Sociodemographic variation in the oral microbiome

Audrey Renson,*[1,2] Heidi E. Jones[1], Francesco Beghini,[3] Nicola Segata,[3] Christine P. Zolnik,[4,5] Mykhaylo Usyk,[4] Thomas U. Moody,[4,6] Lorna Thorpe,[7] Robert Burk,[4,8] Levi Waldron,†[1,9] Jennifer B. Dowd†[1,10]

[1] Graduate School of Public Health and Health Policy, City University of New York, Department of Epidemiology and Biostatistics, New York, NY, USA

[2] New York University (NYU) Langone Hospital – Brooklyn, Department of Clinical Research, Brooklyn, NY, USA

[3] Centre for Integrative Biology, University of Trento, Trento, Italy

[4] Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA

[5] Department of Biology, Long Island University, Brooklyn, NY 11201, USA

[6] Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[7] NYU School of Medicine, Department of Population Health, New York, NY, USA

[8] Departments of Microbiology and Immunology, Epidemiology and Population Health, and Obstetrics, Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx , NY USA

[9] Institute for Implementation Science in Population Health, City University of New York, New York, USA

[10] King's College London, Department of Global Health and Social Medicine, London, UK

* corresponding author
† equal contribution

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

ABSTRACT:
Background: Variations in the human oral microbiome are potentially implicated in health inequalities, but existing studies of the oral microbiome have minimal sociodemographic diversity. This study describes sociodemographic variation of the oral microbiome in a diverse sample of New York City residents.
Methods: Data come from 296 participants, a subsample of the 2013-14 population-based New York City Health and Nutrition Examination Study (NYC-HANES). Mouthwash samples were processed using using 16S v4 rRNA amplicon sequencing. We examined differential abundance of 216 operational taxonomic units (OTUs), in addition to alpha and beta diversity amongst sociodemographic variables including age, gender, income, education, nativity, and race/ethnicity.
Results: A total of 75 OTUs were differentially abundant by any sociodemographic variable (false discovery rate < 0.01), including 27 by race/ethnicity, 23 by family income, 20 by education, and five by gender. Genera differing for more than one sociodemographic characteristic included Lactobacillus, Prevotella, Porphyromonas, and Fusobacterium. Education (p=0.03) and age (p=0.02) were associated with weighted UniFrac distances.
Discussion: In a diverse sample, we identified variations in the oral microbiome 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 understanding the factors that shape population level variation. Oral diseases such as caries, gingivitis, and periodontitis are 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 oral microbiota structure and function 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 (CHD) (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).

Health disparities 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 (CVD) (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. The social environment is associated with a variety of microbial exposures (24, 25). Changes in immune function related to psychosocial stress (26), nutrition (27), smoking (28), or other environmental exposures can alter host interactions with microbes. Differences in microbiome characteristics may also persist via mother-to-child transmission (29, 30) and shared built environments (31). These findings have led researchers to call for investigation into the role of the microbiome in health disparities (32) but limited research has examined sociodemographic associations with the oral microbiome.

The Human Microbiome Project (HMP) collected microbiome samples at nine distinct oral sites on a volunteer sample in the U.S. with minimal race/ethnic variability (approx. 80% white) (33, 34). Nonetheless, the HMP found differentially abundant taxa comparing non-Hispanic white, non-Hispanic black, Asian, Mexican, and Puerto Rican ethnicities (35). In another U.S. volunteer sample, distinct subgingival microbiomes were identified by race/ethnicity, with non-Hispanic blacks having lower microbiome diversity than other groups (36). In a comparison of salivary microbiomes of Cheyenne and Arahapo vs. non-Native individuals in the U.S., strong bacterial species composition clustering, differences in species richness, and numerous differentially abundant taxa were found by ethnicity (37). Several low-throughput studies examining specific periodontal pathogens found significant differences in abundance and/or presence by race/ethnicity (38-40). To our knowledge, only one study has tested associations between SES and the oral 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 (41).

In order to explore the relationship between the oral microbiome and health disparities, population-level sociodemographic associations must be assessed. Our aim was to assess sociodemographic variation in the human salivary microbiome. Specifically, we examined whether bacterial taxa were differentially abundant, and whether variation existed in alpha and beta diversity by sociodemographic characteristics, using high-throughput sequencing data from a population-based sample.

METHODS

Samples came from the 2013-14 New York City Health and Nutrition Examination Survey (NYC HANES-II) previously described (42). Briefly, NYC HANES-II was the second population-representative, cross-sectional survey of adult NYC residents, using a three-stage cluster sampling design. Overall response rate was 36% (n=1524). Eligible participants completed a two-part interview, physical examination, and blood, urine, and oral mouthwash biospecimen collection. Nearly all participants (95%) provided an oral mouthwash specimen. NYC HANES-II was approved by the City University of New York institutional review board, and all subject gave informed consent, including consent to use oral mouthwash specimens for future studies.

Subsample Selection

The current study uses a 296 NYC HANES-II participants selected to examine oral microbiome associations with tobacco use. Eligible participants were sampled based on self-reported and/or serum cotinine-confirmed smoking status as follows: All current hookah, cigar/cigarillo and/or e-cigarette users, regardless of cotinine or cigarette smoking (n=79), self-reported cigarette smokers with the highest cotinine levels (n=90), and a random subsample of: former smokers with no secondhand smoke exposure (cotinine <0.05 ng/mL, n=45), former/never smokers with secondhand smoke exposure (cotinine 1-14 ng/mL, n=38), and never smokers with no secondhand smoke exposure (cotinine <0.05 ng/mL, n=45).

Oral rinse collection and microbiome sample processing

Oral rinse samples were collected by trained field interviewers, and consisted of a 20-second oral rinse using 15 mL of Scope© mouthwash. Samples were transported on dry ice and stored at -80°C at a New York City Department of Health and Mental Hygiene laboratory, and processed at Albert Einstein College of Medicine.

DNA Extraction

All laboratory procedures were performed under a hood (AirClean Systems) to minimize environmental contamination and negative controls were used throughout. From each oral rinse sample, a 1.5 mL aliquot was centrifuged at 750 x g for 5 min and all but 150 µl of supernatant was removed. The pellet was re-suspended in the remaining supernatant and incubated in an enzyme mixture consisting of lysozyme (0.84 mg/ml, Sigma Aldrich), mutanolysin (0.25 U/ml, Sigma Aldrich) and lysostaphin (21.10 U/ml, Sigma Aldrich), at 37°C for 30 minutes. This was followed by incubation at 56°C for 10 minutes in 15 µl proteinase K and 150 µl Buffer AL. Samples were then transferred to screw top tubes with 100 g of 0.1-mm-diameter Zirconia/Silica Beads (BioSpec) and bead beaten using a FastPrep-24 homogenizer (MP Biomedicals) at speed 6.0 for 40 seconds. Tubes were centrifuged at 750 x g for 30 sec and 150 µl of supernatant was added to a new 1.7 ml tube with 150 µl of 100% ethanol and mixed by vortexing for 15 seconds. Supernatant was then added to the spin column from the QIAamp DNA mini kit (QIAGEN) and centrifuged at 6000 x g for 1 minute. Column purification was performed according to the QIAamp DNA mini kit directions starting at the AWI wash step. Final elution was performed in 100 µl of Buffer AE.

16S rRNA Gene Amplification

DNA was amplified for the V4 variable region of the 16S rRNA gene using the primers 16SV4_515F (GTGYCAGCMGCCGCGGTA) and 16SV4_806R (GGACTACHVGGGTWTCTAAT) (43, 44). Each primer had an 8-bp unique Hamming barcode with forward primers containing a 3-bp (TCG) and 4-bp (ACTG) pad on either side, with reverse primers including a 3-bp (GTA) and 4-bp (TC) pad on each side of the barcode (45). PCR reactions were performed with 17.75 µl of nuclease-free PCR-grade water, 2.5 µl of 10X Buffer w/ $MgCl_2$ (Affymetrix, Santa Clara, CA), 1µl of $MgCl_2$ (25 mM, Affymetrix, Santa Clara, California, USA), 0.5 µl of dNTPs (10 mM, Roche, Basel, Switzerland), 0.25 µl of AmpliTaq Gold DNA Polymerase (5 U/µl, Applied Biostystems, Foster City, California), 0.5 µl of HotStart-IT FideliTaq (2.5 U/µl, Affymetrix, Santa Clara, CA), 1µl of each primer (5 µM), and 0.5 µl of DNA extraction template. Thermal cycling conditions consisted of initial denaturation of 95ºC for 5 min, followed by 15 cycles of 95ºC for 1 min, 55ºC for 1 min, and 68ºC for 1 min, followed by 15 cycles of 95ºC for 1 min, 60ºC for 1 min, and 68ºC for 1 min, a final extension for 10 min at 68ºC on a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA).

PCR products were combined before running 100 µl of the pooled products on a 4% agarose gel at 80V for 2 hours. The ~450 bp bands were excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and eluted in 30 µl of elution buffer. Purified PCR products were quantified using a Qubit 2.0 Fluorometic High Sensitivity dsDNA Assay (Life Technologies, Carlsbad, CA).

Library Preparation and Sequencing

Library preparation of the purified PCR products was performed using a KAPA LTP Library Preparation Kit (Kapa Biosystems, Wilmington, MA). The size integrity of the amplicon was validated with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). High-throughput amplicon sequencing was conducted on a MiSeq (Illumina, San Diego, CA) using 2x300 paired-end fragments. The fastq sequences from the Illumina MiSeq were demultiplexed using Novobarcode (Novocrat Technologies, Selangor, Malaysia) and the 5'-pads and primers were trimmed from each read.

Bacterial taxa were determined by clustering the 16S rRNA sequences into operational taxonomical units (OTUs) using 97% similarity, taxonomy was assigned at the genus level using the SILVA 123 (46) database as reference, excluding samples with less than 1000 reads. To assess within-genera differences, we determined oligotypes using 62 positions for Prevotella, 80 for Streptococcus, and 29 for Neisseria, and removed oligotypes present in less than 10 samples (46, 47).

Statistical Analysis

We compared differences in oral microbiome characteristics by seven sociodemographic factors (race/ethnicity, age group, gender, highest level of education completed, income tertile, marital status, nativity) and by four behavioral/health measures (self-reported gum disease, mouthwash use, smoking status and sugar-sweetened beverage consumption).

We examined relative abundance, or the proportion of total OTU counts accounted for by each taxa, at the phylum and genus levels. To assess differential abundance by sociodemographic variables, we used edgeR (48). This technique estimates log-linear generalized linear models (GLMs) with each OTU as the response, and estimates dispersion parameters by sharing information across OTUs by empirical Bayes. Before edgeR, we filtered out OTUs that did not have three or more samples with a count of at least eight, leaving 216 OTUs for analysis, a filter representing an approximate inflection point on the curve of remaining OTUs against the minimum count. We fit edgeR models, both crude and adjusted separately for mouthwash use, sugar sweetened beverage use, smoking status, age and gender. OTUs were considered differentially abundant at false discovery rate (FDR) < 0.01.

We measured alpha diversity using Chao1 richness (49), which we compared by each sociodemographic variable using Kruskal-Wallis tests. Beta diversity was assessed using principal coordinates analysis on weighted UniFrac distances (50). Principal coordinates plots were colored by each sociodemographic variable and permutation multivariate analysis of variance (51) was used to assess group differences.

We performed clustering of samples with respect to OTUs using partitioning around medoids on Bray Curtis, Jenson-Shannon, root-Jenson Shannon, weighted and unweighted UniFrac distances (52). Prediction strength (PS) was calculated for k=2:10 clusters on each distance measure, using PS≥0.9 to signify strong support for k clusters (52)

Statistical analyses were conducted in R version 3.4 (53) for Linux, using 'edgeR' for differential abundance (48), 'phyloseq' for ecological diversity measures (54), 'vegan' for permutation multivariate analysis of variance, (55) and 'fpc' for clustering (56).

RESULTS

Table 1 shows 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), and educational achievement (23.3% less than high school diploma, 30.4% college degree or greater). 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%). (Figure 2).

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)

The overall mean chao1 was 462±118, with no differences by age group (p=0.823), gender (p=0.102), educational achievement (p=0.942), annual family income (p=0.589), marital status (p=0.575), race/ethnicity (p=0.189), or nativity (p=0.655, Figure 3).

Differential Abundance and Oligotyping

Numerous taxa were differentially abundant by race/ethnicity, nativity, marital status, gender, family income tertiles, education, and age groups. Figure 4a displays log fold change (logFC), or coefficient from edgeR log linear models, for each comparison group and all significant OTUs.

A total of 75 OTUs were differentially abundant by any sociodemographic 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. We also found 11 by mouthwash use, five by self-reported gum disease, 53 by smoking status, and 26 by sugar-sweetened beverage consumption. The most frequently differentially abundant were Lactobacillus (all variables), and Prevotella (age, education, family income, marital status, race/ethnicity, nativity, Figure 4a) Differential abundance findings for selected taxa are presented in Table 2.

As numerous associations were present at FDR<0.01, we focus here on findings where the logFC was 2 or greater. 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 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. 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).

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

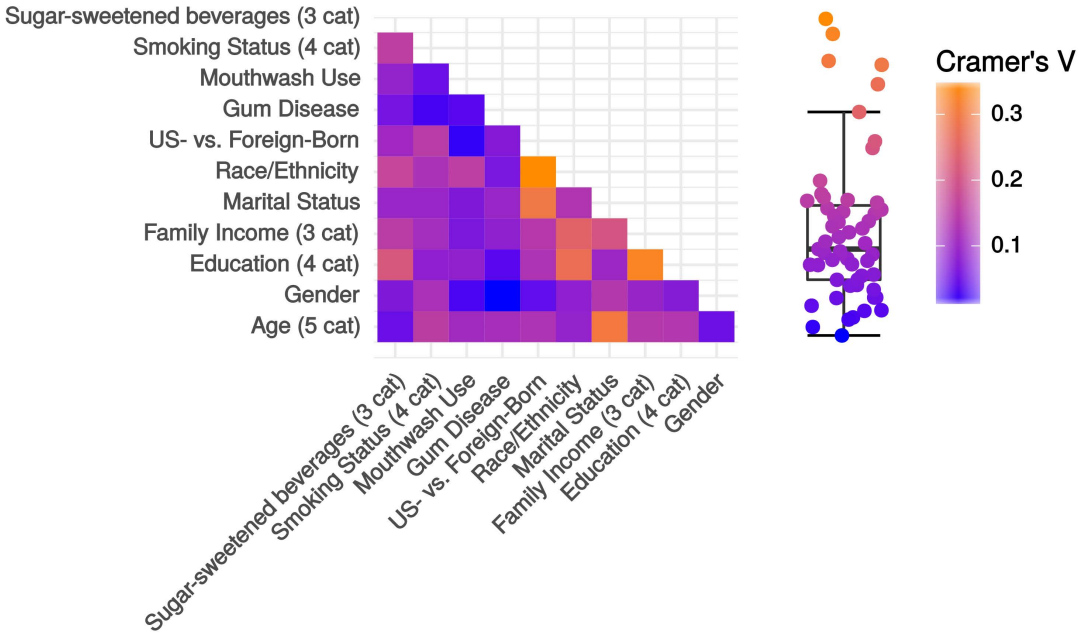

Figure 1. Examining collinearity among sociodemographic variables. Data are absolute value of pairwise Cramer's V correlation coefficient between sociodemographic factor levels. Data are from the full sample (n=1,527) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: cat=categories; US=United States.

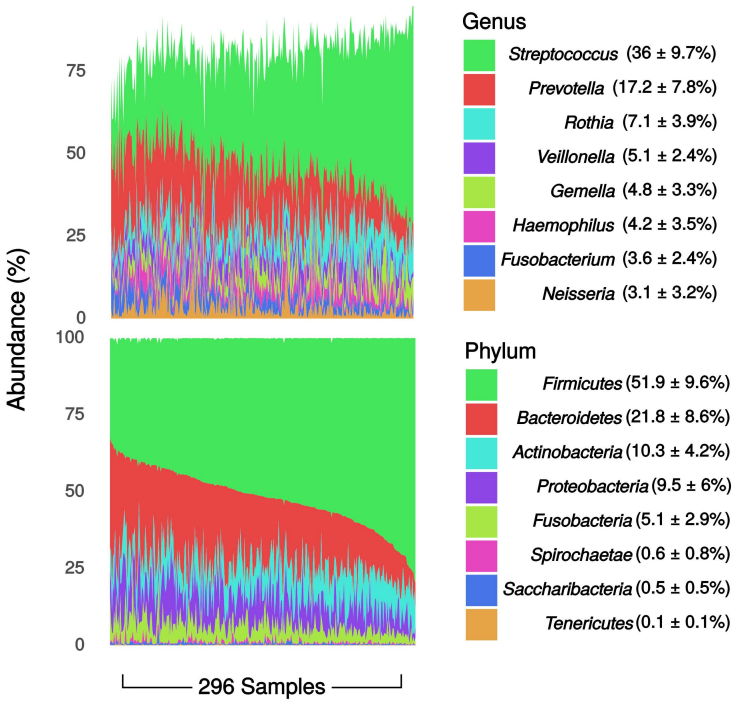

Figure 2. Genus- and phylum-level relative abundances. Data are percent of overall communities within samples, summarized as mean ± standard deviation of percent across samples. Data are from the oral microbiome subsample (n=296) of the New York City Health and Nutrition Examination Survey, 2013-2014.

Figure 5 displays the distribution of logFCs for both crude and adjusted, including all OTUs with FDR <0.01 in crude models. Adjusting for smoking, mouthwash use, age and gender, had a minor effect on crude estimates; however, adjustment for smoking exerts the largest effect on findings for age, income and education.

Analyzing oligotypes of Neisseria, Prevotella, and Streptococcus revealed associations not apparent in the OTU analysis, whereas some associations present in OTU analysis were not apparent in oligotypes (Figure 4). New associations were revealed between Prevotella and gender, Streptococcus and gender, race/ethnicity and nativity, and Neisseria and gender, age, education, marital status, race/ethnicity and nativity. Associations present in OTUs but not in oligotyping were age, education and income in Prevotella, and income in Neisseria.

Oligotype associations within Neisseria for gender, race/ethnicity, and nativity are each for a mutually exclusive set of taxa, and associations with gender in Neisseria, Prevotella, and Streptococcus are all in separate taxa from the associations with other sociodemographic 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.

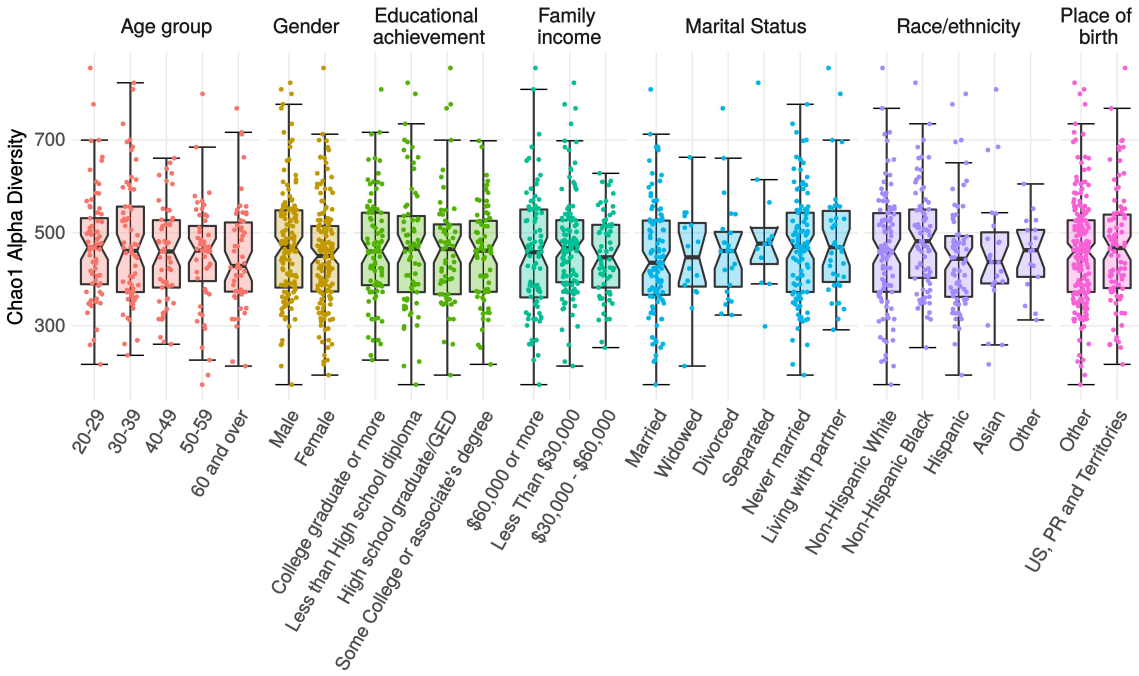

Figure 3. Alpha diversity by Sociodemographic Characteristics. Chao1 alpha diversity of 16S rRNA oral microbiome samples. Measures were compared using a null hypothesis of no difference between groups (Kruskal-Wallis test, $p > 0.1$ for all tests). Data are from the oral microbiome subsample (n=296) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: GED=General equivalency diploma; PR=Puerto Rico; US=United States.

Beta Diversity and Clustering

Figure 6 illustrates between-versus within-group weighted UniFrac distances by each sociodemographic variable. Education (p=0.025, R2=0.020) and age group (p=0.018, R2=0.028) were associated with beta diversity, with no other variables showing between-group variation. Plots of the first two principal coordinates based on weighted UniFrac distances showed little patterning by any variable (not shown). Clustering scores were sensitive to the distance metric used, with Bray-Curtis indicating moderate support for 2 clusters (PS=0.86), and all other measures providing little support for clustering.

DISCUSSION

In a diverse population-based sample, we found that a large number of bacterial taxa were differentially abundant by age group, race/ethnicity, family income, education, nativity, and gender. 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 mouthwash use (11) or gum disease (5), and a similar number of associations were found with sugar-sweetened beverage use (26). Sociodemographic associations were not appreciably diminished by adjustment for these factors. We also found that differential abundance by sociodemographic characteristics differed in oligotyping vs. OTUs, especially for Neisseria. Alpha diversity was similar across groups, and beta diversity explained only a small percent of variance by education (2.0%) and age (2.8%), and less by other variables. We found poor support for clustering of samples by OTUs, and that, similarly to Koren et al. (2013) (52), clustering findings were sensitive to the distance metric employed.

Our relative abundance findings are consistent with those found in oral microbiome samples in the HMP (Figure 1): the most abundant genus was Streptococcus and the most abundant phylum was Firmicutes. (35) Streptococcus appears to be the most frequent genus in oral microbiomes worldwide (57).

Our finding of differentially abundant taxa by race/ethnicity is consistent with existing studies. The HMP Consortium found that, for all body sites, ethnicity was the host phenotypic variable with the most associations (35). For the oral microbiome, a study examining 40 periodontal disease-related taxa found differences among Asian, Hispanic, and Blacks (38). Two lower-throughput studies found differences by race/ethnicity, including greater Prevotella and Porphyromonas prevalence (40), and lower Fusobacterium abundance (39) in blacks vs. whites.

We also identified a number of differentially abundant taxa by SES, measured by family income and education. This is consistent with findings from the Danish Health Examination Survey (DANHES, n=292), which found nine differentially abundant taxa and differential clustering in

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. Where more than one OTU was significant within one genus, the maximum logFC is displayed. Most commonly differential genera in (A) included Prevotella (n=8) and Lactobacillus (n=7). Data are from the oral microbiome subsample (n=296) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: cat=categories; GLM=generalized linear model; logFC=log fold change; OTU=operational taxonomic unit; US=United States.

the oral microbiome by municipal-level SES (41). Many differentially abundant taxa by SES in DANHES were also differentially abundant by SES 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. The difference in direction could be explained by exclusion of participants with oral disease in DANHES, which may have artificially selected a low SES population with more protective oral microbial profiles.

While existing oral microbiome studies are limited, the absence of differences in alpha and beta diversity by demographic groups contrasts with previous findings. A study comparing oral profiles of non-Hispanic Black, non-Hispanic white, Latinos, and Chinese individuals in the U.S. found lower alpha diversity in non-Hispanic blacks, along with ethnicity-based clustering (36). A comparison of oral microbiomes in Cheyenne and Arahapo vs. non-native individuals of primarily Euro-American ancestry found higher alpha diversity in the latter along with ethnicity-based clustering (37).

Many genera found differentially abundant by multiple variables represent taxa that have documented associations with health and disease. Streptococcus, Lactobacillus (58) and Prevotella (59) are implicated in caries, while Fusobacterium and Streptococcus levels reflect different stages in gingivitis (60). Pathogens with potentially causal associations with periodontitis include Prevotella, Porphyromonas, and Fusobacterium spp (61, 62). Further, each of these organisms likely play a role in wide ranging systemic conditions (6). Porphyromonas gingivalis is a key determinant of oral microbiome structure (63) reflective of its ability to disrupt homeostasis of the oral ecosystem (64).

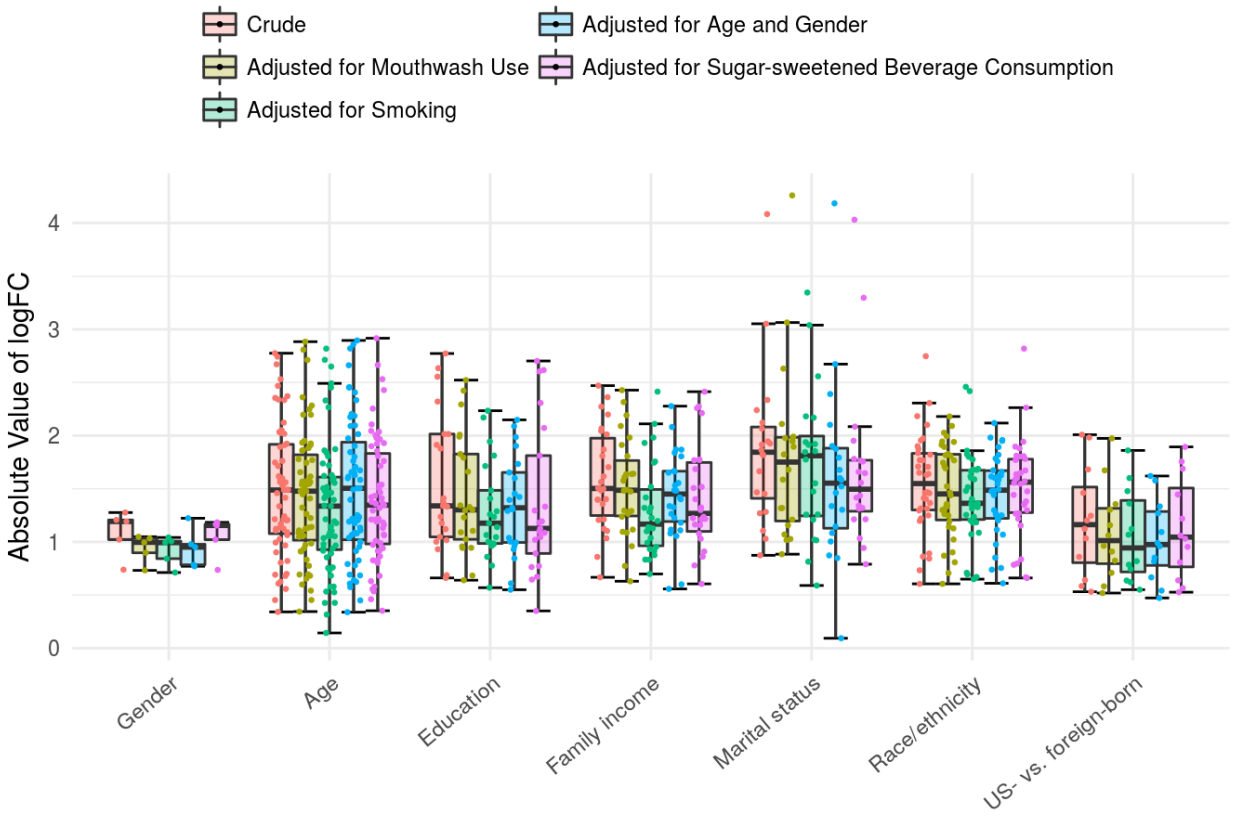

Figure 5. Distribution of absolute values of log-fold change (logFC) in crude and adjusted negative binomial log-linear GLMs edgeR models for each sociodemographic variable. Data are from the oral microbiome subsample (n=296) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: GLM=generalized linear model; logFC=log fold change; US=United States.

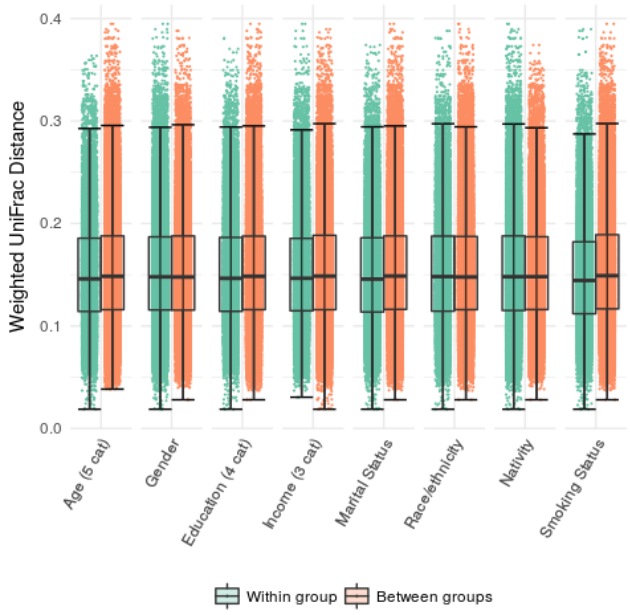

Figure 6. Within and between group beta diversity estimate distributions. Data are from the oral microbiome subsample (n=296) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: cat=category.

P. gingivalis in the oral cavity is hypothesized to mediate an array of systemic pathogenic processes (6), and has associations with stroke (16), CHD (15), a number of cancers (65) including pancreatic (8) and orodigestive (9), and likely plays a role in rheumatoid arthritis (13). In our sample, Porphyromonas 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 compared to those with lower income or education.

Another differentially abundant organism, Fusobacterium nucleatum is increasingly linked to colorectal cancer (66, 67), an association likely mediated by systemic inflammation (68, 69). F. nucleatum is implicated in adverse pregnancy outcomes, CVD and rheumatoid arthritis, via direct virulence and systemic inflammation (69). 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. A few studies also link Prevotella spp. and rheumatoid arthritis (13), Streptococcus anginosis with esophageal cancer (70), and Lactococcus lactis with inflammatory markers (71).

Despite the strength of NYC-HANES as a diverse population-based sample, the cross-sectional design limits in its ability to test the oral microbiome as a mediator in health disparities. Since changes in the oral microbiome may reflect existing disease rather than etiological factors,

prospective studies are needed to test the oral microbiome as a mediator. Additionally, our findings are limited by having primarily genus-level information. P. gingivalis, F. nucleatum, and P. intermedia are best characterized in literature supporting a possible etiologic role in oral and systemic health, but we are unable to confirm differences in the species. There may also be variability in virulence at the species level, as is the case with P. gingivalis (64).

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

ACKNOWLEDGMENTS

The individual author contributions are as follows: HEJ, LW, LT, RB, and JD conceptualized and designed the study; AR, FB, NS, and LW led data analysis and data visualization; RB, CPZ, MU, and TUM led specimen processing and 16S data generation; AR wrote first draft of manuscript; and all authors contributed to editing/revisions on manuscript.
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, et al. 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, et al. 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, 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, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study [electronic article]. Gut. Advance Access: 2016/10/14. (DOI:10.1136/gutjnl-2016-312580).
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, 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, et al. 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, et al. 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, et al. 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, et al. Heart disease and stroke statistics—2017 update: A report from the American Heart Association. Circulation 2017;135(10):e146-e603.
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, 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, 2013:145-57.
26. Bosch JA, Turkenburg M, Nazmi K, et al. 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, et al. Nutritional correlates of human oral microbiome. J Am Coll Nutr 2017;36(2):88-98.
28. Wu J, Peters BA, Dominianni C, et al. Cigarette smoking and the oral microbiome in a large study of American adults. The ISME journal 2016;10(10):2435-46.
29. Corby PM, Bretz WA, Hart TC, et al. Heritability of oral microbial species in caries-active and caries-free twins. Twin Res Hum Genet 2007;10(6):821-8.

30. Li Y, Ismail AI, Ge Y, et al. Similarity of bacterial populations in saliva from African-American mother-child dyads. J Clin Microbiol 2007;45(9):3082-5.
31. Lax S, Smith DP, Hampton-Marcell J, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science 2014;345(6200):1048-52.
32. Findley K, Williams DR, Grice EA, et al. Health disparities and the microbiome. Trends Microbiol 2016;24(11):847-50.
33. Aagaard K, Petrosino J, Keitel W, 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.
34. Segata N, Haake SK, Mannon P, 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.
35. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature 2012;486(7402):207-14.
36. Mason MR, Nagaraja HN, Camerlengo T, et al. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. PloS one 2013;8(10):e77287.
37. Ozga AT, Sankaranarayanan K, Tito RY, et al. Oral microbiome diversity among Cheyenne and Arapaho individuals from Oklahoma. Am J Phys Anthropol 2016;161(2):321-7.
38. Craig RG, Boylan R, Yip J, et al. Prevalence and risk indicators for destructive periodontal diseases in 3 urban American minority populations.pdf. J Clin Periodontal 2001(28):524-35.
39. Schenkein HA, Burmeister JA, Koertge TE, et al. The influence of race and gender on periodontal microflora. J Periodontol 1993;64(4):292-6.
40. Beck JD, Koch GG, Zambon JJ, et al. Evaluation of oral bacteria as risk indicators for periodontitis in older adults. J Periodontol 1992;63(2):93-9.
41. Belstrøm D, Holmstrup P, Nielsen CH, et al. Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. J Oral Microbiol 2014;6.
42. Thorpe LE, Greene C, Freeman A, et al. Rationale, design and respondent characteristics of the 2013-2014 New York City Health and Nutrition Examination Survey (NYC HANES 2013-2014). Prev Med Rep 2015;2:580-5.
43. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America 2011;108 Suppl 1:4516-22.
44. Wang Y, Qian PY. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PloS one 2009;4(10).
45. Hamady M, Walker JJ, Harris JK, et al. Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. Nature methods 2008;5(3):235-7.
46. Glöckner FO, Yilmaz P, Quast C, et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J Biotechnol 2017;261.
47. Eren AM, Maignien L, Sul WJ, et al. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 2013;4(12).
48. 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.
49. Chao A. Nonparametric estimation of the number of classes in a population. Scandinavian Journal of statistics 1984:265-70.
50. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology 2005;71(12):8228-35.
51. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral ecology 2001;26(1):32-46.
52. Koren O, Knights D, Gonzalez A, 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.
53. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2017.
54. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PloS one 2013;8(4):e61217.
55. Oksanen J, Blanchet G, Friendly M, et al. vegan: Community ecology package. 2017.
56. Hennig C. fpc: Flexible procedures for clustering. 2015.
57. Nasidze I, Li J, Quinque D, et al. Global diversity in the human salivary microbiome. Genome Res 2009;19(4):636-43.
58. Takahashi N, Nyvad B. The role of bacteria in the caries process: ecological perspectives. J Dent Res 2011;90(3):294-303.
59. Yang F, Zeng X, Ning K, et al. Saliva microbiomes distinguish caries-active from healthy human populations. ISME J 2012;6(1):1-10.
60. Al-Ahmad A, Wunder A, Auschill TM, 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.
61. Zijnge V, van Leeuwen MBM, Degener JE, et al. Oral biofilm architecture on natural teeth. PloS one 2010;5(2):e9321.
62. Gursoy UK, Könönen E, Uitto V-J, et al. Salivary interleukin-1beta concentration and the presence of multiple pathogens in periodontitis. J Clin Periodontol 2009;36(11):922-7.
63. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. Nat Rev Immunol 2015;15(1):30-44.
64. 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.
65. 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.
66. Flanagan L, Schmid J, Ebert M, 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.
67. Mima K, Nishihara R, Qian ZR, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut 2016;65(12):1973-80.
68. Warren RL, Freeman DJ, Pleasance S, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome 2013;1(1):16.
69. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol 2015;23:141-7.
70. Morita E, Narikiyo M, Yano A, et al. Different frequencies of Streptococcus anginosus infection in oral cancer and esophageal cancer. Cancer Sci 2003;94(6):492-6.
71. Luerce TD, Gomes-Santos AC, Rocha CS, et al. Anti-inflammatory effects of Lactococcus lactis NCDO 2118 during the remission period of chemically induced colitis. Gut Pathog 2014;6:33.