Sociodemographic patterning in the oral microbiome of a diverse sample of New Yorkers

Audrey Renson*[1], Heidi E. Jones,[1] Francesco Beghini,[2] Nicola Segata,[2] Christine Zolnik,[3,4] Mykhaylo Usyk,[3] Lorna Thorpe,[5] Robert Burk,[3] Levi Waldron,†[1] Jennifer B. Dowd†[1,6]

1 City University of New York (CUNY) School of Public Health, Department of Epidemiology and Biostatistics

2 Centre for Integrative Biology, University of Trento, Trento, Italy

3 Albert Einstein College of Medicine

4 Long Island University

5 NYU School of Medicine, Department of Population Health

6 King's College London, Department of Global Health and Social Medicine

* Corresponding author

† Equal contribution

KEYWORDS: Oral microbiome; sociodemographics; health disparities; race/ethnicity, socioeconomic status

ABSTRACT:

Background: Variations in the human oral microbiome are implicated in a range of conditions including diabetes and systemic inflammation, and may potentially mediate health inequalities in these conditions. Existing studies of the oral microbiome are based on samples with minimal sociodemographic diversity, limiting knowledge of underlying population variation and ultimately how differences in microbial profiles may be associated with health outcomes. This study describes sociodemographic patterning of the oral microbiome in a diverse sample of New York City residents.

Methods: Data come from a subsample of 296 respondents from the 2013-14 New York City Health and Nutrition Examination Study (NYC-HANES), a population-based representative study of New York City adults. Oral mouthwash samples were processed by amplicon sequencing targeting the V4 region of the 16S rRNA gene. We examined associations between the structure and diversity of the oral microbiome and sociodemographic variables including age, gender, income, education, nativity, and race/ethnicity.

Results: A total of 79 operational taxonomic units (OTUs) were differentially abundant by any sociodemographic variable (false discovery rate < 0.01), including 52 by age group, 27 by race/ethnicity, 14 by income, 14 by education, 12 by nativity, and 5 by gender. Twenty-one genera differed for more than one sociodemographic characteristic including *Lactobacillus* (7), *Prevotella* (6), *Porphyromonas* (5), and *Fusobacterium* (4). Education (p=0.03) and age (p=0.02) were associated with major shifts in community structure as measured by UniFrac distances.

Discussion: In a diverse population-based sample, we identified important sociodemographic variations in several taxa of the oral microbiome, with associations consistent with health inequalities. Further investigation is warranted into possible mediating effects of the oral microbiome in social disparities in diabetes, inflammation, oral health, and preterm birth.

Introduction

Variations in the human oral microbiome are implicated in a wide range of health outcomes, but large gaps remain in our understanding of the factors that shape variation at the population level. Oral diseases such as caries, gingivitis, and periodontitis are recognized to be favored by states of dysbiosis in the oral cavity,(1, 2) involving complex interactions between microbial communities and host immunity.(3) In addition, a growing body of evidence links differences in structure and function of the oral microbiota to systemic illness, including birth outcomes,(4, 5) orodigestive and other cancers,(6-9) diabetes,(10, 11) rheumatoid arthritis,(12, 13) atherosclerosis,(14) coronary heart disease,(15) and stroke. (16) While causal evidence is limited, hypothesized pathways for such associations include both direct virulence and modulation of systemic immune response.(6)

Disparities in health status by race/ethnicity, socioeconomic status (SES), gender, and other sociodemographic variables have long been observed but their mechanisms remain poorly understood. In particular, racial/ethnic and socioeconomic disparities have been consistently observed in oral health outcomes,(17) cardiovascular disease,(18, 19) diabetes,(20) preterm birth and low birth weight,(21, 22) and rheumatoid arthritis.(23).

A number of mechanisms potentially link social inequality to the microbiome. There is strong evidence that the so-

cial environment is associated with the distribution of a variety of microbial exposures.(24, 25) Changes in immune function as a result of psychosocial stress,(26) nutrition,(27) smoking, or other environmental exposures can alter susceptibility and host interactions with microbes among different sociodemographic groups. Differences in microbiome characteristics may also persist among sociodemographic groups via mother-to-child transmission (28, 29) and shared built environments.(30)

These findings have led researchers to call for investigation into the role of the microbiome in health disparities, (31) but limited research to date has examined sociodemographic associations with the oral microbiome. The Human Microbiome Project (HMP), a small volunteer sample in the U.S. with minimal sociodemographic or race/ethnic variability (approx. 80% white, n=60 racial minority members) sampled at 9 distinct oral sites,(32, 33) nonetheless found numerous differentially abundant taxa by ethnicity, comparing non-Hispanic white, non-Hispanic black, Asian, Mexican, and Puerto Rican ethnicities.(34) In another small volunteer sample, Mason et al. identified distinct subgingival microbiomes by race/ethnicity in the U.S., with Non-Hispanic blacks having lower microbiome diversity than other ethnic groups.(35) In a comparison of salivary microbiomes of Cheyenne and Arahapo vs. non-Native individuals in the U.S. found strong clustering by ethnicity, differences in species richness, and numerous differentially abundant taxa.(36) Several low-throughput studies examining specific periodontal pathogens found significant differences by race/ethnicity.(37-39) To our knowledge, only one published study has tested associations between SES and the microbiome, finding substantial differences (20% of variation) in the bacterial profiles of the oral microbiome by municipal-level SES in the Danish Health Examination Survey (DANHES).(40)

In order to explore the relationship between the oral microbiome and health disparities, the presence of sociodemographic patterning at the population level must be assessed. The aim of the present study was to assess sociodemographic variations in the human salivary microbiome. Specifically, we examined whether taxa were differentially abundant, and whether variation existed in alpha and beta diversity by sociodemographic categories, using high-throughput microbiome data collected from a population-based sample of New York City residents.

Methods

Data

Data come from the 2013-14 New York City Health and Nutrition Examination Survey (NYC HANES), whose design has been previously described.[40] Briefly, NYC HANES 2013-2014 was the second population-representative, cross-sectional survey of adult residents of New York City (NYC), administered by the City University of New York School of Public Health (CUNY SPH) and the NYC Department of Health and Mental Hygiene (DOHMH). NYC HANES used a three-stage cluster sampling design with primary sampling units (PSUs) based on counts of households from the 2010 U.S. Census. Overall response rate was 36% (n=1524). Eligible participants completed a two-part interview (face-to-face and audio computer assisted self-interview), physical examination, and blood, urine, and oral mouthwash biospecimen collection, with nearly all participants providing an oral mouthwash biospecimen (95%).

Subsample Selection

The current study uses data from a subsample of n=296 NYC HANES 2013-2014 participants selected for the purposes of examining oral microbiome associations with tobacco use. Subsample selection methods are reported elsewhere.{Begnini 2017} Briefly, eligible participants were selected for the sub-study if they consented to having their oral rinse samples stored for future research, and groups were sampled based on self-reported and/or serum cotinine-confirmed smoking status. Participants were either current hookah, cigar/cigarillo and/or e-cigarette users, regardless of cotinine or cigarette smoking (n=79), self-reported cigarette smokers with highest cotinine (n=90), former smokers with no secondhand smoke exposure and cotinine <0.05 ng/mL (n=45), former or never smokers with secondhand smoke exposure and cotinine >1 and <14 ng/mL (n=38), or never smokers with no secondhand smoke exposure and cotinine <0.05 ng/mL (n=45).

Oral rinse collection and microbiome sample processing

Oral rinse samples were collected by trained field interviewers, transported on dry ice and stored at -80°C at an NYC DOHMH laboratory, and periodically shipped to Albert Einstein College of Medicine for processing. Oral mouthwash samples were processed for microbiome analysis as follows. DNA was amplified for the V4 variable region of the 16s rRNA gene using PCR and quantified using a Qubit 2.0 Fluorometic High Sensitivity dsDNA Assay (Life Technologies), followed by high-throughput amplicon sequencing using the Illumina MiSeq.

Bacterial taxa examined in the analysis were determined by clustering the 16S rRNA sequences into Operational Taxonomical Units (OTUs). After clustering sequences based on 97% similarity, taxonomy was assigned at the genus or species level using the SILVA 123 (41) database as reference, and samples with less than 1000 reads were removed. Further information about laboratory, sequencing, and data pre-processing procedures is described elsewhere.{Beghini 2017}

Statistical Analysis

Taxa Abundance

We first examined relative abundance, or the proportion of total OTU counts accounted for by each taxa, at the phylum and genus levels. To assess whether taxa were differentially abundant by sociodemographic variables, we used edgeR,(42) a linear modelling technique developed for differential expression in RNA-Seq analysis. EdgeR estimates separate generalized linear models (GLMs) with each OTU as the response variable. The dispersion parameter for each model is estimated by sharing information about the dispersion in all OTUs using an empirical Bayes approach. Before performing differential abundance analysis, we filtered out

any OTUs if they did not meet the criteria of three or more samples having a count of at least eight, which left 216 OTUs for analysis.

We used edgeR to calculate crude negative binomial log-linear GLMs predicting each taxa separately by race/ethnicity, age, gender, highest level of education completed, income, marital status, nativity, gum disease, and self-reported mouthwash use frequency. We fit additional models for each sociodemographic variable separately adjusted for mouthwash, gum disease, smoking status, age and gender. OTUs were considered differentially abundant at false discovery rate (FDR) < 0.01.

Alpha & Beta Diversity

Alpha diversity is a measure of the ecological diversity within a given sample. We measured α-diversity using the Chao1 richness estimator, which is a prediction of the absolute number of OTUs present in a sample based on extrapolating from the number of singleton OTUs, a procedure which accounts for any differences in sampling depth.(43) We examined Chao1 alpha diversity by each sociodemographic variable and assessed group differences using Kruskal-Wallis tests. Beta diversity is a measure of difference in ecological communities between samples. To assess beta diversity, we conducted principal coordinates analysis (PCoA) using weighted UniFrac distances, a measure based on relative phylogenetic relatedness of organisms.(44) We explored PCoA plots colored by each sociodemographic variable and performed permutation multivariate analysis of variance (PERMANOVA) to assess group differences in beta diversity.

We performed clustering of samples with respect to OTUs using partitioning around medoids (PAM) on Bray Curtis, Jenson-Shannon, root-Jenson Shannon, weighted and unweighted UniFrac distances.(45) To assess the presence of discrete clusters, and if present, select the optimal number of clusters, prediction strength (PS) was calculated for k=2:10 PAM clusters on each of the distance measures, using PS≥0.9 to signify strong support for k clusters, as suggested by Koren et al.(45)

All statistical analyses were conducted in R version 3.4 (46) for Linux, using the packages 'edgeR' for differential abundance,(42) 'phyloseq' for alpha and beta diversity measures,(47) 'vegan' for PERMANOVA, (48) and 'fpc' for clustering.(49)

Results

Descriptive statistics for the oral microbiome subsample and for the full sample of the parent NYC HANES study are shown in Table 1. The sample had wide sociodemographic variation with respect to age (median [range]: 42 [20 to 94]), gender (52.4% female), race/ethnicity (34.5% non-Hispanic White, 26.4% non-Hispanic Black, 25.7% Hispanic), annual family income (43% less than $30K, 33.3% $60k or more), educational achievement (23.3% less than high school diploma, 30.4% college degree or greater), and birthplace (32.3% born in the U.S., Puerto Rico, or U.S. territories). Figure 1 displays absolute values of Cramer's V, a correlation coefficient for categorical variables, on all pairwise combinations of sociodemographic variables, indicating minor collinearity (all V<.35).

Relative Abundance and Alpha Diversity

Oral microbiomes were characterized at the phylum level by a gradient between *Firmicutes* and *Bacteroides* abundance, with overall dominance by *Firmicutes* (mean=52±10%). Streptococcus was the most abundant genus (36±10%) followed by *Prevotella* (17±8%) and *Rothia* (7±4%). Relative abundance at the phylum and genus levels are illustrated in Figure 2.

The mean±SD Chao1 for the overall sample was 462±118, and no differences were found by sociodemographic variables, indicating similar numbers of species across subgroups (p>0.05 for all comparisons, using Kruskal Wallis tests). Chao1 alpha diversity distributions by sociodemographics are presented in Figure 3.

Differential Abundance

To determine which taxa were differentially abundant across demographics, we used unadjusted log-linear models fit with edgeR against seven variables: race/ethnicity, U.S. vs. other birthplace, marital status, gender, family income tertiles, education, and age groups.

At FDR<0.01, we found that numerous taxa were differentially abundant by each of the seven sociodemographic variables of interest. Figure 4a displays effect size and direction for all demographics against all significant OTUs, using the log fold change (logFC), or coefficient from log linear models, for each comparison group and all significant OTUs. OTUs are indicated at the genus level, as few OTUs were specified at the species level. Where more than one OTU was significant within one genus, the maximum logFC is displayed.

A total of 75 OTUs were differentially abundant by any variable, including 52 by age group, 27 by race/ethnicity, 23 by family income, 20 by education, 17 by marital status, 12 by nativity, and five by gender. To contextualize effect sizes and number of significant taxa, we also looked at differential abundance by behavioral variables: mouthwash use showed 11 OTUs differentially abundant, self-reported gum disease showed five, smoking status showed 53, and sugar-sweetened beverage consumption showed 26. Many of these OTUs were associated with more than one sociodemographic variable. The most frequent genera to be differentially abundant were *Lactobacillus* (all variables), and *Prevotella* (Age, Education, Family income, Marital status, Race/ethnicity, US- vs. foreign-born). (See figure 4a for other genera.) Differential abundance findings for selected taxa are presented in table 2.

As numerous associations were present at FDR<0.01, we focus here on differential abundance findings where the logFC was 2 or greater, representing a particularly strong association. Compared to individuals aged 20-34, individuals aged 65 and over displayed greater abundance of *Lactobacillus* (logFC 2.8, FDR <0.0001), *Lactococcus* (logFC 2.4, FDR 0.0007), *Bifidobacterium* (logFC 2, FDR 0.0001), and Scardovia (logFC 2, FDR 0.0006). Compared to individuals with a

Table 1. Sample demographics and oral health behavioral characteristics of participants in the NYC HANES smoking and oral microbiome substudy.

	Oral Microbiome Subsample	**Full NYC HANES Sample**
Total	296	1527
Age in years – median [range]	42 [20 to 94]	42 [20 to 97]
Age group (%)		
20-29	73 (24.7)	360 (23.6)
30-39	64 (21.6)	337 (22.1)
40-49	53 (17.9)	252 (16.5)
50-59	52 (17.6)	264 (17.3)
60 and over	54 (18.2)	314 (20.6)
Gender = Female (%)	155 (52.4)	885 (58.0)
Educational achievement (%)		
College graduate or more	90 (30.4)	628 (41.1)
Less than High school diploma	69 (23.3)	316 (20.7)
High school graduate/GED	65 (22.0)	244 (16.0)
Some College or associate's degree	72 (24.3)	337 (22.1)
Missing	0 (0.0)	2 (0.1)
Annual family income (%)		
$60,000 or more	86 (29.1)	429 (28.1)
Less Than $30,000	111 (37.5)	537 (35.2)
$30,000 - $60,000	61 (20.6)	348 (22.8)
Missing	38 (12.8)	213 (13.9)
Marital Status (%)		
Married	100 (33.8)	590 (38.6)
Widowed	16 (5.4)	76 (5.0)
Divorced	24 (8.1)	156 (10.2)
Separated	12 (4.1)	51 (3.3)
Never married	106 (35.8)	511 (33.5)
Living with partner	38 (12.8)	143 (9.4)
Race/ethnicity (%)		
Non-Hispanic White	102 (34.5)	513 (33.6)
Non-Hispanic Black	78 (26.4)	340 (22.3)
Hispanic	76 (25.7)	390 (25.5)
Asian	23 (7.8)	204 (13.4)
Other	17 (5.7)	80 (5.2)
Place of birth (%)		
US, PR and Territories	95 (32.1)	668 (43.7)
Other	199 (67.2)	851 (55.7)
Missing	2 (0.7)	8 (0.5)
Gum disease (self-reported) (%)		
Yes	27 (9.1)	175 (11.5)
No	268 (90.5)	1322 (86.6)
Missing	1 (0.3)	30 (2.0)
Mouthwash use (times per week) (%)		
None	119 (40.2)	591 (38.7)
1 to 5	72 (24.3)	370 (24.2)
6 to 7	105 (35.5)	565 (37.0)
Missing	0 (0.0)	1 (0.1)
Sugar-sweetened beverages (per week) (%)		
0-<1	159 (53.7)	217 (14.2)
1-5	70 (23.6)	313 (20.5)
6 or more	66 (22.3)	227 (14.9)
Missing	1 (0.3)	770 (50.4)
Smoking status (%)		
Never smoker	109 (36.8)	928 (60.8)
Current smoker	119 (40.2)	280 (18.3)
Former smoker	68 (23.0)	316 (20.7)

Table 2. Differential abundance findings for OTUs selected based on clinical relevance. Greater or lower abundance indicates false discovery rate (FDR) <0.01.

Genus	Greater abundance in:	Lower abundance in:
Prevotella	Age group (3 cat) = 35-64 Education (3 cat) = High School Diploma or Less Income (3 cat) = Less Than $30,000 Income (3 cat) = $30,000 - $60,000 Marital Status = Living with partner Marital Status = Widowed Race/ethnicity = Non-Hispanic Black Race/ethnicity = Hispanic Race/ethnicity = Other U.S. vs. foreign-born = Other	Age group (3 cat) = 65 and over Age group (3 cat) = 35-64 Race/ethnicity = Non-Hispanic Black
Lactobacillus	Age group (3 cat) = 65 and over Age group (3 cat) = 35-64 Education (3 cat) = Some College or Associate's Degree Income (3 cat) = $30,000 - $60,000 Marital Status = Divorced Race/ethnicity = Non-Hispanic Black	Gender = Female Marital Status = Never married U.S. vs. foreign-born = Other
Streptococcus	Age group (3 cat) = 65 and over Age group (3 cat) = 35-64 Education (3 cat) = High School Diploma or Less Income (3 cat) = $30,000 - $60,000 Income (3 cat) = Less Than $30,000	Marital Status = Never married Marital Status = Living with partner
Porphyromonas	Age group (3 cat) = 35-64 Education (3 cat) = High School Diploma or Less Income (3 cat) = Less Than $30,000 Race/ethnicity = Hispanic Race/ethnicity = Non-Hispanic Black	Gender = Female Age group (3 cat) = 35-64 Age group (3 cat) = 65 and over
Fusobacterium	Education (3 cat) = High School Diploma or Less Income (3 cat) = Less Than $30,000 Marital Status = Never married	Age group (3 cat) = 35-64 Age group (3 cat) = 65 and over
Lactococcus	Age group (3 cat) = 65 and over Education (3 cat) = Some College or Associate's Degree Race/ethnicity = Non-Hispanic Black	U.S. vs. foreign-born = Other

high school degree or less, those with some college or an associate's degree showed greater abundance of *Lactobacillus* (logFC 2, FDR <0.0001). Individuals with annual family incomes between $30,000 and $60,000 had greater abundance of *Lactococcus* (logFC 2.6, FDR 0.0001), compared to those making less than $30,000.

Marital status showed a large number of strong associations: compared to being married, those living with a partner showed greater abundance of *Atopobium* (logFC 2.2, FDR 0.003), and *Prevotella* (logFC 2.1, FDR 0.003), those separated showed greater abundance of *Leptotrichia* (logFC 3.1, FDR 0.001), and those who were widowed, greater abundance of *Pseudomonas* (logFC 4.1, FDR <0.0001). Compared to non-Hispanic whites, Asians had greater abundance of *Treponema* (logFC 2.3, FDR 0.009), and non-Hispanic Blacks had greater abundance of *Lactococcus* (logFC 2.7, FDR <0.0001), *Atopobium* (logFC 2.2, FDR <0.0001), *Anaeroglobus* (logFC 2.1, FDR 0.0003), *Lactobacillus* (logFC 2, FDR <0.0001), and *Campylobacter* (logFC 2, FDR <0.0001).

To assess the extent to which sociodemographic associations were potentially confounded by smoking, mouthwash use, or age and gender, we fit additional edgeR models adjusted for each of these (age and gender in the same model). Figure 5 displays the distribution of logFCs for both crude and adjusted, including all OTUs with FDR <0.01 in crude models. The effect of adjustment appears to be minor in all cases, although adjustment for smoking exerts a noticeable effect on findings for age, family income and education.

We conducted oligotyping on the 16S samples, which is a method for elucidating diversity of closely related taxa that may be concealed by standard taxonomic assignment and clustering approaches to 16S OTU classification.(50) After filtering out low-abundance taxa, the analysis pipeline revealed nine oligotypes of Neisseria, 19 of Prevotella, and 14 of Streptococcus. Analyzing oligotypes revealed associations that were not apparent in the OTU analysis, whereas some associations present in OTU analysis were not apparent at the oligotype level (figure 4b). Specifically, *Prevotella* oligotypes revealed new associations for gender, *Streptococcus* showed associations for gender, race/ethnicity and US- vs. foreign-born, and *Neisseria*, for gender, age, education, marital status, race/ethnicity and US- vs. foreign-born. Associations that were present in OTU analysis but not in oligotyping were age, education and family income in *Prevotella*, and the only OTU association with *Neisseria*, family income.

Figure 1. Examining collinearity among sociodemographic variables. Data are absolute value of pairwise Cramer's *V* correlation coefficient between sociodemographic factor levels.

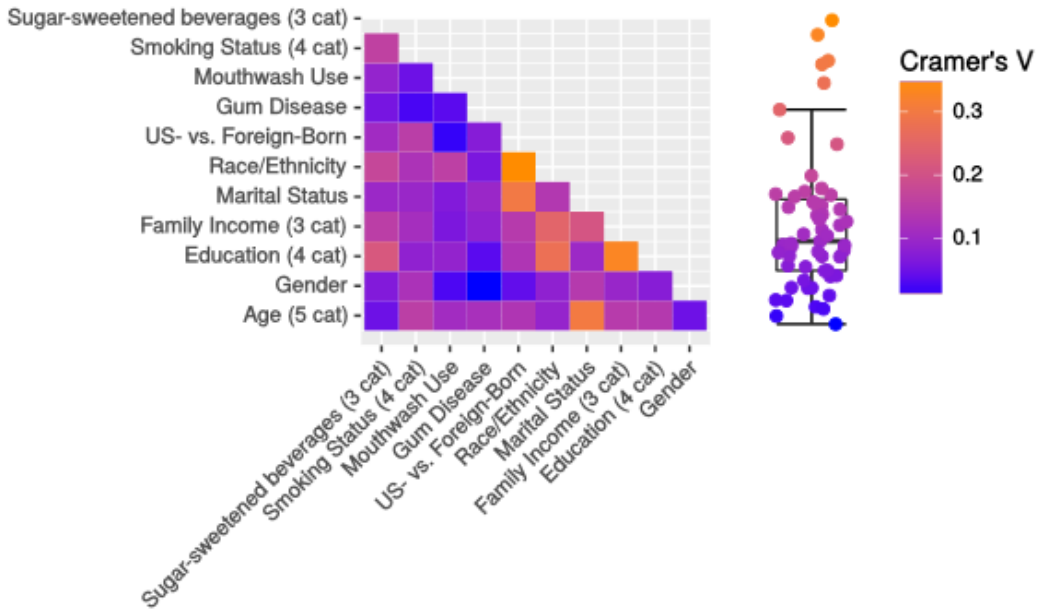


Figure 2. Genus- and phylum-level relative abundances. Data are percent of overall communities within samples, summarized as mean ± standard deviation of percent across samples.

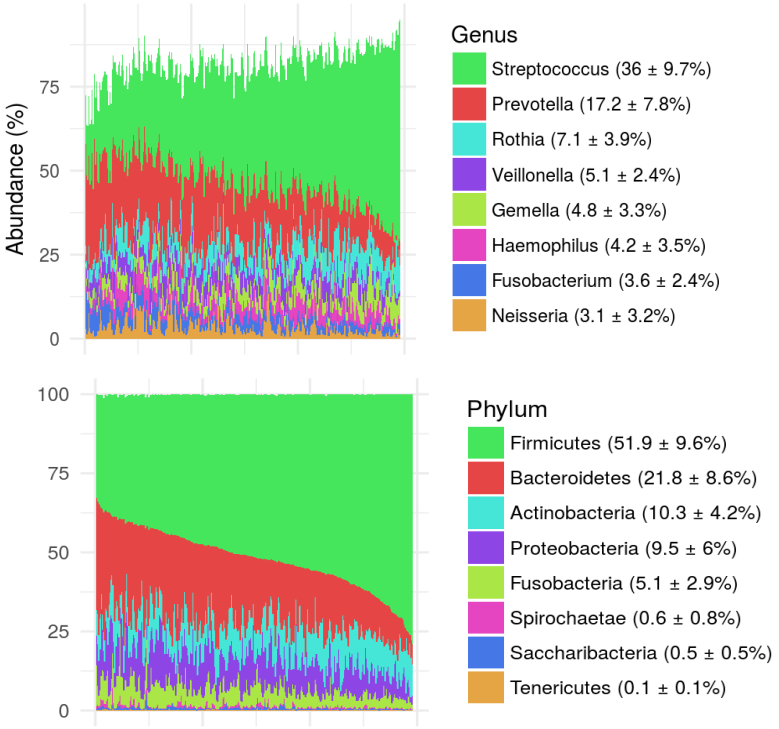
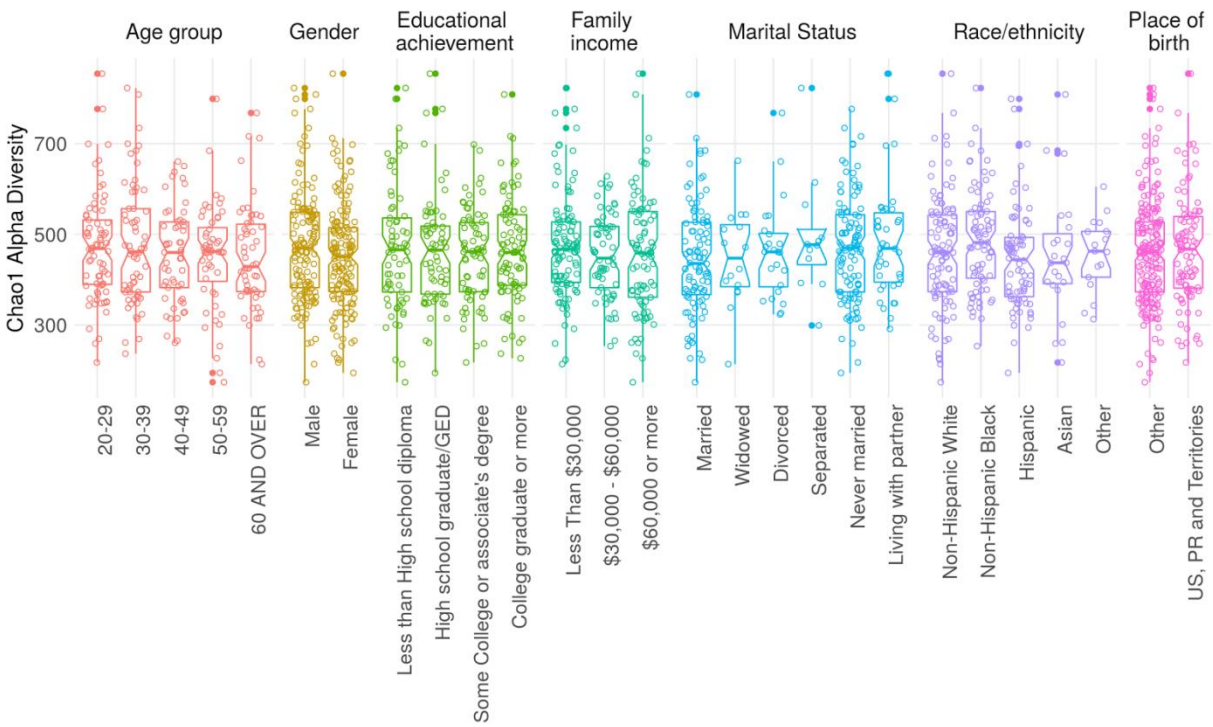


Figure 3. Alpha diversity by Sociodemographic Characteristics. Chao1 alpha diversity of 16S rRNA oral microbiome samples. Measures were compared using a null hypothesis of no difference between groups (Kruskal-Wallis test, p > 0.5 for all tests).



Oligotype associations within *Neisseria* for gender, race/ethnicity, and US- vs. foreign born are each for a mutually exclusive set of taxa. In fact, the associations with gender in *Neisseria*, *Prevotella*, and *Streptococcus* are all in a separate set of taxa than the associations with other socioeconomic variables. Age group and education had unidirectional associations in OTU analysis in *Streptococcus* but bidirectional differential abundance in oligotypes. In *Prevotella*, race/ethnicity and marital status had unidirectional associations in OTU analysis but bidirectional associations in oligotypes.

Clustering and Beta Diversity

Figure 6 illustrates the distribution of between- versus within-group weighted UniFrac distances by each sociodemographics variable. When testing these differences in β-diversity using PERMANOVA, education (p=0.025) and age group (p=0.018) were found to be predictors of beta diversity, with no other variables showing major between-group variation. We additionally explored plots of the first two principle coordinates based on weighted UniFrac distances colored by each sociodemographic variable, which showed very little patterning by any variable (not shown). To assess clustering, we calculated prediction strength for k=2 to 10 clusters on Bray-Curtis dissimilarity, Jensen-Shannon Divergence (JSD), root-JSD, weighted and unweighted UniFrac distances. Scores appeared to be sensitive to the distance metric used, with Bray-Curtis indicating moderate support for 2 clusters (PS=0.86), and all other measures providing little to no support for any number of clusters.

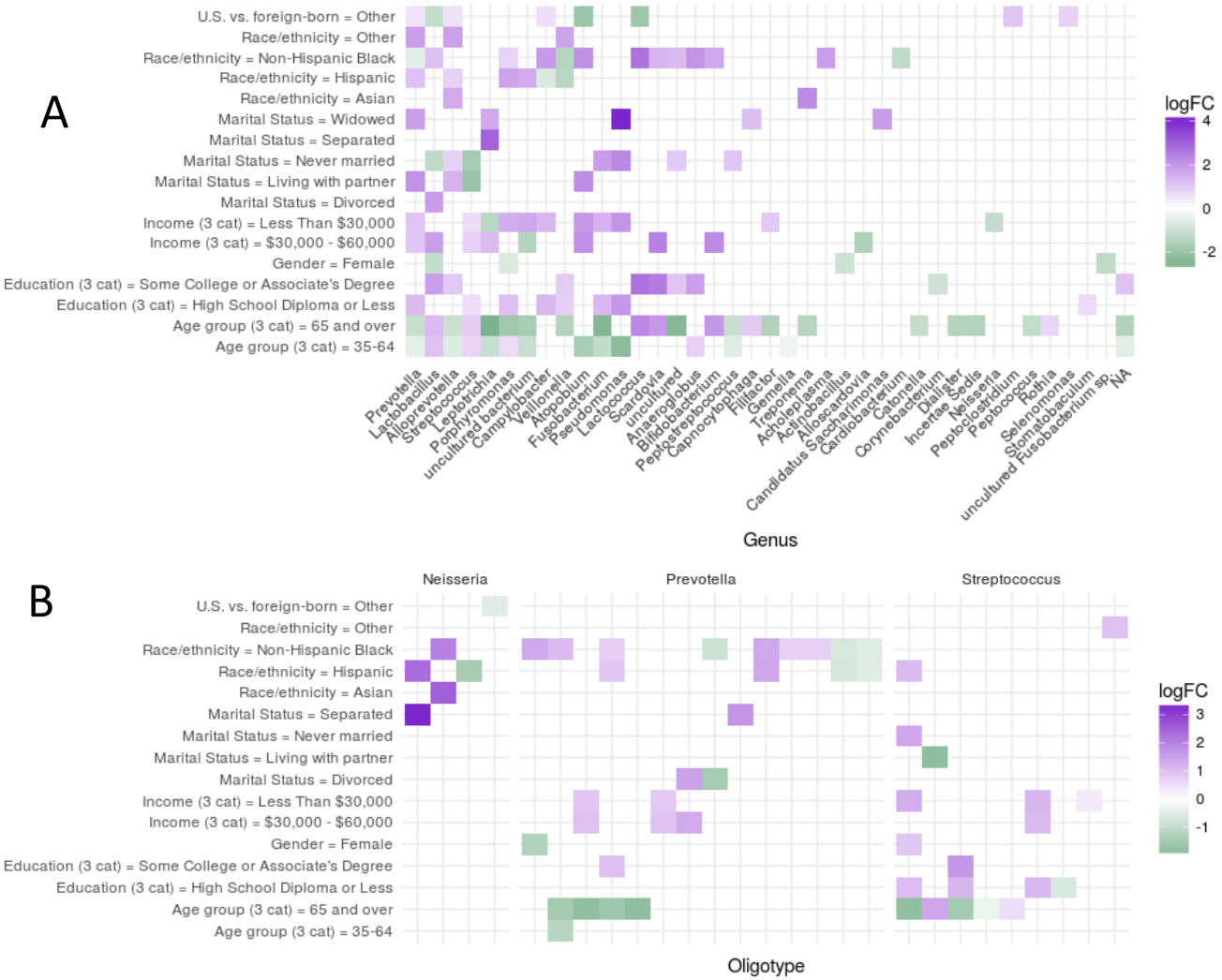
Discussion

This study examined associations of the oral microbiome with sociodemographic characteristics in a diverse population-based sample. We found that a large number of taxa were differentially abundant by age group, race/ethnicity, family income, education, nativity, and gender, indicating important variation in the oral microbiome across sociodemographic groups. Notably, we found a greater number of associations with SES variables (23 by family income, 20 by education) than with gender, marital status or nativity. There were more associations with SES than with mouthwash use (11) or gum disease (5), and a similar number of associations were found with sugar-sweetened beverage use (26). Furthermore, sociodemographic associations were not appreciably diminished by adjustment for smoking, mouthwash use, or sugar-sweetened beverage use.

When examining differential abundance by oligotypes, rather than OTUs, of *Neisseria*, *Prevotella*, and *Streptococcus*, different associations emerged, while some previous associations were not apparent at this level. Some variables, such as gender, had associations with certain oligotypes to the exclusion of other variables, while other variables had oligotype associations that seemingly contradicted one another. Measures of alpha diversity were consistently similar across groups, and similarly, measures of beta diversity and/or clustering did not reveal any major patterns in the oral microbiome by sociodemographic variables.

Our findings are consistent with the distribution of genera and phyla found in the oral microbiome samples in the

Figure 4. Differential abundance by sociodemographic characteristics. OTUs (A) and oligotypes (B) meeting unadjusted FDR < 0.01 in negative binomial log-linear GLMs using edgeR. Filled tiles in (A) indicate the genus had at least one OTU differentially abundant by at least one coefficient contrast within the sociodemographic factor. Most commonly differential genera in (A) included Prevotella (n=8) and Lactobacillus (n=7).



HMP (Figure 1): the most abundant genus across samples was *Streptococcus* and the most abundant phylum was *Firmicutes*. (34) *Streptococcus* appears to be the most frequent genus found in oral microbiomes worldwide.(51)

Our finding of a large number of differentially abundant taxa by race/ethnicity is consistent with existing studies. The HMP Consortium found that, when looking all body sites at once, ethnicity was the host phenotypic variable with the largest number of associations, followed by vaginal pH, age group, and body mass index (BMI).(34) Our findings with respect to race/ethnicity are also consistent with earlier studies evaluating a limited subset of oral bacteria. One study examining 40 pre-selected periodontal disease-related taxa found a number of differences among three racial/ethnic groups studied (Asian, Hispanic, Black).(37) Two lower-throughput studies found differences in *Porphyromonas*, *Fusobacterium*, and prevotella species prevalence by race/ethnicity.(38, 39) A study from a random sample of 663 community dwelling older adults in North Carolina found greater prevalence of A. actinomycetemcomitans, *P. intermedia*, and *P. gingivalis* in blacks vs. whites.(39) Another study from a convenience sample of 167 individuals with periodontal disease found *P. gingivalis* and *Peptostreptococcus anaerobius* were more abundant in Blacks vs. whites, while *Fusobacterium nucleatum* was more abundant in whites.(38)

We also identified a number of differentially abundant taxa by socioeconomic status (SES), measured by individual-level family income and education. This is consistent with findings from the Danish Health Examination Survey (DANHES), a similar-sized population-based sample (n=292). This study found nine 16S taxa in the oral microbiome that were differentially abundant by a municipal-level SES score, along with differences in cluster membership.(40) Many of the taxa differentially abundant by socioeconomic status in DANHES were also differentially

Figure 5. Distribution of absolute values of log-fold change (logFC). Values are from crude and adjusted edgeR models for all OTUs meeting FDR<0.01 in the crude model, for each sociodemographic variable.

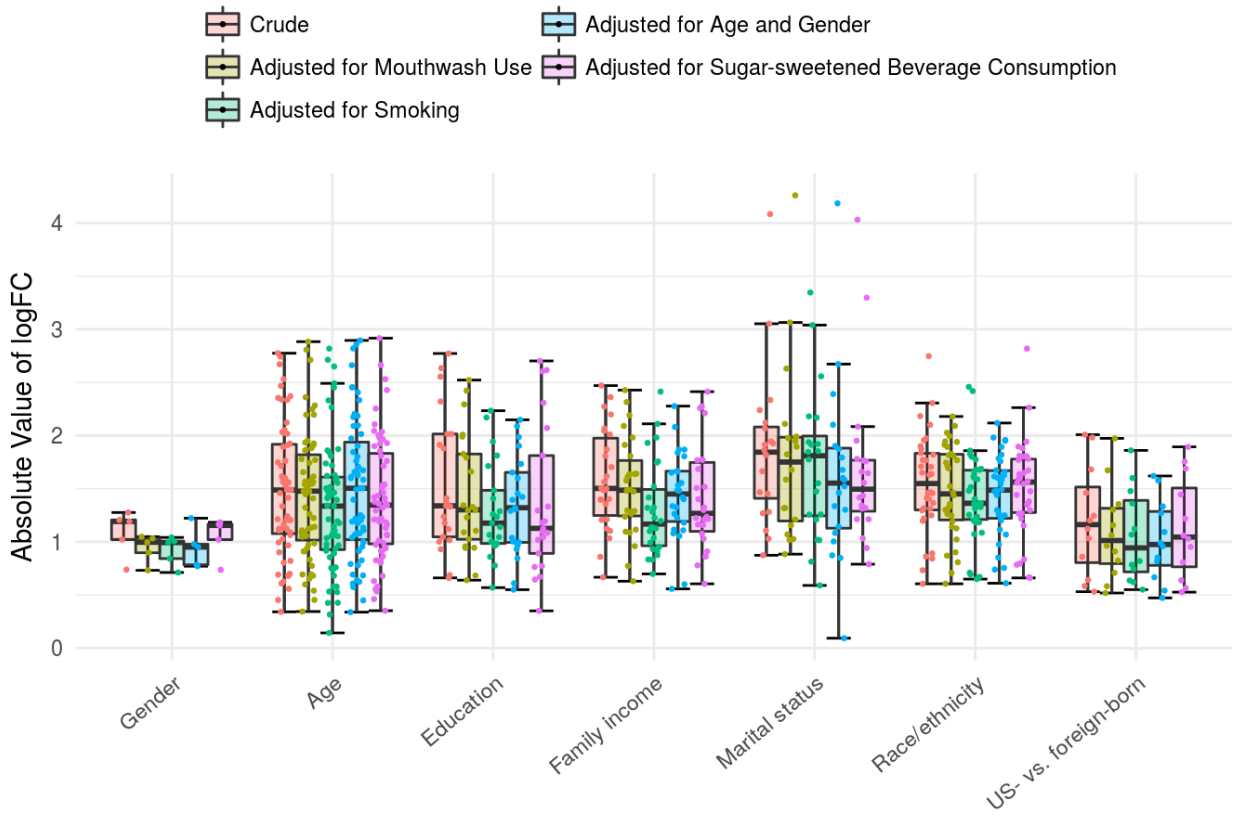
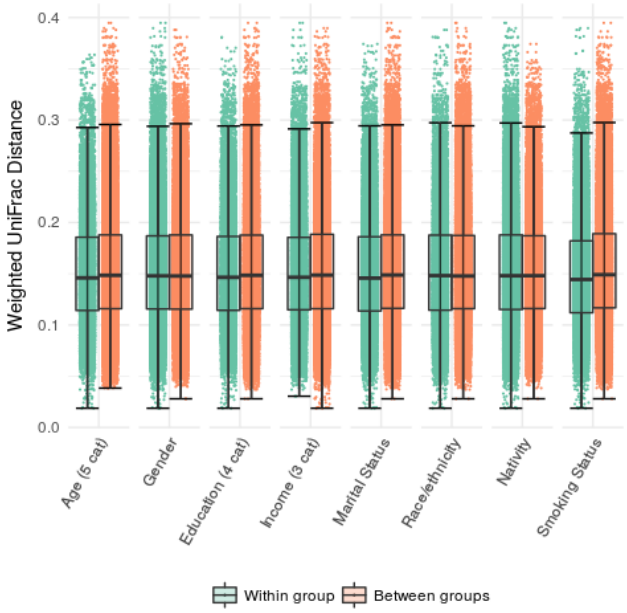


Figure 6. Within and between group beta diversity estimate distributions.



abundant by either income or education in our study, including *Streptococcus*, *Prevotella*, *Fusobacterium* (by education and income) *Leptotrichia* and *Neisseria* (income), and *Veillonella* (education), although many of the associations were in the opposite direction. However, this finding may be explained by the fact that the DAHNES oral microbiome substudy excluded participants with periodontitis and/or caries, while NYC HANES made no such exclusions. This procedure may have artificially selected a low SES population with more protective oral microbial profiles compared to the low SES population as a whole.

While existing oral microbiome studies are limited, the absence of differences in alpha and beta diversity by demographic groups is in contrast to some previous findings. A study comparing 16S oral profiles of non-Hispanic Black, non-Hispanic White, Latinos, and Chinese individuals in the United States found significantly lower alpha diversity in non-Hispanic Blacks, along with significant ethnicity-based clustering.(35) A comparison of 16S oral microbiomes in Cheyenne and Arahapo vs. non-native individuals or primarily Euro-American ancestry found higher alpha diversity in the latter along with differences in cluster membership.(36)

Many of the genera that were differentially abundant by multiple variables represent taxa that have been well characterized in the oral microbiome literature, and which

have documented associations with oral health and disease. *Streptococci* and *Lactobacilli (52)* as well as *Prevotella (53)* appear to be major players in caries, while differing levels of *Fusobacterium* and *Streptococcus* species are indicative of different stages in gingivitis.(54) Pathogens associated with, and thought to have an etiologic role in periodontitis include *Prevotella*, *Porphyromonas*, and *Fusobacterium* spp.(55, 56) Further, each of these organisms also appear to play a role in wide ranging systemic conditions.(6) *Porphyromonas gingivalis* is one of the most extensively studied periodontal pathogens, and appears to be a key structural determinant of the oral microbiome (57) in part because of its unique ability to evade and modulate the host immune response and disrupt homeostasis of the oral ecosystem.(58) In addition to being a major player in periodontal disease, *P. gingivalis* in the oral cavity is hypothesized to mediate an array of systemic pathogenic processes.(6) Abundance of *P. gingivalis* in the oral microbiome has strong associations with a number of cancer outcomes,(59) including pancreatic cancer incidence (8) and orodigestive cancer mortality.(9) Systemic antibody to *P. gingivalis* is associated with incident stroke (16) and CHD.(15) *P. gingivalis* has also been shown to play a role in rheumatoid arthritis.(13) In our sample, *Porphyromonas* is one of the most frequently differentially abundant microbes (in 5 out of 7 variables tested), is more abundant in Hispanics and non-Hispanic Blacks compared to non-Hispanic Whites, and less abundant in people earning $60k or more per year or with some college or more education compared to those with lower income or education.

Another differentially abundant organism, *Fusobacterium nucleatum* is a powerful predictor of colorectal cancer,(60) an association which is likely mediated by systemic inflammation rather than specific virulence factors.(61, 62) *F. nucleatum* is also implicated in adverse pregnancy outcomes, CVD and rheumatoid arthritis, through direct virulence as well as systemic inflammatory pathways.(62) *Fusobacterium* has associations consistent with socioeconomic disparities in our sample—it is decreased in those earning $60k or more and those with a college degree or greater. While knowledge of health-related associations with other oral microorganisms is still emerging, associations have been observed between *Prevotella spp.* and rheumatoid arthritis,(13) *Streptococcus anginosis* with esophageal cancer,(63) and *Lactococcus lactis* with inflammatory markers.(64)

Despite the strength of NYC-HANES as a diverse population-based sample, the cross-sectional design limits in its ability to test the role of the oral microbiome as a mediator in health disparties. Since changes in the oral microbiome may reflect existing disease rather than etiological factors, prospective studies are needed to better identified causal ordering and potential mediation. Additionally, our findings are limited by having primarily genus-level information. For instance, *P. gingivalis*, *F. nucleatum*, and *P. intermedia* are best characterized in literature supporting a possible etiologic role for these organisms in oral and systemic health, but we were unable to confirm differences in the species. There may also be wide variability in virulence even at the species level, as is the case with *P. gingivalis* (58). Given the importance of many of the differentially abundant genera in health and disease, our findings suggest that further investigation into mediation of health disparities by oral microbial factors is warranted. Future investigations should consider use of whole genome shotgun (WGS) or other methods able to reliably classify at the species level, given our finding that oligotypes of *Prevotella* show differential abundance patterns that were not seen using 16S analysis.

AUTHOR INFORMATION

Corresponding Author

Audrey Renson
audrey.renson@nyumc.org

ACKNOWLEDGMENT

We gratefully acknowledge the efforts of the New York City Department of Health and Mental Hygiene in co-leading the parent NYC HANES study. In particular, we wish to thank Sharon Perlman, Carolyn Greene, Claudia Chernov, Amado Punsalang, and the many other staff who helped support data collection.

REFERENCES

1. Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiotia to periodontal health and disease. Cell Host Microbe. 2011;10(4):302-6.
2. Darveau RP. Periodontitis: a polymicrobial disruption of host homeostasis. Nat Rev Microbiol. 2010;8(7):481-90.
3. Devine DA, Marsh PD, Meade J. Modulation of host responses by oral commensal bacteria. J Oral Microbiol. 2015;7:26941.
4. Buduneli N, Baylas H, Buduneli E, Türkoğlu O, Köse T, Dahlen G. Periodontal infections and pre-term low birth weight: a case-control study. J Clin Periodontol. 2005;32(2):174-81.
5. Fardini Y, Chung P, Dumm R, Joshi N, Han YW. Transmission of diverse oral bacteria to murine placenta: evidence for the oral microbiome as a potential source of intrauterine infection. Infect Immun. 2010;78(4):1789-96.
6. Atanasova KR, Yilmaz Ö. Prelude to oral microbes and chronic diseases: past, present and future. Microbes Infect. 2015;17(7):473-83.
7. Hooper SJ, Crean S-J, Fardy MJ, Lewis MAO, Spratt DA, Wade WG, et al. A molecular analysis of the bacteria present within oral squamous cell carcinoma. J Med Microbiol. 2007;56(Pt 12):1651-9.
8. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut. 2016.
9. Ahn J, Segers S, Hayes RB. Periodontal disease, Porphyromonas gingivalis serum antibody levels and orodigestive cancer mortality. Carcinogenesis. 2012;33(5):1055-8.
10. Gurav A, Jadhav V. Periodontitis and risk of diabetes mellitus. J Diabetes. 2011;3(1):21-8.
11. Lalla E, Papapanou PN. Diabetes mellitus and periodontitis: a tale of two common interrelated diseases. Nat Rev Endocrinol. 2011;7(12):738-48.
12. Bingham CO, 3rd, Moni M. Periodontal disease and rheumatoid arthritis: the evidence accumulates for complex pathobiologic interactions. Curr Opin Rheumatol. 2013;25(3):345-53.

13. Ogrendik M. Rheumatoid arthritis is an autoimmune disease caused by periodontal pathogens. Int J Gen Med. 2013;6:383-6.
14. Lockhart PB, Bolger AF, Papapanou PN, Osinbowale O, Trevisan M, Levison ME, et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association?: a scientific statement from the American Heart Association. Circulation. 2012;125(20):2520-44.
15. Pussinen PJ, Jousilahti P, Alfthan G, Palosuo T, Asikainen S, Salomaa V. Antibodies to periodontal pathogens are associated with coronary heart disease. Arterioscler Thromb Vasc Biol. 2003;23(7):1250-4.
16. Pussinen PJ, Alfthan G, Jousilahti P, Paju S, Tuomilehto J. Systemic exposure to Porphyromonas gingivalis predicts incident stroke. Atherosclerosis. 2007;193(1):222-8.
17. Huang DL, Park M. Socioeconomic and racial/ethnic oral health disparities among US older adults: oral health quality of life and dentition. J Public Health Dent. 2015;75(2):85-92.
18. Karlamangla AS, Merkin SS, Crimmins EM, Seeman TE. Socioeconomic and ethnic disparities in cardiovascular risk in the United States, 2001-2006. Ann Epidemiol. 2010;20(8):617-28.
19. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. Circulation. 2017:CIR.0000000000000485.
20. Beckles GL, Chou C-F. Disparities in the Prevalence of Diagnosed Diabetes - United States, 1999-2002 and 2011-2014. MMWR Morb Mortal Wkly Rep. 2016;65(45):1265-9.
21. Crawford S, Joshi N, Boulet SL, Bailey MA, Hood M-E, Manning SE, et al. Maternal Racial and Ethnic Disparities in Neonatal Birth Outcomes With and Without Assisted Reproduction. Obstet Gynecol. 2017;129(6):1022-30.
22. Martinson ML, Reichman NE. Socioeconomic Inequalities in Low Birth Weight in the United States, the United Kingdom, Canada, and Australia. Am J Public Health. 2016;106(4):748-54.
23. McBurney CA, Vina ER. Racial and ethnic disparities in rheumatoid arthritis. Curr Rheumatol Rep. 2012;14(5):463-71.
24. Cohen JM, Wilson ML, Aiello AE. Analysis of social epidemiology research on infectious diseases: historical patterns and future opportunities. J Epidemiol Community Health. 2007;61(12):1021-7.
25. Aiello AE, Dowd JB. Socio-economic Status and Immunosenescence. Immunosenescence: Springer, New York, NY; 2013. p. 145-57.
26. Bosch JA, Turkenburg M, Nazmi K, Veerman ECI, de Geus EJC, Nieuw Amerongen AV. Stress as a determinant of saliva-mediated adherence and coadherence of oral and nonoral microorganisms. Psychosom Med. 2003;65(4):604-12.
27. Kato I, Vasquez A, Moyerbrailean G, Land S, Djuric Z, Sun J, et al. Nutritional Correlates of Human Oral Microbiome. J Am Coll Nutr. 2017;36(2):88-98.
28. Corby PM, Bretz WA, Hart TC, Schork NJ, Wessel J, Lyons-Weiler J, et al. Heritability of oral microbial species in caries-active and caries-free twins. Twin Res Hum Genet. 2007;10(6):821-8.
29. Li Y, Ismail AI, Ge Y, Tellez M, Sohn W. Similarity of bacterial populations in saliva from African-American mother-child dyads. J Clin Microbiol. 2007;45(9):3082-5.
30. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345(6200):1048-52.
31. Findley K, Williams DR, Grice EA, Bonham VL. Health Disparities and the Microbiome. Trends Microbiol. 2016;24(11):847-50.
32. Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J. 2013;27(3):1012-22.
33. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome biology. 2012;13(6):R42.
34. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.
35. Mason MR, Nagaraja HN, Camerlengo T, Joshi V, Kumar PS. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. PloS one. 2013;8(10):e77287.
36. Ozga AT, Sankaranarayanan K, Tito RY, Obregon-Tito AJ, Foster MW, Tallbull G, et al. Oral microbiome diversity among Cheyenne and Arapaho individuals from Oklahoma. Am J Phys Anthropol. 2016;161(2):321-7.
37. Craig Rg BRYJBPKJMDFJIMSSSHAD. Prevalence and risk indicators for destructive periodontal diseases in 3 urban American minority populations. J Clin Periodontal. 2001(28):524-35.
38. Schenkein HA, Burmeister JA, Koertge TE, Brooks CN, Best AM, Moore LV, et al. The influence of race and gender on periodontal microflora. J Periodontol. 1993;64(4):292-6.
39. Beck JD, Koch GG, Zambon JJ, Genco RJ, Tudor GE. Evaluation of oral bacteria as risk indicators for periodontitis in older adults. J Periodontol. 1992;63(2):93-9.
40. Belstrøm D, Holmstrup P, Nielsen CH, Kirkby N, Twetman S, Heitmann BL, et al. Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. J Oral Microbiol. 2014;6.
41. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J Biotechnol. 2017.
42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
43. Chao A. Nonparametric estimation of the number of classes in a population. Scandinavian Journal of statistics. 1984:265-70.
44. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology. 2005;71(12):8228-35.
45. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, et al. A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput Biol. 2013;9(1):e1002863.
46. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017.
47. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PloS one. 2013;8(4):e61217.
48. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2.4-3 ed2017.
49. Hennig C. fpc: Flexible Procedures for Clustering. 2015.
50. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol. 2013;4(12).
51. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. Genome Res. 2009;19(4):636-43.

52. Takahashi N, Nyvad B. The role of bacteria in the caries process: ecological perspectives. J Dent Res. 2011;90(3):294-303.
53. Yang F, Zeng X, Ning K, Liu K-L, Lo C-C, Wang W, et al. Saliva microbiomes distinguish caries-active from healthy human populations. ISME J. 2012;6(1):1-10.
54. Al-Ahmad A, Wunder A, Auschill TM, Follo M, Braun G, Hellwig E, et al. The in vivo dynamics of Streptococcus spp., Actinomyces naeslundii, Fusobacterium nucleatum and Veillonella spp. in dental plaque biofilm as analysed by five-colour multiplex fluorescence in situ hybridization. J Med Microbiol. 2007;56(Pt 5):681-7.
55. Zijnge V, van Leeuwen MBM, Degener JE, Abbas F, Thurnheer T, Gmür R, et al. Oral biofilm architecture on natural teeth. PloS one. 2010;5(2):e9321.
56. Gursoy UK, Könönen E, Uitto V-J, Pussinen PJ, Hyvärinen K, Suominen-Taipale L, et al. Salivary interleukin-1beta concentration and the presence of multiple pathogens in periodontitis. J Clin Periodontol. 2009;36(11):922-7.
57. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. Nat Rev Immunol. 2015;15(1):30-44.
58. Tribble GD, Kerr JE, Wang B-Y. Genetic diversity in the oral pathogen Porphyromonas gingivalis: molecular mechanisms and biological consequences. Future Microbiol. 2013;8(5):607-20.
59. Atanasova KR, Yilmaz O. Looking in the Porphyromonas gingivalis cabinet of curiosities: the microbium, the host and cancer association. Mol Oral Microbiol. 2014;29(2):55-66.
60. Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, et al. Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. Eur J Clin Microbiol Infect Dis. 2014;33(8):1381-90.
61. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome. 2013;1(1):16.
62. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol. 2015;23:141-7.
63. Morita E, Narikiyo M, Yano A, Nishimura E, Igaki H, Sasaki H, et al. Different frequencies of Streptococcus anginosus infection in oral cancer and esophageal cancer. Cancer Sci. 2003;94(6):492-6.
64. Luerce TD, Gomes-Santos AC, Rocha CS, Moreira TG, Cruz DN, Lemos L, et al. Anti-inflammatory effects of Lactococcus lactis NCDO 2118 during the remission period of chemically induced colitis. Gut Pathog. 2014;6:33.