

Concordance between upper and lower airway microbiota in children with Cystic Fibrosis

Sedreh Nassirnia¹, Valentin Scherz¹, Gilbert Greub^{1,2}, Giorgia Caruana¹, Patrick Taffé³, Katia Jaton¹, Sebastien Papis⁴, Klara M. Posfay-Barbe⁴, Anne Mornand⁵, Isabelle Rochat-Guignard⁶, Claire Bertelli^{1*}, and Sandra A. Asner^{2, 7*}

¹ Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

² Infectious Diseases Service, Department of Internal Medicine, Lausanne University Hospital, Lausanne, Switzerland

³ Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

⁴ Pediatric Infectious Diseases Unit, Division of General Pediatrics, Children's Hospital of Geneva, University Hospitals of Geneva, Geneva, Switzerland

⁵ Pediatric Pulmonology Unit, Children's Hospital of Geneva, University Hospitals of Geneva, Geneva, Switzerland

⁶ Pediatric Pulmonology Unit, Department Mother-Woman-Child, Lausanne University Hospital, Lausanne, Switzerland

⁷ Pediatric Infectious Diseases and Vaccinology Unit, Department Mother-Woman-Child, Lausanne University Hospital, Lausanne, Switzerland

* Equal contribution

Corresponding author

Sandra Andrea Asner, MD, MSc
Pediatric Infectious Diseases and Vaccinology Unit
Department Mother-Women-Child
Rue du Bugnon 46; CH-1011 Lausanne, Switzerland
Email: Sandra.Asner@chuv.ch
Phone: +41 21 314 10 79; Fax: +41 21 314 10 33
ORCID iD: 0000-0001-7911-0752

Keywords: cystic fibrosis, pathogen detection, metabarcoding, community profiling, lower respiratory microbiota

Abstract

Background: Sputum is the sample to monitor the lower respiratory tract microbiota in cystic fibrosis (CF), but young patients often cannot expectorate. We hypothesized that throat swabs could reflect lower airway colonization and assessed the concordance of bacterial community composition between paired sputum and throat swab samples from children with CF.

Methods: The prospective longitudinal multicenter MUCOVIB cohort included 379 samples from 61 CF children. Using V3-V4 16S rRNA amplicon metagenomics, we compared bacterial community diversity and composition between sputum and throat swabs in the full cohort and in 11 patients with paired samples from the same visit.

Results: Sputum and Throat swabs exhibited similar bacterial diversity, regardless of the exacerbation status, and presented a substantial agreement for detecting pathogens (Cohen's Kappa: 0.6). Differences in bacterial abundance were observed ($p=0.001$), but not presence/absence ($p=0.098$). Community typing revealed three distinct community types, with 86% of paired samples falling into the same cluster, highlighting the homogeneity between sputum and throat swabs microbiota. Network analysis demonstrated slight, non-random similarities in microbial interactions between sample types ($ARI = 0.08$ and 0.10). The average distance between samples collected from the same visit was shorter ($0.505, \pm 0.056$ 95%CI), compared to sputum ($0.695, \pm 0.017$) or throat swab ($0.704, \pm 0.045$) from the same patient collected during different visits.

Conclusions: Throat swabs can provide representative information on lower respiratory microbiota. Clinicians should collect throat swabs rather than relying on sputum samples from previous visits to guide antibiotic prescriptions in CF children unable to expectorate.

1. Introduction

Cystic fibrosis (CF) is characterized by chronic pulmonary exacerbations, primarily driven by persistent bacterial infections, that lead to progressive lung disease [1–3]. The evolution of CF's respiratory tract microbiome - from an initially diverse population to one increasingly dominated by pathogenic bacteria such as *Staphylococcus aureus*, *Haemophilus influenzae*, and *Pseudomonas aeruginosa* - highlights the interest in precise microbiota monitoring, to identify turning points in disease progression such as *P. aeruginosa* colonization [4,5].

The Climax-Attack Model (CAM) provides insights into the microbial dynamics within the CF lung [6,7]. The model differentiates 'attack' communities present during exacerbations, characterized by transient species colonization that triggers acute immune responses and lower pH through sugar fermentation, from 'climax' communities of more stable, slower-growing microbes resistant to antibiotics, associated with nitrogen waste management and an increased pH during stable periods. This framework underscores the dynamic nature of CF pathology and the importance of understanding microbial shifts for effective disease management [8]. Indeed, guidelines recommend regular microbiological workup to tailor antimicrobial therapy [10].

A significant challenge in CF management is obtaining the appropriate sample for respiratory microbiota analysis [9], especially among non-expectorating infants and young children. Bronchoalveolar lavage (BAL) is the gold standard and the best representative sample of lower respiratory microbiota. Yet, regular bronchoscopy is not recommended due to its invasive nature and the absence of proven clinical benefit [10]. Expectoated sputum stands as the best alternative for adults and older children capable of expectorating [10]. Throat swabs (TS) are the recommended alternative for infants and preschoolers [11]. Yet, culture-independent analyses have questioned the ability of TS to capture the lower respiratory tract microbiota, where chronic infection and inflammation occur [12]. Furthermore, interpreting TS results in children is complicated by age-related variations in the expected healthy microbial within different niches. This variability particularly affects infants and preschoolers, leading to inconsistent results across pediatric age groups and fueling debates about the reliability of TS in these cohorts [13–17].

The **Muc**oviscidosis, **R**espiratory **V**iruses, and **I**ntracellular **B**acteria (MUCOVIB) project, a Swiss prospective longitudinal multicenter cohort, collected TS, sputum, nasopharyngeal swabs, bronchoalveolar lavage, and bronchial aspirate, to detect respiratory bacteria and

viruses in 61 children with CF under 18 years of age [18, 19]. The efficacy of the 16S assay for pathogen detection compared to conventional culture results on respiratory samples showed fair agreement (0.4) based on Cohen's Kappa score [23]. However, the 16S assay was more sensitive in detecting *H. influenzae* and *S. aureus* in respiratory samples compared to standard culture methods.

The present sub-study aims to clarify whether TS is a reliable proxy to sputum samples to reflect the microbial dynamics in the lower respiratory tract. By analyzing 16S rRNA sequence data from paired sputum and TS, we evaluated the concordance between microbial communities in these sample types, and key factors influencing their composition. Through community typing and ecological network analysis, focusing on keystone species and their interactions, we investigated the similarity in microbial community structure between sputum and TS during follow-up visits and pulmonary exacerbations.

2. Materials and methods

2.1. Study population and design

Details of the collection of clinical information, study design, and protocol have been published in the first MUCOVIB studies [18, 19]. Briefly, 61 patients were recruited from the respiratory clinics of Lausanne and Geneva University Hospital Centers and followed routinely by their treating pulmonologists. Culture, quantitative PCRs, and 16S amplicon-based metagenomics were used to assess the presence of viruses and bacteria in the respiratory tract. Overall, 269 TS and 51 sputa samples were collected longitudinally every 3 months from children with CF. The present MUCOVIB substudy evaluates the microbial composition of sputum and TS collected from the entire cohort and a subset of eleven individuals who provided 22 paired sputa and TS samples at the same visit.

2.2. Bioinformatics and statistical analyses

The composition of the microbial community was assessed through analyses of the V3-V4 regions of the 16S rRNA gene, sequenced using the Illumina MiSeq platform, as previously described [18, 19]. Raw sequencing reads are available in the European Nucleotide Archive (ENA) with project number PRJEB41059. The zAMP pipeline, an in-house DADA2-based bioinformatics tool, was used to process the paired-end sequences from the 16S amplicons (<https://github.com/metagenlab/zAMP> release v 0.9.15).

To assess the concordance between sputum and TS for detecting pathogens, Cohen's Kappa statistic was computed using R software (v4.2.2). The `vegan` [20], `mia` [21], and `ggplot2` packages were used for downstream statistical analysis and visualization. Wilcoxon rank-sum test was used to compare bacterial alpha diversity across two sample types, using paired tests for matched samples. Beta diversity was assessed by Bray-Curtis distance, considering the abundance of bacteria, and Jaccard distance, accounting for their presence/absence. Homogeneity of group dispersion was calculated using the `betadisper` test and bacterial composition between sample types was compared using the permutational multivariate ANOVA (PERMANOVA) test with 999 permutations. Adjusted p-values obtained from the BH method and values below 0.05 were considered statistically significant.

To examine the effect of host factors on variation of microbial composition, distance-based redundancy analysis (db-RDA) was performed with Bray-Curtis and Jaccard distances using `capscale` [20]. The species profiles from paired sputum and TS were used for community typing using the Dirichlet Multinomial Mixtures model (DMM) [22] and Laplace approximation, which was applied to identify the optimal number of DMMs. For both analyses, missing metadata for FEV1% (percentage of predicted Forced Expiratory Volume) and LCI (lung clearance index) were imputed using the median, as it is robust against outliers and skewed data.

The `NetCoMi` [23] package was used to construct and compare networks graphically and quantitatively. For this analysis, we applied sparse Correlations for Compositional data (SparCC) correlation coefficients [24], considering species that were present in at least 50% of each sample type and selecting associations with SparCC > 0.5. To identify topology and structural differences between networks of sputum and TS, a differential network was constructed where only differentially associated taxa are connected with only significant ($P < 0.05$) and strongly positive (SparCC > 0.5) correlations.

3. Results

3.1. Characteristics of the study population

The microbial composition of sample types was compared across the entire MUCOVIB cohort, consisting of 61 children (average age: 7.4; IQR [3.7; 12.2]). This cohort was followed for up to 756 days, with a median number of 7 visits (IQR [5;8]), resulting in the

collection of a total of 379 samples. Furthermore, the sub-study focused on 22 paired sputum and TS samples collected during the same visit from 11 children, which represents 18% of the cohort. Among these 11 patients, 36% (4 children) contributed a single pair of samples, while 64% (7 children) contributed two or more pairs.

While the sex ratio was slightly imbalanced towards males (59%) in the entire cohort, in the subset of patients who provided both sample types on the same day of the visit, males and females accounted for 72.7% and 27.3%, respectively (**Table 1**). FEV1% was lower and LCI was increased during exacerbations. Although FEV1% did not show a significant difference between the entire cohort and the paired cohort during follow-up visits, the difference was significant during pulmonary exacerbations. Additionally, LCI values were increased in the cohort of paired samples. Patients in the paired cohort tended to be older, with an average age of 12.7 years (IQR: 9.7- 14.6), and exhibited more severe disease symptoms, as correlated to the observed increase in LCI values.

3.2. Comparability of sputum and throat swabs for pathogen detection

To assess the reliability of TS, we measured pathogen detection rates in paired sputum and TS specimens using 16S assay. Compared to sputum, which was considered the gold standard, the 16S assay on TS showed an overall sensitivity of 78% and a specificity of 98%. **Table 2** details the Cohen's Kappa scores, along with sensitivity and specificity values for the detection of five common CF pathogens. The Kappa score indicates a variable level of agreement in detection rates between paired sputum and TS samples depending on the bacterial species, ranging from substantial agreement for *Achromobacter* (0.69) to no agreement for *Stenotrophomonas maltophilia* (0), likely due to the small sample size. For *S. aureus*, throat samples yielded a sensitivity of 94%, a specificity of 25%, a positive predictive value (PPV) of 81%, and a negative predictive value (NPV) of 50%. In the case of *P. aeruginosa*, TS demonstrated sensitivities, specificities, PPVs, and NPVs of 50%, 83%, 40%, and 88%, respectively. The sensitivity rates varied significantly between the two sample types, with sputum samples yielding higher sensitivities.

3.3. Microbiome diversity and composition in paired sputum and throat swab samples

Microbial diversity was calculated across sample types, using Chao1, indicating ASV richness, and the Shannon diversity index, which accounts for ASV richness and evenness (**Fig. 1A, B**). Overall richness was not significantly different between the two sample types within the entire MUCOVIB cohort (p-value: 0.34) as well as in paired samples (p-values: 0.59). Similarly, comparisons of the Shannon index revealed no significant differences across both sample types. Hence, sputum and TS may offer similar insights into microbial diversity, irrespective of the patient's exacerbation status (**Fig. S1A, B**). Furthermore, bacterial communities across sample types were compared at the ASV level. In the MUCOVIB cohort, variability (dispersion) within microbial communities was significantly different between sputum and TS based on the abundance of bacteria (**Fig. 1C**, Bray-Curtis index, p-value: 0.05), but not based on the species composition (Jaccard index, p-value: 0.9). The significant difference in the overall bacterial community structure for both indices (PERMANOVA p-value: 0.001) indicated distinct microbial compositions between sample types. Similar results were observed within the paired samples, with a significant difference in dispersion using Bray-Curtis index (**Fig. 1D**, p-value: 0.04) but not with the Jaccard index (p-value: 0.41). In both cohorts, a PERMANOVA performed using sample type and visits (regular follow-up or exacerbations) as grouping factors, while stratifying by patient (**Fig. S1C, D**) suggested that both sample types (p-value: 0.001) and clinical status of the patient (p-value: 0.001) independently affect microbial community composition. However, their interaction did not show a significant impact (p-value: 0.64).

Patient-specific signatures in the microbiota were evident from the clustering of sputum and TS samples collected from the same visit observed in both the hierarchical clustering (**Fig. 2A**) and the NMDS (**Fig. 2B**) based on the presence/absence of the bacteria. A multivariate multiple regression model was built to evaluate the correlation between clinical and demographic covariates and the observed microbial communities. A significant proportion (57%, p-value: 0.001) of the variation in species abundances based on the adjusted R² (**Fig. 2C**) was attributed to the patient's visit. Only 15% of the total microbiota variation was explained by the exacerbation status, demographic factors (gender, age), and pulmonary functions (FEV1%, LCI) thereby suggesting that intra-individual variability mostly affected the microbial composition.

3.4. Microbial community typing

We then investigated whether variance in bacterial abundance reflects distinct microbial community types, or if community structures remain consistent despite quantitative variances

across the two sample types. Using DMM, three distinct microbial community types, referred to as clusters, were detected (**Fig. 3A**). We hypothesize that they could reflect host physiological states, rather than random statistical variances in microbial populations between sputum and TS samples, and potentially influence disease severity and clinical outcomes. Eighty-six percent of paired samples (19/22) were grouped into the same cluster, indicating a high degree of homogeneity between microbial communities in sputum and TS regardless of exacerbation status. The relative abundance heatmap of dominant taxa at species level provides a detailed view of the variability within each cluster (**Fig. 3A**).

Cluster 1 is notable for a community composition similar to the oropharyngeal flora, featuring a mix of facultatively anaerobic bacteria such as *Streptococcus* sp. and *Granulicatella adiacens*, alongside the obligate anaerobe *Fusobacterium nucleatum* and aerobic *Neisseria* spp.. This cluster presents a microbiota profile common to both throat and sputum samples, which may reflect a baseline microbial environment in CF patients. While sharing similarity with cluster 1 for the presence of anaerobic bacteria, *Streptococcus* sp. and *Neisseria* sp., cluster 2 is marked by a significant presence of *Veillonella* sp., a common organism from the oral microbiota, *Gemella haemolysans*, and *Haemophilus parainfluenzae*. Additionally, the abundance of bacteria present in cluster 2 is more stable over patient visits, whereas cluster 1 displays more dynamic changes. In contrast, cluster 3 is composed of samples from only three patients dominated by *S. aureus*.

Cluster 1, which contains more pulmonary exacerbation visits, displayed a lower species richness and evenness compared to cluster 2 (**Fig. 3B**). Cluster 3 showed the lowest diversity, maybe owing to the small sample size (n=4) limiting the statistical power. When comparing patient clinical and demographic variables across clusters, there was no significant age difference (**Fig. 3C**). Patients in cluster 1 had a lower FEV1% (**Fig. 3D**) and higher LCI scores (**Fig. 3E**) than cluster 2, indicating a reduced lung function. Clusters 1 and 3 had a higher proportion of patients with the homozygous F508del genotype, known to be associated with more severe disease phenotypes (**Fig. 3F**).

3.5 Network

Expanding on the compositional similarities between sample types revealed by DMM analysis, we built an ecological network to delve deeper into the capacity of TS to capture microbial interactions within the lower respiratory tract microbiota of non-expectorating patients (**Figure 4A, B**). Both networks displayed several clusters of microbial species that

are associated, positively or negatively, within each sample type, exhibiting a mixture of conserved and niche-specific associations. For instance, the preservation of clusters of *Haemophilus*, *Veillonella*, *Porphyromonas*, *Neisseria*, and *Gemella* across both upper and lower respiratory samples in CF patients underlines the stability of these microbial associations despite the differing physical and immunological characteristics of these niches. Quantitative comparisons of global network properties between the two sample types revealed no significant differences in key metrics (**Table S1**). Adjusted Rand Index (ARI) for the whole network and the Largest Connected Component (LCC) indicated a non-random similarity between the two networks (ARI of 0.08 and 0.10, respectively) with significant p-values ($p=0.001$ and 0.008). Whereas the sets of central nodes within the two microbial networks present moderate similarity in centrality measures, hub taxa presented a noticeable divergence ($P(\leq \text{Jacc})=0.03^*$) between the two sample types (**Table S2**). Furthermore, while several species are present in both networks, their roles as hub nodes differ, indicating the varied contributions of these central species to the ecological dynamics within each niche. In the sputum network, hub nodes such as *H. parainfluenzae*, *Solobacterium moorei*, and *Neisseria* sp. play central roles, while in the throat swab network, different taxa emerge as hubs, specifically *Capnocytophaga leadbetteri*, *Bergeyella* DQ241813_s, *Actinomyces* sp., and *Veillonella rogosae*. None of the nodes had significantly different centrality measures (**Table S1**) and no significant differential association was detected after multiple testing adjustments.

To contextualize the ecological networks within the CAM framework and assess the ability of the TS to reflect exacerbation status, we constructed microbial networks at the climax and attack phases. In the climax phase, sputum samples exhibited a substantially larger relative LCC size (0.92) compared to TS (0.62), indicating a more complex and interconnected microbial network (**Table S3**). Significant variation in hub taxa suggests that the species concentrating interactions in the lungs (represented by sputum) may not hold the same influence in the TS (**Table S4**). During the attack phase, there was no substantial difference in the overall topology and the types of connections or graphlets between networks (**Table S5**). However, a significant difference in the proportion of cooperative interactions among taxa (positive edge percentage) was observed between sputum (49.08%) and TS (54.13%). Similarly, network analysis identified significant differences according to exacerbation status (**Table S6**).

3.6. Temporal dynamics of microbial communities across patient visits

Building upon our analyses that demonstrate a concordance between microbial compositions of the upper and lower respiratory tracts in children with CF, we further investigated the temporal stability of these microbial communities. Specifically, we aimed to evaluate the reliability of TS as a proxy for lower respiratory tract microbiota at the time of patient visits, as compared to sputum samples or TS from previous visits. To address this, the Generalized Estimating Equations (GEE) method, which accounts for the intra-individual correlation of repeated measures, was employed to estimate the mean beta-diversity distances using both the Bray-Curtis and Jaccard indices while considering the dependence between observations from the same individual (**Fig. 5A, B**). For the Bray-Curtis index, the GEE model indicated an average distance of 0.51 (95% CI: 0.45-0.57) between sputum and TS from the same visit. In contrast, the average distance between sputum samples from different visits was estimated at 0.57 (95% CI: 0.52-0.62), and the distance between TS samples from different visits was 0.66 (95% CI: 0.64-0.68), suggesting increased variability over time. Similarly, for the Jaccard index, sample types collected from the same visit presented a more similar bacterial composition (0.47; 95% CI: 0.42-0.52) than sputum or TS collected from different visits (0.68; 95% CI: 0.64-0.72 and 0.69; 95% CI: 0.67-0.70, respectively).

4. Discussion

In this study, we compared the microbial landscape of TS and sputum using 16S rRNA amplicon sequencing to address a critical need for non-invasive, reliable sampling methods in CF. By focusing on paired samples collected on the same day and using a culture-independent method, we provided a direct comparison of the two microbiota and showed a reduced distance as compared to samples collected from the same site and the same patient during another visit. Our findings suggest the reliability of TS, collected on the same day of the visit, to mirror the CAM's Climax and Attack phases in CF patients.

TS exhibited comparable microbial alpha-diversity (richness and evenness), with partially similar microbial composition to sputum. This is consistent with Boutin et al. [16], who demonstrated a close microbiota between sputum and TS in non-exacerbating patients. While both sample types harbor a broad range of microbial species, differences in relative abundances likely reflect anatomical and functional differences between the two niches.

The multivariate analysis showed that the predictive value of TS in relation to clinical and demographic covariates, compared to sputum, varies according to the individual-specific nature of the CF microbiota. Similarly, Prevaes et al. [12] noted that the degree of concordance between the microbial communities of different sample types with their corresponding BAL microbiota was variable within each patient when studied by either conventional culture or molecular assays.

Interestingly, community typing revealed that sputum-TS pairs often fall within the same cluster, suggesting consistency in microbial community structure across sample types. This consistency contrasts with the inter-individual variability observed across clusters, likely inherent to host-related factors such as CFTR F508del allele zygosity. Stephanie et al. [26] and Juan de Dios Caballero et al. [27] also observed unique microbiome patterns within each patient over time, despite variations in bacterial populations and without clear associations with exacerbation status or antibiotic treatment, thereby suggesting that personalized monitoring and treatment approaches may be beneficial in CF care.

Additionally, network analysis showed notable consistency in the microbial associations between sputum and TS across different clinical states (climax and attack phases) in our small population of children with exacerbation. The preservation of several anaerobes' interactions, including *Veillonella* and *Prevotella*, between the two networks suggests they play roles in respiratory infections. Their ability to break down carbohydrates and mucins, could contribute to metabolic diversity, influencing immune modulation or biofilm formation within the mucosal environment of CF patients [28, 30, 31, 32]. The presence of these key bacterial taxa in both sample types support the potential utility of TS for monitoring microbial dynamics in young children with CF unable to expectorate and not yet colonized by *P. aeruginosa*. However, hub taxa—key players within each network—remained distinct, indicating differences in microbial roles across the two sample types.

Our 16S amplicon-based study provides insights into bacterial identity rather than function. Hence the taxa presence in TS may not necessarily reflect active biological roles specific to the ecological niche. Throat microbiota may partly be constituted of transient spillover bacteria from the mouth (saliva) and lungs (sputum) rather than a stable, throat-specific entity [29]. The larger Bray-Curtis distances between TS samples within individuals (**Fig. 6**) supported this hypothesis. Additionally, imbalanced sampling of paired samples among patients could introduce bias into the co-occurrence network. Patients with more frequent

sampling may disproportionately influence inferred microbial associations and interactions. Not all multivariate analyses were adjusted to host-related factors such as F508del genotypes and antibiotic exposure, thereby limiting the assessment of specific patient-signatures.

Despite these limitations, the overall trends and key findings remain robust. When sputum samples are not available, the simplified collection of throat swabs has the potential to facilitate our understanding of disease progression and treatment response in young children with CF. Future studies with larger sample sizes including children with exacerbations should employ normalization techniques to weight samples or interactions based on the number of visits per patient. Large-scale longitudinal studies including F508del genotyping are needed to quantify the stability of microbial signatures, exchanges between compartments, and establish the functional profile of microbial communities across sample types.

5. Funding

MUCOVIB study was supported by grants from the Leenaards and Santos-Suarez Foundations, Vifor, and Novartis pharmaceutical industries. This analysis was supported as part of NCCR Microbiomes, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 180575). VS was supported by an SNSF Grant (n° 10531C-170280–F. Taroni, L. Falquet, and G. Greub).

6. Conflict of interest statement

The authors declare no conflict of interest linked to this manuscript.

References

1. Cox, M.J.; Allgaier, M.; Taylor, B.; Baek, M.S.; Huang, Y.J.; Daly, R.A.; Karaoz, U.; Andersen, G.L.; Brown, R.; Fujimura, K.E.; et al. Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. *PLoS One* 2010, 5, e11044, doi:10.1371/journal.pone.0011044.
2. *Cystic Fibrosis*; Marshall, S.G., Borowitz, D., Stanton, B.F., Eds.; Pediatric clinics of North America; Elsevier: Philadelphia, Pennsylvania, 2016; ISBN 978-0-323-45983-9.
3. Rosenfeld, M.; Emerson, J.; Williams-Warren, J.; Pepe, M.; Smith, A.; Montgomery, A.B.; Ramsey, B. Defining a Pulmonary Exacerbation in Cystic Fibrosis. *J Pediatr* 2001, 139, 359–365, doi:10.1067/mpd.2001.117288.
4. Zanin, M.; Baviskar, P.; Webster, R.; Webby, R. The Interaction between Respiratory Pathogens and Mucus. *Cell Host Microbe* 2016, 19, 159–168, doi:10.1016/j.chom.2016.01.001.
5. Thornton, C.S.; Acosta, N.; Surette, M.G.; Parkins, M.D. Exploring the Cystic Fibrosis Lung Microbiome: Making the Most of a Sticky Situation. *Journal of the Pediatric Infectious Diseases Society* 2022, 11, S13–S22, doi:10.1093/jpids/piac036.
6. Soret, P.; Vandenberg, L.-E.; Francis, F.; Coron, N.; Enaud, R.; Avalos, M.; Schaeffer, T.; Berger, P.; Fayon, M.; Thiebaut, R.; et al. Respiratory Mycobion and Suggestion of Inter-Kingdom Network during Acute Pulmonary Exacerbation in Cystic Fibrosis. *Sci Rep* 2020, 10, 3589, doi:10.1038/s41598-020-60015-4.
7. Conrad, D.; Haynes, M.; Salamon, P.; Rainey, P.B.; Youle, M.; Rohwer, F. Cystic Fibrosis Therapy: A Community Ecology Perspective. *Am J Respir Cell Mol Biol* 2013, 48, 150–156, doi:10.1165/rcmb.2012-0059PS.
8. Bevivino, A.; Bacci, G.; Drevinek, P.; Nelson, M.T.; Hoffman, L.; Mengoni, A. Deciphering the Ecology of Cystic Fibrosis Bacterial Communities: Towards Systems-Level Integration. *Trends in Molecular Medicine* 2019, 25, 1110–1122, doi:10.1016/j.molmed.2019.07.008.
9. Lahiri, T.; Hempstead, S.E.; Brady, C.; Cannon, C.L.; Clark, K.; Condren, M.E.; Guill, M.F.; Guillerman, R.P.; Leone, C.G.; Maguiness, K.; et al. Clinical Practice Guidelines From the Cystic Fibrosis Foundation for Preschoolers With Cystic Fibrosis. *Pediatrics* 2016, 137, e20151784, doi:10.1542/peds.2015-1784.
10. Ronchetti, K.; Tame, J.-D.; Paisey, C.; Thia, L.P.; Doull, I.; Howe, R.; Mahenthalingam, E.; Forton, J.T. The CF-Sputum Induction Trial (CF-SpIT) to Assess Lower Airway Bacterial Sampling in Young Children with Cystic Fibrosis: A Prospective Internally Controlled Interventional Trial. *Lancet Respir Med* 2018, 6, 461–471, doi:10.1016/S2213-2600(18)30171-1.
11. Borowitz D, Robinson KA, et al. Cystic Fibrosis Foundation evidence-based guidelines for management of infants with cystic fibrosis. *J Pediatr.* 2009;155(6 Suppl):S73-S93. doi:10.1016/j.jpeds.2009.09.001
12. Prevaes, S.M.P.J.; De Steenhuijsen Piters, W.A.A.; de Winter-de Groot, K.M.; Janssens, H.M.; Tramper-Stranders, G.A.; Chu, M.L.J.N.; Tiddens, H.A.; Van Westreenen, M.; Van Der Ent, C.K.; Sanders, E.A.M.; et al. Concordance between Upper and Lower Airway Microbiota in Infants with Cystic Fibrosis. *Eur Respir J* 2017, 49, 1602235, doi:10.1183/13993003.02235-2016.
13. Kumpitsch, C., Koskinen, K., Schöpf, V., & Moissl-Eichinger, C. (2019). The microbiome of the upper respiratory tract in health and disease. *BMC biology*, 17(1), 87. <https://doi.org/10.1186/s12915-019-0703-z>.
14. Maughan, H.; Wang, P.W.; Diaz Caballero, J.; Fung, P.; Gong, Y.; Donaldson, S.L.; Yuan, L.; Keshavjee, S.; Zhang, Y.; Yau, Y.C.W.; et al. Analysis of the Cystic Fibrosis Lung Microbiota via Serial Illumina Sequencing of Bacterial 16S rRNA Hypervariable Regions. *PLoS ONE* 2012, 7, e45791, doi:10.1371/journal.pone.0045791.
15. Ahmed, B.; Cox, M.J.; Cuthbertson, L.; James, P.L.; Cookson, W.O.C.; Davies, J.C.; Moffatt, M.F.; Bush, A. Comparison of the Upper and Lower Airway Microbiota in Children with Chronic Lung Diseases. *PLoS ONE* 2018, 13, e0201156, doi:10.1371/journal.pone.0201156.
16. Boutin, S.; Graeber, S.Y.; Weitnauer, M.; Panitz, J.; Stahl, M.; Clausnitzer, D.; Kaderali, L.; Einarsson, G.; Tunney, M.M.; Elborn, J.S.; et al. Comparison of Microbiomes from Different Niches

- of Upper and Lower Airways in Children and Adolescents with Cystic Fibrosis. *PLoS ONE* 2015, *10*, e0116029, doi:10.1371/journal.pone.0116029.
17. Fenn, D.; Abdel-Aziz, M.I.; Brinkman, P.; Kos, R.; Neerinx, A.H.; Altenburg, J.; Weersink, E.; Haarman, E.G.; Terheggen-Lagro, S.W.J.; Maitland-van Der Zee, A.H.; et al. Comparison of Microbial Composition of Cough Swabs and Sputum for Pathogen Detection in Patients with Cystic Fibrosis. *Journal of Cystic Fibrosis* 2022, *21*, 52–60, doi:10.1016/j.jcf.2021.08.031.
 18. Pittet, L.F.; Bertelli, C.; Scherz, V.; Rochat, I.; Mardegan, C.; Brouillet, R.; Jaton, K.; Mornand, A.; Kaiser, L.; Posfay-Barbe, K.; et al. *Chlamydia Pneumoniae* and *Mycoplasma Pneumoniae* in Children with Cystic Fibrosis: Impact on Bacterial Respiratory Microbiota Diversity. *Pathogens and Disease* 2021, *79*, ftaa074, doi:10.1093/femspd/ftaa074.
 19. Scherz, V.; Caruana, G.; Taffé, P.; Brouillet, R.; Bertelli, C.; Jaton, K.; Fougère, Y.; Posfay-Barbe, K.M.; Mornand, A.; Rochat-Guignard, I.; et al. Unexpected Associations between Respiratory Viruses and Bacteria with Pulmonary Function Testing in Children Suffering from Cystic Fibrosis (MUCOVIB Study). *Journal of Cystic Fibrosis* 2022, *21*, e158–e164, doi:10.1016/j.jcf.2021.10.001.
 20. Oksanen, J.; Simpson, G.L.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O’Hara, R.B.; Solymos, P. Vegan: Community Ecology Package. Ordination Methods, Diversity Analysis and Other Functions for Community and Vegetation Ecologists.
 21. Ernst F, Shetty S, Borman T, Lahti L (2023). *_mia: Microbiome analysis_*. R package version 1.10.0, <https://github.com/microbiome/mia>.
 22. Holmes, I.; Harris, K.; Quince, C. Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. *PLoS ONE* 2012, *7*, e30126, doi:10.1371/journal.pone.0030126.
 23. Peschel, S.; Müller, C.L.; Von Mutius, E.; Boulesteix, A.-L.; Depner, M. NetCoMi: Network Construction and Comparison for Microbiome Data in R. *Briefings in Bioinformatics* 2021, *22*, bbaa290, doi:10.1093/bib/bbaa290.
 24. Friedman, J.; Alm, E.J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput Biol* 2012, *8*, e1002687, doi:10.1371/journal.pcbi.1002687.
 25. Coburn, B.; Wang, P.W.; Diaz Caballero, J.; Clark, S.T.; Brahma, V.; Donaldson, S.; Zhang, Y.; Surendra, A.; Gong, Y.; Elizabeth Tullis, D.; et al. Lung Microbiota across Age and Disease Stage in Cystic Fibrosis. *Sci Rep* 2015, *5*, 10241, doi:10.1038/srep10241.
 26. Widder, S., Carmody, L.A., Opron, K. *et al.* Microbial community organization designates distinct pulmonary exacerbation types and predicts treatment outcome in cystic fibrosis. *Nat Commun* 15, 4889 (2024). <https://doi.org/10.1038/s41467-024-49150-y>
 27. de Dios Caballero J, Vida R, Cobo M, Máiz L, Suárez L, Galeano J, Baquero FCantón R, del Campo R 2017. Individual Patterns of Complexity in Cystic Fibrosis Lung Microbiota, Including Predator Bacteria, over a 1-Year Period. *mBio* 8:10.1128/mbio.00959-17. <https://doi.org/10.1128/mbio.00959-17>
 28. Quinn, R.A.; Whiteson, K.; Lim, Y.W.; Zhao, J.; Conrad, D.; LiPuma, J.J.; Rohwer, F.; Widder, S. Ecological Networking of Cystic Fibrosis Lung Infections. *npj Biofilms Microbiomes* 2016, *2*, 4, doi:10.1038/s41522-016-0002-1.
 29. Goddard, A. F., Staudinger, B. J., Dowd, S. E., Joshi-Datar, A., Wolcott, R. D., Aitken, M. L., Fligner, C. L., & Singh, P. K. (2012). Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(34), 13769–13774. <https://doi.org/10.1073/pnas.1107435109>
 30. Mirković B, Murray MA, Lavelle GM, Molloy K, Azim AA, Gunaratnam C, Healy F, Slattery D, McNally P, Hatch J, Wolfgang M, Tunney MM, Muhlebach MS, Devery R, Greene CM, McElvaney NG. The Role of Short-Chain Fatty Acids, Produced by Anaerobic Bacteria, in the Cystic Fibrosis Airway. *Am J Respir Crit Care Med*. 2015 Dec 1;192(11):1314-24. doi: 10.1164/rccm.201505-0943OC. PMID: 26266556; PMCID: PMC4731701.
 31. Flynn JM, Niccum D, Dunitz JM, Hunter RC. Evidence and Role for Bacterial Mucin Degradation in Cystic Fibrosis Airway Disease. *PLoS Pathog*. 2016 Aug 22;12(8):e1005846. doi: 10.1371/journal.ppat.1005846. PMID: 27548479; PMCID: PMC4993466.
 32. Welp AL, Bomberger JM. Bacterial Community Interactions During Chronic Respiratory Disease. *Front Cell Infect Microbiol*. 2020 May 14;10:213. doi: 10.3389/fcimb.2020.00213. PMID: 32477966; PMCID: PMC7240048.

It is made available under a [CC-BY 4.0 International license](#) .

Tables

Table 1. Patients' characteristics in MUCOVIB and the subcohort

Clinical parameters	Complete cohort 61 patients	Paired samples subcohort 11 patients	Significant differences (p-value)
Gender			
Female	25 (41.0%)	3 (27.3%)	0.9
Male	36 (59.0%)	8 (72.7%)	0.7
Age at inclusion (Years)	7.4 [3.7, 12.2]	12.7 [9.7, 14.6]	0.02
Comorbidities	55 (90.2%)	11 (100%)	0.04
Pancreatic insufficiency	50 (82.0%)	10 (90.9%)	
Allergic bronchopulmonary aspergillosis	4 (6.6%)	2 (18.8%)	
Diabetes	6 (9.8%)	2 (18.8%)	
Liver disease	5 (8.2%)	2 (18.8%)	
Others	16 (26.2%)	4 (36.6%)	
Follow-up (days)	553 [456, 640]	470 [441, 585]	0.09
Genotype			0.9
Heterozygous for DeltaF508	24 (39.3%)	5 (45.5%)	
Homozygous for DeltaF508	28 (45.9%)	5 (45.5%)	
Other than DeltaF508	9 (14.8%)	1 (9.1%)	
Pulmonary exacerbation (N of events)	32	5	0.9
FEV1%			
Follow-up visits (n=210)	91.0 [81.0, 99.0]	90.0 [73.0, 95.0] (n=5)	0.4
Pulmonary exacerbation (n = 14)	82.0 [58.0, 97.5]	71.0 [58.0, 85.75] (n = 8)	0.001
Lung clearance index (LCI)			
Follow-up visits (n=194)	9.3 [8.2, 11.1]	10.75 [9.87, 13.50]	0.1
Pulmonary exacerbation (n = 9)	9.7 [8.0, 12.8]	10.8 [10.4, 13.67]	0.006

FEV₁: Forced expiratory volume in 1 second

LCI: The lung clearance index

Table 2. Sensitivity and specificity of throat swabs compared to sputum for pathogen detection by 16S amplicon-based metagenomics.

<i>S.pneumoniae</i>	<i>Achromobacter spp.</i> <i>H. influenzae</i> <i>P. aeruginosa</i> <i>S. aureus</i> <i>S. maltophilia</i>						
	EXP	TS					
+	+	3	3	2	17	0	3
+	-	2	4	2	1	1	2
-	+	0	0	3	3	0	0
-	-	17	13	15	1	21	21
Sensitivity		60	56	50	94	0	100
Specificity		100	100	83	25	100	100
Kappa score		0.69	0.59	0.30	0.24	0	1

Figures

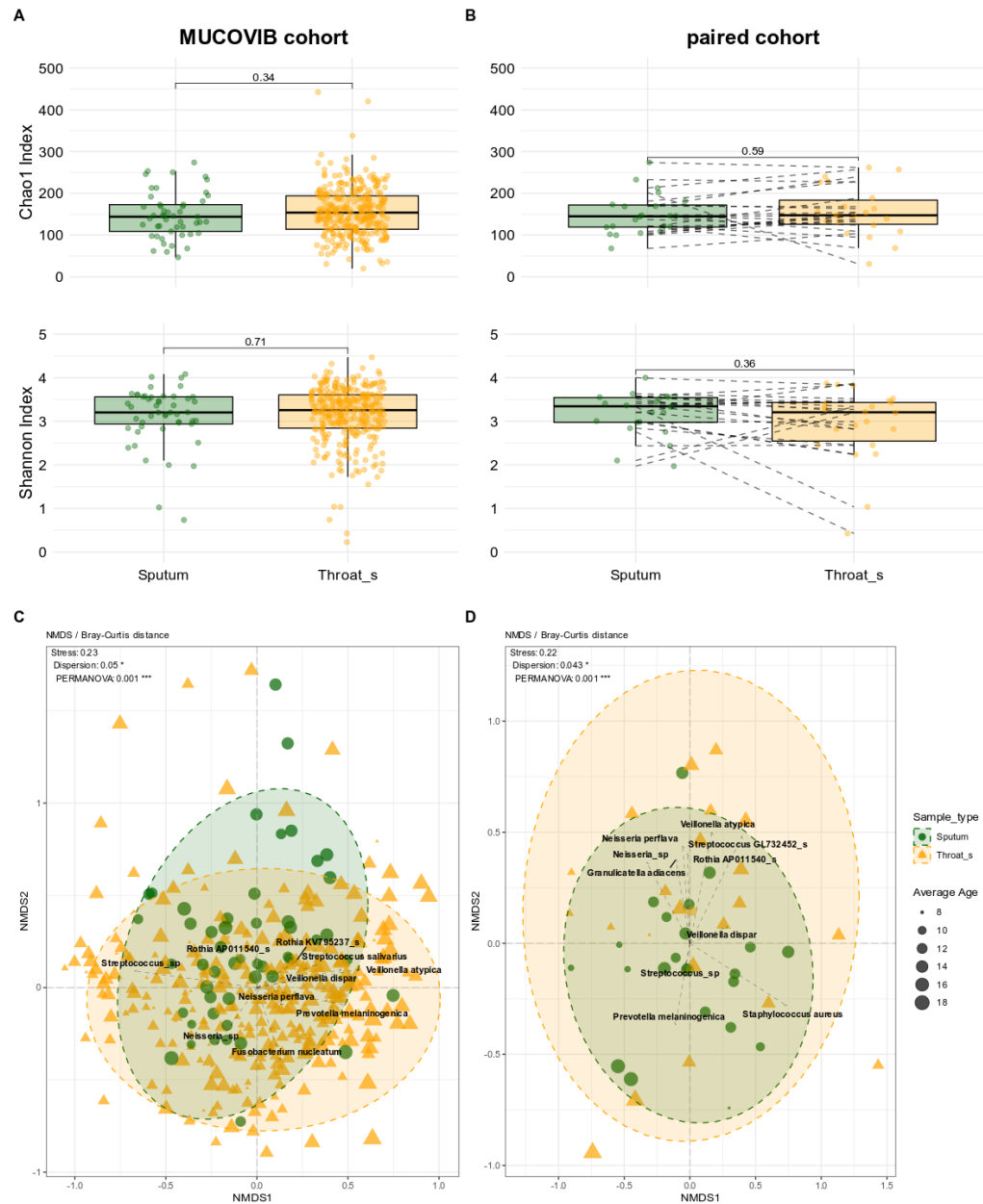


Figure 1. Microbiome diversity and composition in throat and sputum samples. (A, B) The box plot illustrates Chao1 and Shannon alpha diversity indices for each sample type in A. the MUCOVIB and B. paired sputum and throat swabs cohorts. The results of the Wilcoxon test (pairwise for paired samples) are indicated within the graphs. (C, D) Non-metric multidimensional scaling (NMDS) plot with Bray-Curtis distance at the ASV level for the complete and the sub-cohorts, respectively. Sputum samples were illustrated with green circles and throat swabs in orange triangles. Ellipses represent confidence intervals around clusters of sample types. Stress, dispersion, and PERMANOVA p-value indicate the goodness-of-fit, variability within groups, and the statistical significance of the observed clusters, respectively. Ten top abundant species are shown in the plots.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

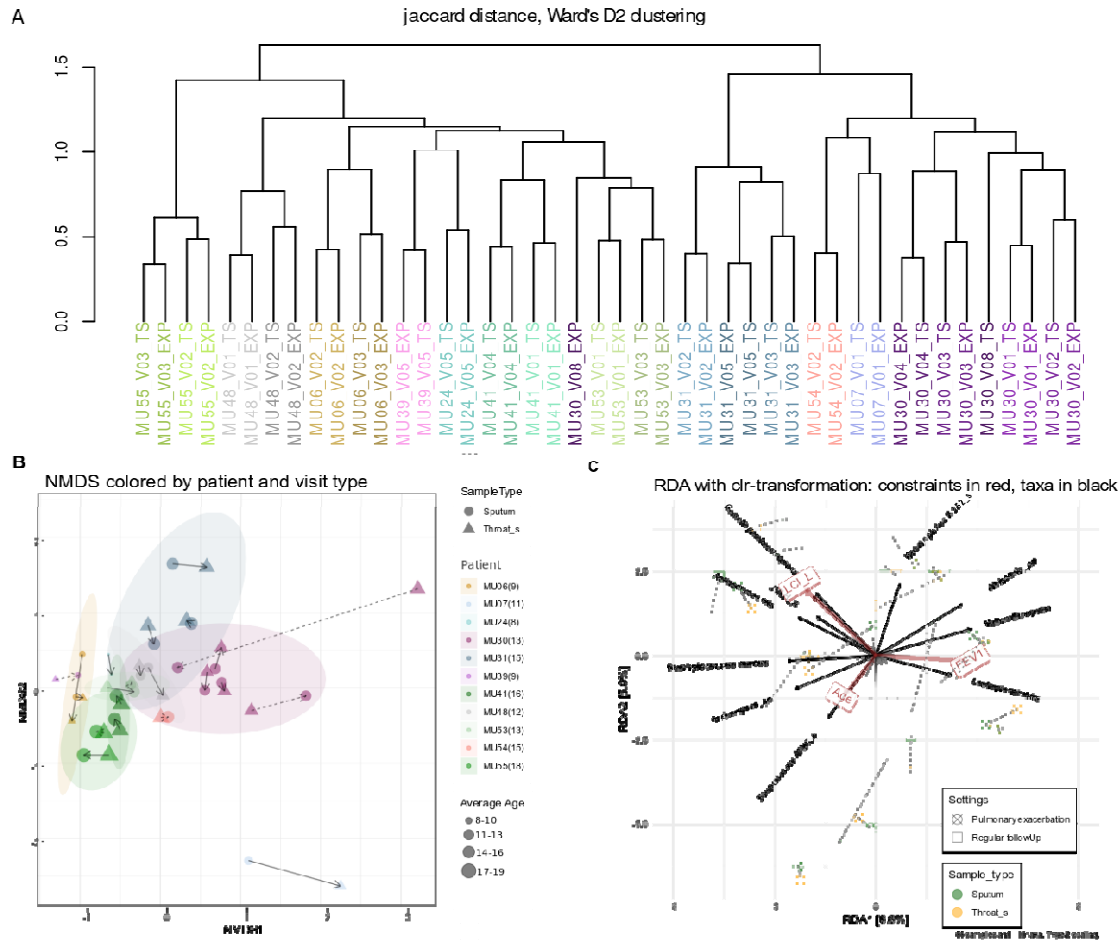
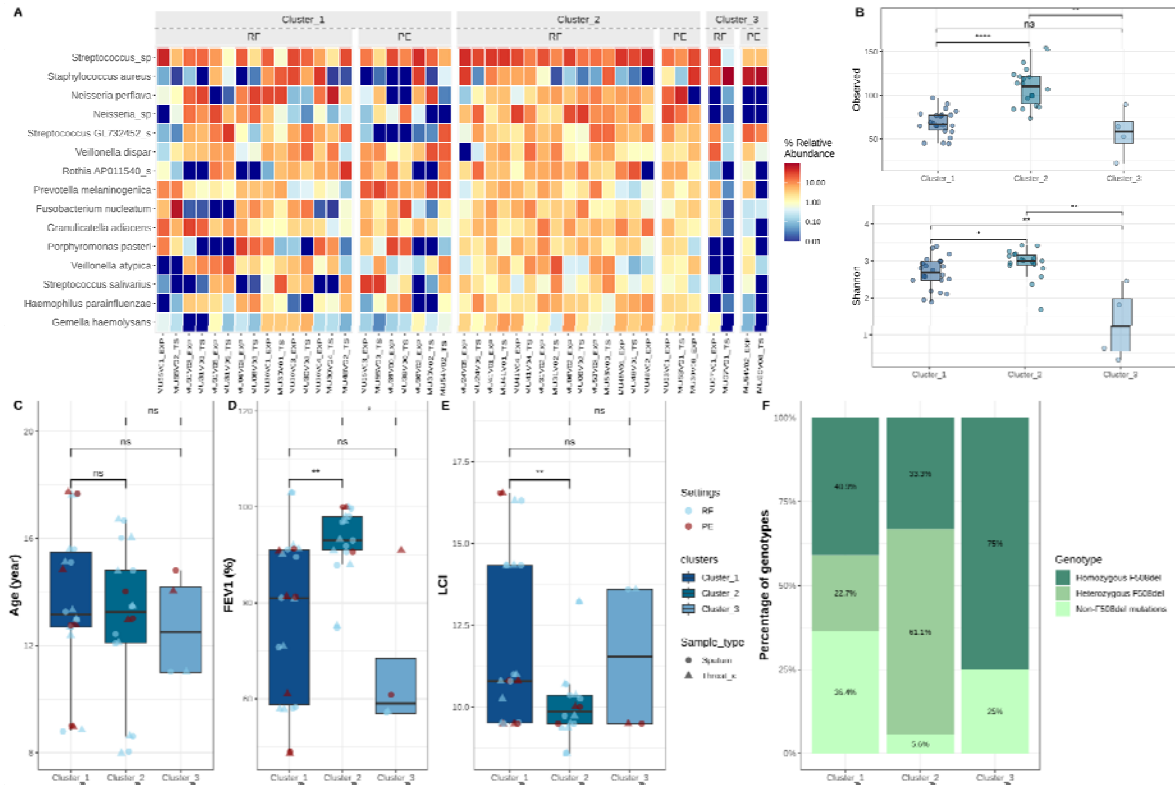


Figure 2. Patient-specific microbial community dynamics across visits. (A) The hierarchical clustering dendrogram based on Jaccard similarities illustrates the similarity in microbial community composition at the ASV level across different patient visits. Each leaf in the dendrogram represents a patient visit, with the color denoting individual patients and the shade variations reflecting different visits. The height of the branches indicates the degree of dissimilarity between clusters. (B) NMDS plot with Jaccard distance at the ASV level. Each point represents a sample colored by a patient and samples from the same visits connected by a line. The type of the line presents the kind of the visit, detailing individual variations in microbial communities from sputum and throat swabs during regular follow-up (solid lines) and pulmonary exacerbation (dashed lines) visits. Points are shaped according to sample types, with sizes representing the average age of patients. This visualization illustrates potential age-related shifts in microbial composition. To ensure anonymity, the age data was shifted by a random factor between 0 and 6 months. (C) Redundancy Analysis (RDA) using Bray-Curtis distance at the Species level. Each sample is represented with a shape indicating the type of the visit and its color represents the sample type. Dashed lines connect paired samples collected from the same visit. Arrows represent constrained variables with their direction and length indicating the strength and direction of their association with the microbial communities.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

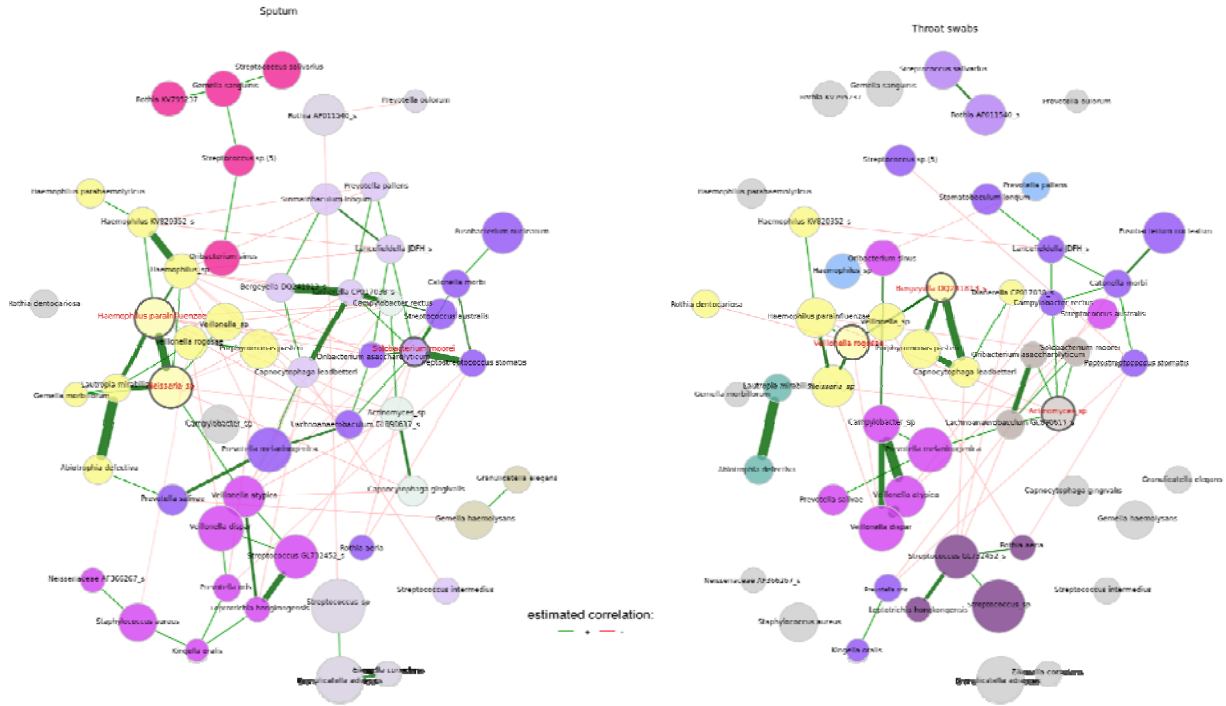


Figure 4. Microbial interactions network of sputum and throat swabs. The estimated correlations using SparCC are transformed into dissimilarities via the “signed” distance metric and the corresponding similarities are used as edge weights. The mclr-transformed abundance was used for defining hubs and scaling node sizes. Clusters were determined using greedy modularity optimization, and nodes were colored based on the clusters. To facilitate the comparison across sample types, the color of clusters differ between networks when the clusters do not share at least two nodes. Green edges correspond to positive estimated bacterial associations and red edges to negative ones. Only nodes that are unconnected in both networks were removed. To enhance visual comparison and highlight distinctions effectively, the layout computed for the sputum network was applied consistently across both networks. Hub species are highlighted with a red label.

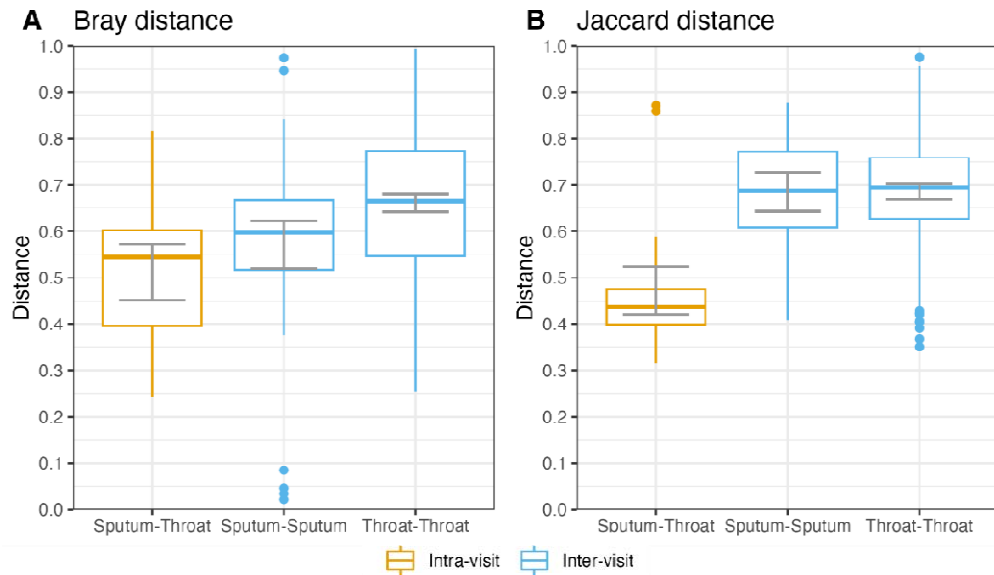


Figure 5. Beta-diversity distance between samples inter- and intra-visit. The average Bray-Curtis (A) and Jaccard (B) distances were calculated using GEE to account for intra-individual correlations. In both panels, orange boxes donate intra-visit comparisons, indicating the distance between throat swabs compared to sputum from the same visit, while blue boxes represent inter-visit comparisons, detailing the distance between samples from different visits. The median of each group is indicated by the central line within each box, with the box edges defining the interquartile range (IQR). The whiskers extend to the furthest points within 1.5 IQR from the box. Gray lines represent the 95% confidence intervals for each distance, providing insight into the variability and statistical certainty of the measurements.