



Published in final edited form as:

*Ann Epidemiol.* 2019 July ; 35: 73–80.e2. doi:10.1016/j.annepidem.2019.03.006.

## Sociodemographic variation in the oral microbiome

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

<sup>1</sup>Graduate School of Public Health and Health Policy, City University of New York, Department of Epidemiology and Biostatistics, New York, NY, USA

<sup>2</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

<sup>3</sup>Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy

<sup>4</sup>Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA

<sup>5</sup>Department of Biology, Long Island University, Brooklyn, NY 11201, USA

<sup>6</sup>Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

<sup>7</sup>NYU School of Medicine, Department of Population Health, New York, NY, USA

<sup>8</sup>Departments of Microbiology and Immunology, Epidemiology and Population Health, and Obstetrics, Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

<sup>9</sup>Institute for Implementation Science in Population Health, City University of New York, New York, USA

<sup>10</sup>King's College London, Department of Global Health and Social Medicine, London, UK

### Abstract

**1.1 Purpose**—Variations in the oral microbiome are potentially implicated in social inequalities in oral disease, cancers, and metabolic disease. We describe sociodemographic variation of oral microbiomes in a diverse sample.

**1.2 Methods**—We performed 16S rRNA sequencing on mouthwash specimens in a subsample (n=282) of the 2013-14 population-based New York City Health and Nutrition Examination Study (NYC-HANES). We examined differential abundance of 216 operational taxonomic units (OTUs), and alpha and beta diversity by age, sex, income, education, nativity, and race/ethnicity. For

---

**Corresponding Author Contact Info:** Audrey Renson, arensong@ad.unc.edu, Phone: 347-743-4927, Permanent Address: 100 South Peak Drive, Carrboro, NC, USA 27510.

<sup>†</sup>equal contribution

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

comparison, we examined differential abundance by diet, smoking status, and oral health behaviors.

**1.3 Results**—69 OTUs were differentially abundant by any sociodemographic variable (false discovery rate < 0.01), including 27 by race/ethnicity, 21 by family income, 19 by education, three by sex. We found 49 differentially abundant by smoking status, 23 by diet, 12 by oral health behaviors. Genera differing for multiple sociodemographic characteristics included *Lactobacillus*, *Prevotella*, *Porphyromonas*, *Fusobacterium*.

**1.4 Conclusions**—We identified oral microbiome variation consistent with health inequalities, more taxa differing by race/ethnicity than diet, and more by SES variables than oral health behaviors. Investigation is warranted into possible mediating effects of the oral microbiome in social disparities in oral and metabolic diseases and cancers.

## Keywords

oral microbiome; health disparities; demographics; social epidemiology

---

## 2 Introduction

Health disparities by race/ethnicity, socioeconomic status (SES), sex, and other sociodemographic factors have long been observed but their mechanisms have yet to be fully elucidated. In particular, racial/ethnic and socioeconomic disparities have been consistently observed in oral health outcomes [1], cardiovascular disease (CVD) [2], diabetes [3], preterm birth and low birth weight [4, 5], and rheumatoid arthritis [6].

Variations in human oral microbiome structure and function have been associated with oral disease [7, 8], as well as a wide range of systemic illnesses including CVD [9–11], diabetes [12, 13], cancers [14–17], birth outcomes [18, 19], and rheumatoid arthritis [20, 21]. Hypothesized pathways for such associations include both direct virulence and modulation of systemic immune response [14], although causal evidence is limited. Also, regardless of their causal role, the microbiota represent potential biomarkers for early disease detection and prognosis.

This combination of findings has led researchers to call for investigation into the role of the microbiome in health disparities [22] but little empirical work has yet been done in this area. A number of mechanisms potentially link social factors to the microbiome [23]. Such mechanisms have been discussed in relation to common pathogens such as CMV and EBV; these may include household crowding, use of public transportation, and differences in susceptibility due to breastfeeding and poor sleep [4, 5], mechanisms which may apply to commensal microbes as well. Changes in immune function related to psychosocial stress [24], nutrition [25], smoking [26], or other environmental exposures can alter host interactions with microbes. Differences in microbiome characteristics may also persist via mother-to-child transmission, as infant microbiomes are seeded from the birth canal and via breastfeeding [27, 28]. Further, assortative social networks and shared built environments may represent reservoirs of shared microbiota membership [29].

So far, 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) [30, 31]. Nonetheless, the HMP found differentially abundant taxa comparing non-Hispanic white, non-Hispanic black, Asian, Mexican, and Puerto Rican ethnicities [32]. 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 [33]. 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 [34]. Several low-throughput studies examining specific periodontal pathogens found significant differences in abundance and/or presence by race/ethnicity [35–37]. To our knowledge, only one study has tested associations between SES and the oral microbiome, finding substantial differences (20% of variation) by municipal-level SES in the Danish Health Examination Survey [38].

In order to explore the relationship between the oral microbiome and health disparities, sociodemographic associations from diverse samples 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 subsample of a population-based sample.

### 3 Methods

#### 3.1 Data Source

Samples came from the 2013-14 New York City Health and Nutrition Examination Survey (NYC HANES-II) previously described [39]. Briefly, the 2013-14 NYC HANES 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 nearly all (95%) provided an oral mouthwash specimen. This study was approved by the institutional review boards of the City University of New York and the New York City Department of Health and Mental Hygiene, and all participants gave informed consent. Participants providing mouthwash specimens in the current sub-study also consented to use these specimens in future studies.

#### 3.2 Subsample Selection

The current study uses 297 NYC HANES participants selected to examine oral microbiome associations with tobacco use, as described elsewhere [CITATION PENDING – Beghini 2018 Companion Paper]. Briefly, we selected 90 self-reported current cigarette smokers with the highest serum cotinine, 45 randomly selected never smokers with cotinine <0.05 ng/mL, 45 randomly selected former smokers with cotinine <0.05 ng/mL, all 38 former and never smokers with serum cotinine between 1 and 14 ng/mL, and 79 participants reporting hookah,

cigar, cigarillo and/or e-cigarette use within 5 days. Table 1 shows descriptive statistics in the subsample and overall NYC HANES sample.

### 3.3 Oral rinse collection and microbiome sample processing

Participants were asked to fast for 9 hours prior to oral rinse collection. A 20-second oral rinse was divided into two 5-second swish and 5-second gargle sessions using 15 mLs of Scope® mouthwash. After each session, participants expectorated into a sterile cup. Timers built into the computer-assisted personal interview program signaled the timing of the swish, gargle and expectoration. Oral rinse specimens were stored cold before delivery to the New York Public Health Laboratory where they were transferred into 50 mL centrifuge tubes, frozen and stored at  $-80^{\circ}\text{C}$ . The oral rinse samples were then transported on dry ice to Albert Einstein College of Medicine, where they were stored at  $-80^{\circ}\text{C}$  until processing.

Specimen processing and sequence analysis methods are described in detail in the appendix. Briefly, we extracted DNA using QIAamp DNA mini kit (QIAGEN), and amplified DNA in the V4 region of the 16S rRNA using primers 16SV4\_515F (GTGYCAGCMGCCGCGGTA) and 16SV4\_806R (GGACTACHVGGGTWTCTAAT) (38,39), followed by amplicon sequencing using a MiSeq (Illumina, San Diego, CA) with  $2\times 300$  paired-end fragments. We merged raw Illumina paired-end reads using the QIIME v1.9.1 (40) command fastq-join (42), and discarded low quality reads (PHRED score  $< 30$ ) when joining split reads (qiime split\_libraries\_fastq.py). We performed open-reference Operational Taxonomic Unit (OTU) picking by clustering using UCLUST at 97% similarity, and assigned taxonomy using the SILVA 123 (43) database. We removed samples with  $< 1000$  reads ( $n=15$ ) and collapsed genera with mean relative abundance  $< 2\times 10^{-4}$  into a category labelled “Other.” [40–43]

### 3.4 Statistical Analysis

We compared oral microbiomes by seven sociodemographic factors (race/ethnicity, age, group, sex, educational attainment, income tertiles, marital status, nativity) and by several behavioral/oral health measures: diet (sugar sweetened beverages, meat, poultry, fish, vegetables, and fruits, recorded as times consumed in the past week); oral health behaviors (mouthwash use, flossing, time since last dental visit) and smoking status (categories defined above). We assessed pairwise correlation between sociodemographic variables using Cramer’s V, a correlation coefficient for nominal variables.

To assess differential abundance (DA) by sociodemographic variables, we used edgeR [44] to estimate a series of log-linear generalized linear models (GLMs) predicting each OTU abundance. OTUs were considered differentially abundant at false discovery rate (FDR)  $< 0.01$ . Before edgeR, we filtered out OTUs for which less than three samples had a count of at least eight, leaving 216 OTUs for analysis. To examine potential mediators, we fit crude models and models adjusted for oral health behaviors, diet, smoking status, and age and sex (when applicable). edgeR was conducted at the taxonomic level of highest specificity allowed, which was the genus in all cases where FDR was less than 1%; therefore DA findings are presented at the genus level.

We measured alpha diversity using Chao1 richness [45], which we compared by each sociodemographic variable using Kruskal-Wallis tests. Beta diversity was assessed using principal coordinates analysis and permutation multivariate analysis of variance (PERMANOVA) [46] on weighted UniFrac distances [47].

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 [48]. 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 [48].

**3.4.1 Sensitivity Analysis**—As described above, the sample was selected based on smoking status, which is plausibly a mediator and/or effect measure modifier for the effect of sociodemographics on the microbiome. In particular, evidence supports the existence of socioeconomic status disparities in smoking prevalence [49, 50], and effects of smoking on the oral microbiome [51, 52][CITATION PENDING – Beghini 2018 Companion Paper]. To characterize the potential bias due to selecting on smoking, we generated sampling weights defined as the inverse of the predicted probability of selection into the substudy, using logistic regression on self-reported smoking status, logarithm serum cotinine, and their interaction, fitted to the entire NYC HANES sample. We also tested models with 2<sup>nd</sup> and 3<sup>rd</sup> order polynomials, and splines for log-transformed cotinine, and conclusions were identical. We applied these weights to edgeR models and compared the logFCs and numbers of OTUs differential at FDR < 0.01

Statistical analyses were conducted in R version 3.4 [53] for Linux.

## 4 Results

### 4.1 Descriptive Statistics

The initial subsample included 297 participants; after removing samples with <1000 reads, there were 282 participants for analysis. Table 1 shows descriptive statistics for sociodemographic characteristics including age (median [range]: 42 [20 to 94]), sex (53.2% female), race/ethnicity (34.4% non-Hispanic White, 26.6% non-Hispanic Black, 25.2% Hispanic), annual family income (42.7% less than \$30K, 33.3% \$60k or more), and educational achievement (23.0% less than high school diploma, 30.9% college degree or greater). Cramer's V on pairwise combinations of sociodemographic variables showed only minor collinearity (all  $V < .35$ ) (Figure A.1), indicating associations with the microbiome for each sociodemographic variable do not merely reflect correlations between sociodemographic variables.

### 4.2 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 1). The mean Chao1 was 462, with no differences by age group (p=0.79), sex (p=0.13), educational achievement (p=0.92), annual family income (p=0.62), marital status (p=0.54), race/ethnicity (p=0.13), or nativity (p=0.97) (Figure A.2).

### 4.3 Differential Abundance

Numerous taxa were differentially abundant (DA) by race/ethnicity, nativity, marital status, gender, family income, education, and age. Figure 2 displays log-2 fold change (logFC), or coefficient from edgeR log-linear models, for each comparison group and all significant OTUs. The logFC can be interpreted as the log-base-2 ratio of relative abundance compared to the reference group, so that e.g. *Lactobacillus* is found to be  $2^{2.5} = 5.7$  times as abundant in family incomes of \$30-60,000 per year, compared to \$60,000 or more. In total, 69 OTUs were DA by any sociodemographic variable, including 56 by age group, 27 by race/ethnicity, 21 by family income, 19 by education, 19 by marital status, seven by nativity, and three by sex. We found 12 unique DA OTUs by oral health behaviors, 49 by smoking status, and 23 by diet variables. The most frequently DA were *Lactobacillus* (all variables), and *Prevotella* (age, education, family income, marital status, race/ethnicity, nativity, Figure 2). DA findings for selected taxa are presented in Table 2 (see table A.1 for all DA findings).

Figure 3 displays boxplots of absolute values of logFCs for crude and adjusted models. The OTUs displayed for all models are those meeting FDR <0.01 in crude models. Comparing adjusted vs. crude boxplots allows visual assessment of effects of adjustment on the entire set of OTUs: a shift towards zero reflects attenuation while a shift away from zero reflects amplification. Over all sociodemographic variables, a minor attenuating effect was observed after adjusting for smoking (mean change in logFC, -3.9%), oral health behaviors (-4.9%), diet (-6.3%), age and sex (-3.3%). Adjustment for oral health had the largest impact on logFCs for age group (-4.0%), sex (-27.4%), and nativity (-13.5%); diet had the strongest impact on logFCs for education (-13.1%) and marital status (-16.9%), smoking had the strongest impact on logFCs for family income (-11.9%), and age and sex had the strongest impact on logFCs for race/ethnicity (-4.2%). Figure A.3 illustrates the effects of the sensitivity analysis with inverse probability of selection weights applied; distributions of logFC estimates were nearly identical and nearly all DA OTUs in unweighted crude analysis were also DA (FDR<0.01) in weighted models. However, weighted models detected a substantially larger number of DA OTUs for every variable, suggesting that selecting on smoking may have biased towards the null for most associations.

### 4.4 Beta Diversity and Clustering

Figure 4 illustrates between-versus-within-group weighted UniFrac distances by each sociodemographic variable. We observed differences in composition by age group ( $p=0.017$ ,  $r^2=0.026$ ), with no other variables showing greater between- than within-group variation, a result which was not changed by adjusting for smoking. Principal coordinates plots 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.

## 5 Discussion

In a diverse subsample of a population-based study, we found that a large number of bacterial taxa in the oral microbiome were DA by age, race/ethnicity, family income, education, nativity, and sex. Notably, we found a greater number of associations with SES



variables (21 by family income, 19 by education) than with sex, marital status or nativity. There were also more associations with SES than oral health behaviors (12). Adjustment for smoking, oral health behaviors, or dietary behaviors did not appreciably diminish sociodemographic associations.

Many genera found DA by multiple variables represent taxa that have documented associations with health and disease. *Streptococcus*, *Lactobacillus* [54], *Prevotella* [55] *Fusobacterium* [56], and *Porphyromonas* [57, 58] are understood to play a role in oral disease. Further, many of these organisms likely play a role in systemic conditions [14]. Specifically, *Fusobacterium* have been linked to colorectal cancer [59, 60], adverse pregnancy outcomes, CVD and rheumatoid arthritis [61]. *Porphyromonas gingivalis* is a key determinant of oral microbiome structure [62], and is hypothesized to mediate multiple systemic pathogenic processes [14], including stroke [10], CHD [11], a number of cancers [16, 17, 63] and rheumatoid arthritis [21].

To our knowledge, our study is the first to examine differences in oral microbiota by individual-level sociodemographic factors in a diverse population-based sample. Our finding of DA taxa by race/ethnicity is consistent with previous studies with small convenience samples. The HMP found that, for all body sites, ethnicity was the host phenotypic variable with the most associations [32]. For the oral microbiome, a study examining 40 periodontal disease-related taxa found differences among Asian, Hispanic, and blacks [35]. Two lower-throughput studies found greater *Prevotella* and *Porphyromonas* [37], and lower *Fusobacterium* [36] in blacks vs. whites. Our finding of differential OTUs by SES variables is also consistent with findings from the Danish Health Examination Survey (DANHES, n=292), which found nine DA taxa by municipal-level SES [38].

Adjustment for smoking, diet, and oral health behaviors each moderately attenuated DA findings across sociodemographic categories. This stands to reason in light of findings by our group [CITATION PENDING - Beghini 2018 Companion Paper] and others [26] that smoking is associated with major shifts in the oral microbiome, along with similar findings for diet [64], and indicates that some portion of observed sociodemographic patterning reflects differences in health habits or access to dental care. However, the finding that differential abundance was not eliminated by adjustments suggests that additional mechanisms underlie sociodemographic variation in the oral microbiome. These may include upstream social factors such as psychosocial stress [24] or features of the built environment [29].

While existing oral microbiome studies are limited, the absence of differences in alpha and beta diversity by race/ethnicity contrasts with two previous studies among non-population-based samples. These found differences in alpha diversity and ethnicity-based clustering in oral microbiomes in non-Hispanic Blacks vs. Whites [33], and in Cheyenne and Arahapo vs. non-native individuals [34]. Differences in alpha and beta diversity can indicate larger-scale shifts in composition; our finding that specific OTUs were differentially abundant but that overall shifts were less present would tend to indicate that, at a population level, sociodemographic patterns in oral microbiome composition are more subtle. An alternative

explanation is that our sampling design attenuated alpha and beta diversity difference estimates.

## 5.1 Limitations

Despite the strength of NYC-HANES as a diverse population-based sample, the selection of the substudy on the basis of smoking means that the microbiome sample was not population-representative. Based on our sensitivity analysis with inverse-probability weighting, this sample selection most likely biased towards the null, yielding conservative estimates of our focal associations. Additionally, our findings are limited by having primarily genus-level information, and in many cases salient differences exist at a greater degree of taxonomic specificity – for example, with *P. gingivalis*, *F. nucleatum*, and *Prevotella intermedia*. There may also be wide variability in virulence even at the species level, as is the case with *P. gingivalis* [65]. Given the importance of many of the differentially abundant genera in health and disease, our findings suggest that further investigation into the role of the oral microbiome in health disparities is warranted. Future investigations should consider use of whole genome shotgun sequencing or other methods able to provide more specific taxonomic classification and describe functional, as well as taxonomic, composition.

## 5.2 Conclusion

Our results lend support to potential role of the social environment in shaping microbiome composition at the population level [23, 66]. The finding of differentially abundant OTUs, many of which are health-relevant, for every sociodemographic variable, suggests that these associations may be important in determining population health patterns. In particular for race/ethnicity and SES, but also for nativity and marital status, the finding that multiple health-relevant microbes are differentially abundant supports a growing hypothesis that the microbiota may partially mediate long-observed social disparities in major disease outcomes. At a minimum, these results highlight that social factors may be important potential confounders in studies of the human oral microbiome and health.

Mechanisms for the observed associations are currently unknown, and one important next step will be to examine the multiple levels of exposures underlying these associations, including macro-level social and health policy, exposure to psychosocial stressors, outdoor and built environment features, and social interactions [23]. Importantly, if the microbiome is a partial mediator of health disparities, then identifying modifiable features of the social environment that are most strongly associated with the microbiome can inform effective interventions to improve population health and reduce health disparities.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

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

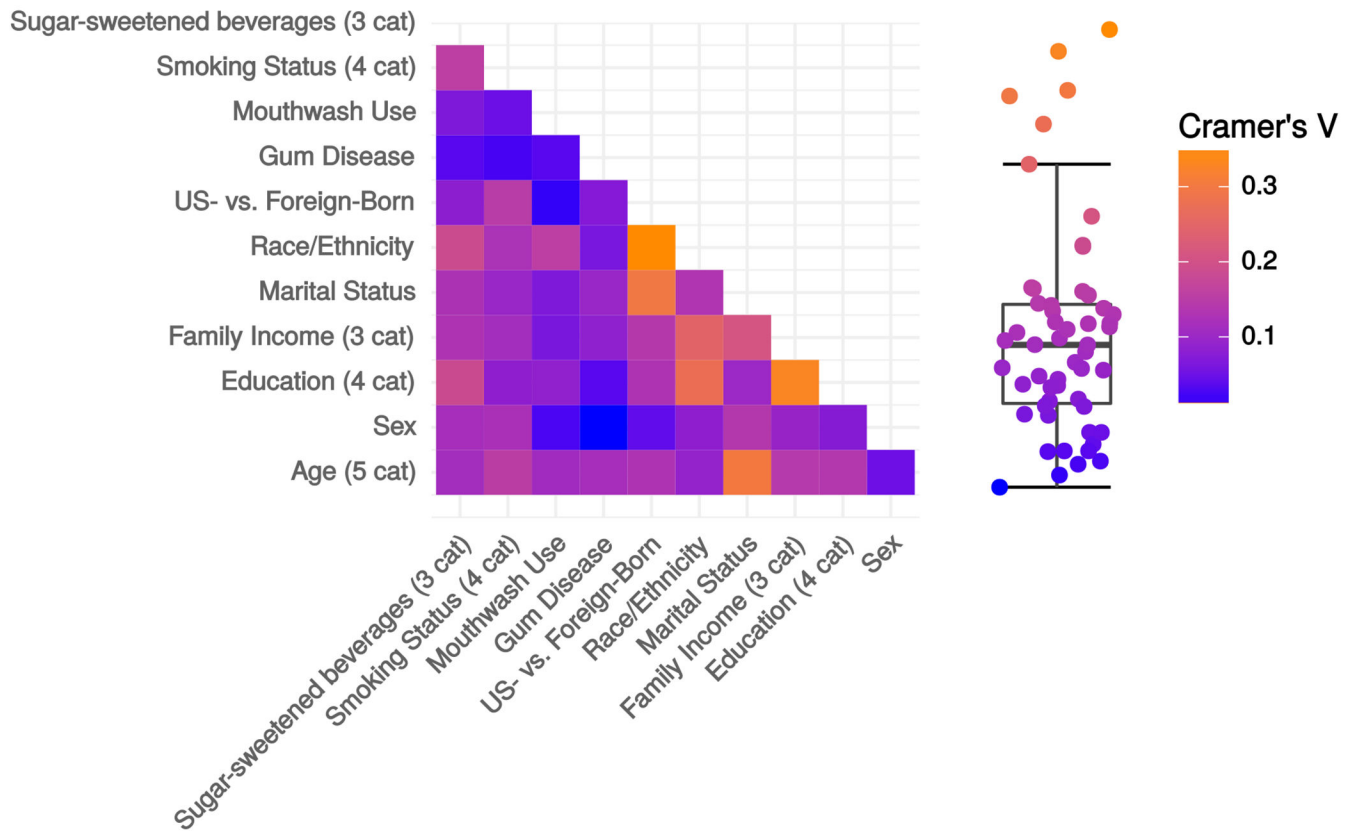


Mental Hygiene in co-leading the parent NYC HANES study. In particular, we wish to thank Sharon Perlman, Carolyn Greene, Claudia Chernov, Amado Punsalang, and the many other staff who helped support data collection.

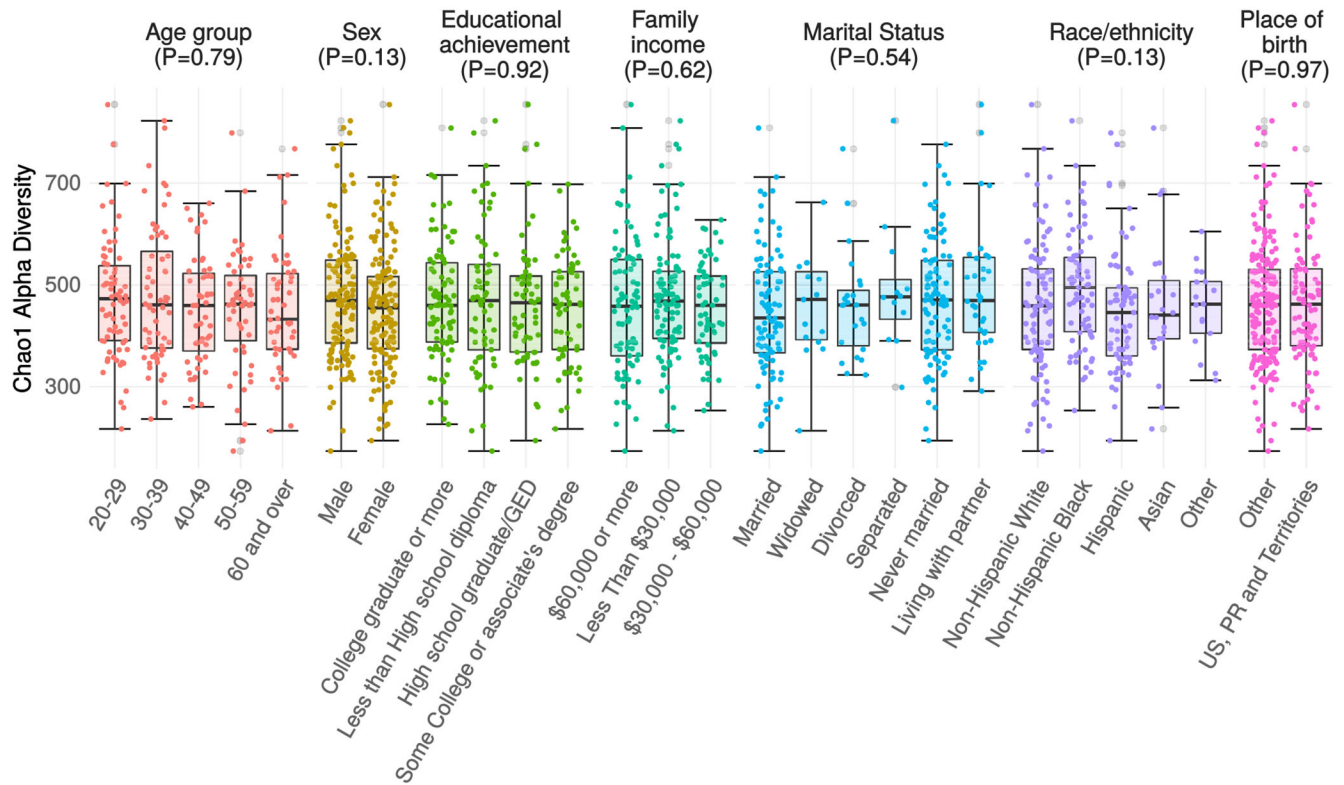
#### 9 Funding

This study was supported by internal funds at the CUNY School of Public Health and Albert Einstein College of Medicine with salary support (JBD, AR, LW) from National Institute of Allergy and Infectious Diseases (1R21AI121784-01).

## Appendix

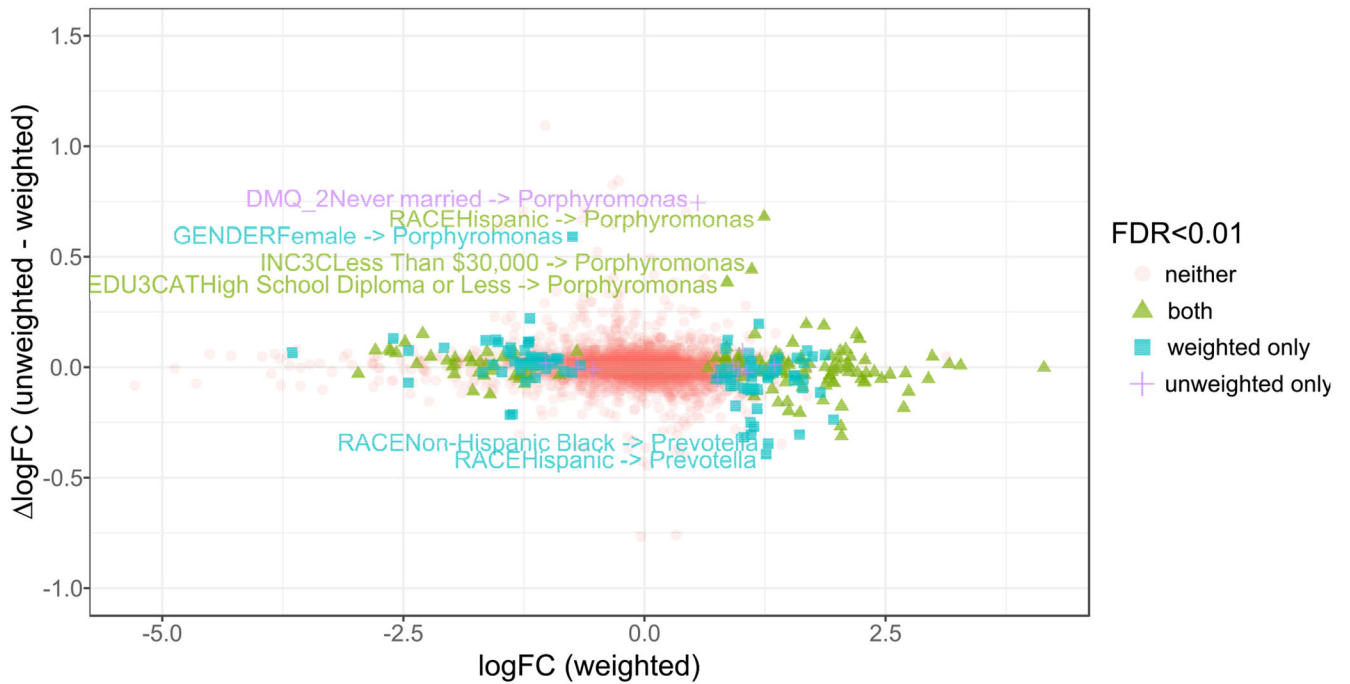


**Figure A1.** Examining collinearity among sociodemographic variables. Data are absolute value of pairwise Cramer’s V correlation coefficient between sociodemographic factor levels. Data are from the full sample (n=1,527) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: cat=categories; US=United States.



**Figure A2.**

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 ( $n=282$ ) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: GED=General equivalency diploma; PR=Puerto Rico; US=United States.



**Figure A3.**

Comparison of log fold change (logFC) estimates between crude unweighted models and models weighted for inverse probability of selection conditional on self-reported smoking status, logarithm cotinine, and their interaction. Differences in logFC estimates between unweighted and weighted models on the y-axis represent an approximation of the bias due to selection on smoking. Estimates are overall fairly concordant, with nearly all (99%) of OTU-variable pairs having an absolute difference in point estimate less than 0.35. Very few ( $n=10$ ) hypotheses that were significant ( $FDR < 0.01$ ) in unweighted analysis were nonsignificant in weighted analysis. Of these, 9/10 had nearly identical point estimates but larger variance in the weighted models. In contrast, a large number of hypothesis tests that were nonsignificant in unweighted analysis were significant in weighted analysis. Specifically, weighting by selection for smoking identified 10 new significant OTUs for gender, 13 for age, 24 for education, 10 for income, 13 for marital status, 26 for race, and 8 for nativity. Where the two models disagreed on significance tests, the vast majority of disagreements were characterized by significance in the weighted model and nonsignificance in the unweighted model. Furthermore, the point estimates from the weighted models were more often further from the null than unweighted models.

### List of abbreviations

<b>SES</b>	socioeconomic status
<b>CHD</b>	coronary heart disease
<b>CVD</b>	cardiovascular disease
<b>NYC HANES</b>	New York City Health and Nutrition Examination Survey

<b>OTU</b>	operational taxonomic unit
<b>FDR</b>	false discovery rate
<b>DA</b>	differential abundance
<b>PS</b>	prediction strength
<b>logFC</b>	log fold change
<b>HMP</b>	Human Microbiome Project

## 10 References

- 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. [PubMed: 25234710]
- Karlamangla AS, Merkin SS, Crimmins EM, Seeman TE. Socioeconomic and ethnic disparities in cardiovascular risk in the United States, 2001–2006. *Annals of epidemiology*. 2010;20(8):617–28. [PubMed: 20609342]
- 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. [PubMed: 27855140]
- 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. [PubMed: 18000122]
- Aiello AE, Dowd JB. *Socio-economic Status and Immunosenescence* Immunosenescence: Springer, New York, NY; 2013 p. 145–57.
- McBurney CA, Vina ER. Racial and ethnic disparities in rheumatoid arthritis. *Curr Rheumatol Rep* 2012;14(5):463–71. [PubMed: 22773376]
- Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiota to periodontal health and disease. *Cell Host Microbe*. 2011;10(4):302–6. [PubMed: 22018230]
- Darveau RP. Periodontitis: a polymicrobial disruption of host homeostasis. *Nat Rev Microbiol* 2010;8(7):481–90. [PubMed: 20514045]
- 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. [PubMed: 22514251]
- Pussinen PJ, Alftan G, Jousilahti P, Paju S, Tuomilehto J. Systemic exposure to *Porphyromonas gingivalis* predicts incident stroke. *Atherosclerosis*. 2007;193(1):222–8. [PubMed: 16872615]
- Pussinen PJ, Jousilahti P, Alftan 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. [PubMed: 12714435]
- Gurav A, Jadhav V. Periodontitis and risk of diabetes mellitus. *J Diabetes*. 2011;3(1):21–8. [PubMed: 20923503]
- Lalla E, Papapanou PN. Diabetes mellitus and periodontitis: a tale of two common interrelated diseases. *Nat Rev Endocrinol* 2011;7(12):738–48. [PubMed: 21709707]
- Atanasova KR, Yilmaz Ö. Prelude to oral microbes and chronic diseases: past, present and future. *Microbes Infect* 2015;17(7):473–83. [PubMed: 25813714]
- 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. [PubMed: 18033835]

16. 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.
17. Ahn J, Segers S, Hayes RB. Periodontal disease, *Porphyromonas gingivalis* serum antibody levels and orodigestive cancer mortality. *Carcinogenesis*. 2012;33(5):1055–8. [PubMed: 22367402]
18. Buduneli N, Baylas H, Buduneli E, Türko lu 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. [PubMed: 15691348]
19. 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. [PubMed: 20123706]
20. 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. [PubMed: 23455329]
21. Ogrendik M Rheumatoid arthritis is an autoimmune disease caused by periodontal pathogens. *Int J Gen Med* 2013;6:383–6. [PubMed: 23737674]
22. Findley K, Williams DR, Grice EA, Bonham VL. Health Disparities and the Microbiome. *Trends Microbiol* 2016;24(11):847–50. [PubMed: 27712950]
23. Dowd JB, Renson A. “Under the Skin” and into the Gut: Social Epidemiology of the Microbiome. *Current Epidemiology Reports*. 2018:1–10. [PubMed: 29503792]
24. 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. [PubMed: 12883111]
25. 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. [PubMed: 27797671]
26. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. *The ISME journal*. 2016;10(10):2435–46. [PubMed: 27015003]
27. 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. [PubMed: 18179393]
28. 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. [PubMed: 17634300]
29. 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. [PubMed: 25170151]
30. 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. [PubMed: 23165986]
31. 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. [PubMed: 22698087]
32. Human Microbiome Project C Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207–14. [PubMed: 22699609]
33. 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. [PubMed: 24194878]
34. 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. [PubMed: 27357925]
35. Craig RG, Boylan R, Yip J, Bamgboye P, Koutsoukos J, Mijares D, et al. Prevalence and risk indicators for destructive periodontal diseases in 3 urban American minority populations. *Journal of clinical periodontology*. 2001;28(6):524–35. [PubMed: 11350519]

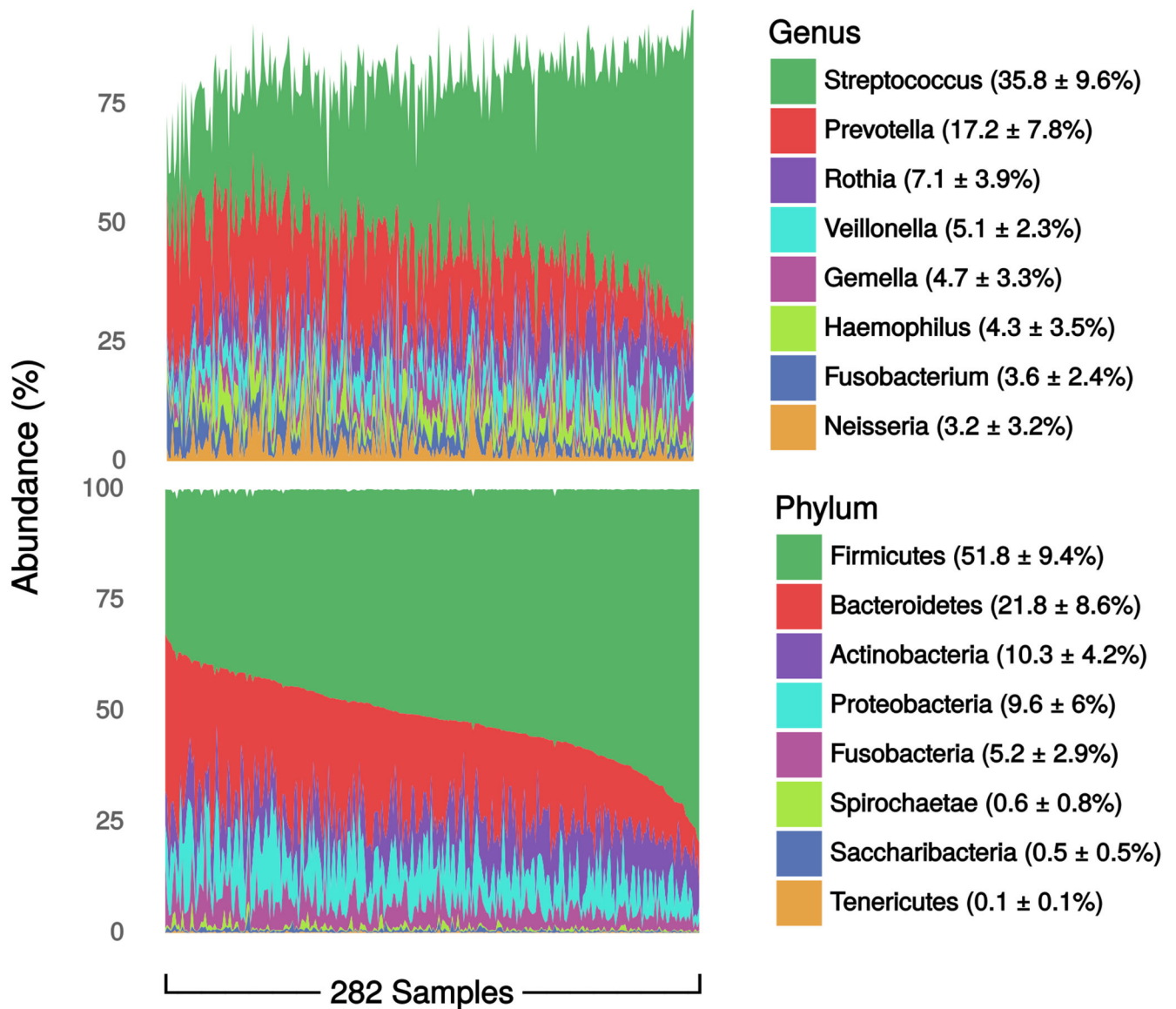


36. Schenkein HA, Burmeister JA, Koertge TE, Brooks CN, Best AM, Moore LV, et al. The influence of race and gender on periodontal microflora. *J Periodontol* 1993;64(4):292–6. [PubMed: 8387107]
37. Beck JD, Koch GG, Zambon JJ, Genco RJ, Tudor GE. Evaluation of oral bacteria as risk indicators for periodontitis in older adults. *J Periodontol* 1992;63(2):93–9. [PubMed: 1313107]
38. 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.
39. Thorpe LE, Greene C, Freeman A, Snell E, Rodriguez-Lopez JS, Frankel M, 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. [PubMed: 26844121]
40. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, 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. [PubMed: 20534432]
41. 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).
42. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. *Nature methods*. 2008;5(3):235–7. [PubMed: 18264105]
43. 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.
44. 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. [PubMed: 19910308]
45. Chao A Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*. 1984;265–70.
46. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral ecology*. 2001;26(1):32–46.
47. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*. 2005;71(12):8228–35. [PubMed: 16332807]
48. 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. [PubMed: 23326225]
49. Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M. Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences*. 2012;1248(1):107–23. [PubMed: 22092035]
50. Garrett BE, Dube SR, Babb S, McAfee T. Addressing the social determinants of health to reduce tobacco-related disparities. *Nicotine & Tobacco Research*. 2014;17(8):892–7. [PubMed: 25516538]
51. Yu G, Phillips S, Gail MH, Goedert JJ, Humphrys MS, Ravel J, et al. The effect of cigarette smoking on the oral and nasal microbiota. *Microbiome*. 2017;5(1):3. [PubMed: 28095925]
52. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. *The ISME journal*. 2016.
53. Team RC. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria 2017.
54. Takahashi N, Nyvad B. The role of bacteria in the caries process: ecological perspectives. *J Dent Res* 2011;90(3):294–303. [PubMed: 20924061]
55. 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. [PubMed: 21716312]
56. 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. [PubMed: 17446294]
57. Zijngje 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. [PubMed: 20195365]
58. 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. [PubMed: 19799718]
59. 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. [PubMed: 24599709]
60. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut*. 2016;65(12):1973–80. [PubMed: 26311717]
61. Han YW. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Curr Opin Microbiol* 2015;23:141–7. [PubMed: 25576662]
62. Hajishengallis G Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* 2015;15(1):30–44. [PubMed: 25534621]
63. 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. [PubMed: 24506890]
64. Hansen TH, Kern T, Bak EG, Kashani A, Allin KH, Nielsen T, et al. Impact of a vegan diet on the human salivary microbiota. *Scientific reports*. 2018;8(1):5847. [PubMed: 29643500]
65. 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. [PubMed: 23642116]
66. Herd P, Palloni A, Rey F, Dowd JB. Social and population health science approaches to understand the human microbiome. *Nature Human Behaviour*. 2018:1.

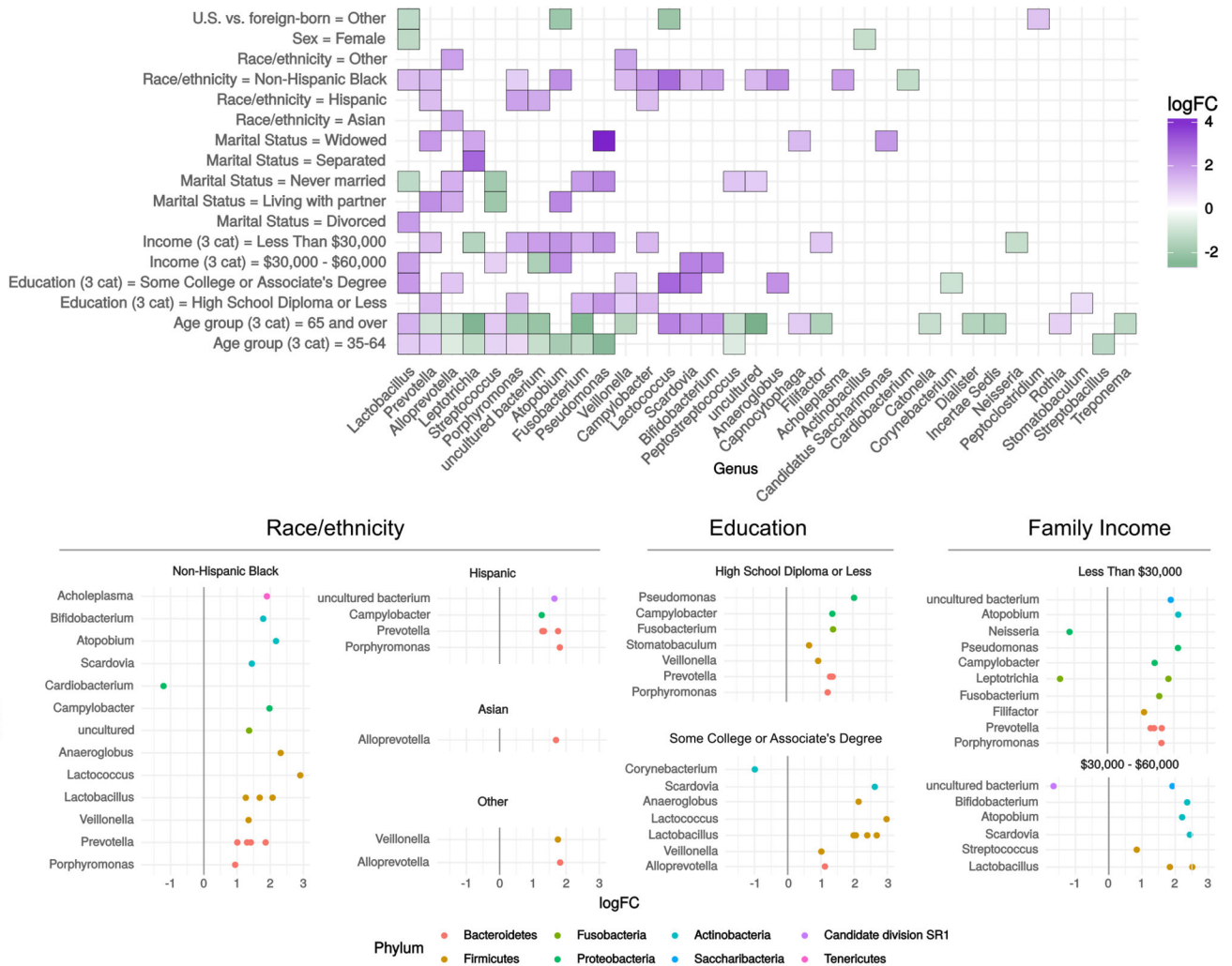
### Highlights

- Oral microbiome studies to date have had limited sociodemographic variability
- We examined the oral microbiome in a subsample of a diverse population-based sample
- Numerous taxa were differentially abundant by every sociodemographic variable
- Differentially abundant taxa included *Porphyromonas*, *Fusobacterium*, and *Prevotella*
- Many differentially abundant taxa are associated with oral and systemic disease

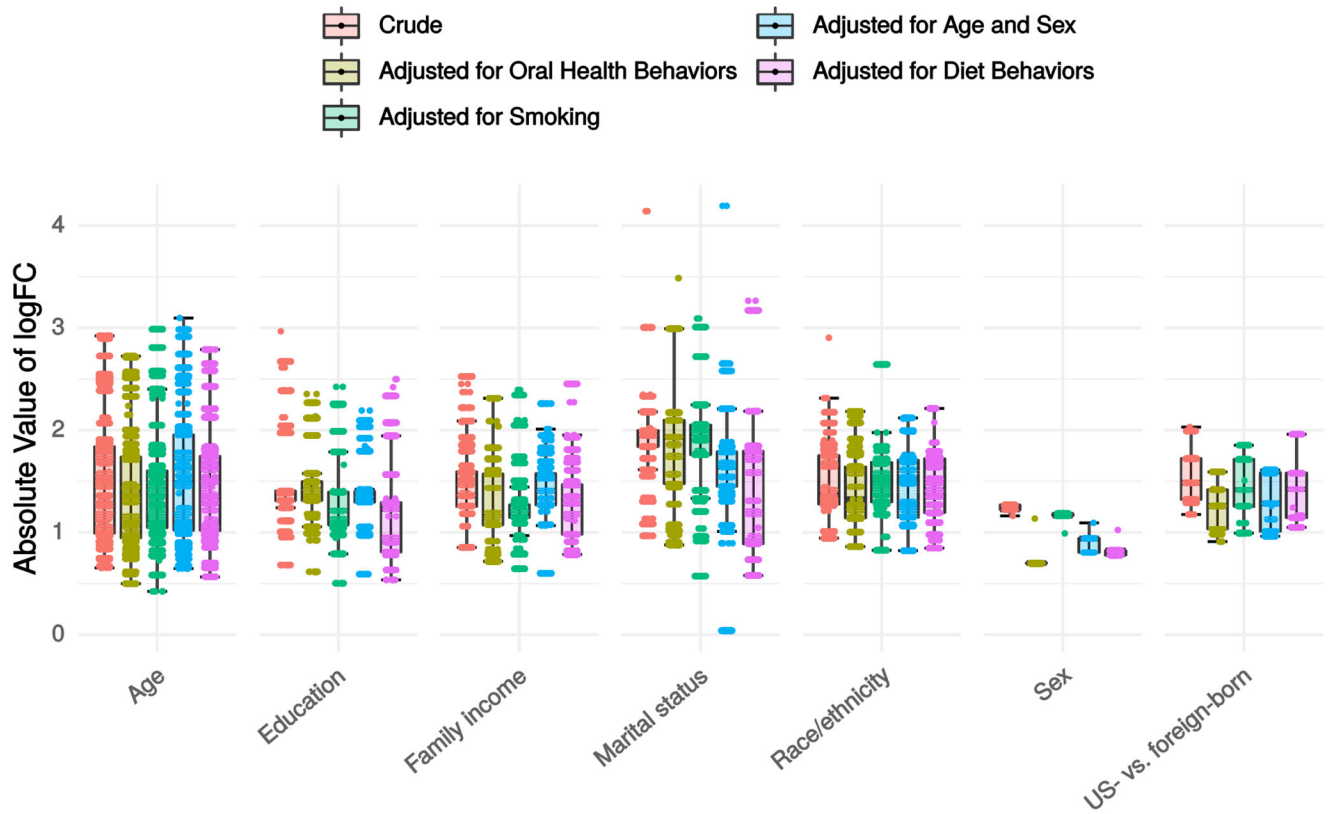


**Figure 1.**

Genus- and phylum-level relative abundances. Data are percent of overall communities within samples, summarized as mean  $\pm$  standard deviation of percent across samples. Data are from the oral microbiome subsample (n=282) of the New York City Health and Nutrition Examination Survey, 2013-2014.

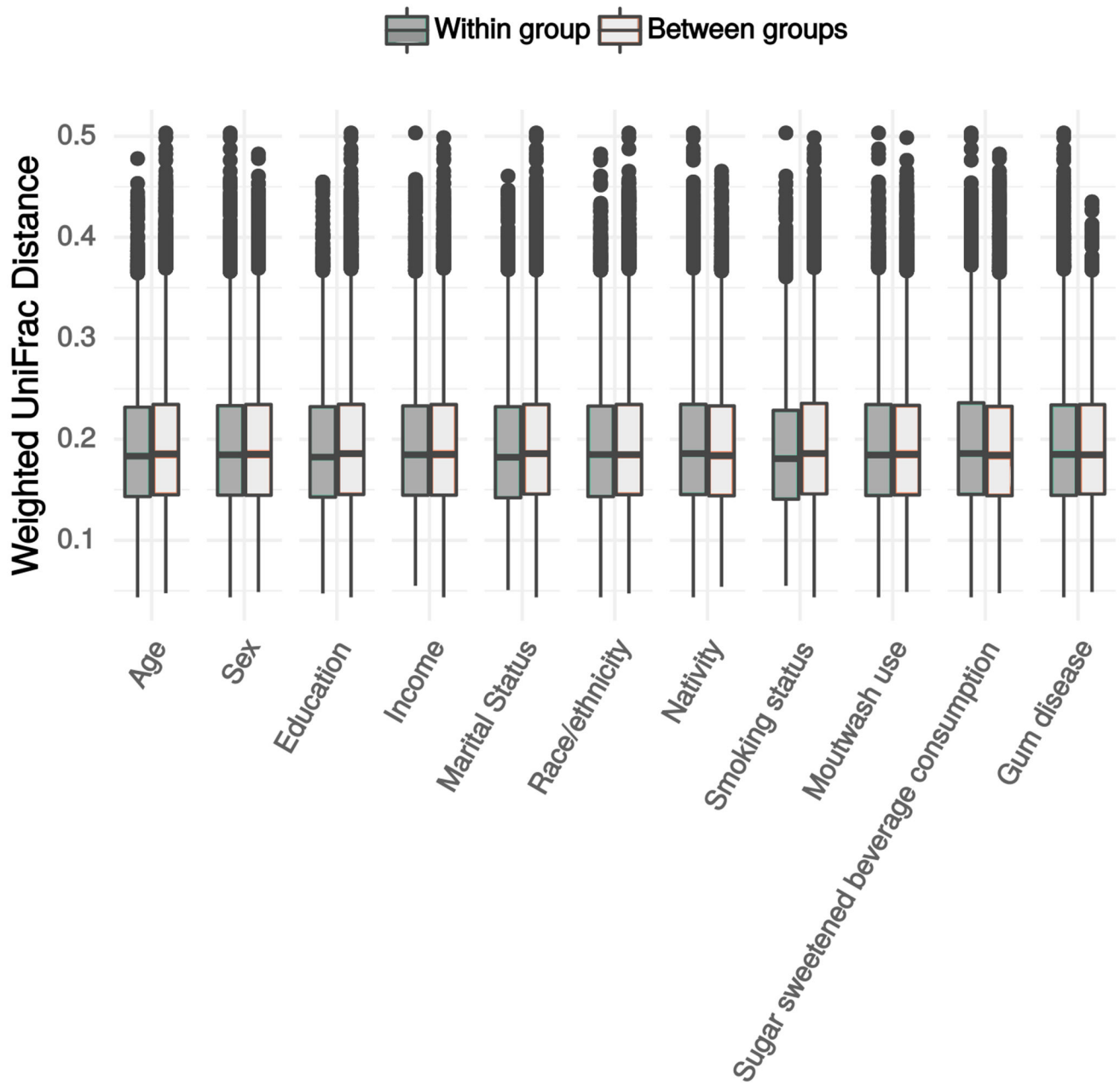


**Figure 2.** Differential abundance by sociodemographic characteristics. OTUs meeting unadjusted FDR < 0.01 in negative binomial log-linear GLMs using edgeR. Data are from the oral microbiome subsample (n=282) of the New York City Health and Nutrition Examination Survey, 2013-2014. Filled tiles in (A) 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 in (A). Reference groups for sociodemographic variables are as follows: Sex: Male, Age: 20-34, Education: College Graduate or More, Family income: \$60,000 or more, Marital status: Married, Race/ethnicity: Non-Hispanic White, US- vs. foreign-born: US-Born, 50 States, DC, PR and Territories. Abbreviations: cat=categories; GLM=generalized linear model; logFC=log fold change; OTU=operational taxonomic unit; US=United States.



**Figure 3.** 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=282) 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 4.** Within and between group beta diversity estimate distributions. Data are from the oral microbiome subsample (n=282) of the New York City Health and Nutrition Examination Survey, 2013-2014. Abbreviations: cat=category.

**Table 1.**

## Demographics

	Oral Microbiome Subsample	Full NYC HANES Sample
Total	282	1527
Age in years – median [range]	42 [20 to 94]	42 [20 to 97]
Age group (%)		
20-29	70 (24.8)	360 (23.6)
30-39	60 (21.3)	337 (22.1)
40-49	51 (18.1)	252 (16.5)
50-59	51 (18.1)	264 (17.3)
60 and over	50 (17.7)	314 (20.6)
Sex = Female (%)	150 (53.2)	885 (58.0)
Educational achievement (%)		
College graduate or more	87 (30.9)	628 (41.1)
Less than High school diploma	65 (23.0)	316 (20.7)
High school graduate/GED	63 (22.3)	244 (16.0)
Some College or associate's degree	67 (23.8)	337 (22.1)
<i>Missing</i>	0 (0.0)	2 (0.1)
Annual family income (%)		
\$60,000 or more	82 (29.1)	429 (28.1)
Less Than \$30,000	105 (37.2)	537 (35.2)
\$30,000 - \$60,000	59 (20.9)	348 (22.8)
<i>Missing</i>	36 (12.8)	213 (13.9)
Marital Status (%)		
Married	96 (34.0)	590 (38.6)
Widowed	15 (5.3)	76 (5.0)
Divorced	23 (8.2)	156 (10.2)
Separated	12 (4.3)	51 (3.3)
Never married	101 (35.8)	511 (33.5)
Living with partner	35 (12.4)	143 (9.4)
Race/ethnicity (%)		
Non-Hispanic White	97 (34.4)	513 (33.6)
Non-Hispanic Black	75 (26.6)	340 (22.3)
Hispanic	71 (25.2)	390 (25.5)
Asian	22 (7.8)	204 (13.4)
Other	17 (6.0)	80 (5.2)
Place of birth (%)		
US, PR and Territories	90 (31.9)	668 (43.7)
Other	190 (67.4)	851 (55.7)
<i>Missing</i>	2 (0.7)	8 (0.5)

	Oral Microbiome Subsample	Full NYC HANES Sample
Gum disease (self-reported) (%)		
Yes	27 (9.6)	175 (11.5)
No	254 (90.1)	1322 (86.6)
<i>Missing</i>	1 (0.4)	30 (2.0)
Mouthwash use (times per week) (%)		
None	115 (40.8)	591 (38.7)
1 to 5	68 (24.1)	370 (24.2)
6 to 7	99 (35.1)	565 (37.0)
<i>Missing</i>	0 (0.0)	1 (0.1)
Sugar-sweetened beverages (per week) (%)		
0-<1	152 (53.9)	985 (64.5)
1-5	67 (23.8)	313 (20.5)
6 or more	62 (22.0)	227 (14.9)
<i>Missing</i>	1 (0.4)	2 (0.1)
Smoking status (%)		
Cigarette	86 (30.5)	215 (14.1)
Never smoker	43 (15.2)	843 (55.2)
Former smoker	43 (15.2)	285 (18.7)
Alternative smoker	72 (25.5)	142 (9.3)
Secondhand	38 (13.5)	42 (2.8)

**Table 2.**

Differential abundance findings for OTUs selected based on clinical relevance. Data are from the oral microbiome subsample (n=282) of the New York City Health and Nutrition Examination Survey, 2013-2014.

	<u>Lactobacillus</u>		<u>Prevotella</u>		<u>Streptococcus</u>		<u>Porphyromonas</u>		<u>Fusobacterium</u>		<u>Lactococcus</u>	
	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR
<b>Race/ethnicity</b>												
Non-Hispanic White	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref
Non-Hispanic Black	2.1	<0.0001	1.9	<0.0001	0.6	0.01	0.9	0.002	0.4	0.5	2.9	<0.0001
Hispanic	0.9	0.1	1.8	<0.0001	0.4	0.1	1.8	<0.0001	1.1	0.04	1.8	0.02
<b>Family income</b>												
\$60,000 or more	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref
\$30,000 - \$60,000	2.5	<0.0001	1.2	0.03	0.9	0.003	0.6	0.2	1.8	0.03	1.7	0.06
Less Than \$30,000	1.1	0.06	1.6	0.003	1.1	0.03	1.6	<0.0001	1.5	0.003	-1.1	0.1
<b>Education</b>												
College Graduate or More	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref	0.	Ref
Some College or Associate's Degree	2.7	<0.0001	0.8	0.1	0.5	0.1	0.8	0.03	1.4	0.1	3.0	<0.0001
High School Diploma or Less	1.0	0.04	1.4	0.006	0.6	0.02	1.2	0.0008	1.4	0.006	1.3	0.07

Abbreviations: logFC, log fold change; FDR, false discovery rate; Ref, reference group.