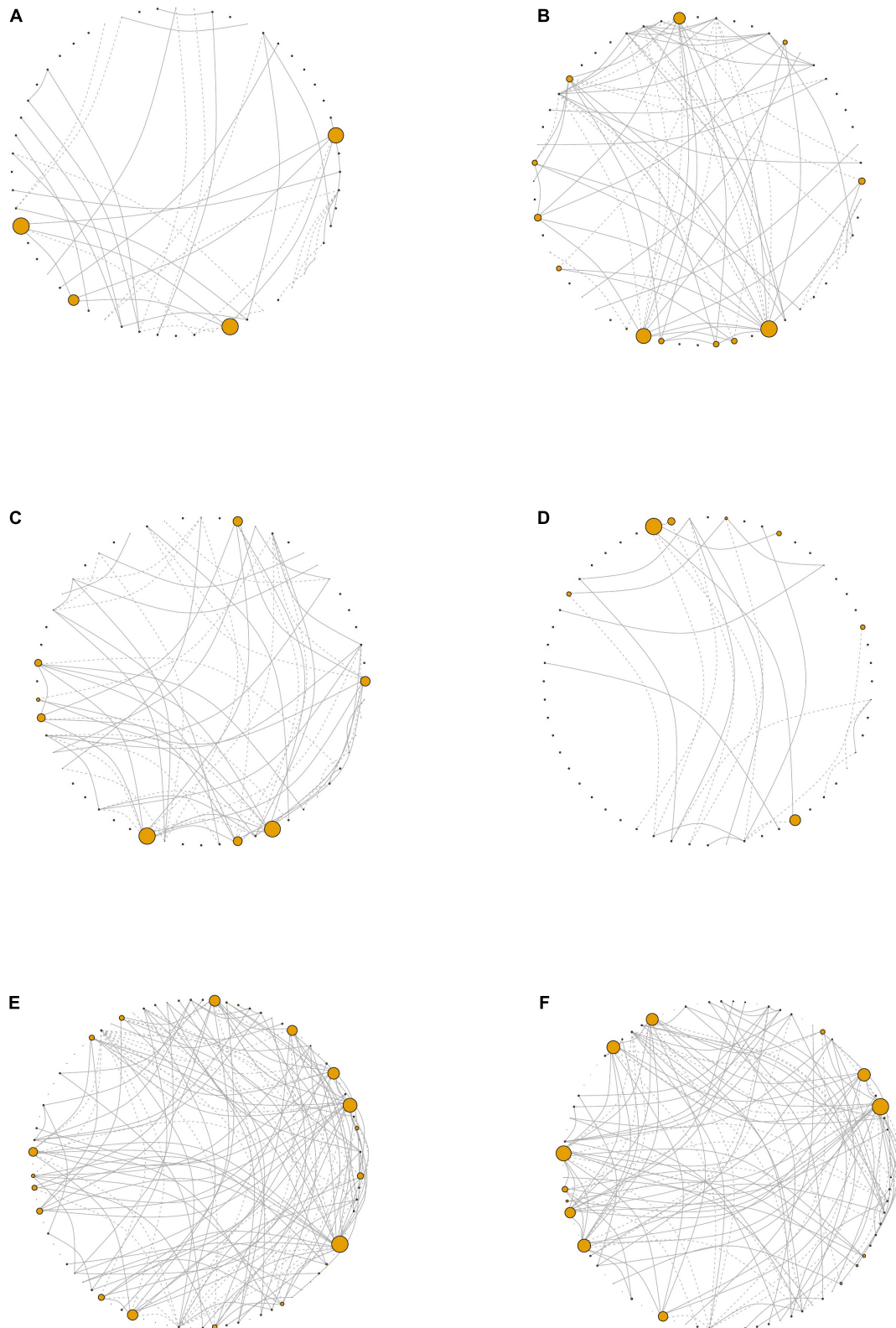


FIGURE 6 | Network visualizations. Network plots derived from the periodontal treatment (PT) study patients whose oral status was designated as **(A)** gingivitis, **(B)** moderate, **(C)** shallow, or **(D)** deep. For the sodium hypochlorite (SHT) study, network plots for biofilms from patients with **(E)** class “A” or **(F)** class “C.” Node size corresponds to eigen centrality. *Dashed lines* represent negative correlations and *solid lines* represent positive correlations.

of both *P. gingivalis* and *T. denticola* synergistically stimulated the production of IL-6 by macrophage-like cells (Tamai et al., 2009). Little is known about the association between *Schwartzia*, periodontal disease, and the immune response, so this genus is a potential new target for investigation. For the SHT study multiomics integration, disease class was able to distinguish integrated 16S and metabolomic samples (Figure 3E); disease class was also the only variable that distinguished between samples in our NMDS analysis, indicating that this discrimination may be largely due to metabolite differences.

The orange complex described by Socransky et al. (1998) as highly intercorrelated and associated with PD includes species from the genera *Prevotella* and *Fusobacterium*, which were both identified in our selbal analysis of microbial balances for the PT and SHT datasets. *Prevotella* relative abundance predicted pretreatment samples in the PT dataset and *Fusobacterium* relative abundance predicted high pocket depth and posttreatment samples that had worsened (Table 2 and Supplementary Figures 6C, 7D). *Fusobacterium* was also predictive of disease class “C” in the SHT study (Table 3 and Supplementary Figure 9C). This aligns well with the original PT studies, where *Fusobacterium* was significantly correlated with pocket depth and a decrease in *Prevotella* after treatment was associated with improvement (Schwarzberg et al., 2014; Califf et al., 2017). Some *Desulfobulbus* species likely play a role in the development of periodontal disease (Camelo-Castillo et al., 2015; Cross et al., 2018), and selbal identified this genus as predictive of pretreatment samples, samples that improved, and shallow pocket depth (Table 2 and Supplementary Figures 6A,D, 9D). Selbal indicated that *Porphyromonas*, the genus that includes a red complex species, was more predictive of deeper pockets in the PT study and disease class “C” in the SHT study (Tables 2 and 3 and Supplementary Figures 8C, 9A) and was found in the original study as correlating with high pocket depth (Califf et al., 2017). The most commonly identified cytokine by selbal, IL-6, was found in the original study to be significantly associated with severe periodontitis (Delange et al., 2018), while we found that IL-6 levels were predictive of pretreatment samples and samples that improved (Table 2 and Supplementary Figures 6B,E). Additionally, selbal identified CRP as predictive of samples that worsened (Table 2 and Supplementary Figure 6E), which is in contrast to the previous study which did not find a strong association between CRP and periodontal disease status (Delange et al., 2018). CRP was also identified by random forest as one of the two top important cytokines for predicting pocket depth (Figure 5B). Random forest also identified *Abiotrophia*, species of which have been isolated from dental plaques (Mikkelsen et al., 2000), as the most stable genus in predicting disease class in the SHT study 16S data, while in the PT study data, *Peptococcus* and *Porphyromonas* were the most important genera in predicting pocket depth (Figure 5).

Analysis of correlation networks can provide insights into the complexity, stability, and function of a microbial community (Barberán et al., 2012). The most striking disparity in overall network connectivity occurred in the PT study. The network

analysis found that, for the PT study, networks of 16S and cytokines for the samples with deep pockets had fewer edges and lower transitivity (Table 5). Fewer inter-nodal connections and a lower overall network connectedness indicate a lack of interdependence of taxa in deeper pockets. Multiple studies have shown that pocket depth is correlated with greater alpha diversity and more pathogenic taxa (Christersson et al., 1992; Stoltenberg et al., 1993; Takeshita et al., 2016). The early stages of periodontal development involve well-known interactions between bacterial species (Nyvad and Kilian, 1987; Diaz et al., 2006; Chalmers et al., 2008), but as biofilms develop and become increasingly anaerobic, more pathogenic species establish within the biofilm (Van Winkelhoff et al., 2002; Vieira Colombo et al., 2016). The fewer connections that we observed in deep pockets could reflect a more random, or less stable, biofilm in the later stages of disease. Additionally, the networks constructed from the samples with shallow pockets had greater transitivity (Table 5), which implies the presence of more inter-nodal interactions within the shallow pocket networks and may be indicative of niche filtering, where similarities rather than differences between species allow the persistence of species in an environment (Röttjers and Faust, 2018). Network analysis also revealed that *Aerococcus* had a high eigen centrality value in the networks for disease classes “A” and “C” (Supplementary Tables 5,6). Higher eigen centralities indicate that nodes are critical for network stability and may point toward keystone species (Bauer et al., 2010; Mandakovic et al., 2018). While *Aerococcus* species have been found in the biofilms of periodontitis patients (Voropaeva et al., 2008), but as little is known about the association between *Aerococcus* and periodontal disease, this may be an interesting future avenue of study.

We should note potential effects of the study population demographics on our data. Most participants from the PT study had an overweight or obese body mass index, and 37% were smokers, both of which could affect the microbiome composition and inflammation levels measured in this study. Additionally, the PT study participants were American Indian/Alaskan Native, on average over a decade younger, and had a higher prevalence of females (66 vs. 44%) than the SHT study participants, so the differences in the results between the two studies could be due to effects of ethnicity, aging, or sex on the oral microbiome and inflammation. Our use of CoDA techniques, which confirmed many of the prior studies’ results and uncovered new findings, shows how this approach is a valuable addition to the current methods of microbiome data analysis for investigating oral disease. We have shown how CoDA approaches are especially useful when integrating multiomics due to the scale-invariance that the clr transformation confers on datasets. The identification of CRP as predictive of pocket depth and samples that worsened is a new finding and an important area of further study. We also identified understudied genera potentially important in periodontal disease (*Schwartzia*, *Rubrivivax*, and *Aerococcus*). Furthermore, the ability of unknown metabolites to discriminate between samples in selbal analyses, and the associations we determined between metabolites and particular taxa, highlights the need to study these compounds in the context of periodontal disease.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study and code for all analyses can be found on zenodo: <https://doi.org/10.5281/zenodo.4604009> further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by USC Health Sciences Campus Institutional Review Board, the SDSU Institutional Review Board (IRB) and the IRB of the Southern California American Indian Health Clinic. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AO-V and MR performed the computations and generated figures with guidance from LS-H and SK. LS-H wrote the initial draft of the manuscript. SK, AO-V, and MR edited the manuscript. All authors discussed the results and contributed to the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.617949/full#supplementary-material>

Supplementary Figure 1 | NMDS ordination plots showing clustering of PT samples by pocket depth. Columns correspond to dataset type; 16S, Cytokine, and Metagenomic datasets are columns one, two, and three, respectively ($n = 60, 104, \text{ and } 22$).

Supplementary Figure 2 | NMDS ordination plots showing clustering of SHT samples by pocket depth. Columns correspond to dataset type; 16S, Metabolites, and Metagenomic datasets are columns one, two, and three, respectively ($n = 209, 153, \text{ and } 24$).

Supplementary Figure 3 | Beta dispersion analysis by periodontal treatment, response to treatment, and disease class. For the PT study, distance from the centroid of beta dispersion is shown for samples pre- and post-treatment for (A) 16S bacterial, (B) metabolite, and (C) metagenomic datasets and for response to

treatment for (D) 16S bacterial, (E) metabolite, and (F) metagenomic datasets. Larger distances indicate samples which are far from the group centroid, small distances indicate samples which are close to the group centroid. For the SHT study, distance from the centroid is shown for samples by disease class for (G) 16S bacterial, (H) metabolite, and (I) metagenomic datasets.

Supplementary Figure 4 | Bacterial-cytokine correlation matrix for post-treatment, worsened samples. All spaces with color are $p\text{-adj} < 0.05$. Blue indicates a positive correlation; Dark orange indicates a negative correlation.

Supplementary Figure 5 | Correlation matrices between comparisons of different SHT datasets (bacterial 16S-metabolite and bacterial 16S-metagenomic). (A) Bacterial-metabolite correlations on a combined dataset from all disease classes; (B) Bacterial-metabolite correlations on disease class "A." All spaces with color are $p\text{-adj} < 0.05$. Blue indicates a positive correlation, Dark orange indicates a negative correlation.

Supplementary Figure 6 | Microbial and cytokine balances for PT dataset at different conditions computed with selbal, where numerator genera are more relatively abundant than denominator genera for higher balance values. Balances are shown for samples before and after periodontal treatment for (A) bacterial 16S data, (B) cytokine, and (C) metagenomic data. Balances are shown for improved and worsened samples of (D) bacteria 16S data, (E) cytokine, and (F) metagenomic data.

Supplementary Figure 7 | Microbial and cytokine balances for PT dataset for improved (A–C) and worsened (D–F) samples computed with selbal. Balances before and after treatment are shown for samples that improved for (A) bacterial 16S data (B) cytokine, and (C) metagenomic data. Balances before and after treatment are shown for samples that worsened for (D) bacterial 16S data, (E) cytokine, and (F) metagenomic data.

Supplementary Figure 8 | Selbal-computed microbial and cytokine balances for PT dataset associated with the sum of all pocket depths. Balances for improved samples by pocket depth (response variable) are shown for (A) 16S bacterial, (B) cytokine, and (C) metagenomic datasets. Balances for worsened samples by pocket depth are shown for (D) 16S bacterial, (E) cytokine, and (F) metagenomic datasets.

Supplementary Figure 9 | Selbal-computed balances for SHT dataset for microbial and metabolite balances by different characteristics. For samples with disease class either "A" or "C," balances shown are for (A) 16S bacterial, (B) metabolite, and (C) metagenomic data. Balances for samples with disease class "A" associated with pocket depth sum are shown for (D) 16S bacterial, (E) metabolite, and (F) metagenomic datasets. Balances for disease class "C" samples associated with pocket depth sum are shown for (G) 16S bacterial, (H) metabolite, and (I) metagenomic datasets.

Supplementary Figure 10 | Network degree distributions. For the PT study for networks with OTUs and cytokines, degree distribution of inferred network from patients who had (A) gingivitis or (B) moderate disease, (C) shallow or (D) deep pockets. For the SHT study, degree distribution of inferred network from patients who had class (E) "A" or (F) "C" disease.

REFERENCES

- Aitchison, J. (1982). The statistical analysis of compositional data. *J. R. Stat. Soc. B* 44, 139–160.
- Apatzidou, D., Lappin, D. F., Hamilton, G., Papadopoulos, C. A., Konstantinidis, A., and Riggio, M. P. (2017). Microbiome associated with peri-implantitis versus periodontal health in individuals with a history of periodontal disease. *J. Oral Microbiol.* 9:1325218. doi: 10.1080/20002297.2017.1325218
- Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119
- Bauer, B., Jordán, F., and Podani, J. (2010). Node centrality indices in food webs: rank orders versus distributions. *Ecol. Complexity* 7, 471–477. doi: 10.1016/j.ecocom.2009.11.006
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.
- Califf, K. J., Schwarzberg-Lipson, K., Garg, N., Gibbons, S. M., Caporaso, J. G., Slots, J., et al. (2017). Multi-omics analysis of periodontal pocket microbial communities pre- and posttreatment. *mSystems* 2:e00016-17. doi: 10.1128/mSystems.00016-17
- Camelo-Castillo, A. J., Mira, A., Pico, A., Nibali, L., Henderson, B., Donos, N., et al. (2015). Subgingival microbiota in health compared to periodontitis and the influence of smoking. *Front. Microbiol.* 6:119. doi: 10.3389/fmicb.2015.00119
- Chalmers, N. I., Palmer, R. J., Cisar, J. O., and Kolenbrander, P. E. (2008). Characterization of a *Streptococcus* sp. *Veillonella* sp. community micromanipulated from dental Plaque. *J. Bacteriol.* 190, 8145–8154. doi: 10.1128/jb.00983-08

- Chávez de Paz, L. E., Molander, A., and Dahlén, G. (2004). Gram-positive rods prevailing in teeth with apical periodontitis undergoing root canal treatment. *Int. Endodontic J.* 37, 579–587. doi: 10.1111/j.1365-2591.2004.00845.x
- Chen, C., Hemme, C., Beleno, J., Shi, Z. J., Ning, D., Qin, Y., et al. (2018). Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *ISME J.* 12, 1210–1224. doi: 10.1038/s41396-017-0037-1
- Christersson, L. A., Fransson, C. L., Dunford, R. G., and Zambon, J. J. (1992). Subgingival distribution of periodontal pathogenic microorganisms in adult periodontitis. *J. Periodontol.* 63, 418–425. doi: 10.1902/jop.1992.63.5.418
- Cross, K. L., Chirania, P., Xiong, W., Beall, C. J., Elkins, J. G., Giannone, R. J., et al. (2018). Insights into the evolution of host association through the isolation and characterization of a novel human periodontal pathobiont, *Desulfobulbus oralis*. *mBio* 9:e02061-17.
- Csardi, G., and Nepusz, T. (2006). *The iGraph Software Package for Complex Network Research: R package. InterJournal.*
- Delange, N., Lindsay, S., Lemus, H., Finlayson, T. L., Kelley, S. T., and Gottlieb, R. A. (2018). Periodontal disease and its connection to systemic biomarkers of cardiovascular disease in young American Indian/Alaskan natives. *J. Periodontol.* 89, 219–227. doi: 10.1002/jper.17-0319
- Diaz, P. I., Chalmers, N. I., Rickard, A. H., Kong, C., Milburn, C. L., Palmer, R. J. Jr., et al. (2006). Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Appl. Environ. Microbiol.* 72, 2837–2848. doi: 10.1128/aem.72.4.2837-2848.2006
- Dzierżewicz, Z., Szczerba, J., Lodowska, J., Wolny, D., Gruchlik, A., Orchel, A., et al. (2010). The role of *Desulfovibrio desulfuricans* lipopolysaccharides in modulation of periodontal inflammation through stimulation of human gingival fibroblasts. *Arch. Oral Biol.* 55, 515–522. doi: 10.1016/j.archoralbio.2010.05.001
- Eke, P. I., Dye, B. A., Wei, L., Thornton-Evans, G. O., and Genco, R. J. (2012). Prevalence of periodontitis in adults in the United States: 2009 and 2010. *J. Dent. Res.* 91, 914–920. doi: 10.1177/0022034512457373
- Faveri, M., Mayer, M. P. A., Feres, M., De Figueiredo, L. C., Dewhirst, F. E., and Paster, B. J. (2008). Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Oral Microbiol. Immunol.* 23, 112–118. doi: 10.1111/j.1399-302x.2007.00397.x
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8:2224. doi: 10.3389/fmicb.2017.02224
- Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., et al. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J.* 6, 1176–1185. doi: 10.1038/ismej.2011.191
- Kroes, I., Lepp, P. W., and Relman, D. A. (1999). Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14547–14552. doi: 10.1073/pnas.96.25.14547
- Kumar, P. S., Griffen, A. L., Moeschberger, M. L., and Leys, E. J. (2005). Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J. Clin. Microbiol.* 43, 3944–3955. doi: 10.1128/jcm.43.8.3944-3955.2005
- Kumar, P. S., Leys, E. J., Bryk, J. M., Martinez, F. J., Moeschberger, M. L., and Griffen, A. L. (2006). Changes in periodontal health status are associated with bacterial community shifts as assessed by quantitative 16S cloning and sequencing. *J. Clin. Microbiol.* 44, 3665–3673. doi: 10.1128/jcm.00317-06
- Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* 11:e1004226. doi: 10.1371/journal.pcbi.1004226
- Lamont, R. J., and Jenkinson, H. F. (1998). Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis*. *Microbiol. Mol. Biol. Rev.* 62, 1244–1263. doi: 10.1128/mmb.62.4.1244-1263.1998
- Loubinoux, J., Bisson-Boutelliez, C., Miller, N., and Le Faou, A. E. (2002). Isolation of the provisionally named *Desulfovibrio fairfieldensis* from human periodontal pockets. *Oral Microbiol. Immunol.* 17, 321–323. doi: 10.1034/j.1399-302x.2002.170510.x
- Mandakovic, D., Rojas, C., Maldonado, J., Latorre, M., Travisany, D., Delage, E., et al. (2018). Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Sci. Rep.* 8:5875.
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139
- Mikkelsen, L., Theilade, E., and Poulsen, K. (2000). Abiotrophia species in early dental plaque. *Oral Microbiol. Immunol.* 15, 263–268. doi: 10.1034/j.1399-302x.2000.150409.x
- Miyamoto, M., Ishihara, K., and Okuda, K. (2006). The *Treponema denticola* surface protease dentilisin degrades interleukin-1 β (IL-1 β), IL-6, and tumor necrosis factor alpha. *Infect. Immun.* 74, 2462–2467. doi: 10.1128/iai.74.4.2462-2467.2006
- Nyvad, B., and Kilian, M. (1987). Microbiology of the early colonization of human enamel and root surfaces in vivo. *Scand. J. Dent. Res.* 95, 369–380. doi: 10.1111/j.1600-0722.1987.tb01627.x
- Offenbacher, S., Barros, S. P., Singer, R. E., Moss, K., Williams, R. C., and Beck, J. D. (2007). Periodontal disease at the biofilm–gingival interface. *J. Periodontol.* 78, 1911–1925. doi: 10.1902/jop.2007.060465
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). *vegan: Community Ecology Package: R package version 2.5-6.*
- Palarea-Albaladejo, J., and Martín-Fernández, J. A. (2015). zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemomet. Intell. Lab. Syst.* 143, 85–96. doi: 10.1016/j.chemolab.2015.02.019
- Paster, B. J., Boches, S. K., Galvin, J. L., Ericson, R. E., Lau, C. N., Levanos, V. A., et al. (2001). Bacterial diversity in human subgingival plaque. *J. Bacteriol.* 183, 3770–3783. doi: 10.1128/jb.183.12.3770-3783.2001
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2012). Scikit-learn: machine learning in python. *arXiv [Preprint]* arXiv:1201.0490v4.
- Quinn, T. P., and Erb, I. (2019). Another look at microbe–metabolite interactions: how scale invariant correlations can outperform a neural network. *bioRxiv [Preprint]* doi: 10.1101/847475
- Quinn, T. P., Erb, I., Richardson, M. F., and Crowley, T. M. (2018). Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* 34, 2870–2878. doi: 10.1093/bioinformatics/bty175
- Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research: R package.* Evanston, IL: Northwestern University.
- Rivera-Pinto, J., Egozcue, J. J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., and Calle, M. L. (2018). Balances: a new perspective for microbiome analysis. *mSystems* 3:e00053-18.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584
- Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: an R package for ‘omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752
- Röttjers, L., and Faust, K. (2018). From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol. Rev.* 42, 761–780. doi: 10.1093/femsre/fuy030
- Schwarzberg, K., Le, R., Bharti, B., Lindsay, S., Casaburi, G., Salvatore, F., et al. (2014). The personal human oral microbiome obscures the effects of treatment on periodontal disease. *PLoS One* 9:e86708. doi: 10.1371/journal.pone.0086708
- Sisk-Hackworth, L., and Kelley, S. T. (2020). An application of compositional data analysis to multiomic time-series data. *Nucleic Acids Res. Genomics Bioinform.* 2:lqaa079.
- Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, R. L. Jr. (1998). Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* 25, 134–144. doi: 10.1111/j.1600-051x.1998.tb02419.x
- Souto, R., Silva-Boghossian, C. M., and Colombo, A. P. (2014). Prevalence of *Pseudomonas aeruginosa* and *Acinetobacter* spp. in subgingival biofilm and saliva of subjects with chronic periodontal infection. *Braz. J. Microbiol.* 45, 495–501. doi: 10.1590/s1517-83822014000200017
- Stoltenberg, J. L., Osborn, J. B., Pihlstrom, B. L., Herzberg, M. C., Aeppli, D. M., Wolff, L. F., et al. (1993). Association between cigarette smoking, bacterial

- pathogens, and periodontal status. *J. Periodontol.* 64, 1225–1230. doi: 10.1902/jop.1993.64.12.1225
- 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.
- Tamai, R., Deng, X., and Kiyoura, Y. (2009). *Porphyromonas gingivalis* with either *Tannerella forsythia* or *Treponema denticola* induces synergistic IL-6 production by murine macrophage-like J774.1 cells. *Anaerobe* 15, 87–90. doi: 10.1016/j.anaerobe.2008.12.004
- Torres, P. J., Thompson, J., Mclean, J. S., Kelley, S. T., and Edlund, A. (2019). Discovery of a novel periodontal disease-associated bacterium. *Microb. Ecol.* 77, 267–276. doi: 10.1007/s00248-018-1200-6
- Van Winkelhoff, A. J., Loos, B. G., Van Der Reijden, W. A., and Van Der Velden, U. (2002). *Porphyromonas gingivalis*, *Bacteroides forsythus* and other putative periodontal pathogens in subjects with and without periodontal destruction. *J. Clin. Periodontol.* 29, 1023–1028. doi: 10.1034/j.1600-051x.2002.291107.x
- Vieira Colombo, A. P., Magalhães, C. B., Hartenbach, F. A. R. R., Martins Do Souto, R., and Maciel Da Silva-Boghossian, C. (2016). Periodontal-disease-associated biofilm: a reservoir for pathogens of medical importance. *Microb. Pathog.* 94, 27–34. doi: 10.1016/j.micpath.2015.09.009
- Vijay Kumar, P. K., Gottlieb, R. A., Lindsay, S., Delange, N., Penn, T. E., Calac, D., et al. (2018). Metagenomic analysis uncovers strong relationship between periodontal pathogens and vascular dysfunction in American Indian population. *bioRxiv* [Preprint] doi: 10.1101/250324
- Voropaeva, E. A., Bairakova, A. L., Bichucher, A. M., D'iakov, V. L., and Kozlov, L. V. (2008). [Protease activity of microflora in the oral cavity of patients with periodontitis]. *Biomeditsinskaia khimiia* 54, 706–711.
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/aem.00062-07
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag New York.
- Ximénez-Fyvie, L. A., Haffajee, A. D., and Socransky, S. S. (2000). Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *J. Clin. Periodontol.* 27, 648–657. doi: 10.1034/j.1600-051x.2000.027009648.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Sisk-Hackworth, Ortiz-Velez, Reed and Kelley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.