Tobacco exposure associated with oral microbiota oxygen utilization in the New York City Health and Nutrition Examination Study

Francesco Beghini,[a] Audrey Renson,[b] Christine P. Zolnik,[c,d] Ludwig Geistlinger,[b,e] Mykhaylo Usyk,[d] Thomas U. Moody,[d] Lorna Thorpe,[f] Jennifer B. Dowd,[b,g] Robert Burk,[d] Nicola Segata,[a] Heidi E. Jones,[b,e]* Levi Waldron[b,e]*

* equal contribution

a. Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy / Via Sommarive, 9 / 38123, Trento, Italy (francesco.beghini@unitn.it, nicola.segata@unitn.it)
b. City University of New York (CUNY) Graduate School of Public Health and Health Policy / 55 W. 125th St. / New York, NY 10027, United States (audrey.o.renson@gmail.com, Ludwig.Geistlinger@sph.cuny.edu; Jennifer.Dowd@sph.cuny.edu; Heidi.Jones@sph.cuny.edu; Levi.Waldron@sph.cuny.edu)
c. Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA (christine.zolnik@liu.edu, mykhaylo.usyk@einstein.yu.edu, tu.moody@gmail.com, robert.burk@einstein.yu.edu)
d. Department of Biology, Long Island University, 1 University Plaza, Brooklyn, NY 11201 (christine.zolnik@liu.edu)
e. Institute for Implementation Science in Population Health, City University of New York / 55 W. 125th St. / New York, NY 10027, United States (audrey.o.renson@gmail.com, Ludwig.Geistlinger@sph.cuny.edu; Jennifer.Dowd@sph.cuny.edu; Heidi.Jones@sph.cuny.edu; Levi.Waldron@sph.cuny.edu)
f. New York University School of Medicine, 650 1st Ave, New York, NY 10016 (lorna.thorpe@nyumc.org)
g. Department of Global Health and Social Medicine, King's College London, The Strand, London WC2R 2LS, United Kingdom (jennifer.dowd@kcl.ac.uk)

Corresponding author:

Heidi E. Jones, PhD MPH
City University of New York (CUNY) Graduate School of Public Health and Health Policy 55 W. 125th St. New York, NY 10027, United States
Heidi.Jones@sph.cuny.edu

ABSTRACT

Purpose: The effect of tobacco exposure on the oral microbiome has not been established.

Methods: We performed amplicon sequencing of the 16S ribosomal RNA gene V4 variable region to estimate bacterial community characteristics in 259 oral rinse samples, selected based on self-reported smoking and serum cotinine levels, from the 2013-14 New York City Health and Nutrition Examination Study. We identified differentially abundant operational taxonomic units (OTUs) by primary and secondhand tobacco exposure, and employed “microbe set enrichment analysis” to assess shifts in microbial oxygen utilization.

Results: Cigarette smoking was associated with depletion of aerobic OTUs (Enrichment Score test statistic ES = -0.75, p = 0.002) with a minority (29%) of aerobic OTUs enriched in current smokers compared to never smokers. Consistent shifts in the microbiota were observed for current cigarette smokers as for non-smokers with secondhand exposure as measured by serum cotinine levels. Differential abundance findings were similar in crude and adjusted analyses.

Conclusion: Results support a plausible link between tobacco exposure and shifts in the oral microbiome at the population level through three lines of evidence: 1) a shift in microbiota oxygen utilization associated with primary tobacco smoke exposure, 2) consistency of abundance fold-changes associated with current smoking and shifts along the gradient of secondhand smoke exposure among non-smokers, and 3) consistency after adjusting for *a priori* hypothesized confounders.

Key words: microbiota; RNA, Ribosomal, 16S; human microbiome; oral health; tobacco; smoking

Highlights

- Cigarette smoke was associated with microbial anaerobiosis in oral rinse specimens

- The microbiome shifts associated with smoking and secondhand exposure were correlated
- Shifts in oral bacterial oxygen utilization may mediate smoking and health outcomes
- We propose “microbe set enrichment analysis” for interpreting shifts in the microbiome

List of abbreviations

CI - Confidence Interval

CUNY - City University of New York

DNA - Deoxyribonucleic Acid

ES - Enrichment Score

FDR - False Discovery Rate

GSEA - Gene Set Enrichment Analysis

GSVA - Gene Set Variation Analysis

HMP - Human MIcrobiome Project

IRB - Institutional Review Board

NHANES - National Health Nutrition and Examination Survey

NYC DOHMH - New York City Department of Health and Mental Hygiene

NYC HANES - New York City Health Nutrition and Examination Survey

OR - Odds Ratio

ORA - Over-representation Analysis

OTU - Operational Taxonomic Unit

PCOA - Principal Coordinates Analysis

PERMANOVA - Permutational Multivariate Analysis of Variance

RNA - Ribonucleic Acid

rRNA - Ribosomal Ribonucleic Acid

Introduction

Dysbiosis of the dental plaque microbiome is a necessary step in the etiology of periodontitis and caries [1], which have been linked to systemic illness, including cardiovascular diseases [2], type 2 diabetes mellitus [3], obesity [4], low birth weight and preterm birth [5], rheumatoid arthritis [6], chronic obstructive pulmonary disease [7], and oral and digestive cancers [8]. Tobacco exposure is a cause of these outcomes [9–14], but whether it causes them through shifts in the general oral microbiome is unknown [15]. If tobacco smoke causes harmful alterations of the oral microbiome, interventions targeting the oral microbiome could mitigate the impact of tobacco exposures. A key aspect of making this distinction lies in establishing whether a range of tobacco exposures, including cigarette smoking, secondhand smoke exposure, hookah and e-cigarette use, cause substantial changes in the structure and function of the general oral microbiome.

Evidence suggests that tobacco smoke exposure causes alterations to the oral microbiome, selecting a community enriched with opportunistic pathogens [16,17] and negatively impacting the resilience and colonization resistance of the sub- and supragingival biofilms [18]. Such alterations may occur directly due to selective toxicity [19], or indirectly via alteration of the host immune system to produce both pro- and anti-inflammatory effects [20–22] which alter the oral biofilm and mucosal microbial habitats. Another potential mechanism by which tobacco smoke reconfigures the oral microbiome is via depletion of oxygen [23], creating a hypoxic oral environment that favors anaerobiosis. Tobacco smoke may also favor anaerobiosis by increasing the amount of free iron [24], and inhibiting oral peroxidase [25]. Anaerobic glycolysis in human salivary cells has been shown to dramatically increase after exposure to tobacco smoke [26], and human experiments show reduction in periodontal pocket oxygen tension [27] and redox potential [23] after smoking cigarettes. Low throughput studies of the oral microbiome have shown greater abundance of the anaerobes *Prevotella intermedia [28]* and *Lactobacillus spp*. [29] in cigarette smokers. Ad hoc findings from high throughput studies have suggested that smokers have greater abundance of anaerobic

microorganisms [30] and depletion of microbial functional pathways related to aerobic respiration [15]. Thus tobacco exposure could plausibly cause changes to the oral microbiome, but available results are limited to laboratory and small-sample studies.

This study suggests a causal link between tobacco exposure and alterations to the saliva microbiome among participants of the 2013-14 New York City Health and Nutrition Examination Study (NYC HANES). It contrasts current smokers to non-smokers with no recent secondhand exposure, and investigates a dose-response relationship among non-smokers with varying degrees of secondhand exposure assessed by quantitative serum cotinine level. It further investigates former smokers and smokers of e-cigarettes and hookah. Multiple lines of causal inference are used to test the hypothesis that tobacco smoke alters the saliva microbiome: controlling for hypothesized confounders, testing for a dose-response relationship, and testing for altered oxygen requirements of the microbial communities associated with tobacco exposure.

Materials and Methods

2013-2014 NYC-HANES

Data for the current study are sub-sampled from the 2013-14 NYC HANES, a population-based study of 1,575 non-institutionalized adults in New York City [31] modeled after the United States Centers for Disease Control and Prevention's National Health and Nutrition Examination Survey (NHANES) [32]. The NYC HANES sample was recruited using a three stage cluster household probabilistic design of all non-institutionalized adults 18 years of age or older. Consenting individuals provided information on smoking, including use of alternative tobacco products such as e-cigarettes and hookahs in the last 5 days, socio-demographic characteristics, and oral hygiene practices through face-to-face interviewing and audio computer assisted interviewing for sensitive questions. Participants also underwent physical examination and provided blood and oral rinse specimens for biomarker analysis. Serum specimens were analyzed for cotinine by liquid chromatography/tandem mass spectrometry [33].

The study was conducted by the City University of New York (CUNY) School of Public Health in collaboration with the New York City Department of Health and Mental Hygiene (NYC DOHMH), with ethical approval from their respective institutional review boards (IRBs). The current sub-study received separate IRB approval from the CUNY School of Public Health.

Tobacco exposure outcome measures and selection of sub-sample

We selected a sub-sample of 297 participants for oral microbiome assessment based on self-reported tobacco use and serum blood cotinine level, classified into five mutually exclusive groups:

- *Current smokers* (n=90) were selected from participants who reported smoking more than 100 cigarettes in their lifetime, smoking a cigarette in the last 5 days, and not using any alternative tobacco product in the last 5 days (the 90 with highest measured serum cotinine were selected).
- *Never smokers* (n=45) were randomly selected from those reporting lifetime smoking of less than 100 cigarettes, no usage of any tobacco product in the last 5 days, and serum cotinine level less than 0.05 ng/mL.
- *Former smokers* (n=45) were randomly selected from participants reporting lifetime smoking of more than 100 cigarettes, but currently not smoking, no use of any tobacco product in the last 5 days, and serum cotinine level less than 0.05 ng/mL.
- *Non-smokers with secondhand exposure* (n=38) comprised all available former or never smokers with serum cotinine between 1 and 14 ng/mL [34]. This group includes four individuals with serum cotinine > 10 ng/mL [35], who may have been light smokers; sensitivity analyses excluding these four had negligible effect (not shown, included in Reproducible Analysis code), so they are included in the secondhand exposure group.
- *"Alternative" smokers* (n=79) were participants with self-reported usage of hookah, cigar, cigarillo and/or e-cigarette in the last 5 days.

For quality control, 5% of samples were randomly selected and sequenced as technical replicates. Results from replicates were used instead of the original sample if the sequencing read count was greater than the original. An additional eight samples failed PCR amplification and were repeated. Fifteen specimens (n=4 current cigarette smokers, n=2 never smokers, n=2 former smokers, n=7 alternative smokers) were discarded for sequencing quality control (below). We excluded an additional 23 participants from the alternative smoker group who also reported smoking cigarettes in the last 5 days, in an attempt to isolate the effect of alternative tobacco exposures, resulting in an analytic sample of 259 participants.

Specimen collection, processing, and sequence analysis

Specimen collection, processing, and sequence analysis methods are described in detail in a companion paper [36]. In brief, participants were asked to fast for 9 hours prior to oral rinse and blood specimen collection. A 20-second oral rinse was divided into two 5-second swish and 5-second gargle sessions using 15 mLs of Scope® mouthwash, which was transported on dry ice and stored at -80˚C. A modified protocol of the QIAamp DNA mini kit (QIAGEN) was used for DNA extraction [36]. DNA was amplified for the V4 variable region of the 16S rRNA gene [37,38]. High-throughput amplicon sequencing was conducted on a MiSeq (Illumina, San Diego, CA) using 2x300 paired-end fragments. 16S read analysis was carried out using QIIME version 1.9.1 [39] and Phyloseq [40]. Paired-end reads were merged with fastq-join [41] and resulting low quality reads (PHRED score < 30) were discarded when joining the split reads (qiime split_libraries_fastq.py). Operational Taxonomic Unit (OTU) picking was performed with an open reference approach by clustering using UCLUST at 97% sequence similarity and taxonomy was assigned using the SILVA 123 [42] database as reference. The QIIME generated OTU table was converted for phyloseq processing. Samples with fewer than 1000 reads (n=15) were removed from the OTU table in the phyloseq preprocessing step. Genera present with a mean relative abundance of less than 2×10^{-4} were collapsed as “Other.”

Unsupervised clustering

We explored differences in beta diversity measures via Principal Coordinates Analysis (PCoA) plots on Weighted UniFrac distances. The grouping of distances by smoking status was tested by PERMANOVA as implemented in the 'vegan' package [43], with 999 permutations.

Oral microbiome measures

We compared oral bacterial community characteristics by tobacco exposure group using four types of oral microbiome measures: 1. alpha (within-sample) diversity of the OTUs present; 2. beta (between-sample) diversity of OTUs; 3. OTU counts at the genus level; 4. enrichment of differential abundance categorized by oxygen requirement. For diversity measures, we estimated Chao1 Index, Shannon Index and observed OTUs, and weighted UniFrac [44] beta diversity using the estimate_richness and distance methods of the phyloseq Bioconductor package [40].

Differential abundance analysis

We performed crude and adjusted negative binomial log-linear regression of tobacco exposure group to identify differentially abundant OTUs using edgeR [45] Bioconductor package. Low-prevalence OTUs, those without 3 or more reads observed in at least 8 samples, were discarded. For adjusted models, *a priori* hypothesized confounders included age, sex, race, self-reported physical activity, education, diabetes status (based on serum HbA1c), and self-reported gum disease. Education [46,47], age, and sex [48–50] are known to be associated with smoking and could plausibly be associated with oral microbiome characteristics. Race was also treated as a possible confounder as studies suggest differences in nicotine metabolism by race/ethnicity [51]. A False Discovery Rate [FDR, 52] less than 0.05 was considered statistically significant. Results from edgeR were compared to results obtained from the application of DESeq2 [53]. Crude and adjusted coefficients were compared to assess which hypothesized confounders had the greatest effect on adjusted analyses.

Microbe set enrichment analysis for oxygen requirements

We categorized genera as aerobic, anaerobic, or facultative anaerobic [54] integrating information from the IMG/MER database [55] and from the Bergey's Manual of Systematics of Archaea and Bacteria [54]. This resulted in three “microbe sets” of OTUs with common oxygen requirements. We applied two concurrent approaches to analyze whether the three microbe sets show coherent changes in abundance of the contained microbes for (i) smokers vs. never smokers (with no recent secondhand smoke exposure), and (ii) among non-smokers with exposure to secondhand smoke. First, over-representation of differentially abundant OTUs in each microbe set was tested based on the hypergeometric distribution (corresponds to a one-sided Fisher’s exact test). Second, Gene Set Enrichment Analysis [GSEA, 56] was used to test whether microbes of a particular microbe set accumulate at the top or bottom of the full OTU vector ordered by direction and magnitude of abundance change between the tested sample groups. Over-representation analysis (ORA) and GSEA were applied as implemented in the EnrichmentBrowser R/Bioconductor package [57]. Application of GSEA incorporated the voom-transformation [58] of OTU counts to concur with GSEA’s assumption of roughly normally distributed data. As the implementations of GSEA and ORA required a binary outcome, serum cotinine levels were binned to contrast the upper tertile (> 4.42 ng/ml) against the lower tertile (< 1.76 ng/ml). We also analyzed serum cotinine level as a continuous measure using Gene Set Variation Analysis [GSVA, 59].

Statement of reproducible research

Analyses were performed in QIIME version 1.9.1 and R version 3.5.1. All results presented in this manuscript are reproducible by installing the package and compiling its associated vignettes provided at https://github.com/waldronlab/nychanesmicrobiome.

Results

A total of 1.4 M reads (mean±sd: 4,758±3,463 reads/sample) were generated from 297 saliva mouthwash specimens [31] of NYC-HANES participants selected based on questionnaire and serum cotinine levels (Table 1, with serum cotinine levels by exposure group shown in Supplementary Figure 1). After quality control and filtering, we retained 91.7% of reads (5,007 mean, 3,491 s.d), which were then classified using the QIIME pipeline [39] into 1291 OTUs with more than 10 reads.

Taxonomic composition of the final analytic sample (n=259) was predominated by *Streptococcus* (36% average relative abundance) and *Prevotella* (17% average relative abundance), which were present in every sample. Other genera commonly associated with the oral cavity like *Rothia, Neisseria, Veillonella* and *Gemella* were also found with average relative abundances less than 10%.

Alpha and beta diversity of the oral microbiome by tobacco exposure

Alpha diversities were not significantly different between the five tobacco exposure groups (Shannon Index p=0.95, Observed OTUs p=0.08, Chao1 Index p=0.26 ANOVA test, Supplementary Figure 2). However, beta diversity differed between current cigarette smokers and never smokers, and was larger than differences by race/ethnicity, age, and other sociodemographic measures (Table 2). The overall microbiome composition and structure in these two classes differed and beta diversity was significantly explained by smoking status (R^2=0.051, p<0.001, PERMANOVA test, Figure 1). Additionally, former smokers were significantly different from current smokers (p=0.001, R^2=0.044, PERMANOVA test), but not from never smokers (p=0.16, R^2=0.018, PERMANOVA test). Within former smokers, we found no evidence of differences between those who quit recently versus longer ago (Supplementary Figure 3).

Proteobacteria **less abundant in the microbiome of smokers**

In crude analyses, 46 differentially abundant OTUs, taxonomically assigned to 28 different genera, were identified between current cigarette smokers and never smokers (Supplementary Figure 4). Relative abundance of OTUs annotated as phylum Proteobacteria (*Neisseria*, *Lautropia*, *Haemophilus* and *Actinobacillus*) and Candidate division SR1 were found to be lower in current cigarette smokers compared to never smokers (Proteobacteria phylum t-test p-value=5e-07, logFC=-0.84, Supplementary Figure 5).

Adjusted differential abundance analysis, accounting for hypothesized confounders, identified fewer (n=21) differentially abundant OTUs between current and never smokers (Figure 2). The phylum Proteobacteria was still identified as less-abundant in current smokers in the adjusted model (t-test p-value=8e-07, logFC=-0.85). Adjusted coefficients were slightly attenuated toward the null compared to crude estimates (Figure 3). Addition of one hypothesized confounder at a time showed that age and education had the strongest impact, resulting in a median decrease in coefficient magnitude of 2 and 3 percent, respectively.

Differences in oxygen utilization in the oral microbiome of smokers compared to never smokers

We functionally annotated the entire set of picked OTUs according to their oxygen requirement: 78 aerobic OTUs, 673 anaerobic OTUs and 395 facultative anaerobic OTUs. We failed to annotate 145 OTUs because their genera was annotated as uncultured bacteria or taxonomic resolution was higher than genus.

A minority of aerobic OTUs (29%) and a majority of anaerobic OTUs (60%) had higher mean abundance in current smokers as compared to never smokers. Facultative anaerobic OTUs were approximately evenly divided, with 51% having higher abundance in current smokers. We accordingly found differentially abundant OTUs between current smokers and never smokers to be over-represented in aerobic OTUs (Hypergeometric test, p = 0.004). Using Gene Set Enrichment Analysis (GSEA) to account for collinearity between OTUs and the direction of the abundance change (up / down), aerobic OTUs

were significantly depleted among current smokers relative to never smokers (Enrichment Score test statistic ES = -0.75, p = 0.002, GSEA permutation test). Anaerobic OTUs were enriched in smokers relative to never smokers but the difference was not statistically significant (ES = 0.36, p = 0.14, GSEA permutation test). We also observed an enrichment of facultative anaerobic OTUs among never smokers compared to current smokers but this result was not statistically significant (ES = -0.29, p = 0.48, GSEA permutation test).

Comparison of those with secondhand smoke exposure to never smokers

To provide independent evidence for causal inference, we compared the coefficients estimated for the contrast of current smokers versus never smokers to the coefficients for serum cotinine level, estimated from a non-overlapping group of self-reported non-smokers (n=38) exposed to secondhand smoke. Consistency was assessed by calculating the Pearson Correlation of the two vectors of coefficients. This correlation is comparable to the Integrative Correlation Coefficient, which was originally proposed to assess the replicability of measurements from independent gene expression studies [60,61]. This correlation was estimated from the intersection of 121 OTUs passing the edgeR filter against low-variance features for both groups. Full differential abundance results for these 121 OTUs are reported in Supplemental File 1. We observed a positive correlation (Figure 4, Pearson's Correlation = 0.40, p = 5e-6); this correlation was slightly attenuated after adjusted for age and educational attainment (Supplementary Figure 6, Pearson's Correlation=0.28, p=0.002). This positive correlation identifies a similarity in the patterns of differential abundance in smokers vs. never smokers when compared to the shifts associated with increasing exposure to secondhand smoke. However, the application of three concurrent approaches for microbe set enrichment analysis (GSEA, GSVA, and ORA) on samples from participants exposed to secondhand smoke, with continuous or dichotomous serum cotinine levels as the response variable, did not identify significant enrichment or depletion of aerobiosis or anaerobiosis, reflecting smaller shifts associated with secondhand smoke exposure.

Comparison of alternative tobacco exposures to never smokers

Differential abundance of OTUs from participants who used e-cigarettes, hookah, and/or cigar/cigarillo but not cigarettes (Table 1) were contrasted to the never smoker group. Phyla Actinobacteria, Firmicutes and Proteobacteria were more abundant in alternative smokers while Bacteroidetes and an uncultured bacterium from Saccharibacteria were more depleted. In those who only smoked hookah (n = 28), genera *Porphyromonas, Leptotrichia, Streptobacillus, Fusobacterium,* and an uncultured bacterium from Saccharibacteria were depleted. No OTUs were identified as differentially abundant among users of e-cigarette (n=11) or cigar/cigarillo (n=23) who did not use any other smoking products. GSEA identified a significant depletion of aerobic OTUs in cigar and cigarillo smokers (ES = -0.697, p = 0.04 GSEA permutation test) and depletion of facultative anaerobic OTUs in e-cigarette and hookah smokers compared to never smokers (e-cigarettes ES = -0.514 p = 0.03 GSEA permutation test; hookah ES = -0.489 p = 0.04 GSEA permutation test).

Discussion

This study analyzes oral mouthwash specimens from a sub-sample of NYC HANES 2013-14 to provide multiple lines of causal inference supporting the hypothesis that tobacco smoke exposure alters the saliva microbiome. We found current smokers to harbour a different microbial composition compared to never smokers and the other tobacco exposure groups in terms of beta diversity, individual OTUs, and oxygen requirements. The microbiome of former smokers was more similar to never smokers than to current smokers. Phyla Candidate division SR1, Bacteroides and Proteobacteria were depleted in smokers with genera *Bergeyella, Porphyromonas, Prevotella, Haemophilus*, *Neisseria, Lautropia* and *Actinobacillus*, consistent with previous studies [62–64]. The depletion of Proteobacteria may be especially important as this depletion has also been found among individuals with periodontal disease compared to healthy controls [65]. Further, Proteobacteria levels in the oral microbiome have been associated with insulin resistance and inflammation [66]. These shifts in genera largely

remained with adjustment for hypothesized confounders.

The large microbiome shifts associated with current smokers compared to never smokers included significant depletion of oxygen-requiring bacteria, and corresponding (but not statistically significant) enrichment of anaerobic bacteria. This finding is consistent with a proposed mechanism [67] by which smoking alters the oxygenation of the oral cavity, depleting oxygen and favouring anaerobic bacteria. Furthermore, the shifts in mean abundance occurring between current smokers and never smokers were positively correlated to those observed among non-smokers with varying levels of secondhand smoke exposure as measured by serum cotinine. This indicates a dose-response relationship for secondhand smoke exposure, and reduces the plausibility of residual confounding as an explanation for the shifts observed in these separate groups of participants.

Reduced aerobiosis and increased anaerobiosis in the oral cavity have implications for oral and systemic health. The Red Complex, a trio of anaerobic bacteria (*Treponema denticola, Porphyromonas gingivalis* and *Tannerella forsythia)* are linked with the development of periodontal disease [68]. While these primarily inhabit the dental plaque, an overall anaerobic oral environment may facilitate colonization. Although this study did not provide species-level resolution to observe the Red Complex, a previous study [69] found increased abundance of *Porphyromonas gingivalis* and *Tannerella forsythia* in the subgingival plaque of smokers compared to non-smokers. Furthermore, oral anaerobiosis could provide greater opportunity for movement of oral bacteria to distant anaerobic environments in the stomach and gut. This study demonstrates how oxygen utilization provides a simplifying measure that can be used by future studies of the oral microbiome and health.

In a mixed group of users of alternative smoking products including e-cigarettes, hookahs, cigars and cigarillos, we found some alterations comparable to those in cigarette smokers (like *Lactococcus* and *Neisseria* genera), while others like *Porphyromonas* had an opposite trend. This small and heterogeneous group of alternative smoking products does not allow robust conclusions, but indicates the possibility that alternative products could alter the oral microbiome composition in ways

similar to cigarette smoke. As e-cigarettes are gaining popularity in use, [70] more research is needed to explore the effect of vaping on the oral microbiome.

This study has a number of limitations. As a cross-sectional study, changes to the oral microbiome in direct response to tobacco exposure were not measured; longitudinal data are needed to directly observe tobacco-induced changes to the oral microbiome. We defined secondhand exposure to smoke using a serum cotinine upper cut-off of 14 ng/mL [71], Four of 38 participants in the secondhand exposure group had serum cotinine higher than the cut-off value of 10 ng/mL [35], and may have misreported recent light cigarette usage. However, removing these four individuals had negligible effect on reported results. We adjusted for self-reported gum disease as a measure of periodontal health; this measure is imperfect and residual confounding by periodontal health may remain. Additionally periodontal health may be a mediator rather than a confounder. Regardless, self-reported gum disease was not strongly associated with beta diversity in our analyses, so adjusting for it should not impact results. Additionally, we did not adjust for differences in dietary habits given the general lack of validity of self-reported diet data [72]; we cannot rule out residual confounding by diet. However, our findings, which show a shift toward anaerobic bacteria among smokers make biological sense in response to smoke exposure, reducing the likelihood that these findings are the result of residual confounding. Finally, the current analysis is based on 16S rRNA gene analyses capturing only genus and higher-level taxonomic information; whole metagenomic sequencing may provide additional important information on shifts to the oral cavity caused by tobacco exposures and functional information.

The strengths of this study are that it included a racially/ethnically diverse group of participants and an array of tobacco exposure groups including self-reported non-smokers exposed to secondhand smoke. The study design allowed multiple, complementary comparisons, as well as biological analysis, to help distinguish causal associations from associations likely to be caused by residual confounding. This study introduces several analyses that are, to the best of our knowledge, novel to epidemiological studies of the human microbiome. These include: 1) the application of Gene Set Enrichment Analysis methods for biological interpretation of gross microbiome

shifts, 2) use of a scatter plot to visualize the comparison of crude vs. adjusted regression coefficients in high-dimensional data, and 3) application of the Integrative Correlation Coefficient [60,61], a method introduced to assess reproducibility of gene expression studies, to make causal inference by comparing regression coefficients from different samples with different measures of tobacco exposure (smokers vs. non-smokers, and dose-response for continuous serum cotinine measurements).

Conclusions

Overall shifts between aerobic and anaerobic microbiota is a relevant simplifying measure that should be considered in future health studies of the oral microbiome. These results support a plausible biological mechanism for population-level shifts in the oral microbiome caused by exposure to tobacco smoke, through three lines of observational evidence: 1) consistency of the microbiome shifts with reduced microbiota oxygen utilization as a biological mechanism for the shifts observed in smokers; 2) consistency of oral microbiome abundance fold-changes in current smokers versus non-smokers with abundance changes along the gradient of secondhand smoke exposure among non-smokers; and 3) tobacco-related associations that are stronger than associations with sociodemographic and health indicators, and that are not meaningfully affected by controlling for hypothesized confounders.

Acknowledgements: We would like to thank Sharon Perlman and Jennifer Rakeman-Cagno from the New York City Department of Health and Mental Hygiene for their collaboration during study implementation and design.

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).

References

[1] Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiotia to periodontal health and disease. Cell Host Microbe 2011;10:302–6.
[2] Blaizot A, Vergnes J-N, Nuwwareh S, Amar J, Sixou M. Periodontal diseases and cardiovascular events: meta-analysis of observational studies. Int Dent J 2009;59:197–209.
[3] Borgnakke WS, Ylöstalo PV, Taylor GW, Genco RJ. Effect of periodontal disease on diabetes: systematic review of epidemiologic observational evidence. J Periodontol 2013;84:S135–52.
[4] Chaffee BW, Weston SJ. Association between chronic periodontal disease and obesity: a systematic review and meta-analysis. J Periodontol 2010;81:1708–24.
[5] Ide M, Papapanou PN. Epidemiology of association between maternal periodontal disease and adverse pregnancy outcomes--systematic review. J Clin Periodontol 2013;40 Suppl 14:S181–94.
[6] Detert J, Pischon N, Burmester GR, Buttgereit F. The association between rheumatoid arthritis and periodontal disease. Arthritis Res Ther 2010;12:218.
[7] Zeng X-T, Tu M-L, Liu D-Y, Zheng D, Zhang J, Leng W. Periodontal disease and risk of chronic obstructive pulmonary disease: a meta-analysis of observational studies. PLoS One 2012;7:e46508.
[8] Fitzpatrick SG, Katz J. The association between periodontal disease and cancer: a review of the literature. J Dent 2010;38:83–95.
[9] Chang K, Yang SM, Kim SH, Han KH, Park SJ, Shin JI. Smoking and rheumatoid arthritis. Int J Mol Sci 2014;15:22279–95.
[10] Chang SA. Smoking and type 2 diabetes mellitus. Diabetes Metab J 2012;36:399–403.
[11] Stallones RA. The association between tobacco smoking and coronary heart disease. Int J Epidemiol 2015;44:735–43.
[12] Laniado-Laborín R. Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21 century. Int J Environ Res Public Health 2009;6:209–24.
[13] Zheng W, Suzuki K, Tanaka T, Kohama M, Yamagata Z, Okinawa Child Health Study Group. Association between Maternal Smoking during Pregnancy and Low Birthweight: Effects by Maternal Age. PLoS One 2016;11:e0146241.
[14] Moura MA de S, Bergmann A, Aguiar SS de, Thuler LCS. The magnitude of the association between smoking and the risk of developing cancer in Brazil: a multicenter study. BMJ Open 2014;4:e003736.
[15] 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. ISME J 2016. doi:10.1038/ismej.2016.37.
[16] Kumar PS, Matthews CR, Joshi V, de Jager M, Aspiras M. Tobacco smoking affects bacterial acquisition and colonization in oral biofilms. Infect Immun 2011;79:4730–8.
[17] Mason MR, Preshaw PM, Nagaraja HN, Dabdoub SM, Rahman A, Kumar PS. The subgingival microbiome of clinically healthy current and never smokers. ISME J 2015;9:268–72.
[18] Joshi V, Matthews C, Aspiras M, de Jager M, Ward M, Kumar P. Smoking decreases structural and functional resilience in the subgingival ecosystem. J Clin Periodontol 2014;41:1037–47.
[19] Macgregor ID. Effects of smoking on oral ecology. A review of the literature. Clin Prev Dent 1989;11:3–7.
[20] Giannopoulou C, Cappuyns I, Mombelli A. Effect of smoking on gingival crevicular fluid cytokine profile during experimental gingivitis. J Clin Periodontol 2003;30:996–1002.
[21] Güntsch A, Erler M, Preshaw PM, Sigusch BW, Klinger G, Glockmann E. Effect of smoking on crevicular polymorphonuclear neutrophil function in periodontally healthy subjects. J

Periodontal Res 2006;41:184–8.
[22] Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. Nat Rev Immunol 2009;9:377–84.
[23] Kenney EB, Saxe SR, Bowles RD. The effect of cigarette smoking on anaerobiosis in the oral cavity. J Periodontol 1975;46:82–5.
[24] Ghio AJ, Hilborn ED, Stonehuerner JG, Dailey LA, Carter JD, Richards JH, et al. Particulate matter in cigarette smoke alters iron homeostasis to produce a biological effect. Am J Respir Crit Care Med 2008;178:1130–8.
[25] Reznick AZ, Klein I, Eiserich JP, Cross CE, Nagler RM. Inhibition of oral peroxidase activity by cigarette smoke: in vivo and in vitro studies. Free Radic Biol Med 2003;34:377–84.
[26] Eichel B, Shahrik HA. Tobacco smoke toxicity: loss of human oral leukocyte function and fluid-cell metabolism. Science 1969;166:1424–8.
[27] Hanioka T, Tanaka M, Takaya K, Matsumori Y, Shizukuishi S. Pocket Oxygen Tension in Smokers and Non-Smokers With Periodontal Disease. J Periodontol 2000;71:550–4.
[28] Eggert FM, McLeod MH, Flowerdew G. Effects of smoking and treatment status on periodontal bacteria: evidence that smoking influences control of periodontal bacteria at the mucosal surface of the gingival crevice. J Periodontol 2001;72:1210–20.
[29] Parvinen T. Stimulated salivary flow rate, pH and lactobacillus and yeast concentrations in non-smokers and smokers. Scand J Dent Res 1984;92:315–8.
[30] Mason MR, Preshaw PM, Nagaraja HN, Dabdoub SM, Rahman A, Kumar PS. The subgingival microbiome of clinically healthy current and never smokers. ISME J 2014;9:ismej2014114.
[31] 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.
[32] NHANES - About the National Health and Nutrition Examination Survey 2017. https://www.cdc.gov/nchs/nhanes/about_nhanes.htm (accessed July 23, 2018).
[33] Perlman SE, Chernov C, Farley SM, Greene CM, Aldous KM, Freeman A, et al. Exposure to Secondhand Smoke Among Nonsmokers in New York City in the Context of Recent Tobacco Control Policies: Current Status, Changes Over the Past Decade, and National Comparisons. Nicotine Tob Res 2016;18:2065–74.
[34] Jarvis MJ, Tunstall-Pedoe H, Feyerabend C, Vesey C, Saloojee Y. Comparison of tests used to distinguish smokers from nonsmokers. Am J Public Health 1987;77:1435–8.
[35] Homa DM, Neff LJ, King BA, Caraballo RS, Bunnell RE, Babb SD, et al. Vital signs: disparities in nonsmokers' exposure to secondhand smoke--United States, 1999-2012. MMWR Morb Mortal Wkly Rep 2015;64:103–8.
[36] Renson A, Jones HE, Beghini F, Segata N, Zolnik CP, Usyk M, et al. Sociodemographic variation in the oral microbiome. bioRxiv 2017:189225. doi:10.1101/189225.
[37] Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol 2015;75:129–37.
[38] Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 2016;18:1403–14.
[39] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6.
[40] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8:e61217.
[41] Aronesty E. Comparison of Sequencing Utility Programs. TOBIOIJ 2013;7:1–8.
[42] Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a

comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35:7188–96.
[43] Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson GL, Solymos P, et al. The vegan package. Community Ecology Package 2008;10.
[44] Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228–35.
[45] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.
[46] Giskes K, Kunst AE, Benach J, Borrell C, Costa G, Dahl E, et al. Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. J Epidemiol Community Health 2005;59:395–401.
[47] Pierce JP, Fiore MC, Novotny TE, Hatziandreu EJ, Davis RM. Trends in cigarette smoking in the United States. Educational differences are increasing. JAMA 1989;261:56–60.
[48] Emmons KM, Wechsler H, Dowdall G, Abraham M. Predictors of smoking among US college students. Am J Public Health 1998;88:104–7.
[49] Chassin L, Presson CC, Rose JS, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood: demographic predictors of continuity and change. Health Psychol 1996;15:478–84.
[50] Pan Y, Wang W, Wang KS, Moore K, Dunn E, Huang S, et al. Age Differences in the Trends of Smoking Among California Adults: Results from the California Health Interview Survey 2001-2012. J Community Health 2015;40:1091–8.
[51] Benowitz NL, Hukkanen J, Jacob P 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. Handb Exp Pharmacol 2009:29–60.
[52] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol 1995;57:289–300.
[53] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
[54] Bergey's Manual of Systematics of Archaea and Bacteria. Wiley; 2015.
[55] Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 2012;40:D115–22.
[56] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
[57] Geistlinger L, Csaba G, Zimmer R. Bioconductor's EnrichmentBrowser: seamless navigation through combined results of set- & network-based enrichment analysis. BMC Bioinformatics 2016;17:45.
[58] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[59] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
[60] Parmigiani G, Garrett-Mayer ES, Anbazhagan R, Gabrielson E. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clin Cancer Res 2004;10:2922–7.
[61] Cope L, Naiman DQ, Parmigiani G. Integrative correlation: Properties and relation to canonical correlations. J Multivar Anal 2014;123:270–80.
[62] 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. ISME J 2016;10:2435–46.
[63] Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care

Med 2013;187:1067–75.
[64] Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. PLoS One 2010;5:e15216.
[65] Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, Yang ZK, et al. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. ISME J 2012;6:1176–85.
[66] Demmer RT, Breskin A, Rosenbaum M, Zuk A, LeDuc C, Leibel R, et al. The subgingival microbiome, systemic inflammation and insulin resistance: The Oral Infections, Glucose Intolerance and Insulin Resistance Study. J Clin Periodontol 2017;44:255–65.
[67] Kenney EB, Saxe SR, Bowles RD. The effect of cigarette smoking on anaerobiosis in the oral cavity. J Periodontol 1975;46:82–5.
[68] Rocas IN, Siqueira JF Jr, Santos KR, Coelho AM. “Red complex” (Bacteroides forsythus, Porphyromonas gingivalis, and Treponema denticola) in endodontic infections: a molecular approach. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2001;91:468–71.
[69] Guglielmetti MR, Rosa EF, Lourencao DS, Inoue G, Gomes EF, De Micheli G, et al. Detection and quantification of periodontal pathogens in smokers and never-smokers with chronic periodontitis by real-time polymerase chain reaction. J Periodontol 2014;85:1450–7.
[70] Department of Health and Human Services US. E-Cigarette use among youth and young adults. A report of the Surgeon General. Of Health and Human Services … 2016.
[71] Benowitz NL, Bernert JT, Caraballo RS, Holiday DB, Wang J. Optimal serum cotinine levels for distinguishing cigarette smokers and nonsmokers within different racial/ethnic groups in the United States between 1999 and 2004. Am J Epidemiol 2009;169:236–48.
[72] Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism 1995;44:18–22.

Table 1. Demographics and characteristics of participants in the 2013-2014 NYC-HANES smoking and oral microbiome study

	Never smoker	Cigarette	Former smoker	Alternative smoker	Secondhand
n	43	86	43	49	38
Sex = Female (%)	28 (65.1)	45 (52.3)	25 (58.1)	22 (44.9)	22 (57.9)
Race/ethnicity (%)					
Non-Hispanic White	13 (30.2)	24 (27.9)	25 (58.1)	19 (38.8)	10 (26.3)
Non-Hispanic Black	13 (30.2)	33 (38.4)	4 (9.3)	9 (18.4)	11 (28.9)
Hispanic	10 (23.3)	19 (22.1)	10 (23.3)	12 (24.5)	14 (36.8)
Asian	3 (7.0)	9 (10.5)	3 (7.0)	3 (6.1)	2 (5.3)
Other	4 (9.3)	1 (1.2)	1 (2.3)	6 (12.2)	1 (2.6)
Educational achievement (%)					
College graduate or more	16 (37.2)	18 (20.9)	21 (48.8)	17 (34.7)	9 (23.7)
Less than High school diploma	8 (18.6)	24 (27.9)	4 (9.3)	10 (20.4)	14 (36.8)
High school graduate/GED	7 (16.3)	24 (27.9)	8 (18.6)	11 (22.4)	10 (26.3)
Some College or associate's degree	12 (27.9)	20 (23.3)	10 (23.3)	11 (22.4)	5 (13.2)
Age in years(mean (sd))	45.42 (16.50)	45.85 (13.07)	55.47 (18.00)	35.59 (16.44)	37.76 (14.70)
Age group (%)					
20-29	7 (16.3)	10 (11.6)	3 (7.0)	26 (53.1)	14 (36.8)
30-39	11 (25.6)	17 (19.8)	7 (16.3)	8 (16.3)	11 (28.9)
40-49	10 (23.3)	25 (29.1)	7 (16.3)	4 (8.2)	3 (7.9)
50-59	6 (14.0)	19 (22.1)	8 (18.6)	7 (14.3)	6 (15.8)
60 and over	9 (20.9)	15 (17.4)	18 (41.9)	4 (8.2)	4 (10.5)
Diabetes (%)					
Yes	5 (11.6)	5 (5.8)	7 (16.3)	3 (6.1)	2 (5.3)
No	38 (88.4)	81 (94.2)	36 (83.7)	38 (77.6)	36 (94.7)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	8 (16.3)	0 (0.0)
Physical activity (%)					
Very active	15 (34.9)	27 (31.4)	11 (25.6)	16 (32.7)	17 (44.7)
Somewhat active	20 (46.5)	37 (43.0)	20 (46.5)	25 (51.0)	15 (39.5)
Not very active/not active at all	8 (18.6)	22 (25.6)	12 (27.9)	8 (16.3)	6 (15.8)

Annual family income (%)					
Less Than $20,000	5 (11.6)	31 (36.0)	8 (18.6)	20 (40.8)	14 (36.8)
$20,000-$49,999	15 (34.9)	20 (23.3)	9 (20.9)	14 (28.6)	9 (23.7)
$50,000-$74,999	6 (14.0)	11 (12.8)	3 (7.0)	6 (12.2)	4 (10.5)
$75,000-$99,999	8 (18.6)	4 (4.7)	6 (14.0)	4 (8.2)	2 (5.3)
$100,000 or More	6 (14.0)	11 (12.8)	13 (30.2)	2 (4.1)	6 (15.8)
Missing	3 (7.0)	9 (10.5)	4 (9.3)	3 (6.1)	3 (7.9)
Serum Cotinine (median [IQR])	0.04 [0.04, 0.04]	271.49 [189.99, 360.99]	0.04 [0.04, 0.04]	10.54 [0.28, 55.36]	3.01 [1.39, 5.48]
Gum disease (self-reported) (%)					
Yes	4 (9.3)	9 (10.5)	5 (11.6)	4 (8.2)	4 (10.5)
No	39 (90.7)	76 (88.4)	38 (88.4)	45 (91.8)	34 (89.5)
Missing	0 (0.0)	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)

Table 2: PERMANOVA analysis on Weighted UniFrac distance measure. Model included smoking status (cigarette smokers/never smokers) and other sociodemographic measures. Df: degrees of freedom, R^2: Coefficient of Determination

	Df	F.Model	R^2	Pr(>F)
Smoking status (Cigarette smokers vs Never Smokers)	1	7.0845	0.05137	0.002
Self reported gum disease	2	0.8649	0.01254	0.496
Race/ethnicity	4	1.6281	0.04723	0.05
Sex	1	2.2591	0.01638	0.06
Age groups	1	2.9418	0.02133	0.014
Physical activity	2	0.8085	0.01173	0.623
Education level	2	0.7274	0.01055	0.708
Diabetes	1	0.304	0.0022	0.937

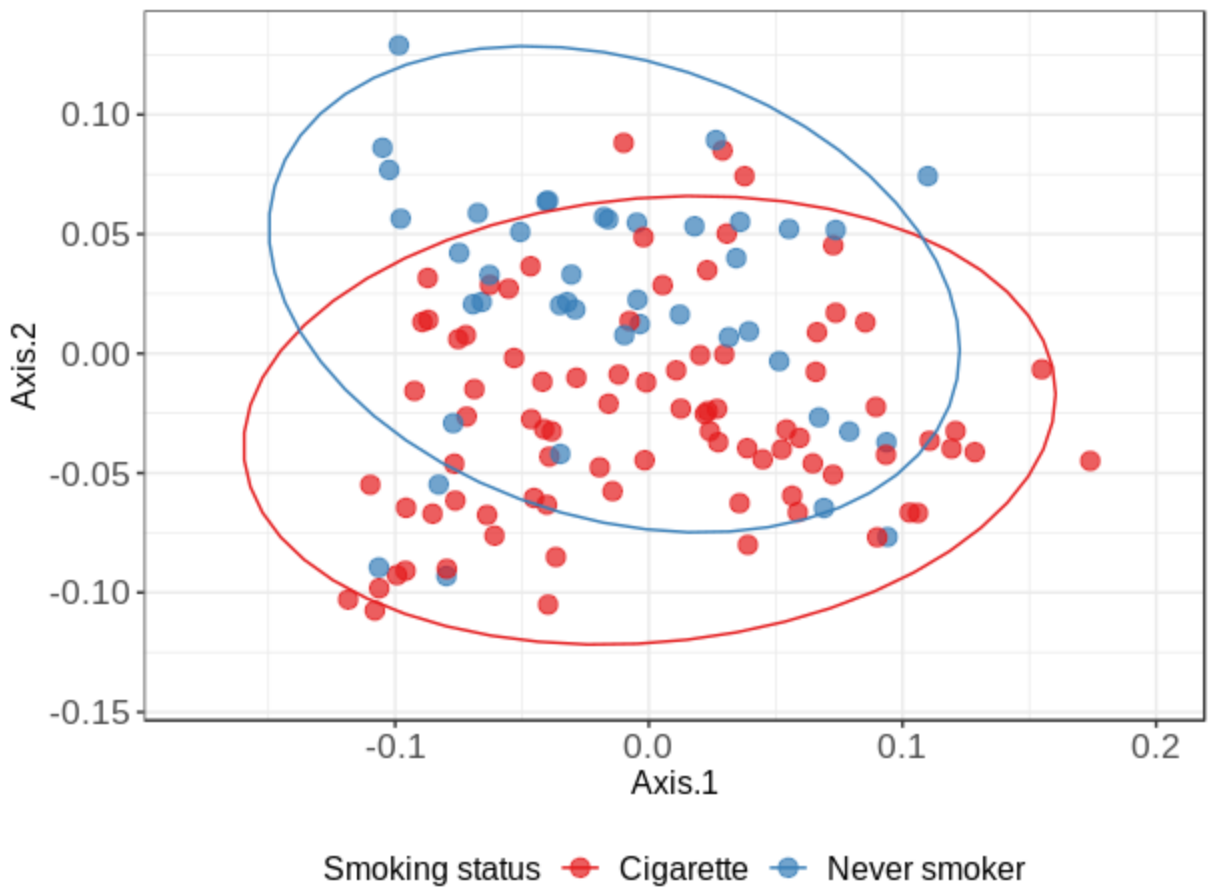


Figure 1: Principal coordinates analysis based on the weighted UniFrac distance. Dots in the ordination plot are samples from never smokers with negligible serum cotinine (blue, n=43) and current cigarette smokers (red, n=86); ellipsis indicating where 95% of observations are expected for each group. A separation between cigarette smokers and never smokers is present and is statistically significant (R^2 = 0.051, PERMANOVA p<0.001). A gradient also exists for the entire sample (n=259) by measured continuous serum cotinine level (R^2 =0.0485, PERMANOVA p = 0.001).

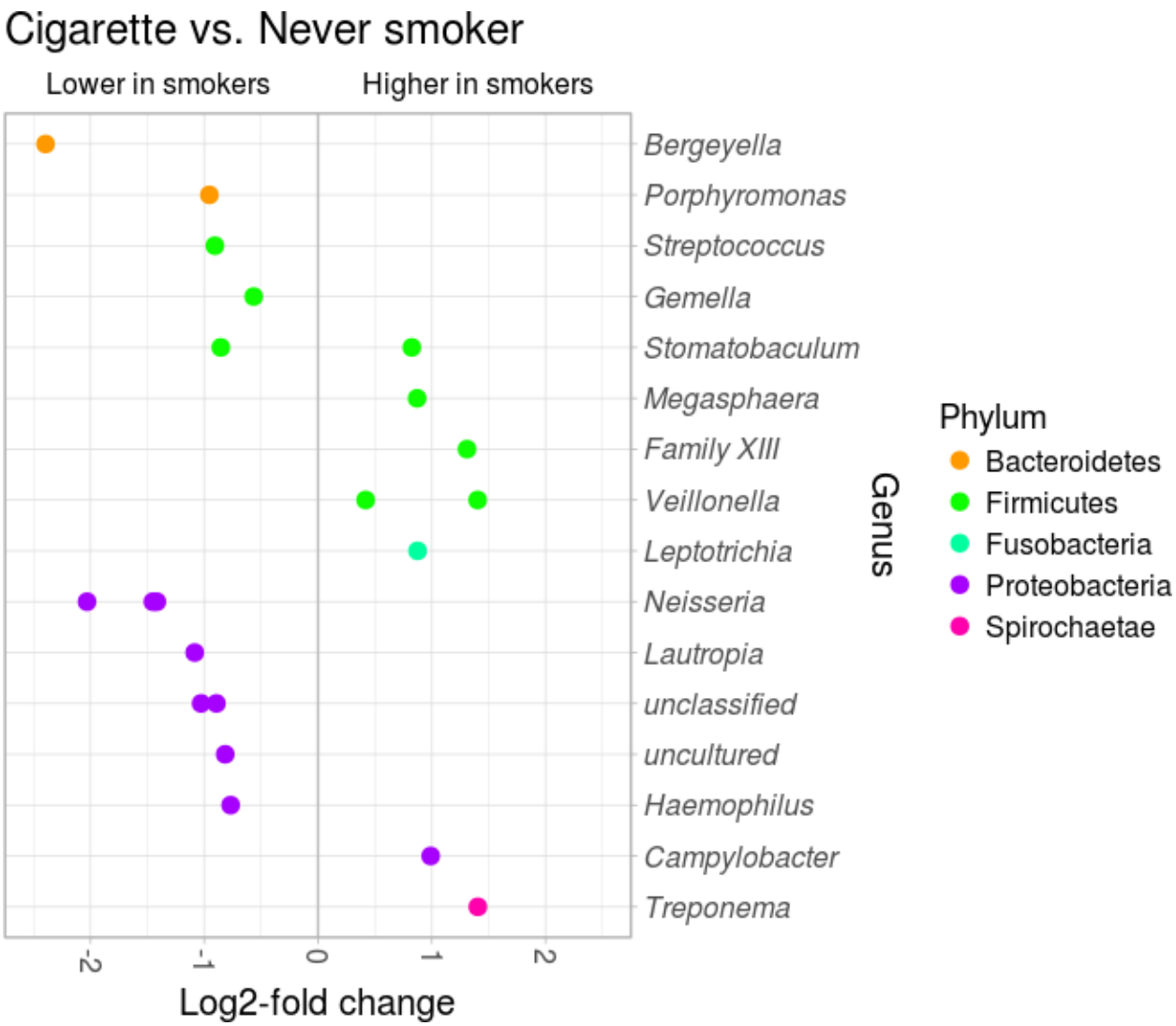


Figure 2: Adjusted multivariate differential analysis between current cigarette smokers (n=86) and never smokers (n=43). Starting from the 46 OTUs identified as differentially abundant from the crude model, adjusting for confounders OTU differentially abundant were reduced to 21.

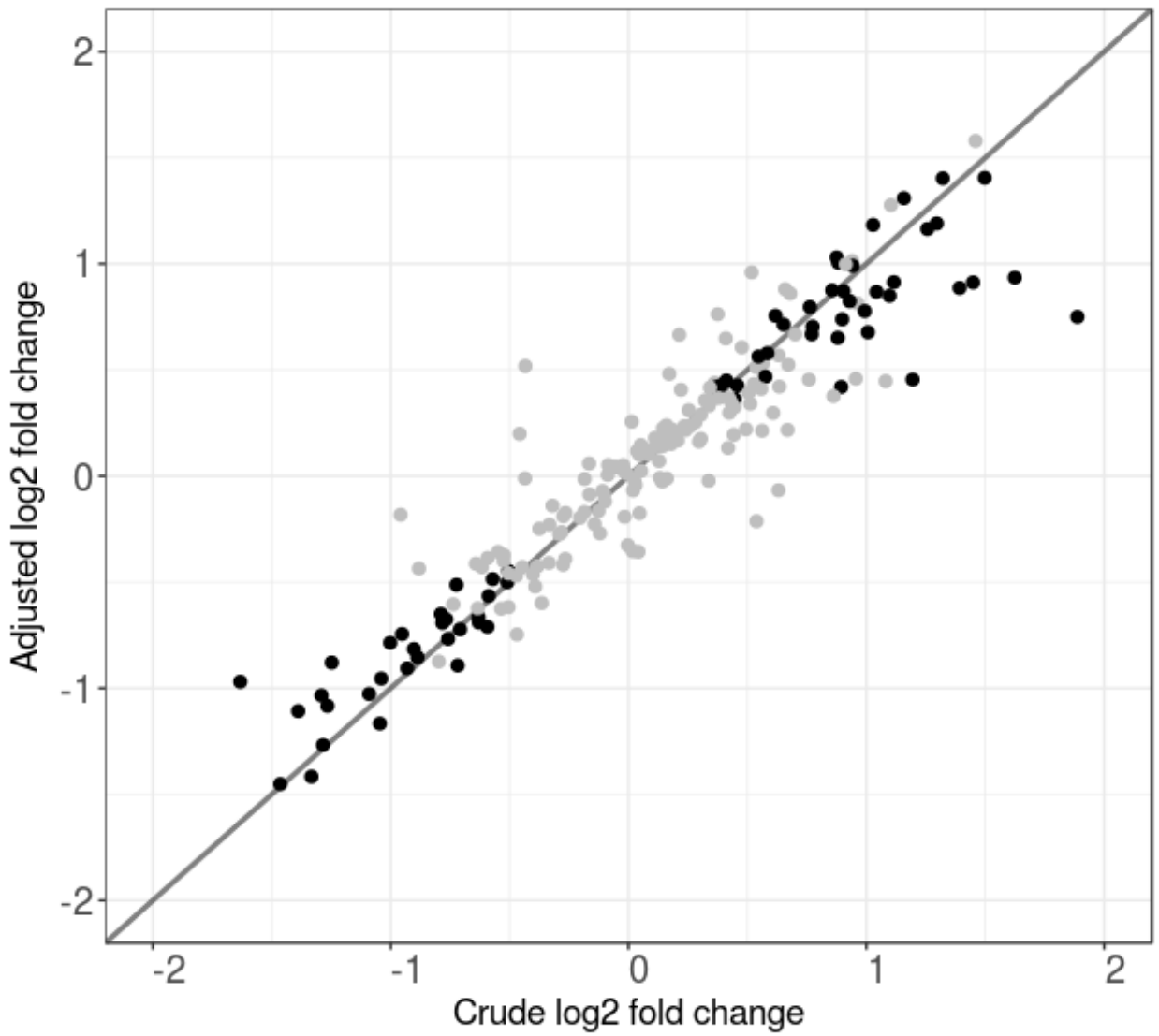


Figure 3: Comparison between 212 coefficients for current (n=86) vs. never smokers (n=43) from crude and adjusted negative binomial log-linear regression (adjusted for age, sex, race/ethnicity, self-reported physical activity, education, diabetes status, self-reported gum disease). Points in the scatter plot represent all differentially abundant OTUs, regardless of statistical significance, with black dot OTUs significant with the Wald test in crude analyses; coordinates are determined by the $\log_2$ fold change resulting from the crude analysis between current and never smokers (x axis) and the $\log_2$ fold change from the adjusted analysis between current and never smokers (y axis).

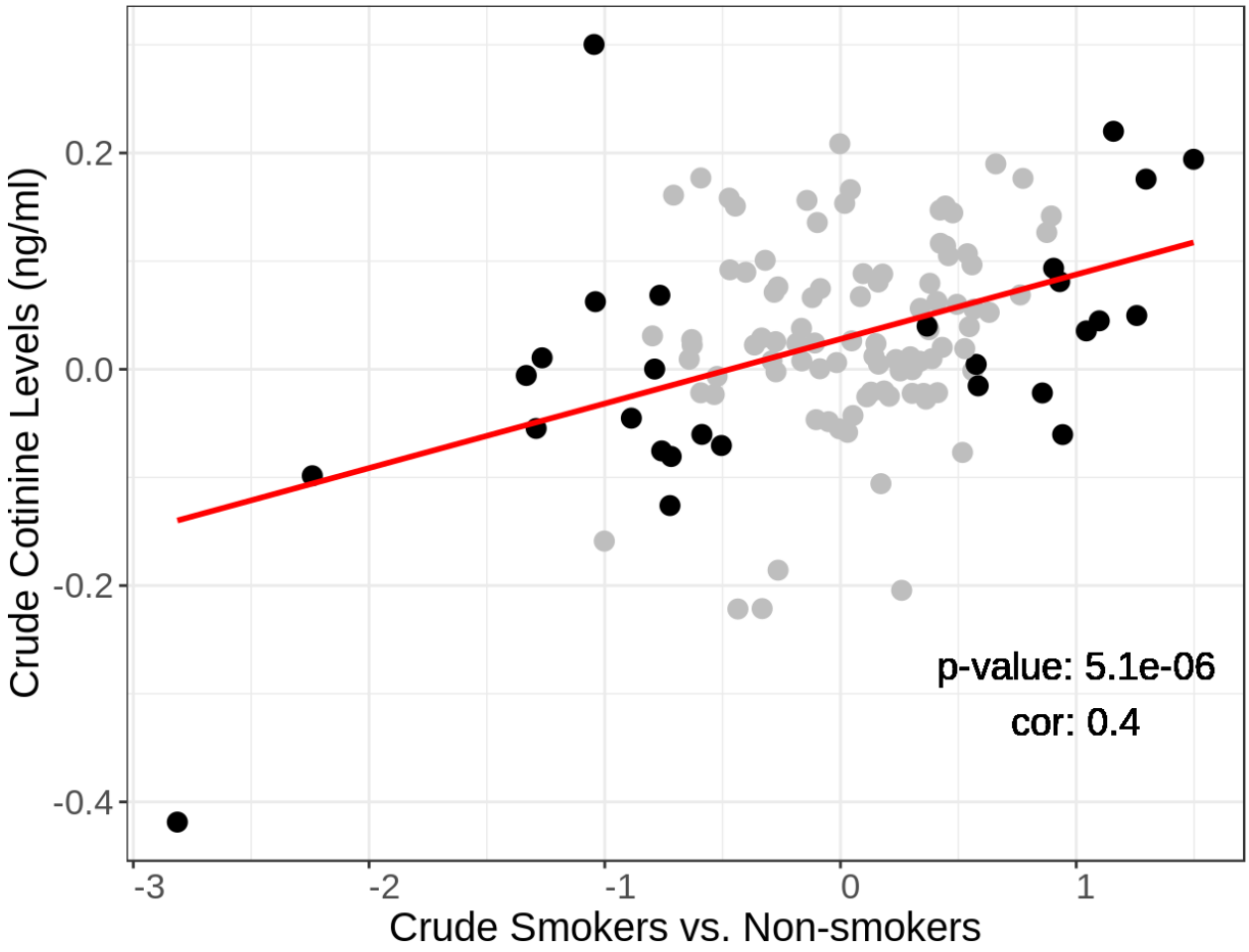
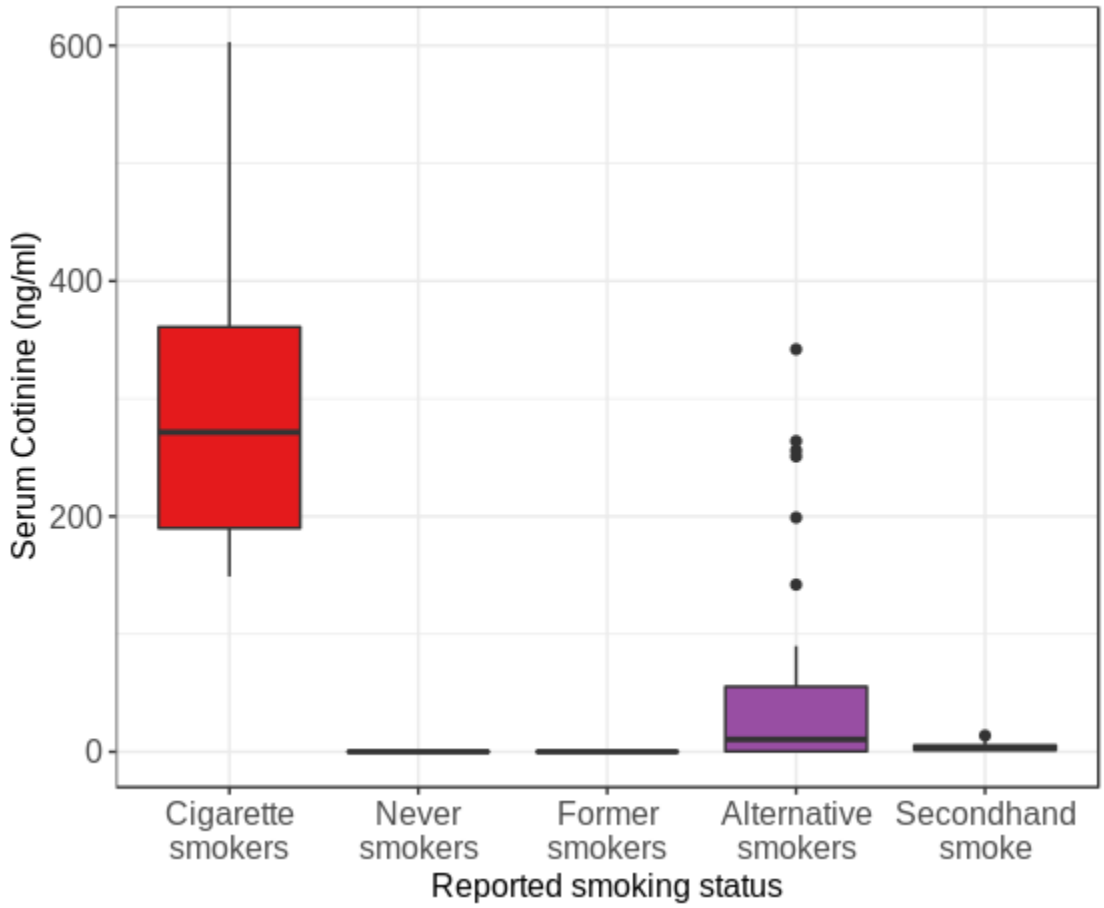
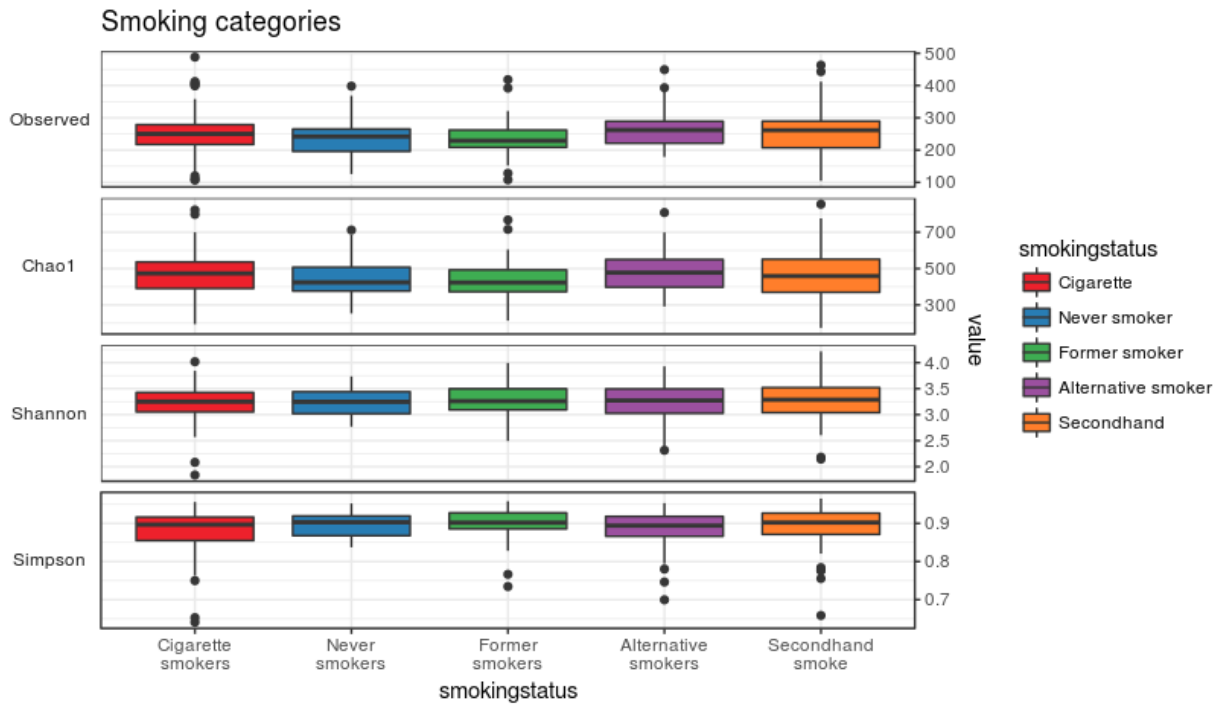


Figure 4: Comparison of $\log_2$ OTU fold changes between a) crude analyses of smokers (n=86) vs non-smokers with no detectable serum cotinine (n=43), and b) analysis of continuous cotinine levels among non-smokers exposed to secondhand smoke (n=38). Plot shows 121 OTUs that passed the edgeR filter for low-variance variables, including 28 OTUs that are differentially abundant in smokers (FDR <0.05, indicated in black). Coefficients for these two contrasts, involving different measures of exposure on different individuals, are positively correlated (Pearson Correlation = 0.58, p = 0.0013 for 28 OTUs; Pearson Correlation = 0.40, p = 5e-6 for all 121 OTUs)

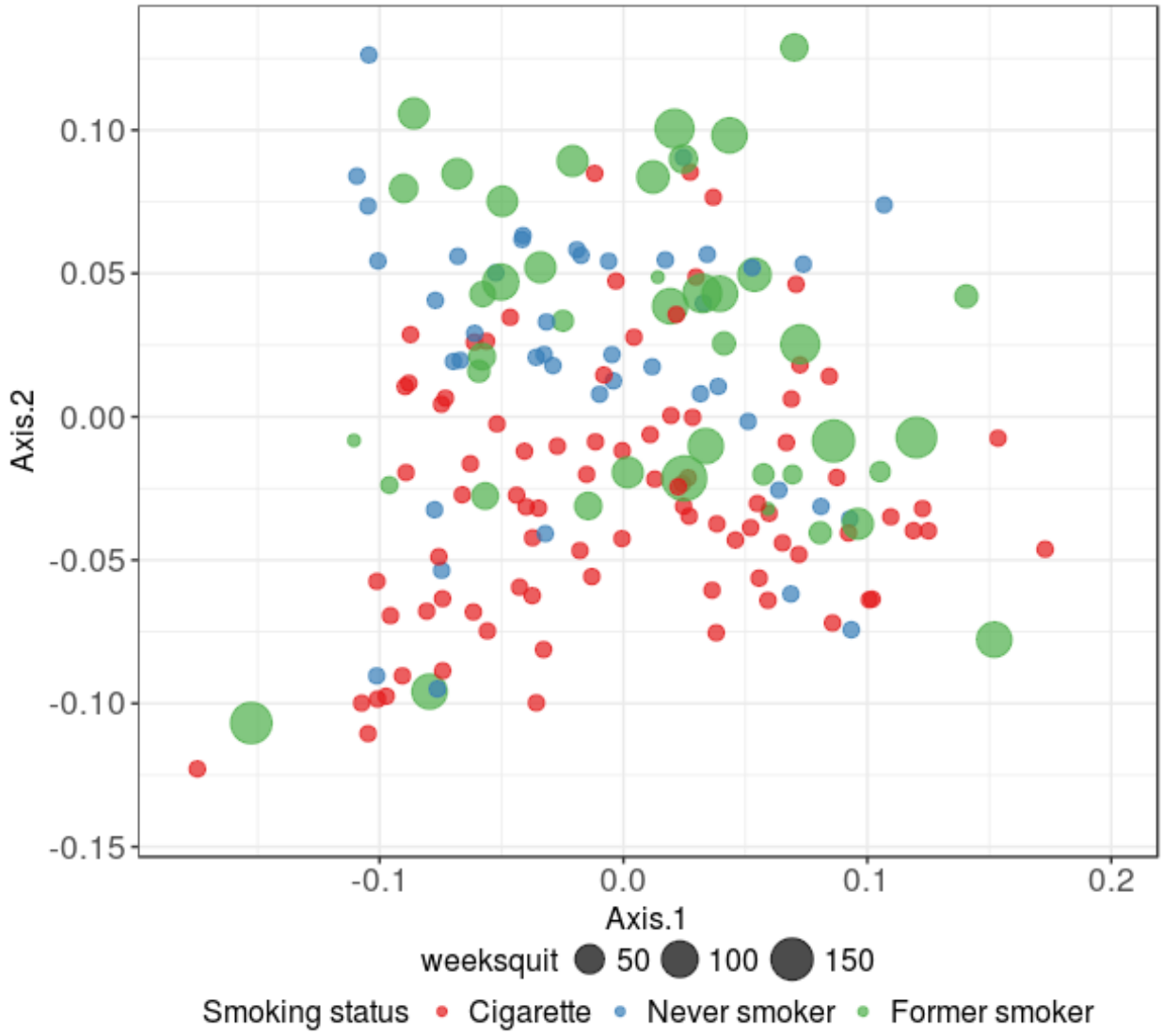
Appendix: Supplementary figures



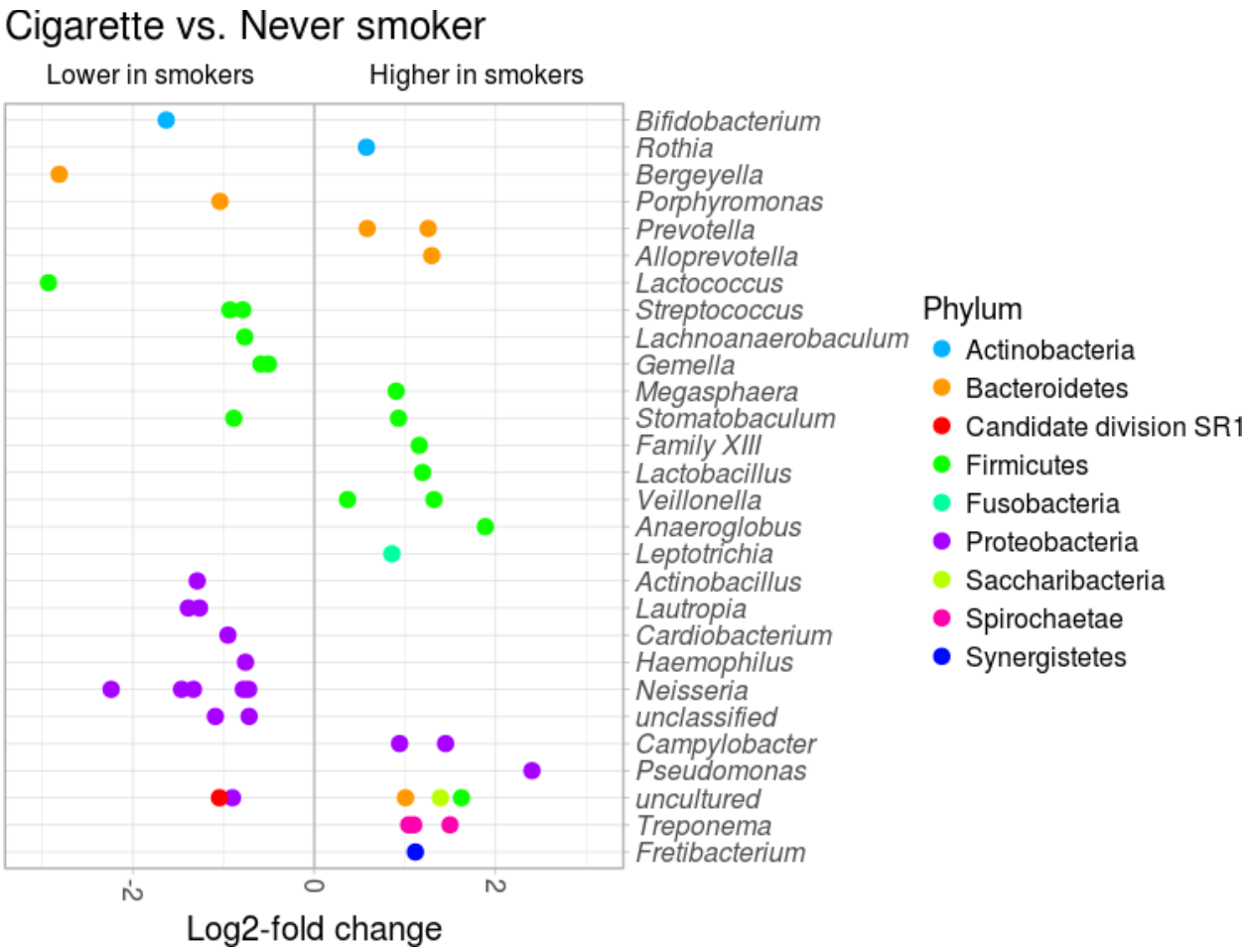
Supplementary Figure 1: Distribution of measured serum cotinine among the five smoke exposure groups with current cigarette smokers followed by smokers of alternative tobacco products showing the highest serum cotinine levels.



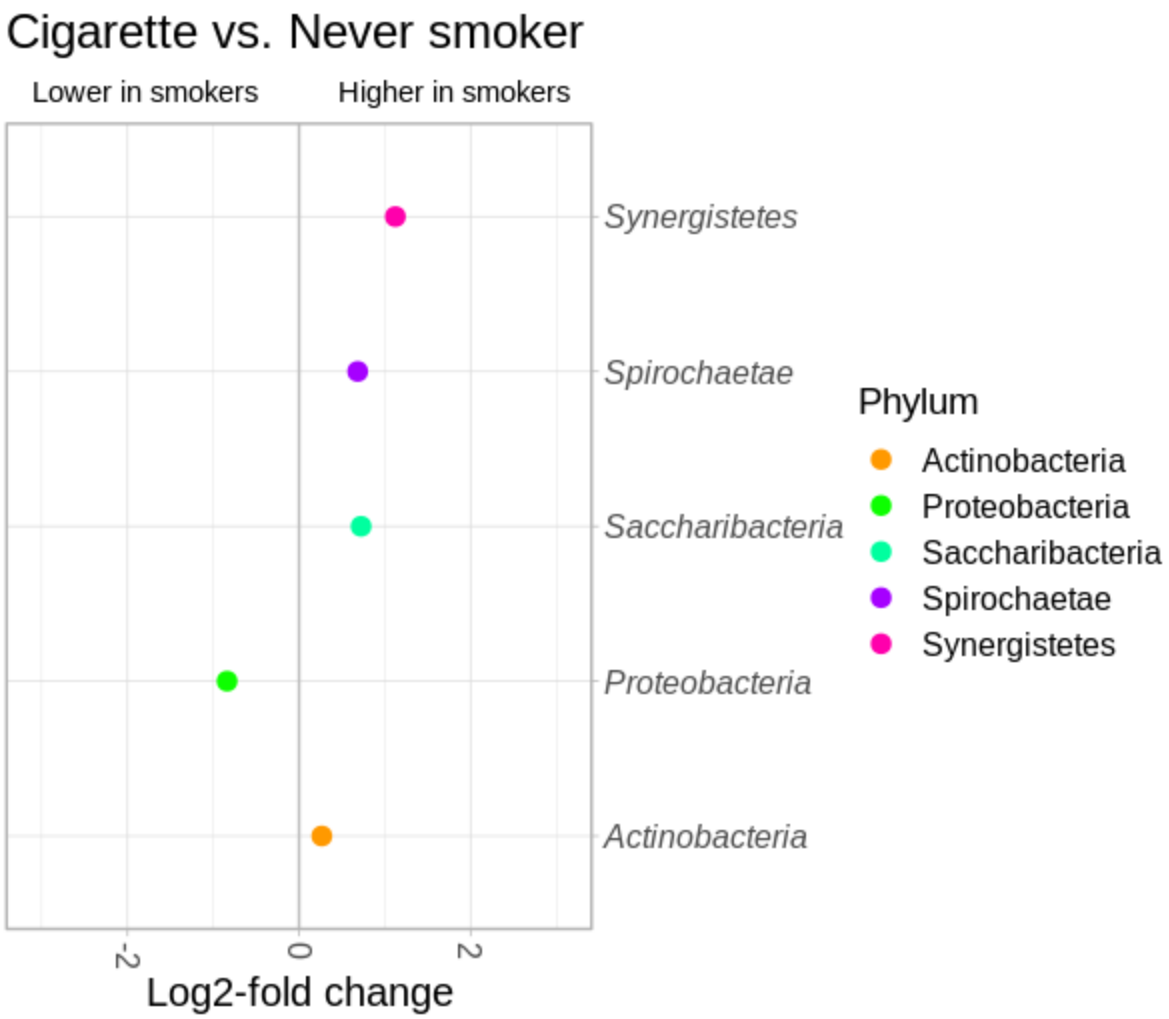
Supplementary Figure 2: Alpha diversity measures between the five tobacco exposure groups. No significant differences were observed between groups with four measures of richness and evenness.



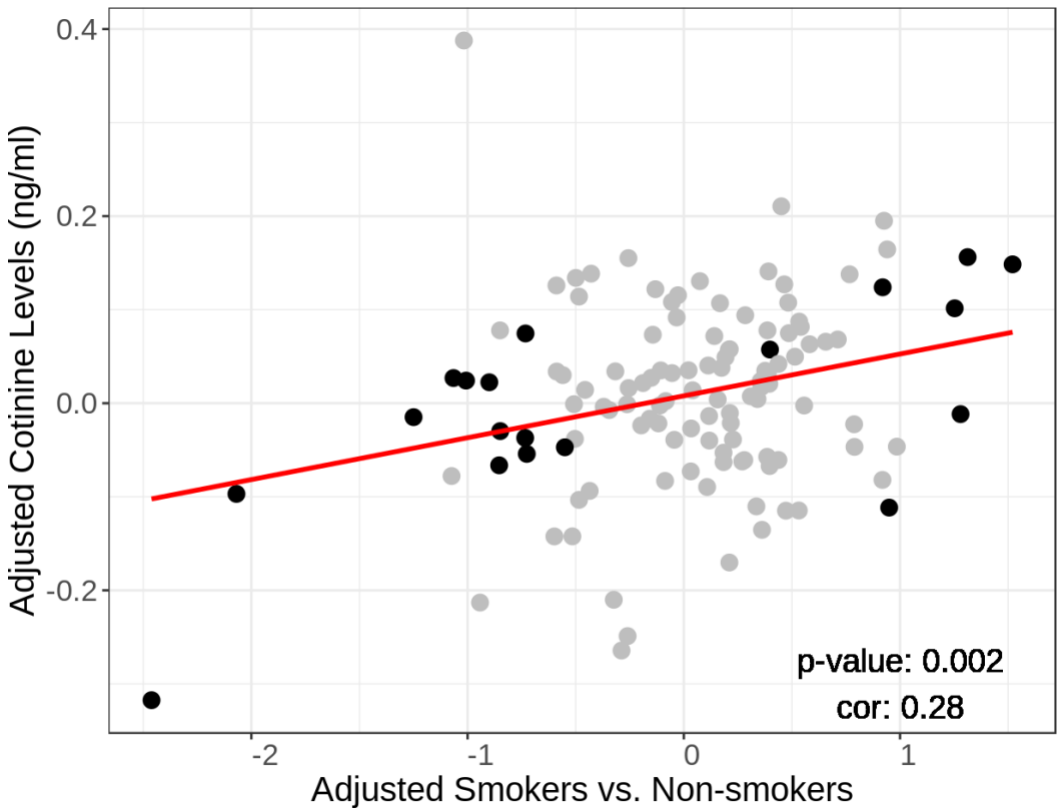
Supplementary Figure 3: Principal coordinates analysis based on the weighted UniFrac distance on samples from cigarette smokers (n=86), never smokers (n=43), and former smokers (n=43) with size of dot indicating how long ago they reported quitting.. A separation on the second axis between cigarette smokers and never smokers is present. We found former smokers were significantly different from current smokers (p=0.002, PERMANOVA test), but not from never smokers (p=0.16, PERMANOVA test).



Supplementary Figure 4: Crude differential analysis between current cigarette smokers (n=86) and never smokers (n=43). Dots in the plot are the 46 OTUs identified as differentially abundant without adjusting for hypothesized confounders and coloured according the taxonomy annotation at the phylum level. Position of the side of the plot is determined by the $\log_2$ fold change of abundance of the OTU.



Supplementary Figure 5: Crude differential analysis between current cigarette smokers (n=86) and never smokers (n=43) performed at the phylum level.



Supplementary Figure 6: Equivalent of Figure 4, with adjustment for age and educational attainment. Comparison of $\log_2$ OTU fold changes between a) adjusted analyses of (x-axis) smokers (n=86) vs non-smokers with no detectable serum cotinine (n=43) and b) adjusted analysis of continuous cotinine levels among non-smokers exposed to secondhand smoke (n=38). Plot shows 121 OTUs that passed the edgeR filter for low-variance variables, including 19 OTUs differentially abundant in smokers (FDR<0.05, indicated in black). Similar to the crude model in Figure 4, a positive correlation was observed (Pearson Correlation = 0.68, p = 0.001 for 19 OTUs; Pearson Correlation = 0.28, p = 0.002 for all 121 OTUs).