

## ORIGINAL ARTICLE

# Cigarette smoking and the oral microbiome in a large study of American adults

Jing Wu<sup>1,2,13</sup>, Brandilyn A Peters<sup>1,13</sup>, Christine Dominianni<sup>1</sup>, Yilong Zhang<sup>3</sup>, Zhiheng Pei<sup>2,4,5</sup>, Liying Yang<sup>6</sup>, Yingfei Ma<sup>4</sup>, Mark P Purdue<sup>7,8</sup>, Eric J Jacobs<sup>9</sup>, Susan M Gapstur<sup>9</sup>, Huilin Li<sup>3</sup>, Alexander V Alekseyenko<sup>10,11,12</sup>, Richard B Hayes<sup>1,2</sup> and Jiyoung Ahn<sup>1,2</sup>

<sup>1</sup>Division of Epidemiology, Department of Population Health, NYU School of Medicine, New York, NY, USA;

<sup>2</sup>NYU Perlmutter Cancer Center, New York, NY, USA; <sup>3</sup>Division of Biostatistics, Department of Population Health, NYU School of Medicine, New York, NY, USA; <sup>4</sup>Department of Pathology, NYU School of Medicine, New York, NY, USA; <sup>5</sup>Department of Veterans Affairs New York Harbor Healthcare System, New York, NY, USA; <sup>6</sup>Division of Translational Medicine, Department of Medicine, NYU School of Medicine, New York, NY, USA; <sup>7</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; <sup>8</sup>Ontario Institute for Cancer Research, Toronto, Ontario, Canada; <sup>9</sup>Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA; <sup>10</sup>Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, USA; <sup>11</sup>Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA and <sup>12</sup>Department of Oral Health Sciences, Medical University of South Carolina, Charleston, SC, USA

**Oral microbiome dysbiosis is associated with oral disease and potentially with systemic diseases; however, the determinants of these microbial imbalances are largely unknown. In a study of 1204 US adults, we assessed the relationship of cigarette smoking with the oral microbiome. 16S rRNA gene sequencing was performed on DNA from oral wash samples, sequences were clustered into operational taxonomic units (OTUs) using QIIME and metagenomic content was inferred using PICRUSt. Overall oral microbiome composition differed between current and non-current (former and never) smokers ( $P < 0.001$ ). Current smokers had lower relative abundance of the phylum Proteobacteria (4.6%) compared with never smokers (11.7%) (false discovery rate  $q = 5.2 \times 10^{-7}$ ), with no difference between former and never smokers; the depletion of Proteobacteria in current smokers was also observed at class, genus and OTU levels. Taxa not belonging to Proteobacteria were also associated with smoking: the genera *Capnocytophaga*, *Peptostreptococcus* and *Leptotrichia* were depleted, while *Atopobium* and *Streptococcus* were enriched, in current compared with never smokers. Functional analysis from inferred metagenomes showed that bacterial genera depleted by smoking were related to carbohydrate and energy metabolism, and to xenobiotic metabolism. Our findings demonstrate that smoking alters the oral microbiome, potentially leading to shifts in functional pathways with implications for smoking-related diseases.**

The ISME Journal (2016) 10, 2435–2446; doi:10.1038/ismej.2016.37; published online 25 March 2016

## Introduction

The human oral cavity is inhabited by over 600 bacterial species, known collectively as the oral microbiome (Dewhirst *et al.*, 2010); these bacteria are involved in a wide variety of functions, and many are important in maintaining oral health (Belda-Ferre *et al.*, 2012). Oral dysbiosis leads locally to periodontitis, dental caries and potentially to head and

neck cancer (Wade, 2013; He *et al.*, 2015). There is also increasing evidence of a role for oral dysbiosis in systemic diseases of the lung (Beck *et al.*, 2012), digestive tract (Ahn *et al.*, 2012) and cardiovascular system (Koren *et al.*, 2011), yet factors that influence the oral microbiome are poorly understood. Cigarette smoke is a source of numerous toxicants (WHO, 2012) that come into direct contact with oral bacteria; these toxicants can perturb the microbial ecology of the mouth via antibiotic effects, oxygen deprivation or other potential mechanisms (Macgregor, 1989). Loss of beneficial oral species due to smoking can lead to pathogen colonization and ultimately to disease; this contention is strongly supported by the well-established role of smoking in the onset and progression of periodontitis

Correspondence: J Ahn, Department of Population Health, NYU School of Medicine, 650 1st Avenue, 5th Floor, New York, 10016 NY, USA.

E-mail: Jiyoung.Ahn@nyumc.org

<sup>13</sup>These authors contributed equally to this work.

Received 27 September 2015; revised 1 February 2016; accepted 10 February 2016; published online 25 March 2016

(Nociti *et al.*, 2015). Previous studies have shown alterations in the abundance of selected oral bacteria in smokers compared with non-smokers (Colman *et al.*, 1976; Ertel *et al.*, 1991; Charlson *et al.*, 2010; Kumar *et al.*, 2011; Hugoson *et al.*, 2012; Morris *et al.*, 2013; Belstrom *et al.*, 2014; Mason *et al.*, 2015); however, results across these studies are largely inconsistent, possibly due to small sample sizes in some, use of different sampling sites in the mouth and use of different laboratory methodologies, some of which impose limitations on bacterial profiling.

To improve our understanding of the influence of smoking on the oral microbiome, we conducted a comprehensive assessment of oral microbiome community composition and individual taxon abundance, by bacterial 16S rRNA gene sequencing, in 1204 individuals from two large US national cohorts. Strengths of our study include the availability of detailed data from both cohorts on individual smoking history and potential demographic confounding factors. In addition, the large sample size available from each cohort provided us with excellent statistical power for discovery in combined analyses, as well as the opportunity to independently replicate results in each cohort.

## Subjects and methods

### *Study population*

Participants were drawn from the National Cancer Institute (NCI) Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial cohort (Hayes *et al.*, 2000) and the American Cancer Society (ACS) Cancer Prevention Study II (CPS-II) Nutrition cohort (Calle *et al.*, 2002), which are described in detail in the above-cited references.

Briefly, the PLCO Cancer Screening Trial is a multicenter trial designed to determine the effects of screening on cancer-related mortality and secondary end points among 55- to 74-year-old subjects (Hayes *et al.*, 2000). In PLCO, 154 901 participants were recruited between 1993 and 2001 at 10 centers across the United States and were randomized to either the screening arm ( $n = 77\,445$ ) or the control arm ( $n = 77\,456$ ) of the trial. Oral wash samples were collected in the control arm only.

The ACS CPS-II Nutrition cohort ( $n = 184\,194$ ) is a prospective study of cancer incidence initiated in 1992. It is a subset of the larger CPS-II cohort ( $n = 1.2$  million participants) recruited by ACS volunteers in 1982 and followed for mortality. At enrollment in the larger cohort in 1982, and in the subcohort in 1992/1993, participants completed self-administered questionnaires that included information on demographics, family characteristics, personal and family history of cancer and other diseases, reproductive history, as well as various behavioral, environmental, occupational and dietary exposures. Beginning in 1997, follow-up questionnaires were sent to cohort members every 2 years to update

exposure information and ascertain newly diagnosed cancers. Oral wash samples were collected by mail from 70 004 CPS-II Nutrition cohort participants.

All subjects included in the present analyses were originally selected from the CPS-II and PLCO cohorts as cases or controls for collaborative nested case-control studies of the oral microbiome in relation to two smoking-related cancers, head and neck cancer and pancreatic cancer. Cases were participants who developed one of these two types of smoking-related cancers at any point after collection of the oral wash samples (time from sample collection to diagnosis ranged up to 12 years). Age- and sex-matched controls were selected by incidence density sampling among cohort members who provided an oral wash sample and had no cancer prior to selection.

Because the oral microbiome assays took place at different times for the pancreas study and the head and neck study, four separate data sets were assembled for this analysis, PLCO-a ( $n = 261$  PLCO participants in the head and neck study), PLCO-b ( $n = 400$  PLCO participants in the pancreas study), CPS-II-a ( $n = 203$  CPS-II participants in the head and neck study), and CPS-II-b ( $n = 340$  CPS-II participants in the pancreas study). All participants provided informed consent and all protocols were approved by the New York University School of Medicine Institutional Review Board.

### *Smoking and other covariate assessment*

Comprehensive demographic and lifestyle information was collected by baseline and follow-up questionnaires in the PLCO and CPS-II Nutrition cohorts. Detailed information on cigarette smoking, including smoking status (never, former, current), smoking dose and smoking duration, was ascertained at baseline and follow-up via questionnaires in both cohorts. These smoking data have been used in previous publications (Hocking *et al.*, 2010; Gaudet *et al.*, 2013).

### *Oral sample collection*

Oral wash samples were collected between 1993 and 2001 from control arm participants in the NCI PLCO cohort (Hayes *et al.*, 2000) and between January 2001 and December 2002 in the ACS CPS-II Nutrition cohort (Calle *et al.*, 2002). Participants in both cohorts were asked to swish vigorously with 10 ml Scope mouthwash (P&G) and were directed to expectorate into a specimen tube. Samples were shipped to each cohort's biorepository, pelleted and stored at  $-80^{\circ}\text{C}$  until use.

### *Oral microbiome assay*

In 2013, we extracted bacterial genomic DNA from oral samples aliquoted from the respective biorepositories using the Mo Bio PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA) with

bead-beating method in the Mo Bio Powerlyzer instrument (Mo Bio Laboratories). As reported previously (Wu *et al.*, 2014), 16S rRNA gene amplicons covering variable regions V3 to V4 were generated using primers (347F-5'-GGAGGCAGCAGTRRGGAAT-3' and 803R-5'-CTACCRGGGTATCTAATCC-3') (Nossa *et al.*, 2010) incorporating barcode sequences as well as Roche 454 FLX Titanium adapters (454 Life Sciences, Branford, CT, USA), and pooled amplicons libraries were sequenced with the 454 Roche FLX Titanium pyrosequencing system following the manufacturer's specifications. Laboratory personnel were blinded to smoking status. Multiplexed and barcoded sequences were deconvoluted. Poor-quality sequences were excluded using the default parameters of the QIIME script *split\_libraries.py* (minimum average quality score = 25, minimum/maximum sequence length = 200/1000 base pairs, no ambiguous base calls and no mismatches allowed in the primer sequence) (Caporaso *et al.*, 2010). From the 1204 oral wash samples, we obtained 14 553 620 quality-filtered 16S rRNA gene sequence reads. Sequences were clustered into *de novo* operational taxonomic units (OTUs) at 97% identity, and representative sequences for each OTU were assigned taxonomy based on fully-sequenced microbial genomes (IMG/GG Greengenes), using the QIIME *pick\_de\_novo\_otus.py* script (Caporaso *et al.*, 2010). Chimeric sequences (identified using ChimeraSlayer; Haas *et al.*, 2011), sequences that failed alignment, and singleton OTUs were removed. The final total data set retained 12 212 734 sequences (mean  $\pm$  s.d.: 10 144  $\pm$  2845 sequences per sample), with similar sequencing depths in the individual four data sets (Supplementary Table S1), and contained 43 435 OTUs.

#### Quality control

Blinded quality control specimens were used in each data set, respectively, across all sequencing batches. The number of quality control subjects and replicates of each sample are shown in Supplementary Table S2. Quality control samples had good reliability: across the four data sets, coefficient of variability ranged from 0.45% to 8.28% for the Shannon Index, and 6.29% to 26% for various phylum relative abundances.

#### Statistical analysis

The relationship between smoking status and overall oral microbiome composition was assessed by analysis of weighted and unweighted UniFrac distances (Lozupone *et al.*, 2011), computed using the QIIME pipeline (Caporaso *et al.*, 2010). Principal coordinate analysis plots were generated using the first two principal coordinates and labeled according to smoking status. Permutational MANOVA ('Adonis' function, vegan package, R) (McArdle and Anderson, 2001) of the weighted UniFrac distance was used to test differences in overall oral

microbiome composition across the categories of smoking status, adjusting for age, sex and data set.

The OTU table of raw counts was normalized to an OTU table of relative abundances, and taxa of the same type were agglomerated at the phylum, class, order, family and genus levels. Our analysis included taxa from the five major phyla of the oral microbiome (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria), and we additionally filtered out taxa present in less than 10% of participants, leaving 13 classes, 20 orders, 40 families and 69 genera in this analysis. Relative abundances of bacterial taxa were compared across the categories of smoking status using the nonparametric Kruskal–Wallis test. We further used polytomous logistic regression for the pairwise comparisons of former vs never smokers and current vs never smokers, adjusting for age and sex. We calculated nominal *P*-values for each individual data set (PLCO-a, PLCO-b, CPS-II-a and CPS-II-b) and report meta-analysis *P*-values based on Z-score methods (Evangelou and Ioannidis, 2013). Kendall's Tau rank coefficient was used to test the association of smoking-related variables (number of cigarettes per day and number of years since quitting) with relative abundances of selected taxa.

In order to investigate whether any of the differentially relative abundant taxa were identified due only to autocorrelation with other taxa, we conducted a parallel analysis using the DESeq function in the DESeq2 R package (Love *et al.*, 2014; McMurdie and Holmes, 2014); this function models raw counts using a negative binomial distribution, taking into account sample library size and the dispersion for each taxon. Using this model we compared current smokers to never smokers, and former smokers to never smokers, adjusting for sex, age and data set, at the phylum through genus levels and at the OTU level.

Bacterial metagenome content was predicted from 16S rRNA gene-based microbial compositions, and functional inferences were made from the Kyoto Encyclopedia of Gene and Genomes (KEGG) catalog, using the PICRUSt algorithm (Kanehisa *et al.*, 2012). A total of 6,909 inferred genes were categorized into 275 KEGG functional pathways; pathways present in <10% of participants were removed, leaving 252 KEGG pathways for analysis. The 'DESeq' function in DESeq2 was used to test for differentially abundant KEGG pathways by smoking status, adjusting for sex, age and data set. Spearman's rank correlation was used to examine associations between pathways significantly associated with smoking status and genera significantly associated with smoking status.

All statistical tests were two-sided, and a *P*-value of less than 0.05 (or false discovery rate (FDR) adjusted *q* less than 0.05) was considered statistically significant. All analyses were carried out using SAS 9.3 (SAS institute, Cary, NC, USA) and R 3.2.0.



## Results

Oral biospecimens, smoking history and other demographic information were collected at baseline for both the CPS-II and PLCO cohorts. Of the 1204 individuals included in the present analyses, 43.3% ( $n=521$ ) were never smokers, 47.4% ( $n=571$ ) were former smokers and 9.3% ( $n=112$ ) were current smokers. Among the former smokers, 17% had quit within 10 years of sample collection, 47% 10–30 years prior to sample collection and 30% over 30 years prior to sample collection (6% were missing this information). Current and former smokers tended to be younger and more often male, compared with never smokers (Table 1). All subjects included in the present analyses were originally selected from the CPS-II and PLCO cohorts as cases and controls for nested case–control studies of the oral microbiome in relation to two smoking-related diseases, head and neck cancer and pancreatic cancer. Therefore, approximately half (49.3%) of the 1204 subjects were individuals who went on to develop one of these cancers at some point after providing an oral sample (ranging from 1 to 12 years after oral sample collection); future cancer status was not a confounder of the observed smoking status–microbiome relationships. Additionally, results were highly similar between cases and controls (data not shown).

To determine whether overall microbiome composition differed according to never, former and current smoking status, we conducted principal coordinate analysis based on UniFrac phylogenetic distances. We found a significant difference in composition between current, former and never smokers ( $P=0.001$ , permutational MANOVA based on weighted UniFrac; Figure 1a), after controlling for age, sex and data set. Former and never smokers overlapped on the principal coordinate analysis plot (Figure 1a); when combining former and never smokers, we observed a significant difference in oral microbiome composition between current and non-current (former+never) smokers ( $P=0.001$ ; Figure 1b). To further support these findings, comparison of within- and between-group distances for all smoking categories indicated that current smokers tended to be more heterogeneous than former or never smokers (Figure 1c), and that never and former smokers are more alike than current smokers ( $P=0.001$  and  $P=0.003$ , respectively; Figure 1d). Results were similar with the unweighted UniFrac distances (data not shown).

We next examined the relative abundance of individual taxa according to smoking status using the Kruskal–Wallis test, and found that relative abundance of the phylum Proteobacteria differed significantly by smoking status (meta-analysis for the four data sets,  $P=2.29 \times 10^{-9}$ , FDR adjusted  $q=1.15 \times 10^{-8}$ ; Table 2), with a clear depletion of Proteobacteria among current smokers (data set-weighted median: 11.7% in never smokers, 4.6% in current smokers). After adjustment for age and sex in polytomous logistic regression, the

lower relative abundance of phylum Proteobacteria among current smokers compared with never smokers was statistically significant ( $q=5.24 \times 10^{-7}$ ), while there was no difference between former and never smokers ( $q=0.60$ ). As cross-validation, we assessed these relationships in the four data sets separately and found highly consistent results (Table 2). Aside from Proteobacteria, we observed that relative abundance of the phylum Actinobacteria was elevated in current smokers compared with never smokers after adjustment for age and sex ( $q=0.04$ ) (Table 2), while relative abundance of the other major phyla did not differ significantly by smoking status (Supplementary Table S3).

Lower-level analysis within Proteobacteria revealed that relative abundances of the major classes, Betaproteobacteria and Gammaproteobacteria, were significantly lower in current smokers (Table 2;  $q=5.05 \times 10^{-15}$  and  $q=3.62 \times 10^{-5}$ , respectively), as were several genera, including *Neisseria*, *Haemophilus* and *Aggregatibacter* (Supplementary Table S4). Several taxa not belonging to Proteobacteria, including the class Flavobacteriia ( $q=0.0003$ ; Table 2) and genus *Capnocytophaga*, were also significantly lower in smokers, as were several genera not altered in class-level analysis, including *Corynebacterium* (Actinobacteria), *Porphyromonas* and *Prevotella* (Bacteroidetes), *Leptotrichia* (Fusobacteria), and *Peptostreptococcus*, *Abiotrophia* and *Selenomonas* (Firmicutes) (Supplementary Table S4). Gram-positive class Coriobacteriia tended to be enriched in current smokers (Table 2). Lower-level analysis indicated that the genus *Atopobium*, belonging to Coriobacteriia, as well as the genera *Bifidobacterium* (Actinobacteria), *Lactobacillus* and *Streptococcus* (Bacilli), were increased in current smokers (Supplementary Table S4).

The majority of the findings in our relative abundance analysis were confirmed in a parallel analysis of raw abundance using DESeq2 (Supplementary Table S5), with some exceptions. Notably, the genus *Lactobacillus* was not identified as differentially abundant in the DESeq analysis, while the Firmicutes phylum was found to be significantly more abundant in current compared with never smokers. At the OTU level, 249 OTUs were identified as differentially abundant between current and never smokers at the  $q<0.05$  cutoff (Supplementary Table S6; Figure 2). These included 42 OTUs within Actinobacteria (generally enriched in current smokers), 95 OTUs from Bacilli (mostly from the *Streptococcus* genus, generally enriched in current smokers), 32 OTUs from Clostridia (generally depleted in current smokers), 25 OTUs from Proteobacteria (generally depleted in current smokers) and 27 unclassified OTUs. In contrast, only 17 OTUs were identified as differentially abundant between former and never smokers.

We further examined whether class-level bacterial relative abundances differed according to number of cigarettes smoked per day and years since smoking cessation. We observed inverse associations of the relative abundances of Betaproteobacteria,

**Table 1** Selected demographic characteristics of study populations from four data sets

	Never smokers	Former smokers	Current smokers	P-value
<i>CPS-II-a</i> (n = 203)				
Age, mean (s.d.)	n = 77 70.53 (6.15)	n = 103 70.71 (6.23)	n = 23 68.82 (6.91)	0.34 <sup>a</sup>
Sex, n (%)				0.02 <sup>b</sup>
Male	47 (61.04)	82 (79.61)	16 (69.57)	
Female	30 (38.96)	21 (20.39)	7 (30.43)	
Race, n (%)				NA
White	77 (100)	103 (100)	23 (100)	
Non-White	0	0	0	
BMI (kg m <sup>-2</sup> ), n (%)				0.57 <sup>c</sup>
Under weight (<18.5)	1 (1.30)	1 (0.97)	0	
Normal weight (18.5–25)	29 (37.66)	38 (36.89)	7 (30.43)	
Overweight (25–30)	33 (42.86)	37 (35.92)	6 (26.09)	
Obese weight (>30)	10 (12.99)	21 (20.39)	3 (13.04)	
Missing	4 (5.19)	6 (5.83)	7 (30.43)	
<i>CPS-II-b</i> (n = 340)				
Age, mean (s.d.)	n = 150 74.80 (5.81)	n = 177 72.97 (5.34)	n = 13 70.77 (6.66)	1.40E – 4 <sup>a</sup>
Sex, n (%)				1.00E – 4 <sup>b</sup>
Male	60 (40.00)	112 (63.28)	8 (61.54)	
Female	90 (60.00)	65 (36.72)	5 (38.46)	
Race, n (%)				0.42 <sup>c</sup>
White	142 (94.67)	171 (96.61)	12 (92.31)	
Non-White	8 (5.33)	6 (3.39)	1 (7.69)	
BMI (kg m <sup>-2</sup> ), n (%)				0.02 <sup>c</sup>
Under weight (<18.5)	5 (3.33)	1 (0.56)	1 (7.69)	
Normal weight (18.5–25)	60 (40.00)	64 (36.16)	8 (61.54)	
Overweight (25–30)	57 (38.00)	81 (45.76)	1 (7.69)	
Obese weight (>30)	26 (17.33)	27 (15.25)	3 (23.08)	
Missing	2 (1.33)	4 (2.26)	0	
<i>PLCO-a</i> (n = 261)				
Age, mean (s.d.)	n = 100 63.45 (5.29)	n = 126 63.71 (5.10)	n = 35 60.77 (4.66)	0.01 <sup>a</sup>
Sex, n (%)				2.00E – 4 <sup>b</sup>
Male	63 (63.00)	107 (84.92)	30 (85.7)	
Female	37 (37.00)	19 (15.08)	5 (14.3)	
Race, n (%)				0.51 <sup>c</sup>
White	96 (96.00)	119 (94.44)	32 (91.4)	
Non-White	4 (4.00)	5 (5.56)	3 (8.5)	
BMI (kg m <sup>-2</sup> ), n (%)				0.33 <sup>c</sup>
Under weight (<18.5)	1 (1.00)	0	0	
Normal weight (18.5–25)	36 (36.00)	36 (28.57)	17 (48.57)	
Over weight (25–30)	45 (45.00)	58 (46.03)	14 (40.00)	
Obese weight (>30)	18 (18.00)	27 (21.43)	4 (11.43)	
Missing	0	4 (3.97)	0	
<i>PLCO-b</i> (n = 400)				
Age, mean (s.d.)	n = 194 64.25 (5.24)	n = 165 63.98 (5.32)	n = 41 61.02 (4.7)	2.00E – 3 <sup>a</sup>
Sex, n (%)				2.00E – 3 <sup>b</sup>
Male	102 (47.42)	109 (66.06)	32 (78.05)	
Female	92 (52.58)	56 (33.94)	9 (21.95)	
Race, n (%)				0.04 <sup>b</sup>
White	179 (92.27)	153 (92.73)	33 (80.49)	
Non-White	15 (7.73)	12 (7.27)	8 (19.51)	
BMI (kg m <sup>-2</sup> ), n (%)				0.29 <sup>c</sup>
Under weight (<18.5)	1 (0.52)	2 (1.21)	0	
Normal weight (18.5–25)	68 (35.05)	41 (24.85)	16 (39.02)	
Over weight (25–30)	77 (39.69)	69 (41.82)	15 (36.59)	
Obese weight (>30)	45 (23.2)	50 (30.30)	9 (21.95)	
Missing	3 (1.55)	3 (1.82)	1 (2.44)	

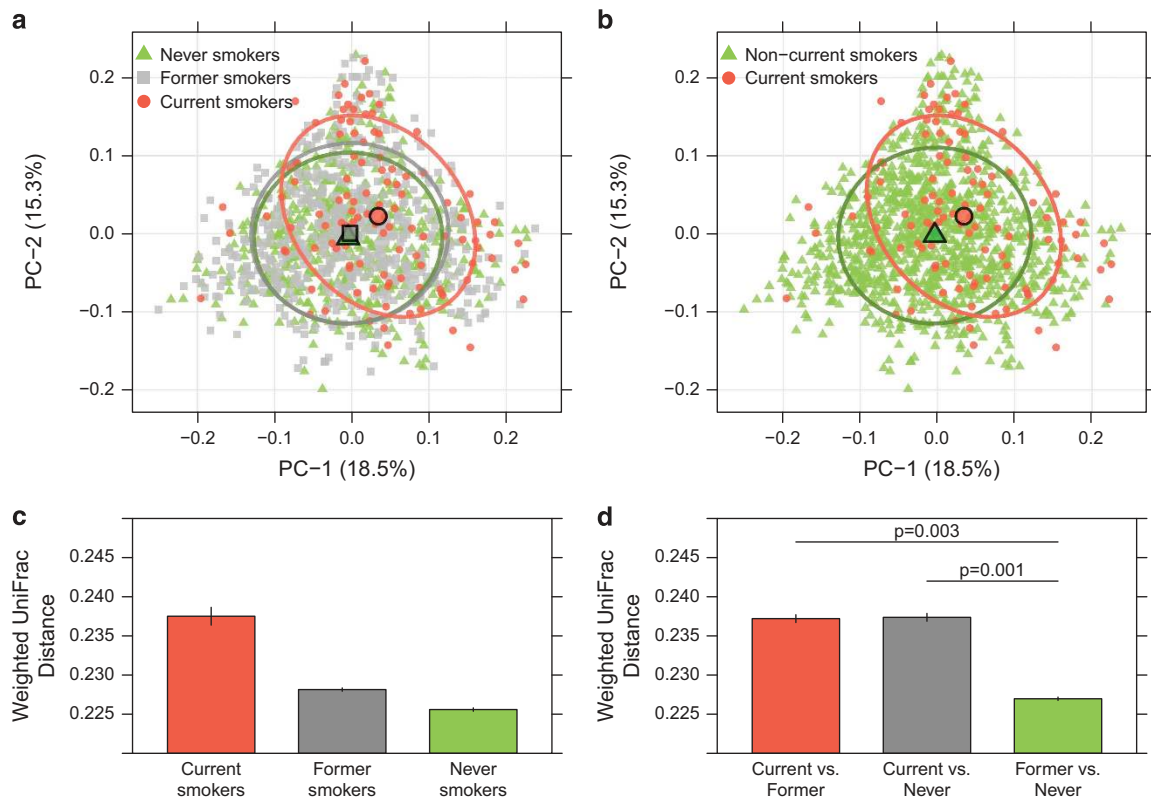
<sup>a</sup>P-values are based on ANOVA test.

<sup>b</sup>P-values are based on Chi-square test.

<sup>c</sup>P-values are based on Fisher's exact test.

Gammaproteobacteria and Flavobacteriia with greater number of cigarettes smoked per day in an analysis including current and never smokers (Figure 3a;  $q = 1.13 \times 10^{-17}$ ,  $q = 4.64 \times 10^{-7}$  and

$q = 1.29 \times 10^{-6}$ , respectively), and positive associations of the relative abundance of Betaproteobacteria, Gammaproteobacteria, and Flavobacteriia with greater number of years since quitting smoking in



**Figure 1** Overall oral microbiome composition according to smoking status (current, former and never) in 1204 individuals. The principal coordinate analysis was conducted based on the weighted UniFrac distance. Sixty-eight percent confidence ellipses were drawn using the panel.ellipse function (Lattice, R), and centroids represent the coordinate mean of the first and second axes. (a) Adjusting for data set, age and sex, there was a significant difference in composition according to smoking status ( $P=0.001$ ). (b) When combining former and never smokers, there was a significant difference in composition between current and non-current smokers ( $P=0.001$ ). Comparison of within (c) and between (d) group distances for all smoking categories indicated that never and former smokers are more alike than are the current smokers.

an analysis including former and current smokers (Figure 3b;  $q=1.96 \times 10^{-10}$ ,  $q=8.13 \times 10^{-4}$  and  $q=8.13 \times 10^{-4}$ , respectively). However, these associations did not remain significant when the cigarettes per day analysis was restricted to current smokers only, and when the years-since-quitting analysis was restricted to former smokers only, indicating that the observed associations were primarily determined by strong differences between current and never smokers, and current and former smokers, respectively.

We also explored microbiota function based on inferred metagenomes using the PICRUST algorithm (Langille *et al.*, 2013). Of 252 KEGG pathways tested, 83 non-human-gene pathways differed in abundance between current and never smokers at  $q<0.05$ , adjusting for sex, age and data set (Supplementary Table S7); these included pathways relating to environmental information processing, carbohydrate and energy metabolism, glycan biosynthesis and metabolism, and xenobiotic biodegradation. Interestingly, pathways related to aerobic metabolism (tricarboxylic acid (TCA) cycle and oxidative phosphorylation) were depleted in current smokers, whereas oxygen-independent pathways (glycolysis, fructose, galactose and sucrose metabolism, and photosynthesis) were

enriched in current smokers. Additionally, abundances of xenobiotic biodegradation pathways were significantly altered in current smokers, including some enriched (polycyclic aromatic hydrocarbon degradation, xylene degradation and drug metabolism) and some depleted (styrene, toluene, nitrotoluene, chlorocyclohexane and chlorobenzene degradation) pathways in current compared with never smokers. Bacterial genera altered in current smokers were related to many of these pathways (Figure 4). For example, genera depleted in current smokers were positively associated with styrene and toluene degradation, the TCA cycle and oxidative phosphorylation, and negatively associated with glycolysis and other carbohydrate metabolism pathways.

## Discussion

In this large meta-analysis of four data sets, we observed that the oral microbiome of current smokers differed substantially from that of never and former smokers. Specifically, at the phylum level we observed a significant depletion of Proteobacteria, and enrichment of Firmicutes and Actinobacteria, in current compared with never smokers. These strong differences at the phylum level resulted

**Table 2** Median relative abundances of selected taxa according to smoking status in four data sets

<i>Taxa</i>	<i>CPS-II-a</i> (n = 203)	<i>CPS-II-b</i> (n = 340)	<i>PLCO-a</i> (n = 261)	<i>PLCO-b</i> (n = 400)	<i>Meta P-value</i> <sup>a</sup> (q-value <sup>b</sup> )	<i>Meta P-value</i> <sup>c</sup> (q-value <sup>d</sup> )	
						<i>Current vs</i> <i>Never</i>	<i>Former vs</i> <i>Never</i>
<i>Proteobacteria (phylum)</i>							
Never	13.18	13.42	10.43	7.4			
Former	10.97	12.55	7.59	7.21			
Current	7.15	2.63	3.62	4.28			
<i>P-value</i> <sup>e</sup>	0.02	0.003	9.29E − 06	0.007	2.29E − 09 (1.15E − 08)	1.05E − 07 (5.24E − 07)	0.27 (0.60)
<i>Betaproteobacteria (class); Proteobacteria (phylum)</i>							
Never	3.15	2.35	2.8	1.86			
Former	2.15	2.22	2.05	2.28			
Current	0.13	0.24	0.28	0.65			
<i>P-value</i> <sup>e</sup>	2.18E − 05	5.07E − 05	2.27E − 08	0.0007	3.88E − 16 (5.05E − 15)	1.02E − 07 (1.02E − 06)	0.17 (0.92)
<i>Gammaproteobacteria (class); Proteobacteria (phylum)</i>							
Never	8.93	8.52	5.23	4.01			
Former	7.02	8.5	3.89	3.96			
Current	6.34	1.87	2.47	2.38			
<i>P-value</i> <sup>e</sup>	0.17	0.03	0.001	0.02	5.57E − 06 (3.62E − 05)	5.73E − 05 (0.0003)	0.52 (0.92)
<i>Flavobacteriia (class); Bacteroidetes (phylum)</i>							
Never	0.46	0.38	0.24	0.29			
Former	0.49	0.39	0.21	0.31			
Current	0.10	0.15	0.08	0.17			
<i>P-value</i> <sup>e</sup>	0.01	0.04	0.001	0.33	6.11E − 05 (0.0003)	0.003 (0.009)	0.31 (0.92)
<i>Actinobacteria (phylum)</i>							
Never	10.59	10.10	15.18	11.62			
Former	11.16	9.30	13.81	11.53			
Current	13.65	13.88	17.46	11.55			
<i>P-value</i> <sup>e</sup>	0.35	0.62	0.36	0.71	0.47 (0.59)	0.02 (0.04)	0.87 (0.87)
<i>Coriobacteriia (class); Actinobacteria (phylum)</i>							
Never	0.36	0.28	0.28	0.33			
Former	0.51	0.33	0.22	0.29			
Current	0.40	0.46	0.51	0.42			
<i>P-value</i> <sup>e</sup>	0.41	0.26	0.02	0.10	0.004 (0.01)	0.01 (0.02)	0.67 (0.92)

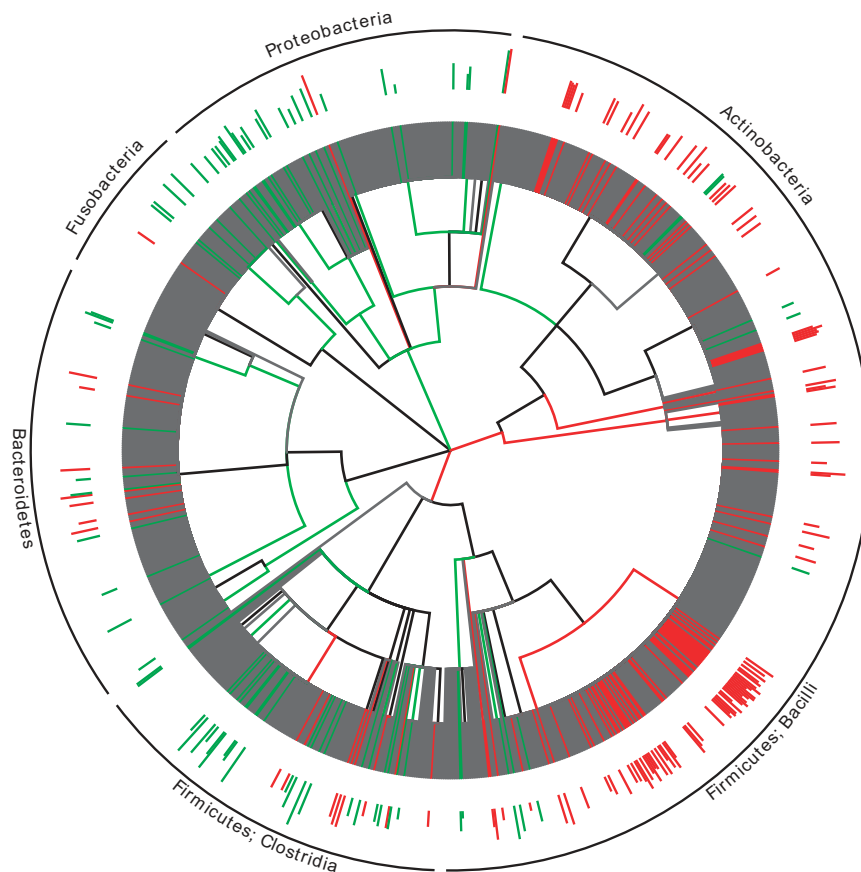
<sup>a</sup>Meta-analysis P-values from Kruskal–Wallis tests within each of the four data sets, calculated using Z-score methods.<sup>b</sup>False discovery rate adjusted q-values were calculated based on the meta-analysis P-values from Kruskal–Wallis tests.<sup>c</sup>Meta-analysis P-values from polytomous logistic regression models within each of the four data sets, calculated using Z-score methods. Age (continuous value) and sex (male, female) were controlled for in the polytomous regression models.<sup>d</sup>False discovery rate adjusted q-values were calculated based on the meta-analysis P-values from polytomous logistic regression models.<sup>e</sup>P-values are based on nonparametric Kruskal–Wallis test of current, former and never smokers.

from phylum-wide differences in OTU abundance between current and never smokers. Analysis of inferred metagenomes indicated that smoking may alter oral microbial ecology through influencing oral oxygen availability, while simultaneously having consequences for microbial degradation of xenobiotics. Finally, we observed that the overall oral microbiome composition of former smokers did not differ from that of never smokers; this is a promising indication that smoking-related changes to the oral microbiome are not permanent. As has been observed with other smoking-related health changes, smoking cessation clearly remains the best practice to restore a healthy phenotype (Godtfredsen and Prescott, 2011).

Given the many toxicants found in cigarette smoke (Rodgman and Perfetti, 2013), it is not surprising that smoking drastically alters the microbial ecology of the mouth. Indeed, several other studies have also

observed effects of smoking on oral bacteria. Early *in vitro* studies using culture-based methods noted that cigarette smoke has a strong inhibitory effect on the growth of *Neisseria* species (Bardell, 1981; Ertel *et al.*, 1991), while *Streptococcus* species were less inhibited by cigarette smoke (Bardell, 1981). Additionally, early studies in humans identified decreased *Neisseria* species on mucosal surfaces of smokers (Colman *et al.*, 1976), and an increased proportion of Gram-positive to Gram-negative bacteria on developing plaques of smokers (Bastiaan and Waite, 1978). Recently, studies with comprehensive oral bacterial profiling in humans have found increased *Streptococcus sobrinus* and *Eubacterium brachy* in the saliva of smokers (Belstrom *et al.*, 2014), decreased *Neisseria*, *Porphyromonas* and *Gemella* in oral wash samples from smokers (Morris *et al.*, 2013), enrichment of *Megasphaera*, *Streptococcus* and *Veillonella*, and depletion of





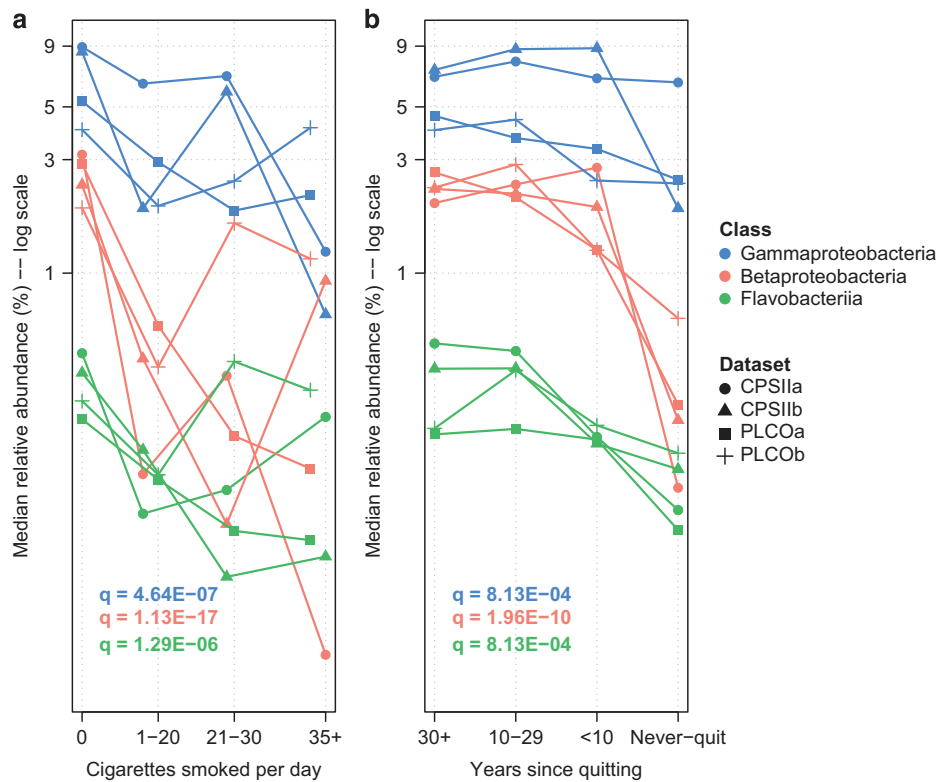
**Figure 2** Cladogram representation of oral microbiome OTUs associated with smoking status. A red branch indicates a taxon or OTU enriched in current smokers and a green branch indicates a taxon or OTU depleted in current smokers, as detected in the DESeq2 analysis. The bars represent log<sub>2</sub> fold changes of counts in current compared with never smokers; red bars indicate positive fold change and green bars indicate negative fold change. A total of 1158 OTUs are included in the cladogram, representing OTUs with at least two sequences in at least 30 subjects in the five major phyla; only OTUs with  $q < 0.05$  are colored. Cladogram was created using EvolView (Zhang *et al.*, 2012).

*Capnocytophaga*, *Fusobacterium* and *Neisseria*, in the oropharynx of smokers (Charlson *et al.*, 2010), and alterations in 172 subgingival plaque OTUs in smokers (Mason *et al.*, 2015). Because of the various sample types used to study the oral microbiome, and the known variation in microbial communities in different parts of the oral cavity (Segata *et al.*, 2012), comparison across studies is difficult. In the current study we have employed Scope mouthwash samples to study the microbiome, which are likely most comparable with studies using saliva or other types of mouthwash samples, and less comparable with plaque samples (Segata *et al.*, 2012). Nevertheless, in similarity with some of the above-mentioned studies, we observed decreases in *Neisseria*, *Porphyromonas* and *Capnocytophaga* and increases in *Veillonella* and *Streptococcus*, in current compared with never smokers. To our knowledge, this is the first study to report phylum- and class-wide associations of bacterial taxa with smoking status; our more robust findings may relate to the large sample size, which provided the power to detect these associations.

There are several potential mechanisms by which smoking may alter microbial ecology, including increasing the acidity of saliva (Parvinen, 1984;

Kanwar *et al.*, 2013), depleting oxygen (Kenney *et al.*, 1975), antibiotic effects (Macgregor, 1989), influencing bacterial adherence to mucosal surfaces (Brook, 2011) and impairing host immunity (Sopori, 2002). Our analysis of inferred metagenomes revealed decreased abundance of aerobic metabolism pathways, including the TCA cycle and oxidative phosphorylation, and increased abundance of glycolysis and other oxygen-independent carbohydrate metabolism pathways, in current smokers compared with never smokers. This finding suggests that cigarette smoke creates an environment favoring strict or facultative anaerobes over strict aerobes. At the genus and OTU level, we observed increased abundance of *Streptococcus* in current smokers; members of the *Streptococcus* genus are facultative or obligate anaerobes (Patterson, 1996) and generally acid tolerant, which may explain their success in the smoking environment. Additionally, we observed smoking-related increases in the anaerobic *Veillonella* genus and Actinobacteria OTUs from anaerobic *Actinomyces* spp., *Rothia mucilaginosa*, *Bifidobacterium longum* and *Atopobium* spp. Conversely, aerobes such as *Neisseria subflava* and *Corynebacterium* were depleted in smokers. Consistent with the oxygen deprivation hypothesis, Mason





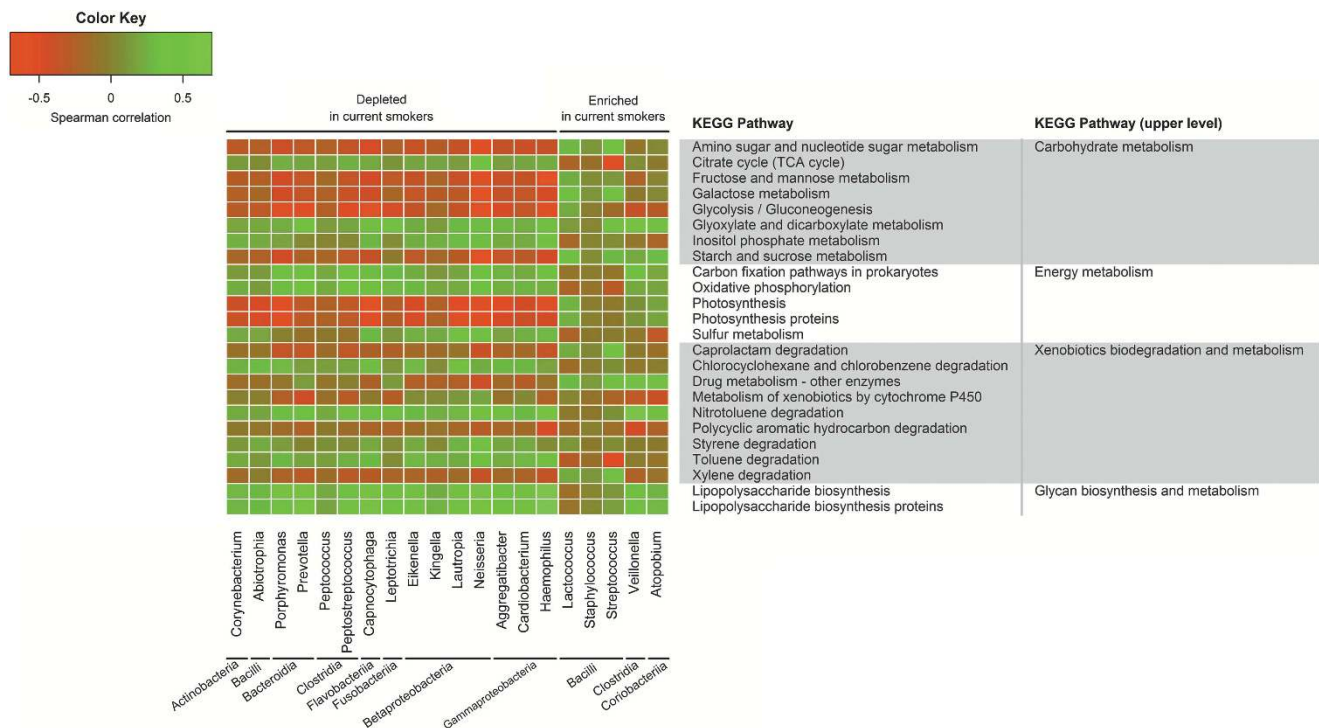
**Figure 3** Median relative abundance of selected taxa according to the number of cigarettes smoked per day and number of years since quitting. Plot (a) includes current and never smokers, while plot (b) includes former and current smokers. False discovery rate adjusted  $q$ -values were calculated based on meta-analysis  $P$ -values of correlations between relative abundance of taxa and number of cigarettes smoked per day or number of years since quitting.

*et al.* observed higher abundance of anaerobes and lower abundance of aerobes in subgingival plaque samples of smokers compared with non-smokers (Mason *et al.*, 2015). Interestingly, we observed depletion of certain anaerobic OTUs in smokers as well, including *Leptotrichia* spp., *Veillonella parvula* and *Peptostreptococcus* sp. It is possible that these bacteria were depleted due to specific antibiotic toxicants in cigarette smoke, or depleted indirectly due to competition for colonization with smoking-enriched bacteria or co-aggregation with smoking-depleted bacteria. Because this is an observational study, we cannot determine which of the altered taxa are directly affected by cigarette smoke or indirectly affected through microbe-microbe interactions.

Aside from creating an anaerobic, acidic and/or selectively toxic environment, smoking is also known to have prominent effects on human immunity (Sopori, 2002), which can in turn influence the host's ability to stave off colonization by pathogens. The chemotactic mobility and phagocytic function of oral polymorphonuclear leukocytes is diminished in smokers (Noble and Penny, 1975; Kenney *et al.*, 1977; Archana *et al.*, 2015); as these cells are crucial to the host defense against pathogens, smoking inherently promotes a more pathogen-friendly oral ecosystem, thus increasing risk for oral disease (e.g. periodontitis). Several of our findings are consistent with progression

towards a diseased state: *Neisseria* and *Eikenella* are depleted in oral mucosa from periodontitis patients (Mager *et al.*, 2003), and *Streptococcus* species are more abundant in periodontal disease-progressing oral sites than healthy oral sites (Yost *et al.*, 2015). However, most of the taxa typically implicated as periodontal pathogens were not affected by smoking in the current study, despite smoking being a strong risk factor for periodontitis. Our use of oral mouthwash samples rather than subgingival samples may account in part for this discrepancy.

The depletion of certain xenobiotic biodegradation pathways in current smokers suggests important functional losses with potential health consequences. The oral bacteria are first to come into contact with cigarette smoke as it enters the human body, and may play an important role in degrading the accompanying toxic compounds. We observed that functional pathways relating to toluene, nitro-toluene, styrene, chlorocyclohexane and chlorobenzene degradation were depleted in current smokers, as was cytochrome P450 xenobiotic metabolism. Conversely, polycyclic aromatic hydrocarbon and xylene degradation were enriched in current smokers. These chemicals are components of cigarette smoke (Rodgman and Perfetti, 2013), and thus alterations in the ability of the oral community to degrade these substances may have toxic



**Figure 4** Bacterial taxa associated with smoking status are related to several gene functional pathways. Bacterial gene functions were predicted from 16S rRNA gene-based microbial compositions using the PICRUST algorithm to make inferences from KEGG annotated databases. Genus and KEGG pathway counts were normalized for DESeq2 size factors and adjusted for data set using the 'removeBatchEffect' function (limma). Spearman's correlation coefficients were estimated for each pairwise comparison of genus counts and KEGG pathway counts, adjusting for age and sex. Only KEGG pathways relating to carbohydrate, energy, xenobiotic and glycan metabolism and selected genera of interest are included in the heatmap; full lists of genera and KEGG pathways associated with smoking can be found in Supplementary Tables S4, S5 and S7.

consequences for the host. It is surprising that some of the xenobiotic degradation pathways are depleted in smokers, given the need for bacterial upregulation of these pathways to detoxify cigarette smoke. This result is also in contrast with a metagenomic study which observed increased cytochrome P450 xenobiotic metabolism in smokers (Boyle *et al.*, 2010). A simple explanation for our finding is that these pathways are carried out in bacteria that were depleted in smokers. Alternatively, the toxic compounds themselves may saturate the enzymes responsible for their degradation, thus killing the bacteria possessing these enzymes. Although the long-term effects of smoking-related oral dysbiosis remain unclear at this time, oral bacteria are known to play an important role in both oral and systemic diseases (Wade, 2013; Olsen, 2015). It is therefore not unreasonable to imagine that changes in the oral bacterial community due to smoking may have detrimental health effects.

We additionally observed that oral bacteria abundances were generally similar between former and never smokers, implying that specific bacteria depleted by smoking may be restored following smoking cessation. Interestingly, a small number of OTUs were identified as differentially abundant between former and never smokers, including a few that were altered in the same direction as in

current smokers. This finding may indicate some minor lingering effects of smoking. We did not observe an association between years since smoking cessation and bacterial class relative abundances in analyses restricted to former smokers. The absence of a clear trend with years since quitting among former smokers may be due to restoration of the oral microbiome occurring relatively quickly following smoking cessation, for example, during the first year or two immediately after quitting. We had insufficient data on the precise timing of smoking cessation to examine potential trends in oral microbiome composition during the period within a few years of quitting. Moreover, the effect of smoking cessation on the oral microbiome would be better studied with longitudinally collected oral wash samples, which would allow for within-person comparison of the oral microbiome pre- and post-smoking cessation. A few studies have examined the effect of smoking cessation on subgingival plaque bacteria longitudinally (Fullmer *et al.*, 2009; Delima *et al.*, 2010); however, these studies were limited by the extent of bacterial profiling. A longitudinal investigation of smoking cessation involving more extensive bacterial profiling of the oral microbiome (i.e. 16S rRNA gene sequencing) will be important to determine which taxa recolonize the oral environment after smoking cessation.

In summary, in this large study of the human oral microbiome, we observed that smoking is related to overall oral microbiome community composition, and to the abundance of many taxa. Smoking may promote an anaerobic oral environment and a bacterial community with reduced xenobiotic degradation capabilities. Strengths of this study include the large sample size, the ability to check replication of findings in four data sets and the control of potential confounders. This study was limited by lack of metagenomic data to determine the actual gene content of bacteria altered by smoking and lack of longitudinal samples pre- and post-smoking cessation. Additionally, due to the elderly age of our study participants, our findings may not be generalizable to younger populations, particularly since the oral microbiome changes with age (Xu *et al.*, 2015). Future studies should investigate the impact of smoking on the metagenomic content of the oral microbiome, and whether smoking-related oral bacterial and/or metagenomic changes mediate the health effects of smoking.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

Research reported in this publication was supported in part by the US National Cancer Institute under award numbers R01CA159036, U01CA182370, U01CA170948-01A1, R01CA164964, R03CA159414, P30CA016087, R21CA183887, and by AACR/Pancreas Cancer Action Network Career Development Award. ZP is a Staff Physician at the Department of Veterans Affairs New York Harbor Healthcare System. The American Cancer Society (ACS) funds the creation, maintenance and updating of the Cancer Prevention Study II cohort. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the US Department of Veterans Affairs or the United States Government.

## References

Ahn J, Chen CY, Hayes RB. (2012). Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes Control* **23**: 399–404.

Archana MS, Bagewadi A, Keluskar V. (2015). Assessment and comparison of phagocytic function and viability of polymorphonuclear leukocytes in saliva of smokers and non-smokers. *Arch Oral Biol* **60**: 229–233.

Bardell D. (1981). Viability of six species of normal oropharyngeal bacteria after exposure to cigarette smoke in vitro. *Microbios* **32**: 7–13.

Bastiaan RJ, Waite IM. (1978). Effects of tobacco smoking on plaque development and gingivitis. *J Periodontol* **49**: 480–482.

Beck JM, Young VB, Huffnagle GB. (2012). The microbiome of the lung. *Transl Res* **160**: 258–266.

Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, Pignatelli M *et al.* (2012). The oral metagenome in health and disease. *ISME J* **6**: 46–56.

Belstrom D, Holmstrup P, Nielsen CH, Kirkby N, Twetman S, Heitmann BL *et al.* (2014). Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. *J Oral Microbiol* **6**: 23609.

Boyle JO, Gumus ZH, Kacker A, Choksi VL, Bocker JM, Zhou XK *et al.* (2010). Effects of cigarette smoke on the human oral mucosal transcriptome. *Cancer Prev Res (Phila)* **3**: 266–278.

Brook I. (2011). The impact of smoking on oral and nasopharyngeal bacterial flora. *J Dent Res* **90**: 704–710.

Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML *et al.* (2002). The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**: 2490–2501.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R *et al.* (2010). Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One* **5**: e15216.

Colman G, Beighton D, Chalk AJ, Wake S. (1976). Cigarette smoking and the microbial flora of the mouth. *Aust Dent J* **21**: 111–118.

Delima SL, McBride RK, Preshaw PM, Heasman PA, Kumar PS. (2010). Response of subgingival bacteria to smoking cessation. *J Clin Microbiol* **48**: 2344–2349.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH *et al.* (2010). The human oral microbiome. *J Bacteriol* **192**: 5002–5017.

Ertel A, Eng R, Smith SM. (1991). The differential effect of cigarette smoke on the growth of bacteria found in humans. *Chest* **100**: 628–630.

Evangelou E, Ioannidis JP. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* **14**: 379–389.

Fullmer SC, Preshaw PM, Heasman PA, Kumar PS. (2009). Smoking cessation alters subgingival microbial recolonization. *J Dent Res* **88**: 524–528.

Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. (2013). Active smoking and breast cancer risk: original cohort data and meta-analysis. *J Natl Cancer Inst* **105**: 515–525.

Godtfredsen NS, Prescott E. (2011). Benefits of smoking cessation with focus on cardiovascular and respiratory comorbidities. *Clin Respir J* **5**: 187–194.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G *et al.* (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**: 494–504.

Hayes RB, Reding D, Kopp W, Subar AF, Bhat N, Rothman N *et al.* (2000). Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin Trials* **21**: 349S–355S.

He J, Li Y, Cao Y, Xue J, Zhou X. (2015). The oral microbiome diversity and its relation to human diseases. *Folia Microbiol* **60**: 69–80.

Hocking WG, Hu P, Oken MM, Winslow SD, Kvale PA, Prorok PC *et al.* (2010). Lung cancer screening in the



- randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *J Natl Cancer Inst* **102**: 722–731.
- Hugoson A, Hellqvist L, Rolandsson M, Birkhed D. (2012). Dental caries in relation to smoking and the use of Swedish snus: epidemiological studies covering 20 years (1983–2003). *Acta Odontol Scand* **70**: 289–296.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109–D114.
- Kanwar A, Sah K, Grover N, Chandra S, Singh RR. (2013). Long-term effect of tobacco on resting whole mouth salivary flow rate and pH: an institutional based comparative study. *Eur J Gen Dent* **2**: 296–299.
- Kenney EB, Kraal JH, Saxe SR, Jones J. (1977). The effect of cigarette smoke on human oral polymorphonuclear leukocytes. *J Periodontol Res* **12**: 227–234.
- Kenney EB, Saxe SR, Bowles RD. (1975). The effect of cigarette smoking on anaerobiosis in the oral cavity. *J Periodontol* **46**: 82–85.
- Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V et al. (2011). Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci USA* **108** (Suppl): 4592–4598.
- Kumar PS, Matthews CR, Joshi V, de Jager M, Aspiras M. (2011). Tobacco smoking affects bacterial acquisition and colonization in oral biofilms. *Infect Immun* **79**: 4730–4738.
- Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–821.
- Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.
- Macgregor ID. (1989). Effects of smoking on oral ecology. A review of the literature. *Clin Prev Dent* **11**: 3–7.
- Mager DL, Haffajee AD, Socransky SS. (2003). Effects of periodontitis and smoking on the microbiota of oral mucous membranes and saliva in systemically healthy subjects. *J Clin Periodontol* **30**: 1031–1037.
- Mason MR, Preshaw PM, Nagaraja HN, Dabdoub SM, Rahman A, Kumar PS. (2015). The subgingival microbiome of clinically healthy current and never smokers. *ISME J* **9**: 268–272.
- McArdle BH, Anderson MJ. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**: 290–297.
- McMurdie PJ, Holmes S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* **10**: e1003531.
- Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL et al. (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* **187**: 1067–1075.
- Noble RC, Penny BB. (1975). Comparison of leukocyte count and function in smoking and nonsmoking young men. *Infect Immun* **12**: 550–555.
- Nociti FH, Casati MZ, Duarte PM. (2015). Current perspective of the impact of smoking on the progression and treatment of periodontitis. *Periodontology 2000* **67**: 187–210.
- Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ et al. (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol* **16**: 4135–4144.
- Olsen I. (2015). From the acta prize lecture 2014: the periodontal-systemic connection seen from a microbiological standpoint. *Acta Odontol Scand* **73**: 563–568.
- Parvinen T. (1984). Stimulated salivary flow rate, pH and lactobacillus and yeast concentrations in persons with different types of dentition. *Scand J Dent Res* **92**: 412–418.
- Patterson MJ. (1996). Streptococcus. In: Baron S (ed). *Medical Microbiology*, 4th edn. University of Texas Medical Branch at Galveston: Galveston, TX.
- Rodgman A, Perfetti TA. (2013). The alphabetical index to components identified in tobacco, tobacco smoke, and tobacco substitute smoke. In: *The Chemical Components of Tobacco and Tobacco Smoke*, 2nd edn. CRC Press: Boca Raton, FL, pp 1713–2064.
- Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D et al. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* **13**: R42.
- Sopori M. (2002). Effects of cigarette smoke on the immune system. *Nat Rev Immunol* **2**: 372–377.
- Wade WG. (2013). The oral microbiome in health and disease. *Pharmacol Res* **69**: 137–143.
- WHO. (2012). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. IARC Monographs vol. 100 (E).
- Wu J, Lin I, Hayes RB, Ahn J. (2014). Comparison of DNA extraction methods for human oral microbiome research. *Br J Med Med Res* **4**: 1980–1991.
- Xu X, He J, Xue J, Wang Y, Li K, Zhang K et al. (2015). Oral cavity contains distinct niches with dynamic microbial communities. *Environ Microbiol* **17**: 699–710.
- Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. (2015). Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial meta-transcriptome analysis. *Genome Med* **7**: 27.
- Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **40**: W569–W572.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)