Host and lifestyle factors associated with the anterior nares' microbiome diversity and composition: A cross-sectional analysis based on the German National Cohort (NAKO)

Sven Kleine Bardenhorst
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 (UKE)
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 Medicine Greifswald
Bianca Klee
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
University of Freiburg
Rafael Mikolajczyk
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 (UKE)
Annette Peters
Helmholtz Center Munich
Pischon Tobias
Max Delbrück Center for Molecular Medicine (MDC)
Tamara Schikowski
IUF - Leibniz Institute for Environmental Medicine
Börge Schmidt
University of Duisburg-Essen
Matthias B. Schulze
German Institute of Human Nutrition Potsdam-Rehbruecke
Julia Six-Merker
Helmholtz Center Munich
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

`nicole.ruebsamen@uni-muenster.de`

University of Münster

Article

Keywords: 16S rRNA gene, nasal microbiota, microbial communities

Posted Date: April 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9116334/v1

Additional Declarations: No competing interests reported.

Abstract

Background

The nasal microbiome plays an important role in respiratory and systemic health, but data from large adult population cohorts remain scarce. We analyzed nasal microbiota from 2,070 adults aged 21–73 years in the population-based German National Cohort (NAKO) to characterize community composition and identify host factors associated with variation in the anterior nares' microbiome. Microbial profiles were obtained using 16S rRNA gene sequencing, and associations with host characteristics—including sex, age, body composition, tobacco smoking, pulmonary function, household context, and self-reported physician diagnoses—were evaluated using a two-stage regression framework, beta diversity analyses, and complementary Latent Dirichlet Allocation (LDA) community modeling.

Results

Despite detecting 358 genera across the cohort, more than 90% of all sequencing reads were assigned to only 15 genera. These core genera showed distinct associations with host physiology and lifestyle. Sex and body mass index were the strongest correlates of alpha and beta diversity, while pulmonary function and tobacco smoking were associated with differences in the relative abundance of several commensal taxa, including *Lawsonella*, *Cutibacterium*, and *Dolosigranulum*. Age was not related to overall diversity but exhibited characteristic shifts in community composition. Antibiotic use within the previous 12 months was associated with lower levels of multiple commensal genera and higher relative abundance of *Staphylococcus*. LDA-derived sub-communities largely mirrored the genus-level associations and indicated that these patterns reflect transitions between recurrent nasal community types rather than isolated changes in single taxa.

Conclusions

This nationwide study provides a detailed characterization of nasal microbiome patterns in the general adult population (up to 73 years) of Germany and highlights that multiple host factors are associated with both taxon-specific and community-level variation. The findings offer a reference for future investigations into the role of the nasal microbiome in respiratory and systemic health.

Introduction

The human nasal microbiome is a complex and dynamic microbial ecosystem that plays a critical role in maintaining respiratory health. As the first barrier of the upper respiratory tract, it contributes to host defense by protecting against the colonization of inhaled pathogens. Several members of the nasal microbial community have been shown to directly influence the growth and virulence of respiratory pathogens [1] while others suppress opportunistic bacteria by competing for ecological niches and nutrients [2], or through the production of inhibitory compounds such as bacteriocins and other antimicrobial molecules [3]. Shifts in nasal microbial composition have been linked to a range of respiratory tract conditions, including chronic rhinosinusitis (CRS) [4], variations in COVID-19 disease severity [5, 6] and general susceptibility to upper respiratory tract infections [7]. Beyond its barrier function, the nasal microbiome also modulates host immune responses. It influences both the onset and progression of allergic and non-allergic inflammation and plays a key role in the development and function of mucosa-associated lymphoid tissue (MALT), including local immunoglobulin A (IgA) production and regulation of effector and regulatory T cells [8]. Emerging evidence suggests that the nasal microbiome may also be involved in the neural system, including potential links to the development and progression of neurodegenerative diseases such as Parkinson's disease [9–11]. Given its accessibility and immunological importance, the nasal microbiome represents a promising source of microbial biomarkers for diagnosing respiratory or neurological conditions [12, 13], and may serve as a potential target for future therapeutic interventions [14, 15].

Despite growing recognition of the nasal microbiome's clinical relevance, most studies to date have focused on small sample sizes or specific sub-populations, such as children, elderly individuals, or individuals with chronic respiratory conditions. Only a limited number of investigations have examined the nasal microbiome in the general adult population, and these have often been constrained by low statistical power and limited metadata. To address this gap, we analyzed nasal microbiome data from one of the largest population-based cohorts to date—and the largest within the German population—leveraging an extensive set of host and environmental variables. This cross-sectional analysis aims at providing a comprehensive overview of relevant host factors shaping nasal microbial diversity and composition in healthy adults.

Methods

This project was conducted with data and biosamples (Application No. NAKO-423) from the German National Cohort (NAKO), a study that aimed to recruit 200,000 adults aged 20–69 over five years through 18 study centers across 16 urban, industrial, and rural regions in Germany [16]. The participants' actual age on the day of the examination ranged from 19 to 74 years. Within the first phase of an extensive research project on the human microbiome, a stratified random sample of 2,400 individuals with triplets of biomaterials (stool, oral and nasal samples) was chosen from the more than 200,000 participants, without missing data for specific interview questions (cancer diagnoses, metabolic diseases, food allergies, gastrointestinal diseases, nutrition, and medication). The sample size for this first phase was chosen due to feasibility.

Assessment of host factors

All host-related variables were assessed at baseline, either through standardized interviews and examinations conducted by trained personnel or via self-administered questionnaires. All clinical diagnoses, including periodontitis, asthma, diabetes, renal insufficiency, depression, allergy, recent respiratory infections (within the last three weeks), elevated blood lipids, and any diagnosed allergy, were based on self-report. All standardized examination procedures as well as all derivations of variables are described in detail in Supplementary Table 1.

Microbial data

Sample collection

Nasal swabs were sampled from both the left and the right side, using one swab for each side. The tip of the swab was first inserted into 0.9% saline solution and then pressed against the side of the tube containing the solution to remove excess solution. The swab was then inserted into one of the nostrils at a depth of 1 to 1.5 cm and rotated three times with light pressure to sample the inner and outer linings of the anterior nares, before thoroughly wiping out the upper and lower corner of the nostril. Both swabs were placed in the same cryotube containing 0.2 mL RNAlater© (QIAGEN, Austin, Texas, USA), a DNA and RNA stabilising solution. The handle of the swab 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.

Laboratory analysis

One of the two swabs was randomly taken for analysis, resulting in a random mixture of swabs from the left or the right nostril for subsequent analyses. DNA was extracted with ZymoBIOMICS 96 DNA Kit (Zymo Research, Irvine, California, USA) including a bead beating step for 5 x 1 min at speed 5.5 on a FastPrep-24 5G Instrument (MP Biomedicals, Santa Ana, California, USA) with a 5 min rest between each step. Further processing was done following the manufacturer's instructions. Amplicon libraries were generated in a two-step PCR protocol as described previously [17]: In the first PCR step, the V3-V4 region of the 16S rRNA gene was amplified in a 20 cycle PCR reaction (total volume 15 µl containing 1x Kapa Hifi Hot Start Ready Mix (Kapa Biosystems Inc., Wilmington, Massachusetts, USA)) and primers Bakt341F and Bakt805R at 0.25 µM each at final concentration and 2 µl of gDNA. The products were then purified using Ampure XP beads (Beckman Coulter, Brea, California, USA) at a ratio of 0.8 according to the manufacturer's instructions. In the second PCR step (15 cycles), barcodes and adapters were attached. The reaction was carried out in a 10 µl volume with 1x Kapa Hifi Hotstart ready mix with primers at final concentration of 0.25 µM each and 3 µl of the purified PCR product from the previous step. Libraries were again purified with Ampure XP beads at ratio 0.8 and eluted in 20 µl of EB Buffer (Qiagen, Hilden, Germany). Libraries were quantified with Qubit dsDNA HS Assay Kit (Invitrogen, Waltham, Massachusetts, USA) using BioTek Synergy HTX plate reader (BioTek by Agilent Technologies, Santa Clara, California, USA). Approximately 450 libraries were pooled equimolar, quantified with Qubits dsDNA Assay Kit and quality checked on D1000 Tape using TapeStation 4200 instrument (Agilent Technologies, Santa Clara, California, USA). Sequencing was performed on Illumina MiSeq Sequencer (Illumina, San Diego, California, USA) using the V3 chemistry (2x300 bp) with a loading concentration of 5.5 pM and the use of 8% of PhiX. Negative-extraction controls (no swab) were included with each extraction and sequencing batch. Sequencing data could be obtained for 2,072 of the 2,400 nasal samples.

Bioinformatics processing

Raw reads were demultiplexed by idemp [18] according to the given barcodes [19]. Libraries were processed including merging the paired-end reads, filtering the low-quality sequences, dereplication to find unique sequence, singleton removal, denoising, and chimera checking using the USEARCH pipeline (version 11.0.667) [20]. In brief, reads were merged by *fastq_mergepairs* command (parameters: maxdiffs 30, pctid 70, minmergelen 200, maxmergelen 400), filtered for low quality with *fastq_filter* (maxee 1) and singletons using *fastx_uniques* command (minuniquesize 2). To predict biological sequences (zOTUs) and filter chimeras we used the *unoise3* command (minsize 10, unoise_alpha 2), following the amplicon quantification using the *usearch_global* command (strand plus, id 0.97, maxaccepts 10, top_hit_only, maxrejects 250). Taxonomic assignment was conducted by Constax (classifiers: rdp, sintax, blast) [21] using the GreenGenes2 database [22] and summarizing into biom-file for import in R via TreeSummarizedExperiments (version 2.14.0) [23] for further downstream statistical analyses. Handling of microbial data in R (version 4.4.3) [24] was mostly performed using mia (version 1.14.0) [25] and microViz (version 0.12.6) [26]. Visualizations were mainly prepared using ggplot2 (version 3.5.1) and ComplexHeatmap (version 2.22.0) [27]. Taxonomically informed color palettes were constructed using the package microshades (version 1.13) [28]. Transformations and beta diversity matrices were calculated using vegan (version 2.6–10) [29] and GUniFrac (version 1.8) [30]. Samples with a sequencing depth below 1000 reads were excluded to ensure sufficient coverage for downstream analyses. For all remaining samples except alpha diversity estimation, zOTUs were filtered based on combined prevalence and detection thresholds: only zOTUs with a prevalence greater than 0.5% across samples at a minimum relative abundance exceeding 0.005% were retained, to remove extremely rare and low-abundant features.

Alpha Diversity Analyses

Alpha diversity was estimated at the zOTU level using the package mia [25] and included measures of both richness and evenness. Specifically, observed richness (total number of zOTUs), the Shannon index, and the Simpson index were calculated to capture complementary aspects of within-sample diversity.

Beta Diversity Analyses

Beta diversity analyses were performed at the genus level to examine inter-individual variation in microbial community composition in relation to host factors. A complementary set of dissimilarity measures was used to capture different dimensions of beta diversity, including Bray-Curtis dissimilarity, unweighted and weighted UniFrac distances (to incorporate phylogenetic relationships), and Aitchison distance (Euclidean distance on robust centered log-ratio (rCLR) transformed abundances). Zero counts were handled using a simple zero-replacement (adding 1 to every count), and the robust CLR transformation was applied accordingly.

Core Microbiota Definition

To characterize host factors associated with the overall structure of the nasal microbiome, a core microbiota was defined based on abundance. zOTUs were ranked by their total read counts across all samples. Starting from the most abundant zOTU, zOTUs were sequentially added to the core set until the cumulative relative abundance accounted for more than 90% of all reads.

Latent Dirichlet Allocation Community Modeling

Latent compositional patterns were inferred by applying Latent Dirichlet Allocation (LDA), a generative probabilistic model that represents each sample as a mixture of communities, where each community is characterized by a distribution over bacterial genera [31]. We fitted LDA models with varying numbers of topics ($K = 1$ to $K = 12$) to genus-level taxonomic abundance data using the variational expectation-maximization (VEM) algorithm implemented in the topicmodels R package (version 0.2–17) [32]. All models were fitted to the same input data matrix of genus counts across samples. The optimal number of topics was determined using a combinatorial approach integrating three complementary strategies. Specifically, we applied the alto R package (version 0.1.0) [33] to align topics across models with different K values using optimal transport methods. This approach tracks how topics split, merge, or persist as model complexity increases, revealing the stability and interpretability of topic structures [34]. For each model, we computed topic coherence (semantic consistency of genera within topics) and topic refinement (degree of splitting from simpler models). Topics with high coherence and meaningful refinement patterns were favored. Visualization of topic alignment revealed that $K = 7$ provided a stable, interpretable solution where topics showed clear patterns without excessive fragmentation. The selected model balanced statistical fit with biological parsimony. Each topic was characterized by examining the posterior probability distribution over genera (β-matrix). Topics were named according to their dominant genus (highest β value), with secondary genera included if their probability exceeded 20% of the top genus. For each sample, the posterior mean topic proportions (γ-matrix) represented the estimated mixture of topics.

Statistical Modeling

Associations between host factors and nasal microbiome features were assessed using a two-stage analytic framework. In the first stage, we screened for associations using univariate linear regression models, applying false discovery rate (FDR) correction ($q < 0.05$) to identify host factors with evidence for a microbiome association. In the second stage, all factors identified during screening were jointly evaluated in multivariable models. This allowed us to assess the independent contribution of each factor while accounting for correlations between them and to reduce the risk of false-positive findings due to shared variance among host characteristics. The same workflow was used for alpha diversity and core-genera analyses. Multivariable alpha diversity models included all host factors associated with at least one diversity metric in the first stage, while multivariable core-genera models included only those factors associated with at least two core genera.

For beta diversity, we used permutational multivariate analysis of variance (PERMANOVA) to estimate the proportion of variation in community composition explained by host factors. To maintain consistency with the taxon-level analyses and to avoid model inflation in PERMANOVA, only host factors with established associations in the core-genera analysis were included in multivariable PERMANOVA models.

To complement these analyses and to establish the robustness of our findings, we additionally modeled associations between host factors and community-level composition using community proportions derived from LDA. While taxon-level analyses focus on individual genera, LDA summarizes co-occurring sets of taxa into probabilistic communities, providing a higher-order representation of community structure. We applied Bayesian Dirichlet regression to these community proportions, which accommodates the compositional constraints of the data and allows simultaneous modeling of mean topic abundances and their variability across individuals. Associations are reported as absolute differences in expected topic proportions. This LDA-based analysis serves as an independent, compositionality-aware confirmation of the patterns observed in the diversity and core-genera models, rather than a separate or competing analysis. The entire workflow is depicted in Fig. 1.

Sensitivity analyses

To further evaluate the robustness of our results, we conducted two sensitivity analyses. First, to assess whether recent antibiotic exposure influenced the observed associations, we repeated all multivariable models after excluding participants who reported antibiotic use within the previous 12 months. Second, to explore potential sex-specific patterns that may be masked in pooled analyses, we performed stratified analyses estimating all associations separately for male and female participants. Together, these sensitivity analyses allowed us to examine the stability of our findings across key sources of biological and methodological variation.

Results

Cohort characteristics

After filtering microbial data (two samples filtered due to sequencing depth < 1000 reads), our study sample comprised 2,070 participants with a mean age of 51.1 years (SD = 11.3), evenly distributed between females (52.9%) and males (47.1%). Body mass index (BMI) and body fat percentage increased progressively with age, with an overall mean BMI of 26.6 kg/m² and a mean body fat percentage of 31.5%. Half of the participants had never smoked (47.3%), while former smokers represented 34.4% and current smokers 17.4%, with decreasing smoking prevalence in older age groups. Allergy diagnosis was reported by 43.2% of participants, and antibiotic usage within the previous 12 months was reported by more than one-third (35.5%) of the participants. Alcohol consumption varied markedly, with most participants consuming alcohol between two and three times per week or month, whereas daily consumption was uncommon (6.4%). Overall, self-reported health status was predominantly positive (90.1% reporting excellent, very good, or good health), despite nearly 20% experiencing respiratory infections within the three weeks prior to assessment (Table 1).

Overall microbiome composition and core microbiota

In total, 18 different phyla (of which 4 were different groups of Firmicutes, considered polyphyletic in the Greengenes2 database, henceforth considered as one phylum) were observed. The phyla *Actinobacteriota*, *Firmicutes*, and *Proteobacteria* were detected in all samples, followed by *Bacteroidota* (89%), *Patescibacteria* (64%), and *Fusobacteriota* (54%). *Actinobacteriota* was the most abundant phylum across all samples, with a mean relative abundance of 42.2% (median: 41.8%, range: 0.05–93.6%). While *Firmicutes* (mean: 11.6%, median: 0.63%) and *Proteobacteria* (mean: 6.66%, median: 1.71%) also showed relatively high mean abundances, their distributions were highly right-skewed, each ranging from 0% to nearly 98% and 95%, respectively. In contrast, *Bacteroidota* and the less prevalent phyla such as *Patescibacteria* and *Fusobacteriota* exhibited much lower mean abundances (0.53%, 0.06%, and 0.10%, respectively), and their median values were close to zero, indicating that they were present in only a subset of the samples at low levels.

Of 358 different detected genera, the genera *Corynebacterium* and *Staphylococcus* were detected in all samples (100% prevalence) and represented the most abundant genera, with mean relative abundances of 29.0% (median: 25.8%) and 24.1% (median: 18.8%), respectively. *Cutibacterium* was also prevalent (99.9%) and showed a moderate abundance (mean: 10.4%, median: 8.0%). In contrast, several genera such as *Dolosigranulum* (80.5% prevalence, mean abundance: 9.5%) and *Anaerococcus* (97.4% prevalence, mean abundance: 4.4%) were frequently detected but with highly skewed distributions, as indicated by low median values (e.g., *Dolosigranulum*: median 0.02%). Other genera such as *Finegoldia*, *Peptoniphilus*, and *Lawsonella* were also common (prevalence > 80%), but had low relative abundances (means < 2.5%). Genera like *Streptococcus* (prevalence: 95%) and *Rothia* (70%) were present in many samples yet contributed only minor fractions to the overall community, highlighting the dominance of a few genera and the long tail of low-abundance taxa.

We defined a core microbiota, focusing on the most abundant taxa, to serve as the basis for subsequent analyses. The core microbiota was defined at the zOTU level but is henceforth presented at the genus level. The top 219 most abundant zOTUs accounted for > 90% of all sequences in the entire sample. The selected zOTUs represent 15 different genera and a group of zOTUs not classified at the genus level (see columns of Fig. 2). The core microbiota accounted for the majority of the microbial community in most samples, with a median relative contribution of 91.1% (IQR: 87.0–93.9%) and a mean abundance of 88.7%. However, there was some variation across samples, with core microbiota comprising as little as 2.8% and as much as 98.7% of the total community, indicating that in a few individuals, the peripheral microbiome dominated.

Table 1
Participant characteristics of German NAKO study participants stratified by age group

	21–29 years (N = 111)	**30–39 years (N = 198)**	**40–49 years (N = 547)**	**50–59 years (N = 676)**	**60–73 years (N = 538)**	**All ages (N = 2070)**
Age (years)						
Mean (SD)	25.8 (2.31)	34.5 (2.80)	45.5 (2.73)	54.1 (2.88)	64.5 (3.05)	51.1 (11.3)
Median	26	34	46	54	64	52
Sex						
female	67 (60.4%)	104 (52.5%)	276 (50.5%)	373 (55.2%)	276 (51.3%)	1096 (52.9%)
male	44 (39.6%)	94 (47.5%)	271 (49.5%)	303 (44.8%)	262 (48.7%)	974 (47.1%)
BMI (kg/m²)						
Mean (SD)	23.8 (4.15)	25.4 (5.36)	26.3 (4.95)	26.8 (4.77)	27.6 (4.93)	26.6 (4.97)
Median [Min, Max]	23.6 [15.8, 40.2]	24.3 [16.7, 55.4]	25.8 [17.2, 50.1]	26.2 [17.0, 54.1]	26.8 [16.9, 48.0]	25.9 [15.8, 55.4]
Missing	2 (1.8%)	1 (0.5%)	9 (1.6%)	11 (1.6%)	8 (1.5%)	31 (1.5%)
Smoking status						
Never smoked	69 (62.2%)	99 (50.0%)	288 (52.7%)	275 (40.7%)	249 (46.3%)	980 (47.3%)
Former smoker	18 (16.2%)	51 (25.8%)	152 (27.8%)	263 (38.9%)	228 (42.4%)	712 (34.4%)
Current smoker	24 (21.6%)	48 (24.2%)	102 (18.6%)	131 (19.4%)	55 (10.2%)	360 (17.4%)
Missing	0 (0%)	0 (0%)	5 (0.9%)	7 (1.0%)	6 (1.1%)	18 (0.9%)

	21–29 years (N = 111)	**30–39 years (N = 198)**	**40–49 years (N = 547)**	**50–59 years (N = 676)**	**60–73 years (N = 538)**	**All ages (N = 2070)**
Age (years)						
Allergy diagnosis						
No	54 (48.6%)	104 (52.5%)	302 (55.2%)	368 (54.4%)	337 (62.6%)	1165 (56.3%)
Yes	56 (50.5%)	94 (47.5%)	244 (44.6%)	302 (44.7%)	198 (36.8%)	894 (43.2%)
Missing	1 (0.9%)	0 (0%)	1 (0.2%)	6 (0.9%)	3 (0.6%)	11 (0.5%)
Antibiotics (previous 12 months)						
No	72 (64.9%)	122 (61.6%)	354 (64.7%)	439 (64.9%)	348 (64.7%)	1335 (64.5%)
1 time	26 (23.4%)	46 (23.2%)	125 (22.9%)	143 (21.2%)	121 (22.5%)	461 (22.3%)
2 times	7 (6.3%)	18 (9.1%)	38 (6.9%)	38 (5.6%)	34 (6.3%)	135 (6.5%)
3–4 times	5 (4.5%)	4 (2.0%)	9 (1.6%)	14 (2.1%)	10 (1.9%)	42 (2.0%)
> 4 times	0 (0%)	3 (1.5%)	2 (0.4%)	9 (1.3%)	5 (0.9%)	19 (0.9%)
Missing	1 (0.9%)	5 (2.5%)	19 (3.5%)	33 (4.9%)	20 (3.7%)	78 (3.8%)
% Body fat						
Mean (SD)	25.2 (8.65)	27.5 (9.06)	29.7 (8.31)	32.6 (8.52)	34.9 (8.54)	31.5 (8.97)
Median [Min, Max]	23.6 [4.10, 47.7]	27.6 [7.40, 58.1]	29.5 [3.40, 55.3]	32.1 [10.4, 54.7]	34.9 [9.50, 56.1]	31.0 [3.40, 58.1]

	21–29 years (N = 111)	30–39 years (N = 198)	40–49 years (N = 547)	50–59 years (N = 676)	60–73 years (N = 538)	All ages (N = 2070)
Age (years)						
Missing	3 (2.7%)	8 (4.0%)	13 (2.4%)	31 (4.6%)	24 (4.5%)	79 (3.8%)
Frequency of alcohol consumption						
Never	5 (4.5%)	15 (7.6%)	37 (6.8%)	48 (7.1%)	24 (4.5%)	129 (6.2%)
>=1 times per month	25 (22.5%)	45 (22.7%)	112 (20.5%)	91 (13.5%)	79 (14.7%)	352 (17.0%)
2–4 times per month	49 (44.1%)	75 (37.9%)	174 (31.8%)	187 (27.7%)	117 (21.7%)	602 (29.1%)
2 or more times per week	32 (28.8%)	63 (31.8%)	219 (40.0%)	342 (50.6%)	312 (58.0%)	968 (46.8%)
Missing	0 (0%)	0 (0%)	5 (0.9%)	8 (1.2%)	6 (1.1%)	19 (0.9%)
Self-reported health status						
Excellent	14 (12.6%)	7 (3.5%)	19 (3.5%)	14 (2.1%)	10 (1.9%)	64 (3.1%)
Very good	54 (48.6%)	94 (47.5%)	240 (43.9%)	202 (29.9%)	99 (18.4%)	689 (33.3%)
Good	39 (35.1%)	87 (43.9%)	239 (43.7%)	382 (56.5%)	365 (67.8%)	1112 (53.7%)
Less good or bad	4 (3.6%)	10 (5.1%)	45 (8.2%)	72 (10.7%)	61 (11.3%)	192 (9.3%)
Missing	0 (0%)	0 (0%)	4 (0.7%)	6 (0.9%)	3 (0.6%)	13 (0.6%)
Respiratory infection in last 3 weeks						
No	85	131	404	476	401	1497

	21–29 years (N = 111)	30–39 years (N = 198)	40–49 years (N = 547)	50–59 years (N = 676)	60–73 years (N = 538)	All ages (N = 2070)
Age (years)						
	(76.6%)	(66.2%)	(73.9%)	(70.4%)	(74.5%)	(72.3%)
Yes	22 (19.8%)	51 (25.8%)	111 (20.3%)	149 (22.0%)	78 (14.5%)	411 (19.9%)
Missing	4 (3.6%)	16 (8.1%)	32 (5.9%)	51 (7.5%)	59 (11.0%)	162 (7.8%)

Main host factors related to nasal microbiota diversity

Across all alpha diversity indices, the variables sex, BMI, forced expiratory volume in one second (FEV_1), and antibiotic use emerged most strongly associated with diversity in first-stage screening and were therefore included in second-stage models. Sex showed the most consistent association with diversity (Fig. 2). Male participants exhibited higher alpha diversity, reflected in less dominance by single genera and greater evenness (Shannon: β = 0.09; 95% CI [0.02, 0.16]). Anthropometric traits were similarly relevant: higher BMI was associated with increased richness (β = 4.59; 95% CI [2.93, 6.26]) and a more even community (Shannon: β = 0.01; 95% CI [0.01, 0.02]; Simpson: β = 0.22; 95% CI [0.08, 0.39]). In contrast, antibiotic use within the last 12 months was linked to reduced richness (β = -24.32; 95% CI [-41.24, -7.40]) and evenness (Shannon: β = -0.09; 95% CI [-0.15, -0.04]; Simpson: β = -1.72; 95% CI [-3.26, -0.18]). No association between age and alpha diversity was observed.

Beta diversity results were concordant across all four metrics of dissimilarity, again highlighting sex and BMI as relevant correlates of overall community structure. Although age did not affect alpha diversity, it explained a comparatively large share of beta diversity variation, indicating age-related compositional shifts without an overall loss or gain in within-sample diversity. Sociodemographic variables, particularly education and employment status, were also consistently associated with beta diversity (Supplementary Table 2).

Core Genera are Consistently Associated with Host and Lifestyle Factors

In first-stage screening, most of the 15 core genera were associated with sex (12 genera) and BMI (11), while fewer were linked to body fat (6) or age (5) (Fig. 2). Based on the prespecified selection rule, the second-stage model included age, sex, BMI, lifetime cigarette consumption per day, employment status, allergy diagnosis, depression, susceptibility to infections, antibiotic use, education, household size, and FEV_1 (Supplementary Fig. 1). As expected, the number of associations attenuated in multivariable models, but several robust patterns remained.

Sex differences were particularly pronounced. *Lawsonella* (β = 0.89; 95% CI [0.59, 1.2]), *Finegoldia* (β = 0.61; 95% CI [0.34, 0.88]), *Anaerococcus* (β = 0.24; 95% CI [0.01, 0.47]), *Peptoniphilus* (β = 0.70; 95% CI [0.41, 0.99]), and *Staphylococcus* (β = 0.28; 95% CI [0.09, 0.48]) were more abundant in males, whereas females showed higher abundances of *Streptococcus* (β = -0.44; 95% CI [-0.71, -0.17]), *Rothia* (β = -0.34; 95% CI [-0.50, -0.18]), *Moraxella_c_651924* (β = -0.29; 95% CI [-0.47, -0.11]), *Moraxella_c_651731* (β = -0.29; 95% CI [-0.47, -0.11]), and unclassified genera from *Enterobacteriaceae* (β = -0.33; 95% CI [-0.49, -0.17]) and *Neisseriaceae* (β = -0.53; 95% CI [-0.84, -0.22]). *Corynebacterium* was also higher in males (β = 0.26; 95% CI [0.08, 0.44]), but this contrast did not persist after excluding recent antibiotic users.

BMI showed broad associations overlapping with the sex-linked group: *Peptoniphilus* (β = 0.27; 95% CI [0.16, 0.38]), *Finegoldia* (β = 0.21; 95% CI [0.11, 0.31]), and *Anaerococcus* (β = 0.21; 95% CI [0.12, 0.29]) increased with BMI. Higher BMI was additionally associated with lower *Corynebacterium* (β = -0.14; 95% CI [-0.21, -0.07]), lower unclassified *Enterobacteriaceae* (β = -0.09; 95% CI [-0.15, -0.03]), reduced *Moraxella_c_651731* (β = -0.09; 95% CI [-0.16, -0.02]) and *Cutibacterium* (β = -0.06; 95% CI [-0.12, 0]). Several additional BMI associations (e.g., with *Campylobacter_b*, *Dolosigranulum*, *Streptococcus*, *Staphylococcus*) were not robust to exclusion of antibiotic users.

Age was negatively associated with *Lawsonella* (β = -0.46; 95% CI [-0.60, -0.32]) and *Cutibacterium* (β = -0.10; 95% CI [-0.17, -0.03]), and positively with *Dolosigranulum* (β = 0.30; 95% CI [0.07, 0.52]) and *Corynebacterium* (β = 0.15; 95% CI [0.07, 0.23]) (Supplementary Fig. 1). Weaker age-related associations (e.g., *Neisseriaceae*, *Finegoldia*) attenuated in the no-antibiotics analysis.

Tobacco smoking showed clear genus-level shifts despite no alpha-diversity signal. *Finegoldia* (β = 0.23; 95% CI [0.13, 0.33]), *Peptoniphilus* (β = 0.24; 95% CI [0.13, 0.35]), *Anaerococcus* (β = 0.16; 95% CI [0.07, 0.25]), and *Campylobacter* (β = 0.11; 95% CI [0.03, 0.19]) increased with smoking, whereas *Dolosigranulum* (β = − 0.31; 95% CI [−0.50, − 0.12]), *Lawsonella* (β = − 0.15; 95% CI [−0.27, − 0.03]) and unclassified *Enterobacteriaceae* (β = − 0.08; 95% CI [−0.14, − 0.02]) decreased.

Higher FEV$_1$ was positively associated with *Lawsonella* (β = 0.16; 95% CI [0, 0.32]), *Anaerococcus* (β = 0.17; 95% CI [0.04, 0.29]), *Peptoniphilus* (β = 0.16; 95% CI [0.01, 0.32]) and *Cutibacterium* (β = 0.10; 95% CI [0.02, 0.19]), but these associations were not robust after excluding antibiotic users.

Further, sociodemographic background was associated with the core genera. Compared to basic education, secondary education was associated with lower *Campylobacter_b* (β = − 0.27; 95% CI [−0.50, − 0.05]) and *Streptococcus* (β = − 0.32; 95% CI [−0.61, − 0.02]) and higher *Corynebacterium* (β = 0.20; 95% CI [0.00, 0.40]). Upper secondary education showed a similar pattern and additionally related to lower *Anaerococcus* (β = − 0.27; 95% CI [−0.50, − 0.05]). Employment was associated with lower *Finegoldia* (β = − 0.27; 95% CI [−0.54, 0.00]) and higher *Staphylococcus* (β = 0.33; 95% CI [0.13, 0.53]).

Genera Associated with Respiratory Health, Mental Health and Household Environment

Only *Moraxella_c_651731* was more abundant in participants reporting elevated susceptibility to respiratory infections (β = 0.26; 95% CI [0.05, 0.48]); this association was not evident among participants without recent antibiotic use. The related *Moraxella_c_651924* increased with household size (β = 0.30; 95% CI [0.17, 0.43]), whereas *Dolosigranulum* decreased (β = 0.30; 95% CI [0.12, 0.48]); a negative association for *Streptococcus* did not persist in the no-antibiotics sample.

Allergy diagnosis was associated with higher *Staphylococcus* (β = 0.21; 95% CI [0.06, 0.35]) and lower *Dolosigranulum* (β = − 0.56; 95% CI [−0.92, − 0.21]) and *Corynebacterium* (β = − 0.15; 95% CI [−0.28, − 0.02]). The positive association with *Rothia* (β = 0.13; 95% CI [0.02, 0.25]) was not robust to exclusion of antibiotic users.

Antibiotic use within the previous 12 months was linked to lower abundances of *Dolosigranulum* (β = − 0.48; 95% CI [−0.85, − 0.10]), *Lawsonella* (β = − 0.40; 95% CI [−0.64, − 0.16]) and *Peptoniphilus* (β = − 0.26; 95% CI [−0.48, − 0.03]), alongside higher *Staphylococcus* (β = 0.33; 95% CI [0.17, 0.48]), *Rothia* (β = 0.15; 95% CI [0.02, 0.27]) and unclassified *Enterobacteriaceae* (β = 0.20; 95% CI [0.08, 0.32]).

A history of depression was associated with lower *Staphylococcus* (β = − 0.27; 95% CI [−0.50, − 0.04]), but this did not persist among non-users of antibiotics.

Anterior nares community types associated with host factors

LDA identified seven recurrent co-occurrence communities that captured major compositional patterns in the microbiome (Fig. 3A). Community mixtures explained substantial variation in alpha diversity (richness Adj. $R^2 = 0.117$; Shannon Adj. $R^2 = 0.393$; Simpson Adj. $R^2 = 0.374$) and in Bray-Curtis dissimilarity, jointly accounting for 62.9% of compositional variation (Fig. 3B–C).

Among the detected communities, several showed distinct and characteristic compositions, such as those dominated by *Corynebacterium*, *Staphylococcus*, *Moraxella_c_651924*, or by the combination of *Dolosigranulum* and *Corynebacterium*. Other communities displayed more transient or mixed structures, including the *Peptoniphilus*/ *Anaerococcus*-dominated community defined by gram-positive anaerobic cocci (GPAC), and the *Streptococcus*/ *Moraxella_c_651731* community, which represented a more diverse collection of taxa (Fig. 3A). It is important to note that, as each sample is a mixture of these communities, the relative community (topic) abundances can be used for downstream analyses similarly to taxa-level relative abundances.

Dirichlet regression on topic proportions largely supported genus-level results by summarizing them into coordinated community shifts. Sex contrasts were fully concordant: males had higher proportions of *Staphylococcus* and *Corynebacterium* communities and lower proportions of communities dominated by *Cutibacterium/Lawsonella*, *Dolosigranulum*, and *Moraxella* (all matching the core-genera associations). BMI and smoking similarly mapped to a shift toward the GPAC-dominated community (*Peptoniphilus/Anaerococcus/Finegoldia*) and away from *Dolosigranulum*- and *Corynebacterium*-rich communities, again mirroring genus-level effects. Age-related restructuring was also consistent, showing increasing *Corynebacterium* and *Dolosigranulum* communities and declining *Cutibacterium/Lawsonella* and *Staphylococcus* communities (Fig. 4).

Notable differences were limited and mainly reflected effects that were weak or antibiotic-sensitive at the genus level. For example, FEV_1 showed clearer community-level shifts (higher GPAC community and lower *Staphylococcus* community), whereas individual FEV_1–genus associations were modest and not robust after excluding antibiotic users. Likewise, some BMI- and education-related community effects persisted weakly after excluding antibiotic users, even when several single-genus associations attenuated. Overall, LDA/Dirichlet results support the core-genera findings as a higher-order confirmation of the same host-linked compositional gradients, while suggesting that a few marginal genus-level signals may reflect broader community rebalancing rather than isolated taxon changes.

Discussion

The aim of this study was to characterize how host factors relate to diversity and composition of the anterior nares' microbiome in an adult population-based cohort. Our results demonstrate that nasal microbiota diversity is primarily associated with sex and physiological traits such as BMI and pulmonary function (FEV_1). Beta diversity and core microbiome analyses further revealed that host factors are associated with distinct shifts in microbial community composition. Although the nasal cavity harbors a highly diverse bacterial community, the majority of reads (> 90%) are accounted for by just 15 core genera. The abundance of these core genera is predominantly associated with host physiology and lifestyle-related factors. In addition, a small number of genera were identified as associated with respiratory health, showing consistent associations with clinical and environmental variables.

We showed that the nasal cavity is a diverse microbial habitat, with a total of 358 different bacterial genera detected across the study population. Despite this richness, the community composition was dominated by a limited number of taxa, with only 15 genera accounting for the majority of the total microbial reads. Interestingly, some of these core genera, such as *Moraxella*, exhibited relatively low prevalence but high relative abundance, suggesting the presence of distinct communities characterizing sub-populations. The LDA framework complements the genus-level analyses by capturing co-occurrence structure — i.e., recurrent combinations of taxa that form meaningful sub-communities — rather than treating genera as independent features. In our data, seven communities summarized the dominant anterior nares community types (dominated by either *Corynebacterium*; *Staphylococcus*; *Moraxella_c_651924*; *Dolosigranulum/Corynebacterium*; *Cutibacterium/Lawsonella*; *Streptococcus/Moraxella_c_651731*; or GPAC in the cluster *Peptoniphilus/Anaerococcus*). This representation aligns with the observed dominance of a small core set of genera and explains a substantial fraction of variation in diversity and overall structure, indicating that host-associated differences in the nasal microbiome often reflect shifts between ecotypes rather than isolated changes in single genera.

Taken together, the LDA-derived communities largely mirrored the genus-level findings while providing a higher-order ecological perspective on host–microbiome associations. Most host factors that shaped individual genera also shifted the balance between characteristic community types, indicating that these associations reflect coordinated transitions between stable anterior nares ecotypes rather than isolated taxon changes. Notably, LDA helped differentiate genera with multiple contexts (such as *Moraxella*) by revealing their distinct community environments, and it highlighted a small number of community-level signals—particularly for BMI, FEV_1, and education—that were detectable only at the community level. Thus, the LDA framework primarily serves as a robustness check, confirming the major host-related gradients observed in core genera while sharpening interpretation where genus-level signals were complex or ambiguous.

Differential abundance analyses revealed that *Moraxella* abundance was positively associated with household size and the number of children living in the household. These findings are consistent with results from a Dutch population cohort [35]. In the community framework, these patterns mapped to a *Moraxella_c_651924*-enriched community that co-varied with household context, while a partially overlapping *Streptococcus/Moraxella_c_651731* community captured a more mixed ecotype—clarifying why "*Moraxella*" signals at the genus level can reflect distinct underlying community states.

Host sex emerged as a major factor associated with both alpha diversity and the composition of the nasal core microbiota. While differences between male and female participants may partly reflect variations in lifestyle or environmental exposures, biological mechanisms such as hormonal influences are also likely contributors. Prior research has shown that sex hormones can shape microbial communities across multiple body sites, including the gastrointestinal tract [36–39]. In our study, female participants exhibited higher relative abundances of *Moraxella*, a genus also associated with household size and the number of children living in the household. As it is shown that microbiota are exchanged across household members [40], this may suggest a link between caregiving roles and increased exposure or transmission of *Moraxella*, which is commonly found in children [41]. In contrast, male participants showed higher abundances of several genera belonging to the family *Peptoniphilaceae*, including *Peptoniphilus_a*, *Anaerococcus*, and *Finegoldia*. These gram-positive anaerobic cocci (GPAC) are closely related and are generally considered opportunistic pathogens [42], with *Finegoldia* in particular recognized as the most pathogenic species within this group [43], while *Peptoniphilus spp.* were shown to be associated with necrotizing infections [44] and reduced wound healing in diabetic foot ulcers [45]. Aligning with these genus-level patterns, the LDA analysis indicated that males had higher proportions of *Staphylococcus* and GPAC-enriched communities and lower proportions of *Dolosigranulum* and *Moraxella_c_651924* communities, whereas females showed the opposite trend; sex-stratified community models further suggested that the same host factors can operate through different community types in women and men.

BMI was identified as the second most important factor associated with nasal microbiome diversity, with higher BMI associated with increased alpha diversity and a more even microbial community structure. This is particularly notable given that, in contrast, higher BMI is often linked to reduced gut microbial diversity, although findings in the gut microbiome literature remain heterogeneous [46]. In our study population, increasing BMI was associated with elevated relative abundances of the GPAC cluster of *Peptoniphilus_a*, *Anaerococcus*, *Finegoldia*, and *Campylobacter_b*, alongside reduced levels of *Moraxella_c_651731*. These shifts overlap with patterns observed in smokers, suggesting shared microbial signatures across adiposity and smoking behavior. Notably, *Anaerococcus* has also been reported as enriched in gut microbiota of individuals with obesity [46], supporting a possible role of this genus in obesity-related microbial profiles. Consistently, community-level models confirmed BMI-related shifts toward *Staphylococcus* and GPAC communities and away from *Corynebacterium*/ *Dolosigranulum* communities, reinforcing a link of adiposity to broader community transitions rather than isolated genera.

Smoking behavior, assessed through average lifetime cigarette consumption per day, showed similar associations with increased abundance of the same GPAC genera from the family *Peptoniphilaceae*. At the same time, smoking was strongly associated with reduced levels of genera such as *Lawsonella*, *Dolosigranulum*, and an unclassified genus from the family *Enterobacteriaceae*. These taxa have been described as beneficial or health-associated members of the nasal microbiota [47–51] with *Lawsonella* only recently characterized as a commensal of the nasal cavity [52]. The observed reductions suggest that both smoking and excess adiposity may be linked to depletion of beneficial nasal taxa, potentially increasing susceptibility to inflammation or infection. In parallel, LDA-derived communities showed smoking-associated enrichment of the GPAC community (*Peptoniphilus/Anaerococcus*) and reductions in *Dolosigranulum*-enriched communities, supporting a coherent community-level signature of smoking.

Pulmonary function (FEV_1) was positively associated with nasal microbial diversity. While its impact on the composition of individual core genera was less pronounced, higher FEV_1 was associated with increased abundance of commensal genera such as *Lawsonella* and *Cutibacterium*, both of which were found to decline with age. Assuming FEV_1 serves as a proxy for fitness levels [53], these findings suggest that higher physical activity mitigates an age-associated loss of beneficial nasal bacteria. In community terms, higher FEV_1 related to reduced *Staphylococcus* community proportions and modest increases in GPAC-independent communities, suggesting that pulmonary fitness is linked to community balance beyond single-taxon effects.

While age-related differences in microbial diversity have been consistently reported in other biological niches, such as the gut [54], and in specific populations including children or individuals with chronic rhinosinusitis, we did not observe a significant association between age and alpha diversity in this adult general population cohort. However, age was significantly associated with the composition of the nasal microbiome at the genus level, particularly when adjusting for sex and other covariates. The most prominent age-related trends included a steady decline in the relative abundances of *Lawsonella*, *Staphylococcus*, *Streptococcus*, and *Cutibacterium*. In contrast, the abundances of other health-associated commensals such as *Corynebacterium* and *Dolosigranulum* increased with age. Taken together with the absence of a difference in overall diversity, these findings suggest an age-related compositional shift in the nasal microbiome — from a community dominated by *Cutibacterium* and

Lawsonella in younger adults to one more heavily composed of *Corynebacterium* and *Dolosigranulum* in older individuals. The LDA results mirrored this trajectory, with age associated with higher proportions of *Corynebacterium*/ *Dolosigranulum* communities and lower *Staphylococcus*/ *Cutibacterium_Lawsonella* communities, reinforcing that aging reflects a transition between stable communities rather than a simple gain or loss of diversity.

Antibiotic use within the previous 12 months was associated with marked shifts in the composition of the nasal microbiome. Most notably, participants who reported antibiotic use showed significantly reduced relative abundances of *Dolosigranulum* and *Lawsonella*. This reduction in *Dolosigranulum* is consistent with previous findings in children, where antibiotic exposure has been shown to disrupt nasal microbial communities and reduce levels of health-associated taxa [50, 51, 55]. Interestingly, the magnitude of the observed shifts appeared to be linearly related to the reported frequency of antibiotic use over the previous 12 months, suggesting a frequency-dependent effect of antibiotic use. In parallel, an increase in the relative abundance of *Staphylococcus* was observed in antibiotic users. Given prior evidence that *Dolosigranulum* can inhibit the growth of *Staphylococcus* species [47, 48], inverse patterns observed between the two taxa suggests that reduction of *Dolosigranulum* may facilitate the growth of potentially pathogenic *Staphylococcus* strains in the nasal niche. Concordantly, the community analysis indicated a shift toward *Staphylococcus*-dominated communities and away from *Corynebacterium*/ *Dolosigranulum* communities in antibiotic users; importantly, excluding recent antibiotic users unmasked an association between susceptibility to respiratory infections and opposing *Corynebacterium* (lower) versus *Staphylococcus* (higher) community proportions, suggesting that antibiotic exposure can obscure clinically relevant community signals.

This study provides insights based on data from a large, population-based cohort of adult participants, a group that remains underrepresented in nasal microbiome research. However, several limitations should be considered when interpreting the findings. All health-related diagnoses were self-reported, which may lead to misclassification due to the absence of clinical verification. Moreover, some variables—particularly self-reported health outcomes—may be less reliable and valid than others. The analyses were based on observational data and exploratory in nature and, thus, do not support causal conclusions. While we identified associations between host factors and nasal microbiome features, establishing more intricate, potentially causal relationships and mechanisms will require longitudinal data collection and more detailed investigations of the complex interactions between host physiology, environmental exposures, and microbial communities. Additionally, a relatively high proportion of missing data, especially for early life factors, may have limited the robustness of analyses using these variables. Overall, the presented findings elucidate host factors shaping the microbial composition of the nasal cavity in the adult general population and serve as a foundation for more targeted hypothesis-driven research.

Conclusion

This study provides a comprehensive overview of how host factors shape the diversity and composition of the nasal microbiome in a nationwide general adult population sample up to 73 years old from Germany. By analyzing data from a large population-based cohort, we identified sex, body composition, pulmonary function, smoking, and recent antibiotic use as key associate factors of both overall microbial diversity and the abundance of core nasal genera. Extending these genus-level findings, an LDA-based community analysis showed that host factors predominantly shift mixtures among a small set of recurrent community types. Together, these results serve as a reference for future work on anterior nares' microbiome dynamics in health and disease and provide a foundation for targeted, hypothesis-driven studies of ecotype-mediated mechanisms in respiratory and systemic health.

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.

Availability of data and material

The raw reads of the nasal swabs can be publicly assessed 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/kleineba/nakonasal.

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.

Authors' contributions

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.

Authors’ information

The authors wish to be known that, in their opinion, the last two authors (MV and NR) should be regarded as joint last authors.

Acknowledgements

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.

References

1. Drigot ZG, Clark SE. Insights into the role of the respiratory tract microbiome in defense against bacterial pneumonia. Curr Opin Microbiol. 2024;77:102428. https://doi.org/10.1016/j.mib.2024.102428.
2. Zhao Y, Bitzer A, Power JJ, Belikova D, Torres Salazar BO, Adolf LA, et al. Nasal commensals reduce *Staphylococcus aureus* proliferation by restricting siderophore availability. ISME J. 2024;18:wrae123. https://doi.org/10.1093/ismejo/wrae123.
3. Dimitri-Pinheiro S, Soares R, Barata P. The Microbiome of the Nose—Friend or Foe? Allergy Rhinol. 2020;11:215265672091160. https://doi.org/10.1177/2152656720911605.
4. Mahdavinia M, Keshavarzian A, Tobin MC, Landay AL, Schleimer RP. A comprehensive review of the nasal microbiome in chronic rhinosinusitis (CRS). Clin Exp Allergy. 2016;46:21–41. https://doi.org/10.1111/cea.12666.
5. Candel S, Tyrkalska SD, Álvarez-Santacruz C, Mulero V. The nasopharyngeal microbiome in COVID-19. Emerg Microbes Infect. 2023;12:e2165970. https://doi.org/10.1080/22221751.2023.2165970.
6. Rueca M, Fontana A, Bartolini B, Piselli P, Mazzarelli A, Copetti M, et al. Investigation of Nasal/Oropharyngeal Microbial Community of COVID-19 Patients by 16S rDNA Sequencing. Int J

Environ Res Public Health. 2021;18:2174. https://doi.org/10.3390/ijerph18042174.

7. Dubourg G, Edouard S, Raoult D. Relationship between nasopharyngeal microbiota and patient's susceptibility to viral infection. Expert Rev Anti Infect Ther. 2019;17:437–47. https://doi.org/10.1080/14787210.2019.1621168.
8. Salzano FA, Marino L, Salzano G, Botta RM, Cascone G, D'Agostino Fiorenza U, et al. Microbiota Composition and the Integration of Exogenous and Endogenous Signals in Reactive Nasal Inflammation. J Immunol Res. 2018;2018:1–17. https://doi.org/10.1155/2018/2724951.
9. Pereira PAB, Aho VTE, Paulin L, Pekkonen E, Auvinen P, Scheperjans F. Oral and nasal microbiota in Parkinson's disease. Parkinsonism Relat Disord. 2017;38:61–7. https://doi.org/10.1016/j.parkreldis.2017.02.026.
10. Rei N, Grunho M, Mendes JJ, Fonseca J. Microbiota Orchestra in Parkinson's Disease: The Nasal and Oral Maestros. Biomedicines. 2024;12:2417. https://doi.org/10.3390/biomedicines12112417.
11. Heintz-Buschart A, Pandey U, Wicke T, Sixel-Döring F, Janzen A, Sittig-Wiegand E, et al. The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. Mov Disord. 2018;33:88–98. https://doi.org/10.1002/mds.27105.
12. Tan X, Liu H, Qiu W, Li Z, Ge S, Luo Y, et al. The nasal microbiota is a potential diagnostic biomarker for sepsis in critical care units. Microbiol Spectr. 2024;12:e03441-23. https://doi.org/10.1128/spectrum.03441-23.
13. Flynn M, Dooley J. The microbiome of the nasopharynx. J Med Microbiol. 2021;70. https://doi.org/10.1099/jmm.0.001368.
14. Mahdavinia M. The nasal microbiome: opening new clinical research avenues for allergic disease. Expert Rev Clin Immunol. 2018;14:645–7. https://doi.org/10.1080/1744666X.2018.1500177.
15. Cope EK, Lynch SV. Novel Microbiome-Based Therapeutics for Chronic Rhinosinusitis. Curr Allergy Asthma Rep. 2015;15:9. https://doi.org/10.1007/s11882-014-0504-y.
16. Peters A, Greiser KH, Göttlicher S, Ahrens W, Albrecht M, Bamberg F, 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.
17. Kircher B, Woltemate S, Gutzki F, Schlüter D, Geffers R, Bähre H, et al. Predicting butyrate- and propionate-forming bacteria of gut microbiota from sequencing data. Gut Microbes. 2022;14:2149019. https://doi.org/10.1080/19490976.2022.2149019.
18. Wu YH. idemp: Barcode demultiplex for Illumina I1, R1, R2 fastq.gz files. 2017. https://github.com/yhwu/idemp.
19. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci. 2011;108 supplement_1:4516–22. https://doi.org/10.1073/pnas.1000080107.
20. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics. 2015;31:3476–82. https://doi.org/10.1093/bioinformatics/btv401.

21. Liber JA, Bonito G, Benucci GMN. CONSTAX2: improved taxonomic classification of environmental DNA markers. Bioinformatics. 2021;37:3941–3. https://doi.org/10.1093/bioinformatics/btab347.
22. 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. 2024;42:715–8. https://doi.org/10.1038/s41587-023-01845-1.
23. Huang R, Soneson C, Ernst FGM, Rue-Albrecht KC, Yu G, Hicks SC, et al. TreeSummarizedExperiment: a S4 class for data with hierarchical structure. F1000Research. 2021;9:1246. https://doi.org/10.12688/f1000research.26669.2.
24. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
25. Borman T, Ernst FGM, Shetty SA, Lahti L. mia. https://doi.org/10.18129/B9.BIOC.MIA.
26. Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. J Open Source Softw. 2021;6:3201. https://doi.org/10.21105/joss.03201.
27. Gu Z. Complex heatmap visualization. iMeta. 2022;1:e43. https://doi.org/10.1002/imt2.43.
28. Dahl EM, Neer E, Bowie KR, Leung ET, Karstens L. *microshades*: An R Package for Improving Color Accessibility and Organization of Microbiome Data. Microbiol Resour Announc. 2022;11:e00795-22. https://doi.org/10.1128/mra.00795-22.
29. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. 2001;:2.7-2. https://doi.org/10.32614/CRAN.package.vegan.
30. Chen J, Zhang X, Yang L, Zhang L. GUniFrac: Generalized UniFrac Distances, Distance-Based Multivariate Methods and Feature-Based Univariate Methods for Microbiome Data Analysis. 2012;:1.9. https://doi.org/10.32614/CRAN.package.GUniFrac.
31. Blei DM, Ng AY, Jordan M. Latent dirichlet allocation. J Mach Learn Res. 2003;3:993–1022.
32. Grün B, Hornik K. topicmodels: An *R* Package for Fitting Topic Models. J Stat Softw. 2011;40. https://doi.org/10.18637/jss.v040.i13.
33. Sankaran K, Symul L, Fukuyama J. alto: Aligns Topics across LDA Models. 2023.
34. Fukuyama J, Sankaran K, Symul L. Multiscale Analysis of Count Data through Topic Alignment. 2021. https://doi.org/10.48550/ARXIV.2109.05541.
35. 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;187:4571–4585.e15. https://doi.org/10.1016/j.cell.2024.07.008.
36. Neuman H, Debelius JW, Knight R, Koren O. Microbial endocrinology: the interplay between the microbiota and the endocrine system. FEMS Microbiol Rev. 2015;39:509–21. https://doi.org/10.1093/femsre/fuu010.
37. Shin J-H, Park Y-H, Sim M, Kim S-A, Joung H, Shin D-M. Serum level of sex steroid hormone is associated with diversity and profiles of human gut microbiome. Res Microbiol. 2019;170:192–201. https://doi.org/10.1016/j.resmic.2019.03.003.

38. Tramice A, Paris D, Manca A, Guevara Agudelo FA, Petrosino S, Siracusa L, et al. Analysis of the oral microbiome during hormonal cycle and its alterations in menopausal women: the “AMICA” project. Sci Rep. 2022;12:22086. https://doi.org/10.1038/s41598-022-26528-w.
39. Duran-Pinedo AE, Solbiati J, Frias-Lopez J. The effect of the stress hormone cortisol on the metatranscriptome of the oral microbiome. Npj Biofilms Microbiomes. 2018;4:25. https://doi.org/10.1038/s41522-018-0068-z.
40. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, et al. Cohabiting family members share microbiota with one another and with their dogs. eLife. 2013;2:e00458. https://doi.org/10.7554/eLife.00458.
41. Goldstein EJC, Murphy TF, Parameswaran GI. *Moraxella catarrhalis*, a Human Respiratory Tract Pathogen. Clin Infect Dis. 2009;49:124–31. https://doi.org/10.1086/599375.
42. 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. Manual of Clinical Microbiology. 1st edition. Wiley; 2011. p. 803–16. https://doi.org/10.1128/9781555816728.ch48.
43. Boyanova L, Markovska R, Mitov I. Virulence arsenal of the most pathogenic species among the Gram-positive anaerobic cocci, Finegoldia magna. Anaerobe. 2016;42:145–51. https://doi.org/10.1016/j.anaerobe.2016.10.007.
44. Brown K, Church D, Lynch T, Gregson D. Bloodstream infections due to Peptoniphilus spp.: report of 15 cases. Clin Microbiol Infect. 2014;20:O857–60. https://doi.org/10.1111/1469-0691.12657.
45. Min KR, Galvis A, Baquerizo Nole KL, Sinha R, Clarke J, Kirsner RS, et al. Association between baseline abundance of Peptoniphilus, a Gram-positive anaerobic coccus, and wound healing outcomes of DFUs. PLOS ONE. 2020;15:e0227006. https://doi.org/10.1371/journal.pone.0227006.
46. Pinart M, Dötsch A, Schlicht K, Laudes M, Bouwman J, Forslund SK, et al. Gut Microbiome Composition in Obese and Non-Obese Persons: A Systematic Review and Meta-Analysis. Nutrients. 2021;14:12. https://doi.org/10.3390/nu14010012.
47. Brugger SD, Eslami SM, Pettigrew MM, Escapa IF, Henke MT, Kong Y, et al. Dolosigranulum pigrum Cooperation and Competition in Human Nasal Microbiota. mSphere. 2020;5:e00852-20. https://doi.org/10.1128/mSphere.00852-20.
48. Stubbendieck RM, Hurst JH, Kelly MS. Dolosigranulum pigrum: A promising nasal probiotic candidate. PLOS Pathog. 2024;20:e1011955. https://doi.org/10.1371/journal.ppat.1011955.
49. Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijser BJF, et al. Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children. Am J Respir Crit Care Med. 2014;190:1283–92. https://doi.org/10.1164/rccm.201407-1240OC.
50. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development. Cell Host Microbe. 2015;17:704–15. https://doi.org/10.1016/j.chom.2015.03.008.

51. Pettigrew MM, Laufer AS, Gent JF, Kong Y, Fennie KP, Metlay JP. Upper Respiratory Tract Microbial Communities, Acute Otitis Media Pathogens, and Antibiotic Use in Healthy and Sick Children. Appl Environ Microbiol. 2012;78:6262–70. https://doi.org/10.1128/AEM.01051-12.
52. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems. 2018;3:10.1128/msystems.00187 – 18. https://doi.org/10.1128/msystems.00187-18.
53. Jakes RW. Physical Inactivity Is Associated with Lower Forced Expiratory Volume in 1 Second: European Prospective Investigation into Cancer-Norfolk Prospective Population Study. Am J Epidemiol. 2002;156:139–47. https://doi.org/10.1093/aje/kwf021.
54. Thomas S, Izard J, Walsh E, Batich K, Chongsathidkiet P, Clarke G, et al. The Host Microbiome Regulates and Maintains Human Health: A Primer and Perspective for Non-Microbiologists. Cancer Res. 2017;77:1783–812. https://doi.org/10.1158/0008-5472.CAN-16-2929.
55. Prevaes SMPJ, De Winter-de Groot KM, Janssens HM, De Steenhuijsen Piters WAA, Tramper-Stranders GA, Wyllie AL, et al. Development of the Nasopharyngeal Microbiota in Infants with Cystic Fibrosis. Am J Respir Crit Care Med. 2016;193:504–15. https://doi.org/10.1164/rccm.201509-1759OC.

Figures

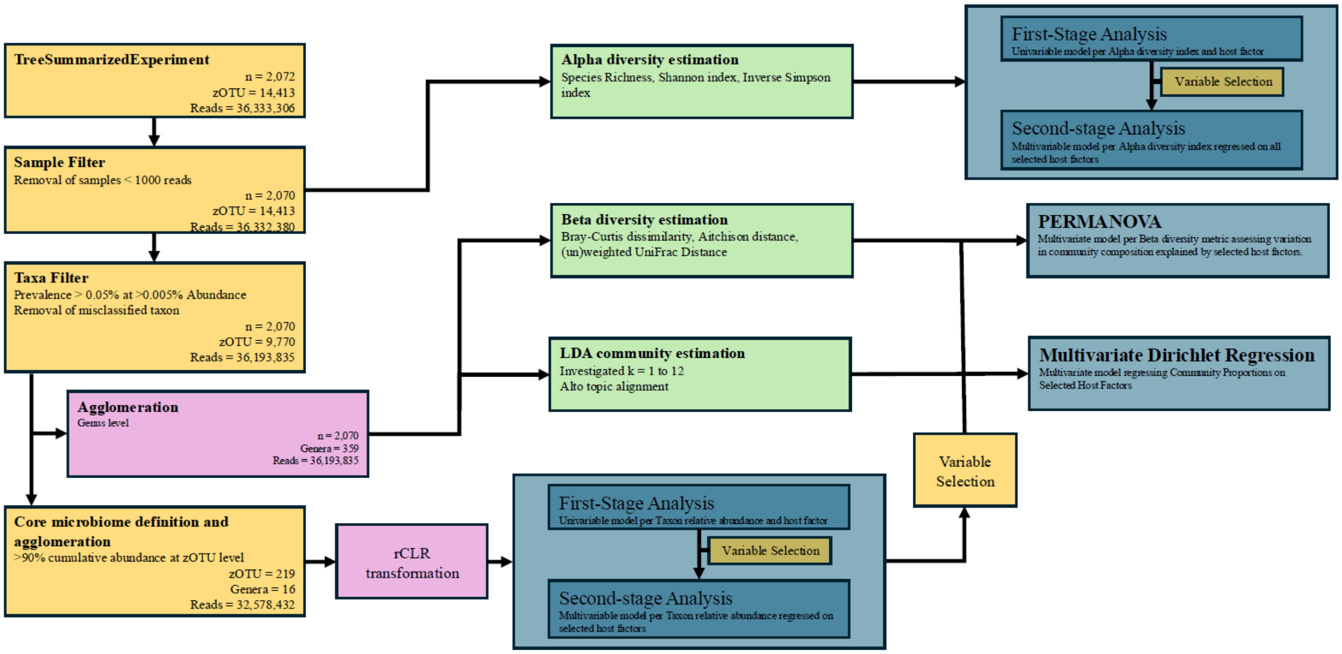

Figure 1

Overview of processing and analysis workflow. Workflow illustrating sample and taxon filtering (yellow), transformations (purple), estimations (green) and statistical inference (blue). After quality filtering and prevalence-based taxon reduction, alpha and beta diversity metrics were computed alongside core microbiota identification. Latent Dirichlet Allocation (LDA) was used to infer community types, followed by Dirichlet regression, PERMANOVA, and a two-stage regression framework for assessing univariate and multivariable associations between host factors and both diversity measures and taxon-level abundances.

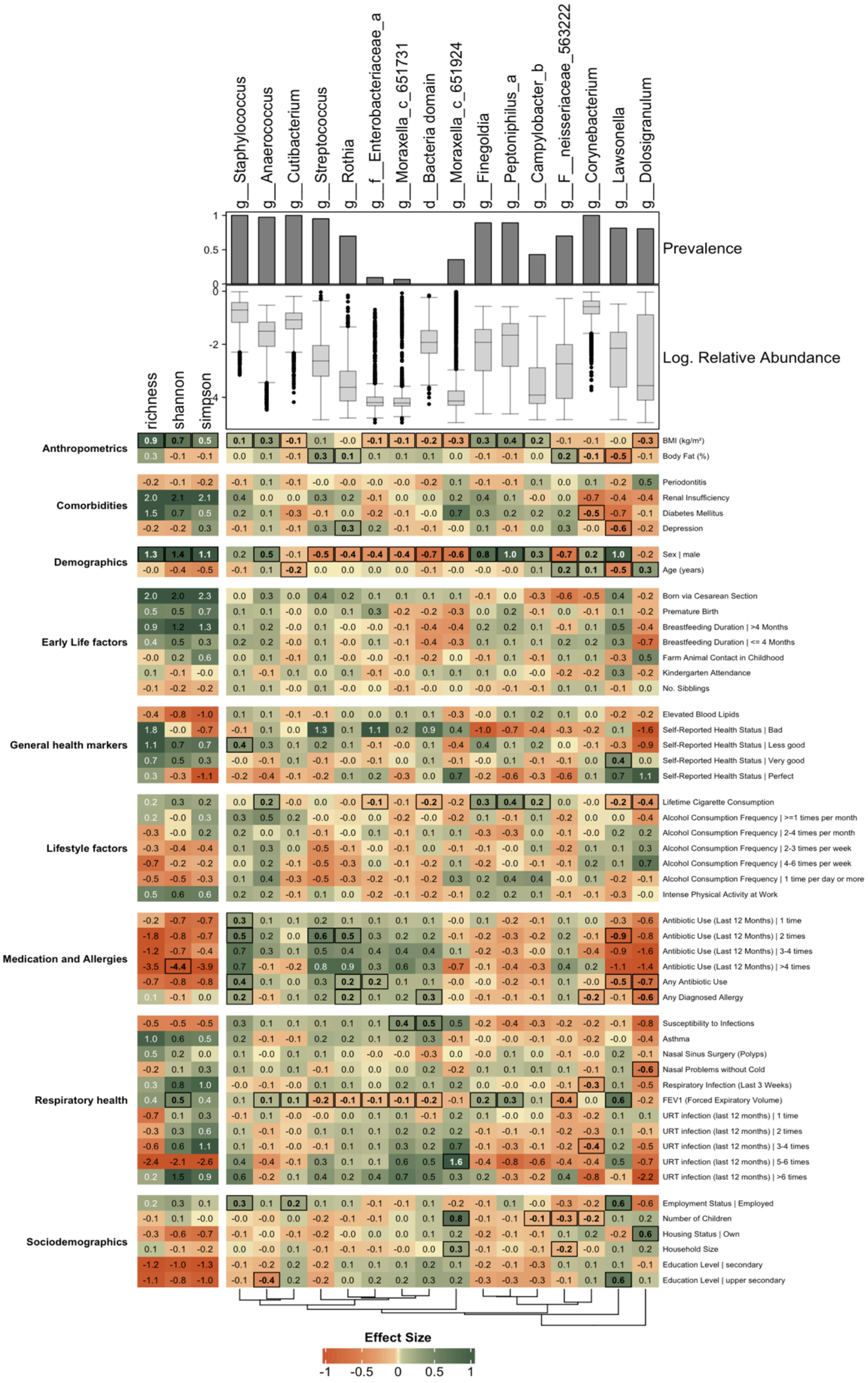

Figure 2

Heatmap of stage-one (univariate) regression coefficients. Coefficients are regression coefficients of host factors regressed on each taxon or diversity estimate, respectively. Bordered boxes indicate q-value < .05 (FDR-corrected). Numbers are displayed in white or black for easier visibility. Boxplots represent log-transformed relative abundances and bar charts represent prevalence across the entire sample.

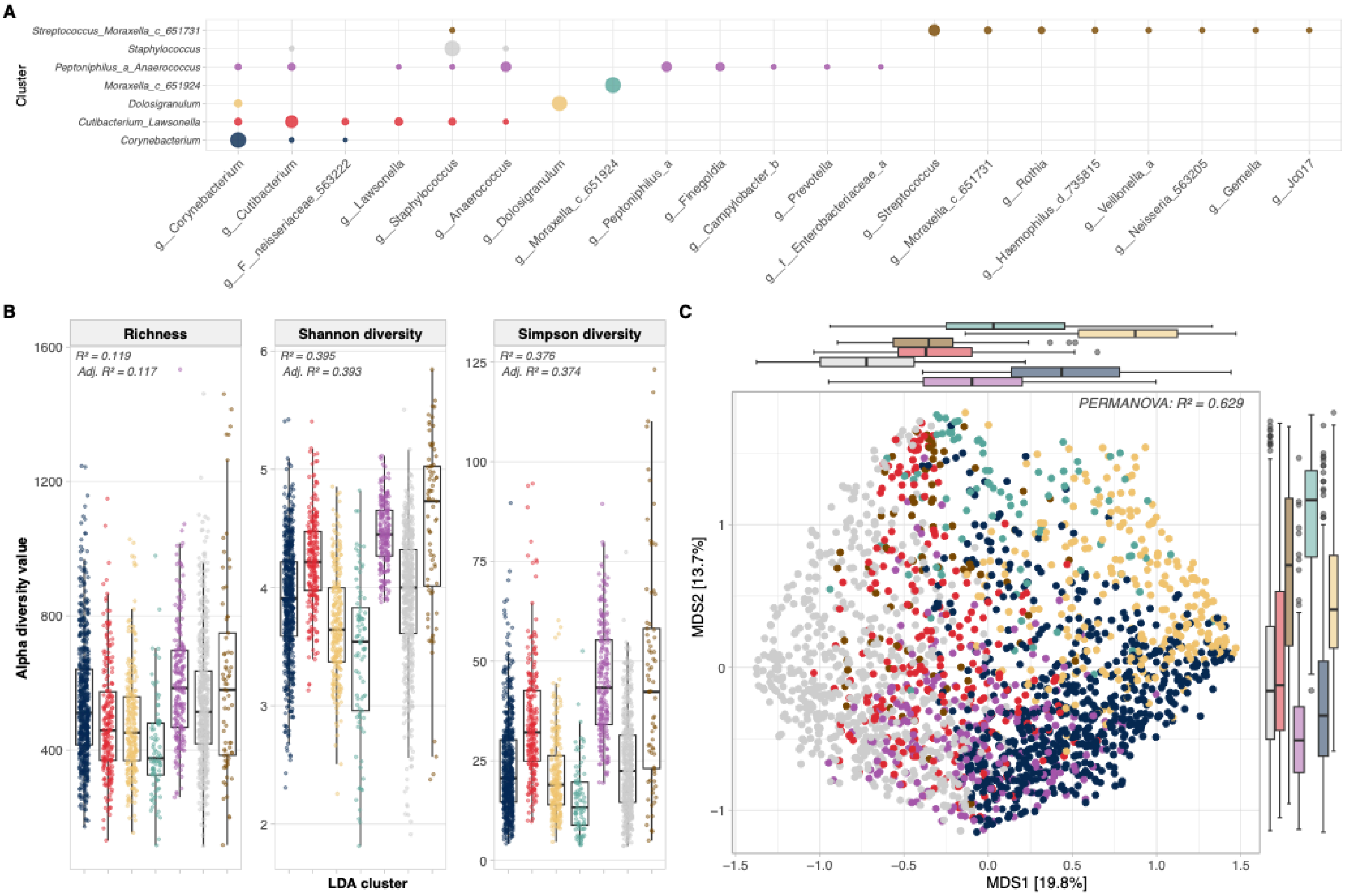

Figure 3

Taxonomic composition and integration of LDA-derived anterior nares community types. (A) Relative contributions of individual genera to each Latent Dirichlet Allocation (LDA) community, illustrating the dominant or characteristic taxa defining the seven inferred community types. (B) Associations between community types and alpha diversity metrics (richness, Shannon, Simpson), showing that LDA communities differ markedly in within-sample diversity. Explained variation was estimates uring multivariable regression analysis regressing diversity metrics on relative abundance of commuty types. (C) Ordinations of Bray-Curtis dissimilarities, colored by community type, demonstrating clear separation of sub-communities and indicating that LDA community structure explains a substantial proportion of overall compositional variation as determined by multivariable PERMANOVA models.

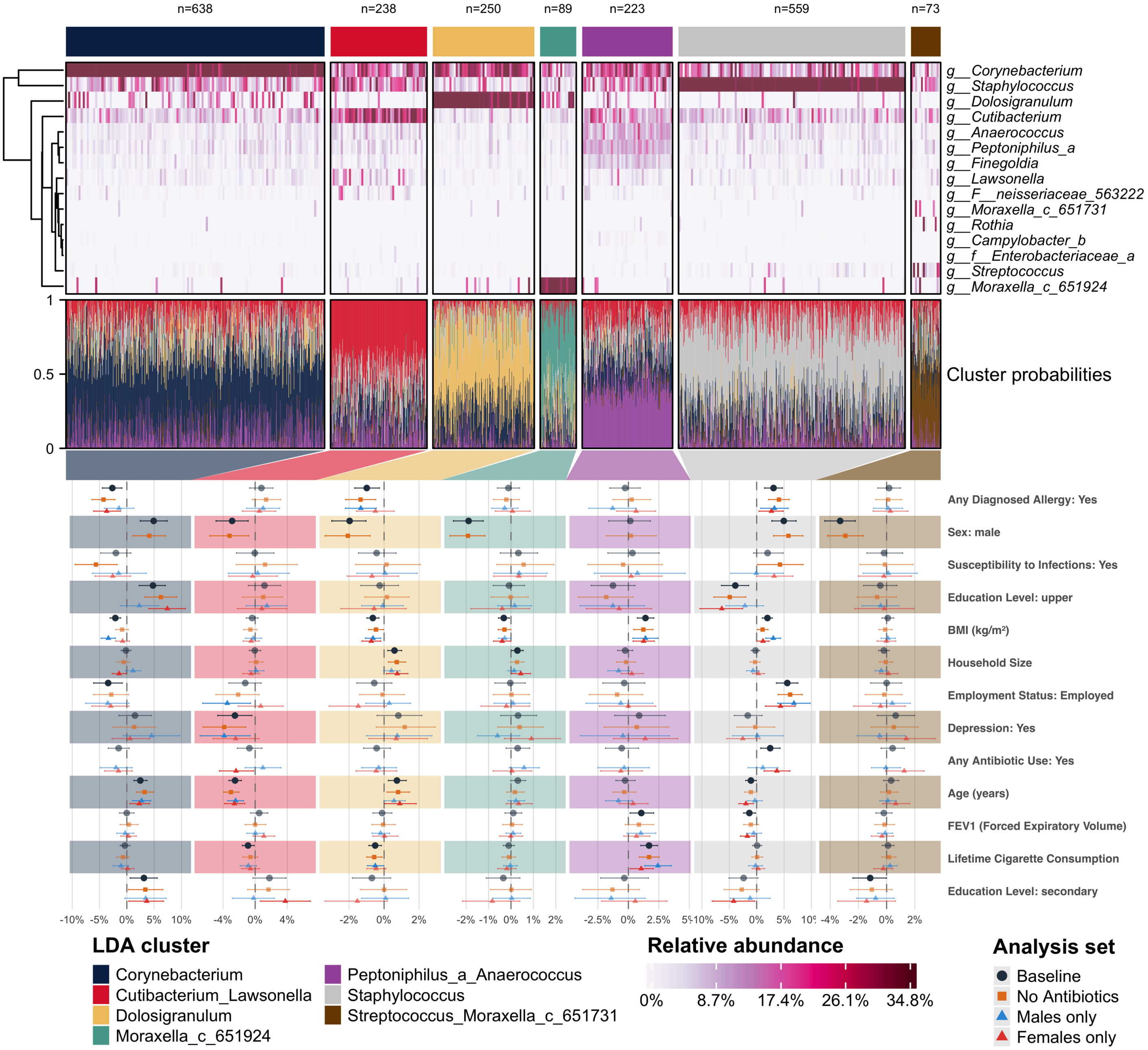

Figure 4

Latent Dirichlet Allocation (LDA) community structure and host-associated differences in community proportions. LDA identified seven characteristic anterior nares community types, shown in the upper panel as taxon-specific abundance heatmaps and sample-level community probability profiles (middle panel). The lower panel displays estimated associations between host factors and community proportions, based on Bayesian Dirichlet regression models. Points represent mean differences in community proportion, with 95% credible intervals shown for the baseline analysis, antibiotic-free subset, and sex-stratified models. Community labels are based on the dominant or defining genera.

Supplementary Files

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

- SUPPLnasalmicrobiomedescriptive20260216.docx