The effect of smoking on microbiome composition of the anterior nares

Angelika Wassmuth
University of Münster
Hansjörg Baurecht
University of Regensburg
Heiko Becher
University Hospital Heidelberg
Klaus Berger
University of Münster
Hermann Brenner
German Cancer Research Center (DKFZ)
Kathrin Günther
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Volker Harth
University Medical Center Hamburg-Eppendorf
Bernd Holleczek
Krebsregister Saarland
Rudolf Kaaks
German Cancer Research Center (DKFZ)
André Karch
University of Münster
Thomas Keil
Charité – Universitätsmedizin Berlin
Yvonne Kemmling
Helmholtz Centre for Infection Research
Muhammad N. K. Khattak
University of Greifswald
Bianca Klee
Interdisciplinary Center for Health Sciences, Medical School of the Martin-Luther University Halle-Wittenberg
Carolina Klett-Tammen
Helmholtz Centre for Infection Research
Lilian Krist
Charité – Universitätsmedizin Berlin

Berit Lange
Helmholtz Centre for Infection Research
Michael Leitzmann
University of Regensburg
Wolfgang Lieb
Christian-Albrechts-University Kiel
Karin B. Michels
Faculty of Medicine and Medical Center, University of Freiburg
Rafael Mikolajczyk
Interdisciplinary Center for Health Sciences, Medical School of the Martin-Luther University Halle-Wittenberg
Matthias Nauck
University Medicine Greifswald
Katharina Nimptsch
Max Delbrück Center for Molecular Medicine (MDC)
Nadia Obi
University Medical Center Hamburg-Eppendorf
Annette Peters
Helmholtz Center Munich, German Research Center for Environmental Health (GmbH)
Tobias Pischon
Max Delbrück Center for Molecular Medicine (MDC)
Tamara Schikowski
IUF - Leibniz Institute for Environmental Medicine
Börge Schmidt
University Hospital of Essen, University of Duisburg-Essen
Matthias B. Schulze
German Institute of Human Nutrition Potsdam-Rehbruecke
Julia Six-Merker
Helmholtz Center Munich, German Research Center for Environmental Health (GmbH)
Kerstin Wirkner
Leipzig University
Maike Wolters
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Sabrina Woltemate
Hannover Medical School
Marius Vital
Hannover Medical School
Nicole Rübsamen
University of Münster

Sven Kleine Bardenhorst

kleineba@uni-muenster.de

University of Münster

Research Article

Keywords: 16S rRNA gene, nasal microbiota, smoking

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10428906/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Smoking is the single largest risk factor for ill-health and premature death globally. While direct toxic effects on the respiratory epithelium are well understood, smoking also changes the microbial environment, resulting in an increased risk of infections. Previous analyses of the nasal microbiome suggested an increased abundance of anaerobic, potentially pathogenic taxa, but were limited by small participant numbers and unclear comparison groups. In this study we use data from the German National Cohort (NAKO) to investigate the association between smoking and the anterior nares' microbiome in a general population sample.

Methods

We analysed participants' nasal swabs from 1,948 participants (16.8% current and 35.6% former smokers) at the NAKO baseline examination with 16S rRNA sequencing of the V3-V4 hypervariable regions. The USEARCH pipeline was applied for taxonomic profiling. Linear regression was performed for the association between smoking habits and species richness and evenness. PERMANOVA was used to assess the explained microbiome variance due to smoking behaviours. The pairwise Wilcoxon rank sum test, MaAsLin3 and a Random Forest model were employed to identify differentially abundant taxa. Latent Dirichlet allocation was utilised to identify groups of taxa potentially sharing biologically similar functions and their topic distribution was compared between smoking statuses.

Results

The most prevalent genera were *Corynebacterium* (100%), *Staphylococcus* (100%), and *Cutibacterium* (99.9%). Current smokers exhibited a higher diversity compared to never smokers (Shannon index: Cohen's d = 0.20, 95% CI [0.07–0.32]; Simpson index: Cohen's d = 0.23, 95% CI [0.11–0.36]). Differential abundance analysis identified a higher relative abundance of anaerobic pathobionts in smokers (current and former), most strongly for the taxa *Anaerococcus, Finegoldia, and Peptoniphilus (a)*. Latent Dirichlet allocation identified six topics whose distributions varied between smoking status and corresponded to the results from the differential abundance analysis. Former smokers showed patterns consistent with partial normalization.

Conclusions

We found in our population-based study on the association between smoking and the anterior nares' microbiome higher microbial diversity and anaerobic abundance in smokers, compatible with a higher

susceptibility for respiratory infections. Our results can inform the underlying mechanisms of why smokers are more susceptible to infection.

Background

Smoking

Smoking is the leading risk factor for premature mortality and morbidity globally, resulting in an estimated 7.69 million deaths and 200 million disability-adjusted life-years in 2019 [1–3]. This is due to smoking causing or being a contributing factor in the development of cardiovascular disease, various forms of cancer, particularly lung cancer, chronic obstructive pulmonary disease (COPD), and respiratory tract infections (RTI) [2, 4]. In Germany, the prevalence of smoking was 29.6% in 2019, with an age-standardized prevalence of 25.1% among women and 34.2% among men [5]. Several factors influence smoking habits, including age, sex, SES, ethnicity, and seasonality [1, 3, 5–10].

Associations between smoking and the nasal microbiome

The nasal cavity and its microbiome play an integral role in the immune and respiratory systems, and is called the “gatekeeper” of the respiratory tract [11, 12]. The nasal microbiome is characterised by being dominated by one of three taxa (*Staphylococcus*, *Corynebacterium* or *Cutibacterium*), with the remaining composition being highly individual and influenced by several host factors such as age, sex, antibiotic use, and smoking [13]. However, most evidence on how smoking affects the nasal microbiome originate from small cross-sectional studies and does not sufficiently capture the heterogeneity in microbiome variation that exists in the general population. Charlson et al. [14] identified a higher abundance of Gram-positive anaerobic microorganisms in the nasopharynx of current smokers compared to never smokers. Moreover, Pfeiffer et al. [15] identified an enrichment of *Staphylococcus epidermidis* in smokers compared to never-smokers, and positive correlations between the number of years of smoking and the relative abundance of *Cutibacterium*. A higher number of packyears was associated with a decrease in *Corynebacterium*, and several anaerobic pathobionts were positively correlated with smoking habits defined as packyears, nicotine, or cotinine levels [15]. These studies suggest that smoking changes the oxygen levels in the nose, thus encouraging the growth of anaerobic, potentially pathogenic taxa [2]. Both studies utilised 16S ribosomal ribonucleic acid (rRNA) sequencing, though Charlson et al. [14] utilised the V1-V2 hypervariable regions and Pfeiffer et al. [15] the V4 region for taxonomic profiling. These regions differ in their taxonomic resolution, limiting their comparability. Further, both studies had modest sample sizes (n = 62 and n = 40) limiting their statistical power.

The Dutch population-based cohort study by Odendaal et al. [16] used both smoking status and the weekly number of cigarettes smoked to identify an association on the nasopharyngeal microbiome in 2,485 participants. A higher abundance of *Peptoniphilus, Finegolida, Corynebacterium*, and *Campylobacter* was identified in smokers (current and former) with a positive association between *Peptoniphilus* and *Finegoldia* and the weekly number of cigarettes smoked.

All previous studies investigated samples obtained from the nasopharynx. However, to further elucidate the association of smoking on the nasal microbiome, particularly at the site, which is most influenced by environmental factors, the associations with the anterior nares' microbiome need to be assessed.

Aim

The aim of the present study is to elucidate the associations between smoking and the anterior nares' microbiome in the German National Cohort (NAKO). This facilitates a better understanding of the underlying mechanisms which make smokers more prone to infections.

Methods

Sample

The study sample originates from the German National Cohort (NAKO; Application No. NAKO-423). A random sample of 2,400 participants was drawn from the 205,000 participants of the NAKO [17] baseline examination across all study centres. To be selected as part of the sample, participants needed to have triplets of biomaterials (stool, oral, and nasal samples) and specific questionnaires on food allergies, dietary habits, diseases, and cancer diagnosis needed to be complete. The sample size was chosen due to feasibility [13]. For 2,072 participants, sequencing data from the nasal swabs was available. A complete case analysis was performed based on the completeness of smoking status and the confounder variables identified via directed acyclic graph (DAG; see section Statistical Analysis). This led to 1,948 participants being included in the current analysis.

Smoking behaviour

Smoking behaviour was assessed via touch screen questionnaire, with participants self-reporting their smoking status as never, current, or former smoker. For current and former smokers, the lifetime number of cigarettes was used to assess a dose-response by estimating the mean number of daily cigarettes smoked throughout the participants' smoking career.

Anterior nares' sampling

Anterior nares' sampling was performed as described in detail by Kleine Bardenhorst et al. [13]. Briefly, nasal swabbing was performed at baseline assessment with one swab for each nostril after insertion into 0.9% saline solution. Swabbing was performed at a depth of 1 to 1.5 cm. The swabs were placed in the same cryotube containing 0.2 ml of RNAlater® (QIAGEN, Austin, Texas, USA) and the swab handle was broken off at the breakpoint. The samples were stored at 4°C until the end of the day and transferred to -80°C storage until DNA isolation.

DNA isolation and 16S rRNA sequencing

As described in more detail [13], one of the two swabs of each participant was randomly chosen and used for DNA extraction using the ZymoBIOMICS 96 DNA Kit (Zymo Research, Irvine, California, USA). Mechanical lysis was performed by bead beating. Amplicon libraries targeting the V3 to V4 hypervariable region of the 16S rRNA gene were generated by performing polymerase chain reaction (PCR) using the 341F and 805R primers. PCR products were purified and dual indices and Illumina adapters were attached in a second PCR. Purified libraries were pooled and sequenced on an Illumina MiSeq platform Sequencer (Illumina, San Diego, California, USA) using V3 chemistry (2x 300 bp, 5.5 pM loading concentration, 8% PhiX).

Bioinformatics

Demultiplexing of the raw sequencing reads was performed using idemp [18]. The USEARCH pipeline, version 11.9.667 [19, 20], was used to process the sequencing reads into biological sequences (zero-radius operational taxonomic units; zOTUs) and their respective counts. The paired-end reads were merged, low quality reads, and singletons were filtered out. The GreenGenes2 database [21] was used with Constax [22] to assign taxonomy and summarise the data into a biom-file. See Kleine Bardenhorst et al.[13] for more details.

Statistical analysis

A complete case analysis was performed, with smoking status (never, current, and former smokers) or the lifetime number of cigarettes, only for current and former smokers, as exposure variable. Age, sex, season, and SES were confounder variables. Confounder variables were identified via a directed acyclic graph (DAG; Supplementary File Fig. 1). The ISEI-08 (International Socioeconomic Index of Occupational Status) [23] was used to assess the SES. It is calculated based on the domains education, occupation, and household income, and is scored on a continuous scale ranging from 16 to 90, with higher scores indicating a higher SES.

Abundance counts of the zOTUs were transformed into centred log-ratio (clr) counts and a pseudocount of 0.5 was applied. All taxa which were unclassified at phylum level were filtered out, and a 0.5% prevalence filter was applied at a relative abundance of 0.005%. For all analyses, Benjamini-Hochberg correction was applied to account for multiple testing [24].

The association between smoking habits and each of the five alpha diversity metrics—observed richness, Chao1 index, Shannon index, Simpson index, and Faith's phylogenetic diversity (Faith's PD)—was assessed with a linear regression model with never smokers as reference category for the dummy coded categories current and former smokers, and for the lifetime number of cigarettes in ever smokers. The alpha diversity metrics were calculated at zOTU level. The models were adjusted for the

confounders as well as the sequencing depth to account for different library sizes of samples. Cohen's d was calculated to estimate effect sizes.

Permutational multivariate analysis of variance (PERMANOVA) was performed using the *adonis2* function [25] with Aitchison distance, Bray-Curtis dissimilarity, unweighted and weighted UniFrac distances. PERMANOVA models were performed as univariate models and an omnibus test for the exposure and confounders. Homogeneity of variance was assessed with the *permutest* [25].

Three differential abundance analysis (DAA) methods were performed on genus level after applying a 10% prevalence threshold. First, the Wilcoxon rank sum tests were calculated for pairwise comparison to identify differentially abundant taxa between current, former, and never smokers. Second, Microbiome Multivariable Associations with Linear Models 3 (MaAsLin3) [26] was applied to the abundance matrix with total sum scaling and log transformation. Z-score standardisation was performed for continuous variables. MaAsLin3 was performed with never smokers as reference category for the dummy coded categories current and former smokers and for the lifetime number of cigarettes in ever smokers. Third, a Random Forest model was applied with a classification model for non-dummy coded smoking status and a regression model for the lifetime number of cigarettes. The data was split into a training and test set (80% vs 20%), stratified by the smoking status. The Ranger engine was used to train 500 decision trees [27]. The Boruta algorithm [28] was applied for variable selection, and model performance was evaluated using 5-fold cross-validation. The final model was chosen by maximising accuracy during the tuning process.

Taxa were classified as differentially abundant if they were detected by all three methods. The results from MaAsLin3 were used to assess the strength of association.

Latent Dirichlet Allocation (LDA) was performed to identify groups of genera potentially sharing similar biological functions. The optimal number of clusters (k) between 2 and 15 was identified by minimising the mean cosine distance between topics [29] and maximising the Kullback-Leibler divergence [30]. Additionally, k was optimised for topics to have high coherence and refinement [31]. The distribution of LDA topics was assessed for the three smoking statuses by applying the same three DAA methods as described to the LDA-gamma matrix. To perform MaAsLin3, the LDA-gamma matrix was multiplied by the sample's sequencing depth, to account for different library sizes of samples.

Sensitivity analyses

A diagnosis or worsening of symptoms of asthma or chronic bronchitis/ COPD could change an individual's smoking status. For this scenario, an alternative DAG was drawn which then includes age, sex, season, SES, chronic bronchitis/ COPD, and asthma as minimally sufficient adjustment set (MSAS; Supplementary File Fig. 2). Importantly, this DAG only holds true if body mass index (BMI) and respiratory tract infections (RTIs) are not responsible for the worsening of symptoms or a diagnosis of chronic bronchitis/ COPD or asthma. The analysis was performed as described but adjusted for the new minimally sufficient adjustment set.

To assess whether the country of birth influenced the results of the main analysis, a sensitivity analysis was performed which included only participants with Germany as their country of birth as surrogate of ethnicity. The analysis was conducted in the same way as the main analysis.

The use of antibiotics alters the nasal microbiome [13, 16]. A third sensitivity analysis was therefore performed including only participants who had not taken any antibiotics within the previous 12 months. The analysis was conducted in the same way as the main analysis.

Results

Descriptive results:

1,948 participants were included in the current study, with the majority (47.6%) being never smokers (Table 1). A total of 14,412 zOTUs were identified across all samples. Following the filtering of unclassified taxa at phylum level and applying the prevalence and detection filters, 5,968 zOTUs were retained. At the phylum level, the most prevalent taxa were *Actinobacteriota* (100%), *Firmicutes d* (100%), *Proteobacteria* (99.9%), *Firmicutes a* (99.5%), and *Bacteroidota* (89.7%). At the genus level, the most prevalent taxa were *Corynebacterium* (100%), *Staphylococcus* (100%), *Cutibacterium* (99.9%), *Anaerococcus* (97.4%), and *Streptococcus* (94.9%). At the genus level, the five most abundant taxa were *Corynebacterium, Staphylococcus, Cutibacterium, Dolosigranulum*, and *Anaerococcus*. The most prevalent or abundant taxa remained unaltered in the sensitivity analyses.

Table 1
Characteristics of the study population stratified by smoking status.

	Never Smoker (N = 927)	Former Smoker (N = 693)	Current Smoker (N = 328)	Overall (N = 1948)
Age				
Mean (SD)	50.6 (11.4)	53.6 (10.0)	48.9 (10.3)	51.4 (10.9)
Median [Min, Max]	50.0 [22.0, 73.0]	54.0 [22.0, 72.0]	50.0 [21.0, 70.0]	52.0 [21.0, 73.0]
Sex				
m	393 (42.4%)	373 (53.8%)	149 (45.4%)	915 (47.0%)
f	534 (57.6%)	320 (46.2%)	179 (54.6%)	1033 (53.0%)
SES				
Mean (SD)	52.5 (15.2)	49.6 (14.6)	48.4 (14.6)	50.8 (15.0)
Median [Min, Max]	51.0 [16.0, 89.0]	49.0 [16.0, 89.0]	48.0 [18.0, 89.0]	49.0 [16.0, 89.0]
Chronic bronchitis/ COPD				
Yes	20 (2.2%)	33 (4.8%)	15 (4.6%)	68 (3.5%)
No	901 (97.8%)	658 (95.2%)	313 (95.4%)	1872 (96.5%)
Missing	6 (0.6%)	2 (0.3%)	0 (0%)	8 (0.4%)
Asthma				
Yes	64 (6.9%)	54 (7.8%)	20 (6.1%)	138 (7.1%)
No	857 (93.1%)	639 (92.2%)	306 (93.9%)	1802 (92.9%)
Missing	6 (0.6%)	0 (0%)	2 (0.6%)	8 (0.4%)
Country of origin				
Germany	854 (92.1%)	636 (91.8%)	296 (90.2%)	1786 (91.7%)
Not Germany	73 (7.9%)	57 (8.2%)	32 (9.8%)	162 (8.3%)
Antibiotic use during the last 12 months				
Never	624 (69.0%)	450 (66.7%)	196 (62.6%)	1270 (67.1%)
Once	198 (21.9%)	165 (24.4%)	74 (23.6%)	437 (23.1%)
Twice	59 (6.5%)	39 (5.8%)	28 (8.9%)	126 (6.7%)
More than three times	23 (2.5%)	21 (3.1%)	15 (4.8%)	59 (3.1%)
Missing	23 (2.5%)	18 (2.6%)	15 (4.6%)	56 (2.9%)

	Never Smoker (N = 927)	Former Smoker (N = 693)	Current Smoker (N = 328)	Overall (N = 1948)
Age				
Lifetime Number of Cigarettes				
Mean (SD)	0 (0)	13.2 (9.86)	10.9 (7.79)	6.52 (9.17)
Median [Min, Max]	0 [0, 0]	10.0 [0, 60.0]	10.0 [0, 40.0]	0.490 [0, 60.0]
Season				
Spring	95 (10.2%)	88 (12.7%)	39 (11.9%)	222 (11.4%)
Summer	49 (5.3%)	29 (4.2%)	18 (5.5%)	96 (4.9%)
Autumn	71 (7.7%)	54 (7.8%)	16 (4.9%)	141 (7.2%)
Winter	712 (76.8%)	522 (75.3%)	255 (77.7%)	1489 (76.4%)

COPD = chronic obstructive pulmonary disease, f = female, m = male, Max = maximum, Min = minimum, SES = socioeconomic status, SD = standard deviation.

Species richness and evenness:

The linear regression models revealed higher Shannon and Inverse Simpson index for current smokers compared to never smokers with small effect sizes (0.20 and 0.23, respectively; Table 2). There were no associations between former smokers' status (Supplementary File Table 1) or the lifetime number of cigarettes (Supplementary File Table 2) and species richness or evenness. The sensitivity analyses revealed no changes to the identified associations (Supplementary File Table 5–6, Supplementary File Table 11–12, Supplementary File Table 17–18).

Table 2
Linear regression model for the association between current smokers and five alpha diversity metrics.

Alpha diversity index	*Estimate*	*95% CI*	*Cohen's d*	*Cohen's d 95 CI*	*p-value*	*adjusted p-value*
Observed richness	11.14	[-7.54, 29.82]	0.06	[-0.04, 0.16]	0.242	0.323
Chao1 index	12.15	[-12.33, 36.64]	0.05	[-0.05, 0.16]	0.330	0.413
Shannon index	0.11	[0.04, 0.18]	0.20	[0.07, 0.32]	0.002	0.010
Inverse Simpson index	3.29	[1.52, 5.06]	0.23	[0.11, 0.36]	< 0.001	< 0.001
Faith's PD	0.07	[-0.11, 0.25]	0.04	[-0.07, 0.16]	0.461	0.542

Results of the linear regression model for current smokers compared to never smokers. Adjusted for age, sex, season, SES, and sequencing depth. CI = confidence interval, SES = socioeconomic status, Faith's PD = Faith's phylogenetic diversity. Adjustment with Benjamini Hochberg correction.

PERMANOVA

The PERMANOVA revealed that smoking status alone explained 0.4% to 1.1% of variance in the microbiome composition (Table 3), and the lifetime number of cigarettes explained 0.2% to 0.3% of variance (Table 4). The sensitivity analyses showed only minor changes (Supplementary File Table 7–8, Supplementary File Table 13–14, Supplementary File Table 19–20).

Heterogeneity of variance was identified for smoking status and the confounders across the four beta diversity metrics, except for age as assessed with weighted UniFrac distance, season as assessed with Aitchison distance, (Supplementary File Table 3, Supplementary File Figs. 3–22) and the lifetime number of cigarettes as assessed with weighted UniFrac distance (Supplementary File Table 4, Supplementary File Figs. 23–26). However, PERMANOVA with large sample sizes is robust against violations of homogeneity assumptions.

Table 3
PERMANOVA results for smoking status across four beta diversity metrics.

Variables	***R2***	***p-value***	***adjusted p-value***
Aitchison Distance			
Smoking Status	0.007	0.001	0.001
Age	0.004	0.001	0.001
Sex	0.015	0.001	0.001
SES	0.001	0.001	0.001
Season	0.003	0.001	0.001
Smoking Status + Age + Sex + SES + Season	0.073	0.001	0.001
Bray-Curtis Dissimilarity			
Smoking Status	0.007	0.001	0.002
Age	0.006	0.001	0.002
Sex	0.009	0.001	0.002
SES	0.001	0.021	0.021
Season	0.003	0.003	0.004
Smoking Status + Age + Sex + SES + Season	0.065	0.001	0.002
Unweighted UniFrac distance			
Smoking Status	0.004	0.001	0.001
Age	0.002	0.001	0.001
Sex	0.009	0.001	0.001
SES	0.001	0.014	0.014
Season	0.003	0.001	0.001
Smoking Status + Age + Sex + SES + Season	0.063	0.001	0.001
Weighted UniFrac distance			
Smoking Status	0.011	0.001	0.002
Age	0.002	0.001	0.002
Sex	0.006	0.001	0.002
SES	0.001	0.034	0.036

Variables	*R2*	*p-value*	*adjusted p-value*
Aitchison Distance			
Season	0.004	0.002	0.003
Smoking Status + Age + Sex + SES + Season	0.064	0.036	0.036

PERMANOVA results using the Adonis2 test performed as an omnibus test for smoking status, age, sex, season, and SES across. SES = socioeconomic status. Adjustment with Benjamini Hochberg correction.

Table 4
PERMANOVA results for the lifetime number of cigarettes across four beta diversity metrics.

Variables	***R2***	***p-value***	***adjusted p-value***
Aitchison Distance			
Lifetime Number of Cigarettes	0.002	0.006	0.009
Age	0.004	0.001	0.002
Sex	0.015	0.001	0.002
SES	0.002	0.013	0.013
Season	0.004	0.008	0.010
Lifetime Number of Cigarettes + Age + Sex + SES + Season	0.079	0.001	0.002
Bray-Curtis Dissimilarity			
Lifetime Number of Cigarettes	0.002	0.017	0.026
Age	0.006	0.001	0.002
Sex	0.009	0.001	0.002
SES	0.002	0.056	0.056
Season	0.004	0.042	0.050
Lifetime Number of Cigarettes + Age + Sex + SES + Season	0.074	0.001	0.002
Unweighted UniFrac distance			
Lifetime Number of Cigarettes	0.002	0.002	0.002
Age	0.003	0.001	0.002
Sex	0.011	0.001	0.002
SES	0.001	0.045	0.045
Season	0.005	0.001	0.002
Lifetime Number of Cigarettes + Age + Sex + SES + Season	0.073	0.001	0.002
Weighted UniFrac distance			
Lifetime Number of Cigarettes	0.003	0.012	0.021
Age	0.003	0.009	0.021
Sex	0.008	0.001	0.006

Variables	*R2*	*p-value*	*adjusted p-value*
Aitchison Distance			
SES	0.002	0.068	0.082
Season	0.006	0.014	0.021
Lifetime Number of Cigarettes + Age + Sex + SES + Season	0.067	0.161	0.161

PERMANOVA results using the Adonis2 test performed as an omnibus test for the lifetime number of cigarettes, age, sex, and SES across four beta diversity metrics. SES = socioeconomic status. Adjustment with Benjamini Hochberg correction.

Differential abundance analysis

After applying the 10% prevalence filter, 105 genera were retained. 29 unique taxa were identified as differentially abundant between smoking statuses by at least one DAA method. Each method identified between 13 and 19 taxa as differentially abundant. Six taxa were identified as differentially abundant by two methods and 16 taxa by only one of the three methods (Supplementary File Figs. 27–28). The seven taxa which were identified as being differentially abundant by all three methods were *Anaerococcus, Campylobacter (b), Finegoldia, Lawsonella, Neisseria (563205), Peptoniphilus (a)*, and *Prevotella* (Fig. 1). For these seven taxa, the results from MaAsLin3 indicate a higher abundance of *Anaerococcus, Campylobacter (b), Finegoldia, Peptoniphilus (a)*, and *Prevotella* and a lower abundance of the remaining taxa in current smokers compared to never smokers. *Anaerococcus, Peptoniphilus (a)*, and *Finegoldia* showed the same direction of the association, but at a weaker magnitude in former smokers compared to never smokers (Table 5).

Table 5
MaAsLin3 results for the eight taxa identified as differentially abundant for current and former smokers.

Smoking Status	*Feature*	*Coefficient*	*95% CI*	*p-value*	*q-value*
Current Smokers					
	g__Finegoldia	1.69	[1.33; 2.06]	< 0.001	< 0.001
	g__Anaerococcus	1.15	[0.85; 1.45]	< 0.001	< 0.001
	g__Peptoniphilus_a	1.45	[1.06; 1.85]	< 0.001	< 0.001
	g__Lawsonella	-1.27	[-1.69; -0.85]	< 0.001	< 0.001
	g__Prevotella	0.85	[0.43; 1.26]	< 0.001	0.002
	g__Neisseria_563205	-0.89	[-1.35; -0.42]	< 0.001	0.004
	g__Campylobacter_b	0.89	[0.42; 1.36]	< 0.001	0.004
Former Smokers					
	g__Anaerococcus	0.57	[0.33; 0.80]	< 0.001	< 0.001
	g__Peptoniphilus_a	0.71	[0.40; 1.02]	< 0.001	< 0.001
	g__Finegoldia	0.68	[0.39; 0.97]	< 0.001	< 0.001

MaAsLin3 results for the taxa identified as differentially abundant by the three Differential Abundance Analysis methods Wilcoxon Rank sum test, MaAsLin3, and Random Forest classification. Never smokers serve as the reference category. CI = confidence interval. Adjustment with Benjamini Hochberg correction.

For the lifetime number of cigarettes, no taxa were identified by both MaAsLin3 and the Random Forest regression model (Supplementary File Fig. 29).

All three sensitivity analyses identified *Anaerococcus, Campylobacter* (*b*), *Finegoldia, Peptoniphilus (a)*, and *Prevotella* as differentially abundant between smoking statuses with similar associations (Supplementary File Table 9, Supplementary File Fig. 33; Supplementary File Table 15, Supplementary File Fig. 34; Supplementary File Table 21, Supplementary File Fig. 35).

LDA clustering

For LDA, a model with six topics showed the best fit (Supplementary File Figs. 30–32). The three DAA methods identified a higher topic proportion of the *Peptoniphilus/ Anaerococcus* topic in current smokers as compared to never smokers (estimate: 2.04, 95% CI [1.47; 2.62], q = 0.035), The sensitivity analyses with the extended MSAS and the subpopulation with Germany as country of birth identified the same association (Supplementary File Table 10, Supplementary File Table 16). Only the sensitivity analysis of participants without antibiotic use in the previous 12 months identified no topic as differentially abundant between the smoking statuses. However, the distribution of the *Peptoniphilus/ Anaerococcus*

topic displayed a similar direction of the association for current smokers compared to never smokers (estimate: 2.06, 95 CI [1.32; 2.79], q = 0.103).

Discussion

Summary of results

This is the first German, population-based study in which the association between smoking habits and the nasal microbiome was assessed by using data from the baseline examination of NAKO [17]. Current smokers exhibited higher diversity and no differences in observed richness compared to never smokers. Several anaerobic pathobionts, including *Anaerococcus, Peptoniphilus*, and *Finegoldia*, were more abundant in both current and former smokers, though the lifetime number of cigarettes showed no association with the abundance of these taxa. Community-level clustering and sensitivity analyses support these findings.

Smoking is associated with higher richness and evenness

The Shannon and Inverse Simpson indices were slightly elevated among current smokers compared to never smokers. Further, there was no evidence for a dose-response relationship when considering the association between the lifetime number of cigarettes and diversity, suggesting that the diversity is largely influenced by whether the individual smokes rather than the quantity of cigarettes. The result for former smokers indicates that the higher diversity could diminish after smoking cessation.

Previous studies [11] have reported a similar direction of association between smoking and Shannon index, though the effect appears to be subtle and difficult to capture in smaller cohorts or when sampling from different nasal sites. The substantially larger cohort in the present study, combined with targeted sampling from the anterior nares, allowed us to more reliably characterize this association. Together, these findings suggest that smoking exerts a modest but consistent influence on microbial diversity in the anterior nares—one that may be overlooked in studies with limited sample sizes or alternative sampling strategies.

Anaerobic pathobionts identified as differentially abundant between smoking status

Three different DAA methods were used to ensure the robustness of the results with eight taxa being identified as differentially abundant between smoking statuses by all methods. These taxa are characterised by being pathobionts that can cause a variety of infections, including sinusitis, gastroenteritis, and skin and soft tissue infections [32–35]. Most of these taxa are anaerobic or microaerophilic, except for *Neisseria*, which is an aerobic organism [15, 32, 35–37]. The higher abundance of anaerobic and microaerophilic taxa, combined with the lower abundance of the aerobic *Neisseria*, suggests that cigarette smoke alters the oxygen levels in the anterior nares which increases the growth of anaerobic and microaerophilic taxa. This is in line with the findings of Pfeiffer et al. [15].

For the lifetime number of cigarettes, no taxa were identified by MaAsLin3 and the Random Forest regression. This indicates again, that smoking rather than the quantity of cigarettes smoked alters the microbial composition.

The higher abundance of these pathobionts is in line with the results from the alpha diversity model, indicating a shift from a few dominant commensals to a group of anaerobic taxa resulting in a less dominated community due to smoking. Further, the microbiome of former smokers showed patterns consistent with partial normalization. However, a longitudinal assessment of individuals who quit smoking is necessary to ensure a reliable characterisation of a potential recovery effect. Importantly, a nasal microbiome enriched for *Finegoldia, Peptoniphilus*, and *Anaerococcus* is associated with an increased susceptibility to infection or inflammation in the host [38]. The identification of the differentially abundant taxa is also in line with the results from the LDA clustering, in which current smoking status is associated with a higher probability of the *Peptoniphilus/ Anaerococcus* topic.

The altered abundances of these specific taxa or topics could aid in identifying individuals at higher risk of infections. However, a follow up of infections in the current participants as part of later examination rounds within NAKO would be necessary. This would also aid in assessing whether individuals with certain topic proportions were indeed at greater risk of developing infections or diseases and how quickly a recovery effect is identifiable following smoking cessation.

Limitations

One limitation of this study is that second-hand smoking was not assessed. An alteration in the composition of the nasal microbiome is not restricted to former and current smokers if individuals are exposed to it, but could also be present in never smokers, albeit to a lesser degree. The lack of information on passive smoking may have caused an underestimation in the impact of smoking if some never smokers were consistently exposed to second-hand smoke.

Residual confounding cannot be excluded, particularly with respect to environmental and behavioural factors such as recent infections or unmeasured exposures.

Despite NAKO being a cohort, the current study only employs a cross-sectional design, which limits causal inference. Therefore, the next examination rounds of NAKO should be used for follow-up, to ensure the validity of the identified effects of smoking.

Conclusion

This study is the first to comprehensively assess the association between smoking and the anterior nares' microbiome in a large German population-based cohort. Key findings include a higher microbial evenness among current smokers which is characterised by a higher abundance of anaerobic pathobionts. Though these taxa are present in an asymptomatic nasal microbiome [39], their heightened abundance could place smokers at a higher risk of developing respiratory infections and diseases.

Normalization patterns were detected in former smokers. Suggesting, that smoking cessation could be beneficial to restore a microbiome profile similar to that of never smokers, which is characterised by a low abundance of pathobionts and might place individuals at a lower risk of infections.

Abbreviations

BMI Body mass index

CI Confidence interval

Clr Centred log-ratio

COPD Chronic obstructive pulmonary disease

DAA Differential abundance analysis

DAG Directed acyclic graph

Faith's PD Faith's phylogenetic diversity

ISEI International Socioeconomic Index of Occupational Status

LDA Latent Dirichlet Allocation

MaAsLin Microbiome Multivariable Associations with Linear Models

MSAS Minimally sufficient adjustment set

NAKO German National Cohort

PCR Polymerase chain reaction

PERMANOVA Permutational multivariate analysis of variance

rRNA Ribosomal ribonucleic acid

RTI Respiratory tract infection

SES Socioeconomic status

zOTU Zero-radius operational taxonomic units

Declarations

Ethics approval and consent to participate

This study involves human participants and the ethics committees of the 18 NAKO study centres across Germany approved the study using their own reference number (e.g., Ethik-Kommission Westfalen-Lippe #2013-134-b-S for study centre Münster). NAKO is in accordance with national law and with the Declaration of Helsinki of 1975 (in the current, revised version). Written informed consent was obtained from all participants. Clinical trial number: DRKS00037328 (registered on 11. September 2025).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

SKB and NR were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 514199519 and by the program Innovative Medical Research of the University of Münster Medical School (RÜ122010, to NR). The NAKO is funded by the Federal Ministry of Research, Technology and Space (BMFTR) [project funding reference numbers: 01ER1301A/B/C, 01ER1511D, 01ER1801A/B/C/D and 01ER2301A/B/C], Federal States of Germany and the Helmholtz Association, the participating universities and the institutes of the Leibniz Association. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author Contribution

AW, SKB and NR contributed to writing the manuscript, formal analysis, funding acquisition, validation, visualization, software, and methodology. SW and MV contributed to data curation, investigation and resources. HBa, HBe, KB, HBr, KG, VH, BH, RK, AK, TK, YK, MNKK, BK, CKT, LK, BL, ML, WL, KBM, RM, MN, KN, NO, AP, TP, TS, BS, MBS, JSM, KW and MW contributed to funding acquisition, investigation, data curation, project administration, conceptualization, supervision, substantial review and editing of the manuscript.

Acknowledgement

This project was conducted with data (Application No. NAKO-423) from the German National Cohort (NAKO) (www.nako.de). We thank all participants of NAKO and the staff of this research initiative.

Members and affiliations of the NAKO Investigator Consortium can be accessed via www.nako.de/principal-investigators.

Data Availability

The raw reads of the nasal swabs can be publicly accessed at the European Nucleotide Archive (PRJEB107435). NAKO study data can be obtained via an electronic application portal: www.nako.de/transferhub. The repository that contains the analysis pipeline is available at https://zivgitlab.uni-muenster.de/clinical-epi/nasal_microbiome_smoking.

References

1. West R. Tobacco smoking: Health impact, prevalence, correlates and interventions. Psychol Health. 2017;32:1018–36. https://doi.org/10.1080/08870446.2017.1325890.
2. Huang C, Shi G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med Engl. 2019;17:225. https://doi.org/10.1186/s12967-019-1971-7.
3. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet Lond Engl. 2021;397:2337–60. https://doi.org/10.1016/S0140-6736(21)01169-7.
4. Brook I, Gober AE. Recovery of Potential Pathogens in the Nasopharynx of Healthy and Otitis Media—Prone Children and Their Smoking and Nonsmoking Parents. Ann Otol Rhinol Laryngol. 2008;117:727–30. https://doi.org/10.1177/000348940811701003.
5. Robert Koch Institute. Prevalence of obesity and smoking among adults in Germany – trends from 2003 to 2023. Robert Koch Institute; 2025. [cited 2025 July 15]. https://doi.org/10.25646/13038.
6. Allen AM, Scheuermann TS, Nollen N, Hatsukami D, Ahluwalia JS. Gender Differences in Smoking Behavior and Dependence Motives Among Daily and Nondaily Smokers. Nicotine Tob Res Off J Soc Res Nicotine Tob. 2016;18:1408–13. https://doi.org/10.1093/ntr/ntv138.
7. Reidpath DD, Davey TM, Kadirvelu A, Soyiri IN, Allotey P. Does one cigarette make an adolescent smoker, and is it influenced by age and age of smoking initiation? Evidence of association from the U.S. Youth Risk Behavior Surveillance System (2011). Prev Med. 2014;59:37–41. https://doi.org/10.1016/j.ypmed.2013.11.011.
8. Veldhuizen S, Zawertailo L, Ivanova A, Hussain S, Selby P. Seasonal Variation in Demand for Smoking Cessation Treatment and Clinical Outcomes. Nicotine Tob Res. 2021;23:976–82. https://doi.org/10.1093/ntr/ntaa214.
9. Colwell B, Ramirez N, Koehly L, Stevens S, Smith D, Creekmur S. Seasonal variations in the initiation of smoking among adolescents. Nicotine Tob Res. 2006;8:239–43. https://doi.org/10.1080/14622200600576503.

10. Cigarette smoking among adults - GOV. UK Ethnicity facts and figures [Internet]. [cited 2025 July 28]. https://www.ethnicity-facts-figures.service.gov.uk/health/alcohol-smoking-and-drug-use/adult-smokers/latest/. Accessed 28 July 2025.
11. Hickman E, Roca C, Zorn BT, Rebuli ME, Robinette C, Wolfgang MC, et al. E-Cigarette Use, Cigarette Smoking, and Sex Are Associated With Nasal Microbiome Dysbiosis. Nicotine Tob Res Off J Soc Res Nicotine Tob Engl. 2024;27:114–24. https://doi.org/10.1093/ntr/ntae176.
12. Yan M, Pamp SJ, Fukuyama J, Hwang PH, Cho D-Y, Holmes S, et al. Nasal Microenvironments and Interspecific Interactions Influence Nasal Microbiota Complexity and S. aureus Carriage. Cell Host Microbe Elsevier BV. 2013;14:631–40. https://doi.org/10.1016/j.chom.2013.11.005.
13. Kleine Bardenhorst S, Baurecht H, Becher H, Berger K, Brenner H, Günther K et al. Host and lifestyle factors associated with the anterior nares' microbiome diversity and composition: A cross-sectional analysis based on the German National Cohort (NAKO) [Internet]. In Review; 2026 [cited 2026 May 13]. https://doi.org/10.21203/rs.3.rs-9116334/v1
14. 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. Heimesaat MM, editor. PLoS ONE. 2010;5:e15216. https://doi.org/10.1371/journal.pone.0015216
15. Pfeiffer S, Herzmann C, Gaede KI, Kovacevic D, Krauss-Etschmann S, Schloter M. Different responses of the oral, nasal and lung microbiomes to cigarette smoke. Thorax. 2022;77:191–5. https://doi.org/10.1136/thoraxjnl-2020-216153.
16. Odendaal M-L, De Steenhuijsen Piters WAA, Franz E, Chu MLJN, Groot JA, Van Logchem EM, et al. Host and environmental factors shape upper airway microbiota and respiratory health across the human lifespan. Cell. 2024;187:4571–e458515. https://doi.org/10.1016/j.cell.2024.07.008.
17. Peters A, German National Cohort (NAKO) Consortium, Peters A, Greiser KH, Göttlicher S, Ahrens W, et al. Framework and baseline examination of the German National Cohort (NAKO). Eur J Epidemiol. 2022;37:1107–24. https://doi.org/10.1007/s10654-022-00890-5.
18. GitHub - yhwu/idemp. Barcode demultiplex for Illumina I1, R1, R2 fastq.gz files [Internet]. [cited 2025 July 18]. https://github.com/yhwu/idemp. Accessed 18 July 2025.
19. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics. Volume 31. Oxford University Press (OUP); 2015. pp. 3476–82. https://doi.org/10.1093/bioinformatics/btv401.
20. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461.
21. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, et al. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol Springer Sci Bus Media LLC. 2024;42:715–8. https://doi.org/10.1038/s41587-023-01845-1.
22. Liber JA, Bonito G, Benucci GMN. CONSTAX2: improved taxonomic classification of environmental DNA markers. In: Marschall T, editor. Bioinformatics. Volume 37. Oxford University Press (OUP); 2021. pp. 3941–3. https://doi.org/10.1093/bioinformatics/btab347.

23. Ganzeboom HBG, De Graaf PM, Treiman DJ. A standard international socio-economic index of occupational status. Soc Sci Res. 1992;21:1–56. https://doi.org/10.1016/0049-089X(92)90017-B.
24. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Stat Methodol. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
25. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR et al. vegan: Community Ecology Package [Internet]. 2025. https://CRAN.R-project.org/package=vegan
26. Nickols WA, Kuntz T, Shen J, Maharjan S, Mallick H, Franzosa EA, et al. MaAsLin 3: refining and extending generalized multivariable linear models for meta-omic association discovery. Nat Methods Nat Publishing Group. 2026;23:554–64. https://doi.org/10.1038/s41592-025-02923-9.
27. Wright MN, ranger Ziegler A. A Fast Implementation of Random Forests for High Dimensional Data in *C++* and *R*. J Stat Softw [Internet]. 2017 [cited 2025 Aug 15];77. https://doi.org/10.18637/jss.v077.i01
28. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw [Internet]. 2010 [cited 2025 Aug 11];36. https://doi.org/10.18637/jss.v036.i11
29. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing. 2009;72:1775–81. https://doi.org/10.1016/j.neucom.2008.06.011.
30. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Doc Numér. 2014;17:61–84. https://doi.org/10.3166/dn.17.1.61-84.
31. Fukuyama J, Sankaran K, Symul L. Multiscale analysis of count data through topic alignment. Biostatistics. 2023;24:1045–65. https://doi.org/10.1093/biostatistics/kxac018.
32. Nour SI, Khodadadi RB, Schuetz AN, Patel R, Abu Saleh OM. Lawsonella clevelandensis, a case series of vascular graft infections caused by a rare pathogen. IDCases. 2023;31:e01735. https://doi.org/10.1016/j.idcr.2023.e01735.
33. Cobo F, Navarro-Marí JM. First description of Anaerococcus octavius as cause of bacteremia. Anaerobe. 2020;61:102130. https://doi.org/10.1016/j.anaerobe.2019.102130.
34. Brook I. Microbiology of Sinusitis. Proc Am Thorac Soc. 2011;8:90–100. https://doi.org/10.1513/pats.201006-038RN.
35. Neumeister B, Geiss HK, Braun RW, Kimmig P, editors. Mikrobiologische Diagnostik [Internet]. 2nd ed. Stuttgart: Georg Thieme; 2009. [cited 2026 May 13]. https://doi.org/10.1055/b-002-19462. p. b-002-19462.
36. Webb KA, Olagoke O, Baird T, Neill J, Pham A, Wells TJ, et al. Genomic diversity and antimicrobial resistance of Prevotella species isolated from chronic lung disease airways. Microb Genomics [Internet]. 2022. https://doi.org/10.1099/mgen.0.000754. [cited 2025 Aug 12];8.
37. Song Y, Finegold SM. Peptostreptococcus, *Finegoldia, Anaerococcus, Peptoniphilus, Veillonella*, and Other Anaerobic Cocci. In: Versalovic J, Carroll KC, Funke G, Jorgensen JH, Landry ML, Warnock DW, editors. Man Clin Microbiol [Internet]. 1st ed. Wiley; 2011 [cited 2026 June 3]. pp. 803–16. https://doi.org/10.1128/9781555816728.ch48

38. Sarkar A, Harty S, Lehto SM, Moeller AH, Dinan TG, Dunbar RIM, et al. The Microbiome in Psychology and Cognitive Neuroscience. Trends Cogn Sci. 2018;22:611–36. https://doi.org/10.1016/j.tics.2018.04.006.

39. Camarinha-Silva A, Jáuregui R, Chaves-Moreno D, Oxley APA, Schaumburg F, Becker K, et al. Comparing the anterior nare bacterial community of two discrete human populations using I llumina amplicon sequencing. Environ Microbiol. 2014;16:2939–52. https://doi.org/10.1111/1462-2920.12362.

Figures

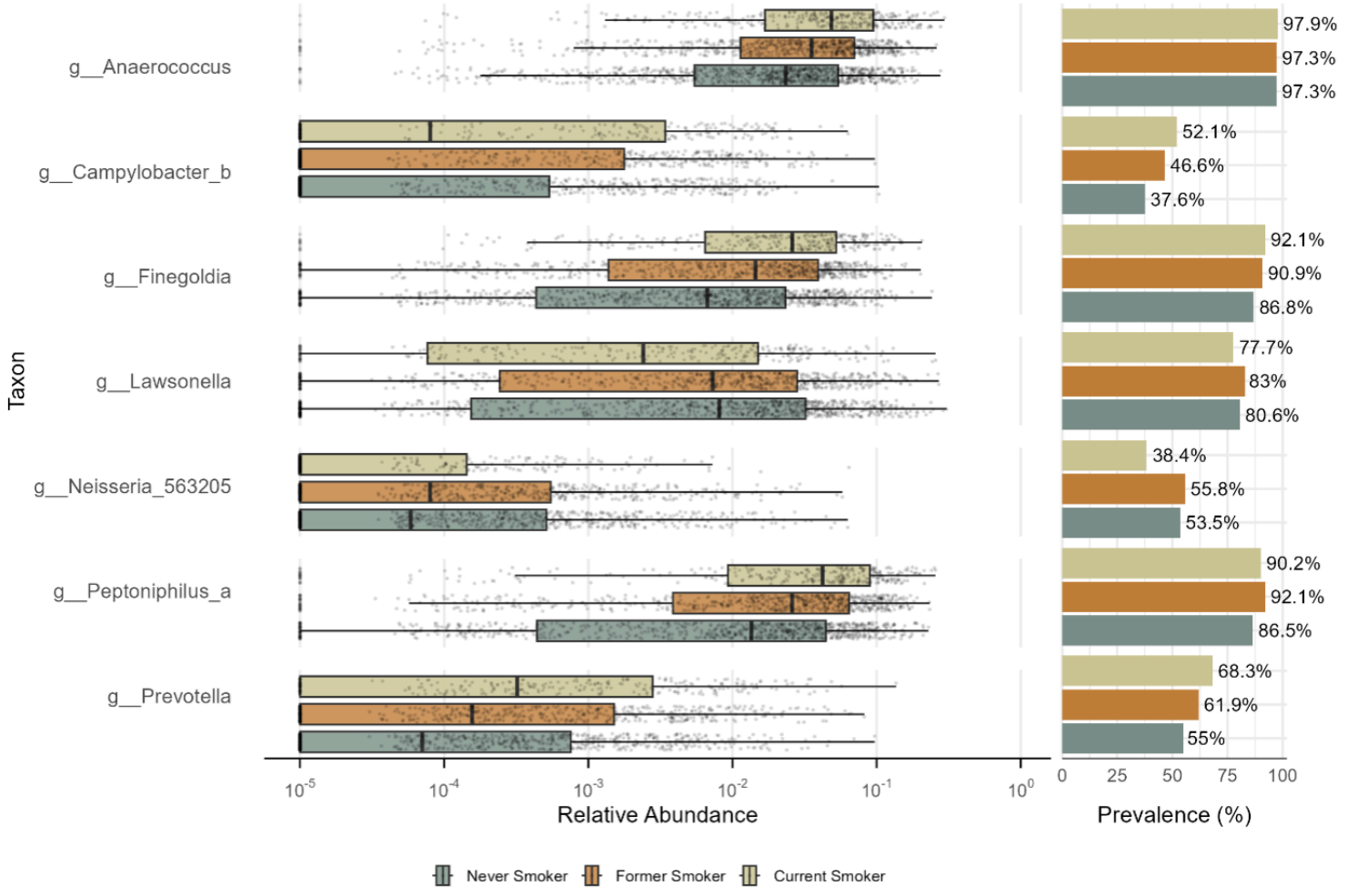

Figure 1

Relative abundance and prevalence of differentially abundant taxa by smoking status.

Taxa identified as differentially abundant by MaAsLin3, the Random Forest classifier and the Wilcoxon rank sum test, their relative abundance and prevalence by smoking status. DAA = differential abundance analysis.

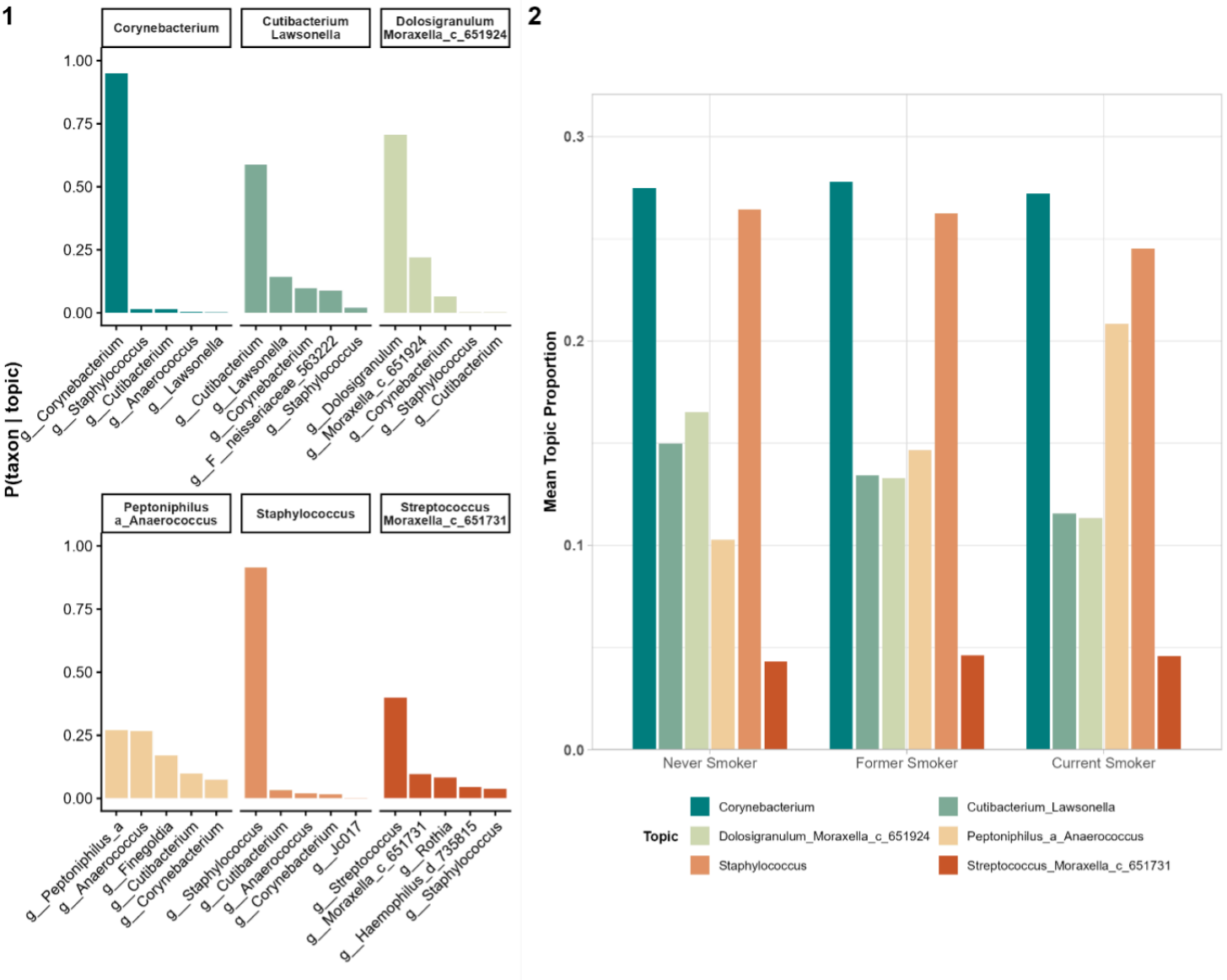

Figure 2

Results from the latent Dirichlet allocation.

Displaying (1) the probability of the taxa belonging to this topic for the five taxa with the highest probability per topic and (2) the mean topic proportion for each smoking status

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile.docx