University of Dundee

Sputum metagenomics of people with bronchiectasis

Rosenboom, Ilona; Thavarasa, Ajith; Richardson, Hollian; Long, Merete B.; Wiehlmann, Lutz; Davenport, Colin F.

Published in:
ERJ Open Research

DOI:
10.1183/23120541.01008-2023

Publication date:
2024

Licence:
CC BY-NC

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA):
Rosenboom, I., Thavarasa, A., Richardson, H., Long, M. B., Wiehlmann, L., Davenport, C. F., Shoemark, A., Chalmers, J. D., & Tümmler, B. (2024). Sputum metagenomics of people with bronchiectasis. *ERJ Open Research*, *10*(2). https://doi.org/10.1183/23120541.01008-2023

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

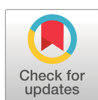
Download date: 04. Jul. 2024

Sputum metagenomics of people with bronchiectasis

Ilona Rosenboom [1], Ajith Thavarasa[1], Hollian Richardson[2], Merete B. Long[2], Lutz Wiehlmann[3], Colin F. Davenport[3], Amelia Shoemark [2], James D. Chalmers[2] and Burkhard Tümmler [1,4]

[1]Department of Paediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany. [2]Division of Molecular and Clinical Medicine, University of Dundee, Dundee, UK. [3]Research Core Unit Genomics, Hannover Medical School, Hannover, Germany. [4]Biomedical Research in Endstage and Obstructive Lung Disease (BREATH), German Centre for Lung Research, Hannover, Germany.

Corresponding author: Ilona Rosenboom (Rosenboom.ilona@mh-hannover.de)



Shareable abstract (@ERSpublications)
The sputum metagenome of people with bronchiectasis segregates into four clusters dominated by either pathogenic or commensal bacteria. Personalised communities depleted of commensals present in healthy airways are seen in patients with severe disease. https://bit.ly/3RXr9U7

Cite this article as: Rosenboom I, Thavarasa A, Richardson H, *et al*. Sputum metagenomics of people with bronchiectasis. *ERJ Open Res* 2024; 10: 01008-2023 [DOI: 10.1183/23120541.01008-2023].

Copyright ©The authors 2024

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 13 Dec 2023
Accepted: 8 Jan 2024

Abstract

Background The microbiota in the sputum of people with bronchiectasis has repeatedly been investigated in cohorts of different geographic origin, but so far has not been studied to the species level in comparison to control populations including healthy adults and smokers without lung disease.

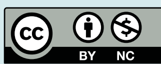
Methods The microbial metagenome from sputa of 101 European Bronchiectasis Registry (EMBARC) study participants was examined by using whole-genome shotgun sequencing.

Results Our analysis of the metagenome of people with bronchiectasis revealed four clusters characterised by a predominance of *Haemophilus influenzae*, *Pseudomonas aeruginosa* or polymicrobial communities with varying compositions of nonpathogenic commensals and opportunistic pathogens. The metagenomes of the severely affected patients showed individual profiles characterised by low alpha diversity. Importantly, nearly 50% of patients with severe disease were grouped in a cluster characterised by commensals. Comparisons with the sputum metagenomes of healthy smokers and healthy nonsmokers revealed a gradient of depletion of taxa in bronchiectasis, most often *Neisseria subflava*, *Fusobacterium periodonticum* and *Eubacterium sulci*.

Conclusion The gradient of depletion of commensal taxa found in healthy airways is a key feature of bronchiectasis associated with disease severity.

Introduction

Bronchiectasis (BE) is a common chronic lung disease with a wide range of underlying causes, including infections, genetic conditions, as well as autoimmunity and hypersensitivity disorders. Patients suffer from chronic bronchitic symptoms and frequent respiratory infections, and the disease is characterised by permanent abnormal dilatation of the bronchi [1, 2]. Airways in BE are predisposed to chronic infection with micro-organisms that have been linked to disease progression [2, 3]. Culture-based studies of sputum show a clear association between infection with *Pseudomonas aeruginosa* and poor outcomes, including increased mortality. However, culture is recognised as being limited in its ability to characterise the complexity of chronic lung infections. Using sputum as a surrogate for bronchial secretions and lung fluids, microbiota have been examined in more than a dozen studies through 16S rDNA amplicon sequencing [2, 4–6]. Consistent with culture-based diagnostics, the low-diversity microbiomes of people with BE were often found to be dominated by either *Haemophilus*, *Staphylococcus*, *Pseudomonas* or *Streptococcus*. Whole-genome shotgun sequencing provides an additional level of taxonomic information by resolving the microbiota to species or even strain levels [7]. Correspondingly, the airway metagenome of people with BE has been explored in the context of exacerbation and the gut–lung axis [8, 9]. Studies to date largely confirm the relationship between classical pathogens such as *P. aeruginosa* and *Haemophilus influenzae* and worse outcomes in BE, but further studies are needed to understand the contribution of less abundant taxa, which may play an important role in the BE microbiome. For example, a recent study



demonstrated that the commensal *Rothia mucilaginosa* had anti-inflammatory effects including inhibition of NF-κB activation in airway cells and its presence could modulate the harmful effects of *P. aeruginosa* [10]. These data suggest that the depletion of commensals found in the healthy airway metagenome may be a key factor in BE disease severity.

Here we report on the cross-sectional analysis of the sputum metagenome of 101 people with BE. After taxonomic profiling with our publicly available Wochenende pipeline [11], the metagenome datasets of people with BE segregated into four clusters of divergent bacterial composition that all are depleted of commensals from the healthy airway metagenome.

Materials and methods

Study participants

The European Bronchiectasis Registry (EMBARC) is a prospective longitudinal cohort study of patients with clinically significant BE confirmed by computed tomography (CT). The registry collects detailed clinical data and, in a subset of patients at selected sites, linked biological samples are collected. Data collection was approved by the research ethics committee (EMBARC 14/SS/1101). For this analysis, 101 patients enrolled in the EMBARC study provided a spontaneous sputum sample during clinical stability, defined as 4 weeks free from antibiotic and/or corticosteroid treatment and without symptoms of an exacerbation. Inclusion criteria also included high resolution CT-confirmed BE, clinical symptoms consistent with BE and the ability to give informed consent. For this analysis, all participants were enrolled from Ninewells Hospital, Dundee, UK. To compare the airway metagenome profiles in non-cystic fibrosis BE with lung health, control groups of healthy smokers (n=8) and healthy nonsmokers (n=88) were recruited [12]. Clinical metadata are available from the Supplementary Material (Tables S1, S2, S3). Induced sputum from healthy nonsmokers was collected in sterile tubes (Sarstedt #62.547.004 or Eppendorf #0030118405) according to the Cystic Fibrosis Foundation Therapeutics Development Network standard operating procedure 530.00 by autogenic drainage after up to four cycles of 3 min inhalation of 5.85% (w/v) hypertonic saline. The samples were immediately stored at −80 °C until further use.

Sputum sampling from BE patients and healthy smokers in Dundee was approved by the research ethics committee (approval numbers 16/NW/0101 and 17/LO/1961). The study regarding healthy nonsmokers was approved by the ethics committee of Hannover Medical School (No. 9299_BO_K_2020). Written informed consent was obtained from all participants. Sputum samples from BE and healthy controls were processed and sequenced at Hannover Medical School through the same wet-lab and bioinformatics pipelines.

DNA isolation, library preparation and sequencing

DNA was isolated as previously described [13] (see also extended methods in the Supplementary Material). The Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific, Karlsruhe, Germany; #Q32854) was applied to measure DNA concentration. 10–50 ng of DNA was sheared into 300 bp fragments using a Covaris S220 instrument [13]. Fragment libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Frankfurt am Main, Germany; #E7645L) and NEBNext Multiplex Oligos for Illumina (#E6440) in six to eight PCR cycles. Purification of fragment libraries was performed with 0.9× AMPure Beads (Beckman Coulter, Krefeld, Germany; #A63881). The Illumina NextSeq 550 platform was applied for single-ended 75 bp short-read sequencing (Illumina, Cambridge, UK; #20024906). Negative controls were processed and sequenced in parallel with the patient samples.

Bioinformatic processing and statistical analysis

Quality filtering, adapter trimming and alignment of the short reads to a reference database consisting of bacterial, fungal, viral and archaeal genomes (n=2800) was performed by the public pipeline Wochenende [11], which additionally normalised the read counts. Additional information and documentation regarding Wochenende is retrievable from GitHub (https://github.com/MHH-RCUG/nf_wochenende/wiki). The reference database used is available for download (https://drive.google.com/drive/folders/1q1btJCxtU15XXqfA-iCyNwgKgQq0SrG4). The microbial reads were normalised to an ideal genome length of one million base pairs and a sequencing depth of one million sequencing reads accounting for both variations in chromosome length and sequencing depth (reads per million base pairs values). The raspir tool was used to filter out false-positive species characterised by nonuniform read distributions across their reference genome [14]. Taxonomic profiling was also performed by MetaPhlAn3 [7] and KrakenUniq [15]. Additional alpha and beta diversity analyses were conducted to compare results from the three metagenome classifiers Wochenende, MetaPhlAn3 and KrakenUniq (Supplementary Figure S1).

Whole-genome shotgun sequencing and further processing with Wochenende generated a median number of 8.4 million high-quality sequencing reads of 75 bp length per BE sputum sample. The mean relative

percentage of microbial reads was 7%, corresponding to about 600 000 reads. Negative controls had a median number of 102 reads and species patterns were highly divergent from those of patient samples (Supplementary Figure S2). Coverage of nonhuman reads was assessed by the tool Nonpareil 3 (Supplementary Figure S3) [16].

Statistical analysis was conducted with R in RStudio version 2022.07.2. Diversity indices and nonmetric multidimensional scaling (NMDS) were calculated with the *vegan* package (R function distance, metaMDS). Comparisons of metagenome datasets between subcohorts of individuals were performed with MaAsLin2 (Microbiome Multivariable Association with Linear Models) applying the suggested default parameters [17]. A permutation test was employed to fit clinical metadata and microbial ecology parameters in the ordination (*vegan*, envfit). Group centroids were evaluated by the betadisper function (*vegan*). Dirichlet multinomial mixtures (DMMs) were applied for microbial community typing [18] (*DirichletMultinomial* package) with parameter π as a measure of weight and θ as a measure of variance. A Laplace approximation was applied to determine the number of components in the Dirichlet mixture. The R package metacoder was used to generate heat trees [19]. Statistical significance was evaluated using the Mann–Whitney or Kruskal–Wallis tests, with a Dunn test for *post hoc* analysis and Benjamini–Hochberg correction. Fisher's exact test was applied to detect significant associations between clinical metadata and disease severity. Please see the Supplementary Material for more details about the selected methods.

The metagenome datasets from people with BE and healthy smokers have been uploaded to the European Nucleotide Archive (Project PRJEB65368). The metagenome data sets from healthy nonsmokers can also be retrieved from the European Nucleotide Archive (project PRJEB52822). Coding scripts are available from GitHub (https://github.com/irosenboom/sputumMetagenomics_bronchiectasis).

Results

Characteristics of the study population

The study cohort consisted of 101 people with BE having a median age of 68 years (range 21–91 years) (table 1, Supplementary Table S1). The population had slightly more males than females (57%). 37 patients were ex-smokers and eight were current smokers. The most common aetiologies were idiopathic and post-infective disease with COPD being the third most common cause, consistent with the European Bronchiectasis Registry aetiologies as a whole [20]. 34 BE patients were taking long-term antibiotics, predominantly long-term azithromycin prophylaxis.

Airway metagenome analyses

Sputum samples from all study participants (n=101) underwent high-throughput shotgun microbial genome sequencing. First, the metagenome datasets were analysed by ecology parameters differentiated by the patient's bronchiectasis severity index (BSI) [21]. The group of mildly and moderately affected individuals had significantly higher diversity measures for species number and Shannon, Simpson and Pielou diversity indices than the severely affected patients (figure 1a,b,c,d). Next, we compared the differential abundance of taxa in patients with mild, moderate and severe BSI by MaAsLin 2 analysis [17]. Figure 1e demonstrates the moderate and strong depletion of commensals in the individuals with moderate and severe BSI, respectively, compared to the patient cohort with mild BSI. Beta diversity of the datasets was assessed by NMDS of Bray–Curtis dissimilarity indices (figure 2a). Sputa collected from patients with severe BSI were significantly more distant from the average sample of the dataset implying individual signatures (figure 2b). Community profiling by DMMs [18] segregated the datasets into four clusters (figure 2c). Cluster DMM1 (n=40) was populated by similar numbers of samples from patients with mild and severe BSI, whereas DMM2 (n=35) predominantly consisted of samples from individuals with mild and moderate BSI (figure 2d). Samples mainly from patients with moderate or severe BSI contributed to DMM3 (n=16) and DMM4 (n=10).

DMM3 and DMM4 consisted of *H. influenzae* and *P. aeruginosa* as primary species, respectively (figure 3). Even though all other species made only minor contributions, the alpha diversity of the metagenomes dominated by *H. influenzae* was significantly higher in samples from individuals with mild and moderate BSI than from those with severe BSI (Supplementary Figure S4). In contrast to the pathogen-driven and less diverse DMM3 and DMM4, clusters DMM1 and DMM2, where commensal species predominated, differed substantially in terms of their parameter θ values, serving as a measure of variance (figure 3a). Cluster DMM1 was characterised by personalised metagenomes as its major constituents, including *Moraxella catarrhalis*, *Staphylococcus aureus*, *H. influenzae* and members of the genus *Streptococcus* (figure 3c). In contrast, DMM2 was characterised by low variance and high diversity. Shannon diversity was significantly different between the four clusters in all pairwise comparisons, except for DMM3 and DMM4 (figure 3b).

TABLE 1 Clinical metadata of 101 individuals with bronchiectasis stratified by bronchiectasis severity index [21]

Cohort characteristics	Mild (n=26)	Moderate (n=40)	Severe (n=35)
Age, years (median 68, 63–75 IQR)			
<50*	6 (23%)	2 (5%)	2 (6%)
50–70*	17 (65%)	19 (48%)	11 (31%)
70–80**	3 (12%)	15 (38%)	17 (49%)
>80	0 (0%)	4 (10%)	5 (14%)
Gender			
Female	9 (35%)	21 (53%)	13 (37%)
Male	17 (65%)	19 (48%)	22 (63%)
BMI (median 26.5, 22.6–30.4 IQR)			
<18.5	0 (0%)	1 (3%)	3 (9%)
18.5–25*	8 (31%)	11 (28%)	20 (57%)
>25–30*	9 (35%)	16 (40%)	5 (14%)
>30	9 (35%)	12 (30%)	7 (20%)
Smoking history			
Ex-smoker	6 (23%)	15 (38%)	16 (46%)
Smoker	2 (8%)	4 (10%)	2 (6%)
Never-smoker	18 (69%)	21 (53%)	17 (49%)
FEV_1 % pred (median 67, 52–87 IQR)			
>80***	19 (73%)	12 (30%)	2 (6%)
50–80	7 (27%)	20 (50%)	16 (46%)
30–50***	0 (0%)	8 (20%)	13 (37%)
<30*	0 (0%)	0 (0%)	4 (11%)
Comorbidities			
Cancer	2 (8%)	2 (5%)	4 (11%)
Cardiovascular disease	8 (31%)	11 (28%)	12 (34%)
Chronic renal failure	0 (0%)	5 (13%)	3 (9%)
Depression*	1 (4%)	5 (13%)	10 (29%)
Diabetes*	0 (0%)	9 (23%)	5 (14%)
Osteoporosis*	1 (4%)	6 (15%)	10 (29%)
Aetiologies			
Idiopathic*	13 (50%)	23 (58%)	10 (29%)
Post-infective	6 (23%)	5 (13%)	2 (6%)
COPD	0 (0%)	5 (13%)	6 (17%)
ABPA	1 (4%)	1 (3%)	4 (11%)
Asthma	2 (8%)	1 (3%)	2 (6%)
Primary ciliary dyskinesia	2 (8%)	1 (3%)	2 (6%)
Inflammatory bowel disease	0 (0%)	1 (3%)	3 (9%)
Nontuberculous mycobacteria	1 (4%)	1 (3%)	1 (3%)
Rheumatoid arthritis	0 (0%)	0 (0%)	3 (9%)
Immunodeficiency	1 (4%)	1 (3%)	0 (0%)
Alpha-1 antitrypsin deficiency	0 (0%)	0 (0%)	1 (3%)
Aspiration	0 (0%)	1 (3%)	0 (0%)
CFTR-related disorder	0 (0%)	0 (0%)	1 (3%)
Treatment			
Inhaled corticosteroids	14 (54%)	21 (53%)	26 (74%)
Long-term antibiotics	7 (27%)	10 (25%)	17 (49%)

Fisher's exact test was applied to detect significant associations of clinical metadata with the severity of disease. ABPA: allergic bronchopulmonary aspergillosis; BMI: body mass index; CFTR: cystic fibrosis transmembrane conductance regulator; FEV_1: forced expiratory volume in 1 s; IQR: interquartile range. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Clusters DMM1 and DMM2 were predominantly populated by commensal species rather than by conventional pathogens. Hence, we wanted to compare the composition of these clusters with those derived from sputa of healthy nonsmokers and healthy smokers. The heat trees in figure 4 point to substantial differences in the airway metagenomes of healthy and BE patients, even in the absence of classical pathogens such as *H. influenzae* or *P. aeruginosa*. Of the 10 most common species in a healthy lung, the major anaerobes, *i.e. Veillonella* and *Prevotella* species, were strongly depleted. The heat trees additionally resolved that cluster DMM1 contained infrequently occurring, personalised pathogens such as *Serratia liquefaciens*, *Neisseria meningitidis* and *Elizabethkingia miricola*. To resolve the different

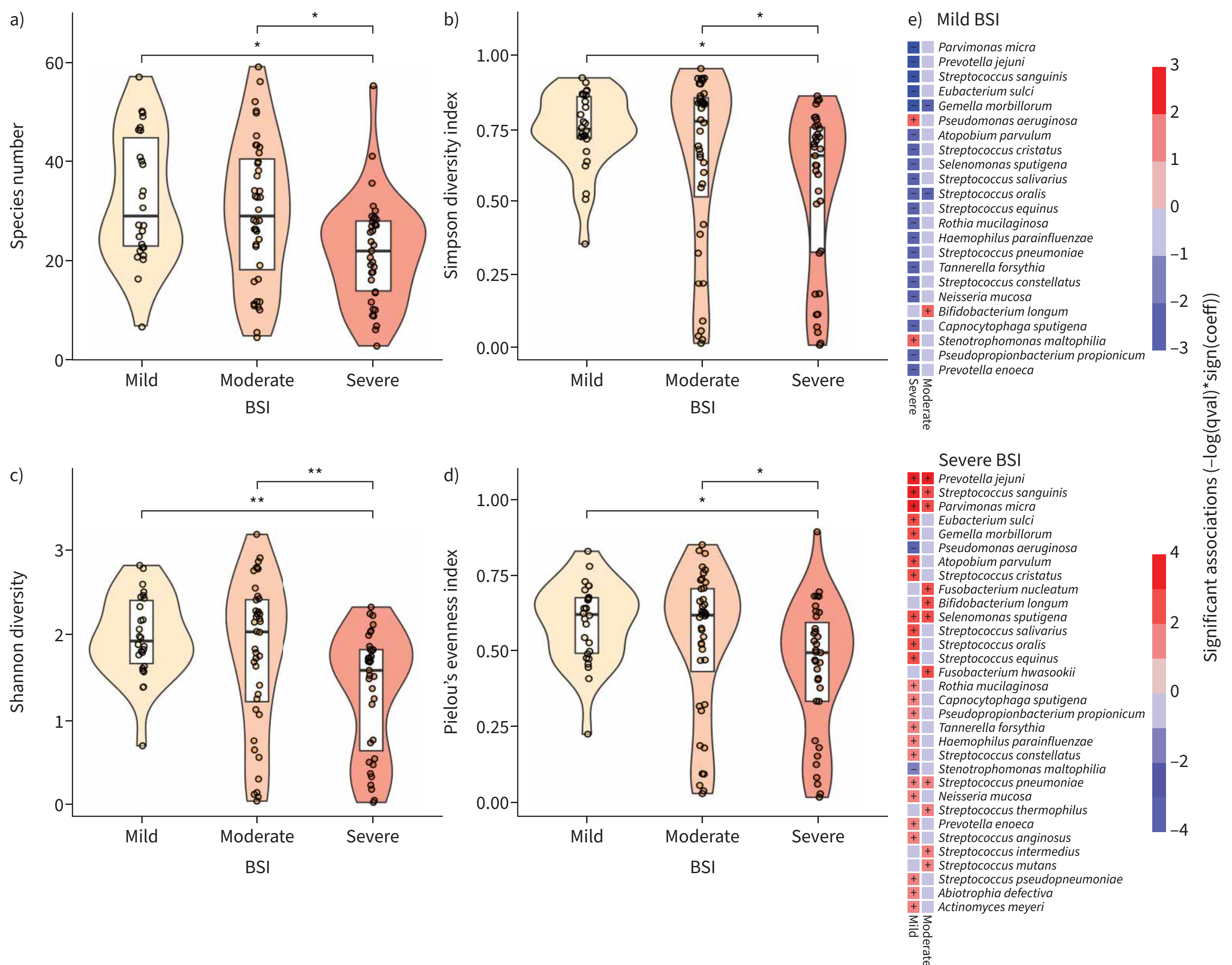


FIGURE 1 Alpha diversity measures of the bronchiectasis cohort stratified by bronchiectasis severity index (BSI). a) Species number was significantly reduced in severe bronchiectasis (Kruskal–Wallis p=0.005; η^2=0.09, 95% CI 0.01–0.24). b) The Simpson diversity index was significantly reduced in severe bronchiectasis (Kruskal–Wallis p=0.008; η^2=0.08, 95% CI 0.005–0.22). c) Shannon diversity was significantly reduced in severe bronchiectasis (Kruskal–Wallis p=0.002; η^2=0.10, 95% CI 0.02–0.25). (d) Pielou's evenness index was significantly reduced in severe bronchiectasis (Kruskal–Wallis p=0.02, η^2=0.06, 95% CI 0.003–0.2). e) Heatmap of the intergroup comparisons of mild, moderate and severe BSI by MaAsLin2 analysis [17]. The reference disease category is given as a header. Note: p-values are depicted with * p<0.05 and ** p<0.01. Boxplots are integrated into the violin plots depicting the median and interquartile range. The four microbial ecology measures reflect different aspects of community heterogeneity. Species number indicates richness of samples. The Simpson index is the measure of the degree of concentration when individuals are classified into types. Shannon's diversity index quantifies the uncertainty in predicting the species identity of an individual that is taken at random from the dataset. Pielou's evenness is Shannon diversity normalised by the observed richness.

abundances in more detail, figures 5a and 5b compare the quantitative abundances of randomly selected pairs of the four age quartiles in the four clusters with that of the average metagenome of healthy nonsmokers aged 20–59 years [12]. Each BE sample demonstrated a substantial and individual depletion of the taxa present in healthy airways, whereby more taxa were absent in specimens of DMM3 and DMM4 (figure 5b). Even though samples from cluster DMM2 were more diverse than those from cluster DMM1, they both were strongly depleted in terms of *Veillonella* spp., *Prevotella* spp., *N. subflava*, *Eubacterium sulci* and *Fusobacterium periodonticum* (figure 5a).

Figure 6a depicts the differential taxonomic profiles of our metagenome datasets from healthy and BE cohorts analysed by the MaAsLin2 [17] software package. As visualised before in the heatmaps (figure 5), numerous commensal species present in healthy lungs were significantly depleted in the sputa of

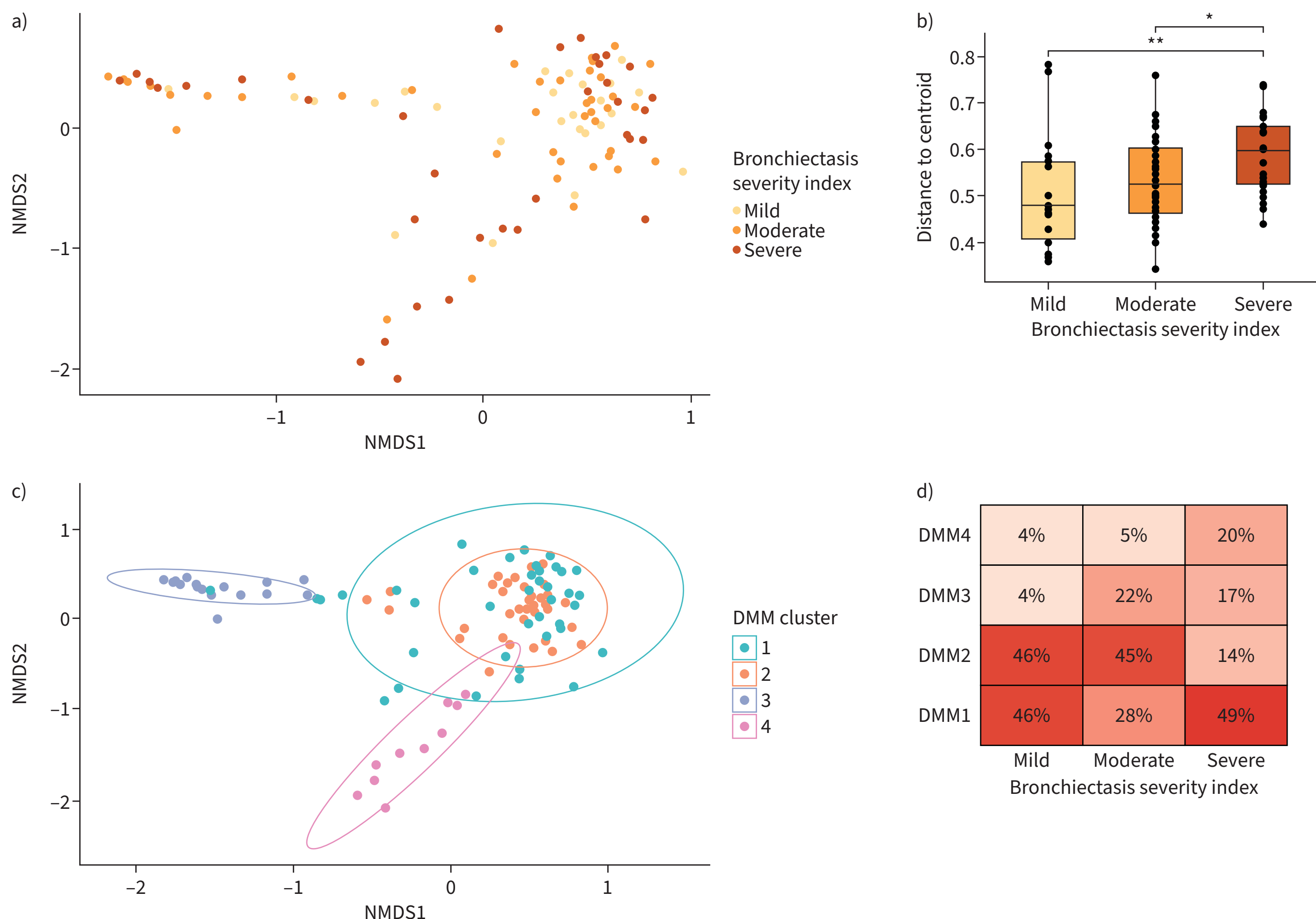


FIGURE 2 Beta diversity and community typing by Dirichlet multinomial mixtures (DMMs). a) Nonmetric multidimensional scaling (NMDS) was applied to Bray–Curtis dissimilarity indices of the bronchiectasis cohort (stress=0.15, k=3). The colour of data points represents the patients' bronchiectasis severity index. b) Multivariate homogeneity testing of group variances reported significant differences for severe bronchiectasis (Kruskal–Wallis $p<0.001$; $\eta^2=0.12$, 95% CI 0.03–0.28). c) DMMs were applied for community typing of the bronchiectasis cohort. A Laplace approximation indicated the presence of four clusters. Samples in the NMDS plot are coloured according to their predicted DMM community type. d) Distribution of disease severity among the four DMM clusters normalised by bronchiectasis severity index group showed a skew in community type distribution. Mild bronchiectasis was predominantly found among clusters DMM1 and DMM2. Moderate bronchiectasis was located in DMM2, whereas half of the patients with severe bronchiectasis was classified in cluster DMM1.

individuals with BE. Of the total 102 significantly differentially abundant taxa (Supplementary Figure S5), taxa of the genera *Fusobacterium*, *Eubacterium*, *Prevotella*, *Leptotrichia*, *Neisseria*, *Haemophilus*, *Actinomyces*, *Campylobacter* and *Veillonella* were most strongly depleted (figure 6a). Besides the pathogens well known to colonise BE lungs, such as *H. influenzae*, *P. aeruginosa* and *M. catarrhalis*, three *Streptococcus* species (*S. salivarius*, *S. equinus* and *S. mutans*) are present in BE lungs and less common in healthy lungs (Supplementary Figure S5, Supplementary Table S5). The NMDS plots in figures 6b and 6c compare the beta diversity of the total datasets of the metagenomes from people with BE, healthy smokers and healthy nonsmokers highlighting the individual signatures of BE samples and partly overlapping profiles of the healthy controls. In other words, samples from smokers are shifted in the second NMDS dimension because *F. periodonticum*, *N. subflava*, *H. parainfluenzae* and *R. mucilaginosa* are at least partially depleted (figure 6c, Supplementary Table S4).

Discussion

This is the first study to characterise the metagenome of sputum from patients with BE compared to control populations including healthy adults and smokers without lung disease. Our analysis of the metagenome of people with BE revealed four clusters characterised by predominances of *Haemophilus*,

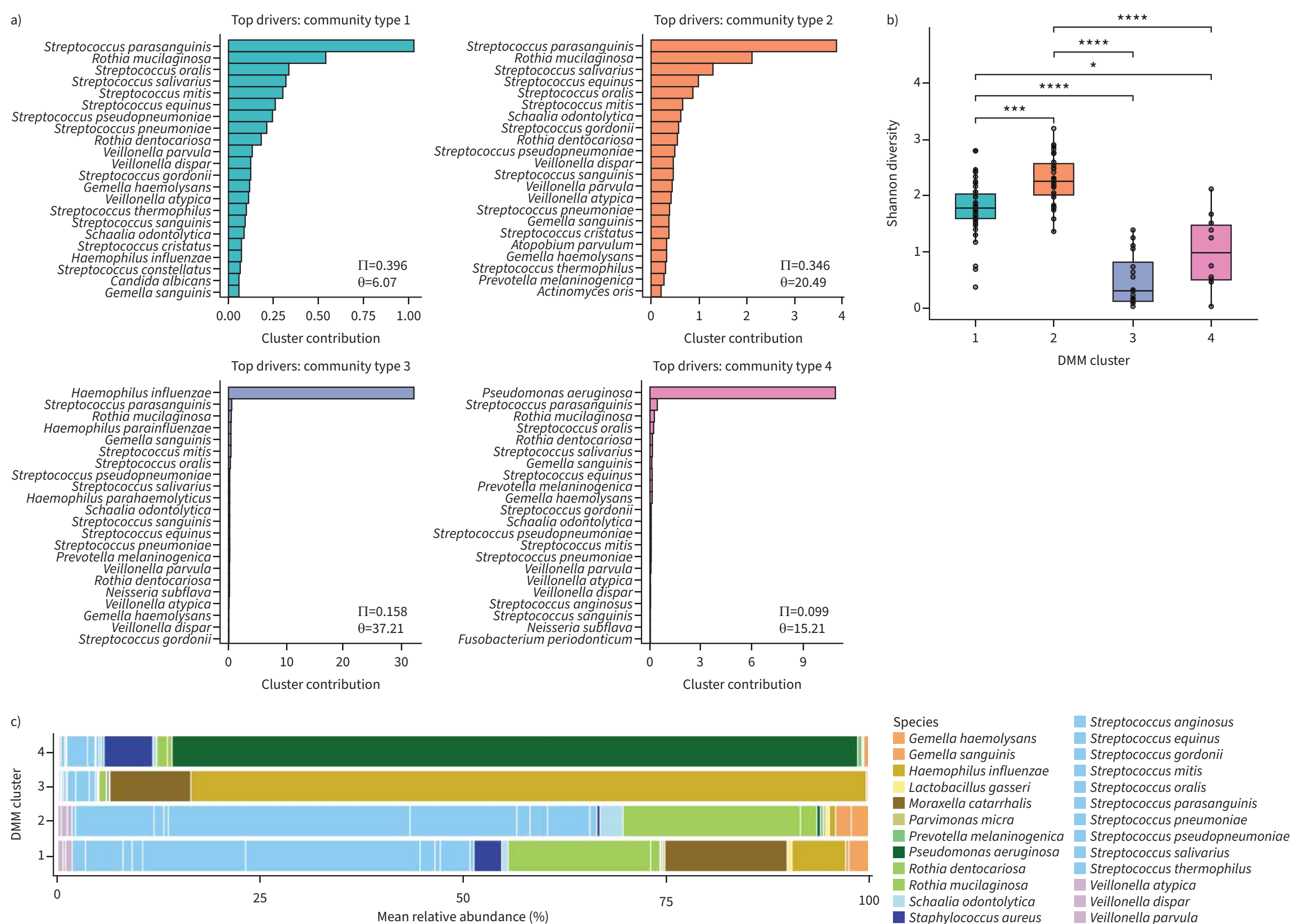


FIGURE 3 Microbial communities characterising the Dirichlet multinomial mixture (DMM) clusters. a) The DMMs were fitted on species RPMM values (reads per million base pairs). The *x*-axis displays the contribution of species to each component based on absolute Dirichlet component vector estimates. The most important drivers for DMM3 and DMM4 were *Haemophilus influenzae* and *Pseudomonas aeruginosa*, respectively. DMM1 and DMM2 were mainly driven by commensal species, whereby DMM1 additionally displayed *H. influenzae* and *Candida albicans* among the species most important for clustering. The four inserts give the π-values as a measure of mixture weight of the cluster and the θ-values as a measure of variability of the cluster. Low π-values indicate that the community is skewed to few dominant species. The lower the values for the variance θ, the higher the variability of the community. For example, clusters DMM1 and DMM2 share similar π-values but are distinct in their θ-values, indicating personalised metagenomes as major constituents of cluster DMM1. b) The distribution of Shannon diversity was significantly different among the clusters (Kruskal–Wallis $p<0.001$; $\eta^2=0.59$, 95% CI 0.45–0.71). c) Stacked barplot of mean relative abundance across the four DMM clusters. The colours represent taxonomic classification at genus level. Please note that *Staphylococcus aureus* and *Moraxella catarrhalis* were identified with high abundance in a minority of cluster DMM1 samples.

Pseudomonas or polymicrobial communities with a differential composition of nonpathogenic commensals and opportunistic pathogens. The metagenomes of the severely affected patients showed individual profiles of low alpha diversity. Importantly, 49% of patients with severe disease were in cluster DMM1, which was largely populated by commensals. Our study demonstrates that far from representing "healthy" metagenomes, these profiles are depleted of key commensals when compared to healthy airways. Comparisons with the sputum metagenomes of healthy smokers and healthy nonsmokers revealed a gradient of depletion of taxa in BE.

Our study confirms previous analyses by culture-dependent diagnostics and amplicon sequencing of the 16S rDNA gene that *P. aeruginosa*, *H. influenzae*, *S. aureus* and *Candida albicans* are the major pathogens in the sputum of people with BE [4, 6, 22–26]. Consistent with our beta diversity analysis (figure 2), microbiota in BE sputum can be assigned to three groups, the first dominated by *P. aeruginosa*, the second by *H. influenzae* and the third negatively defined as non-*P. aeruginosa*/non-*H. influenzae* dominated.

https://doi.org/10.1183/23120541.01008-2023

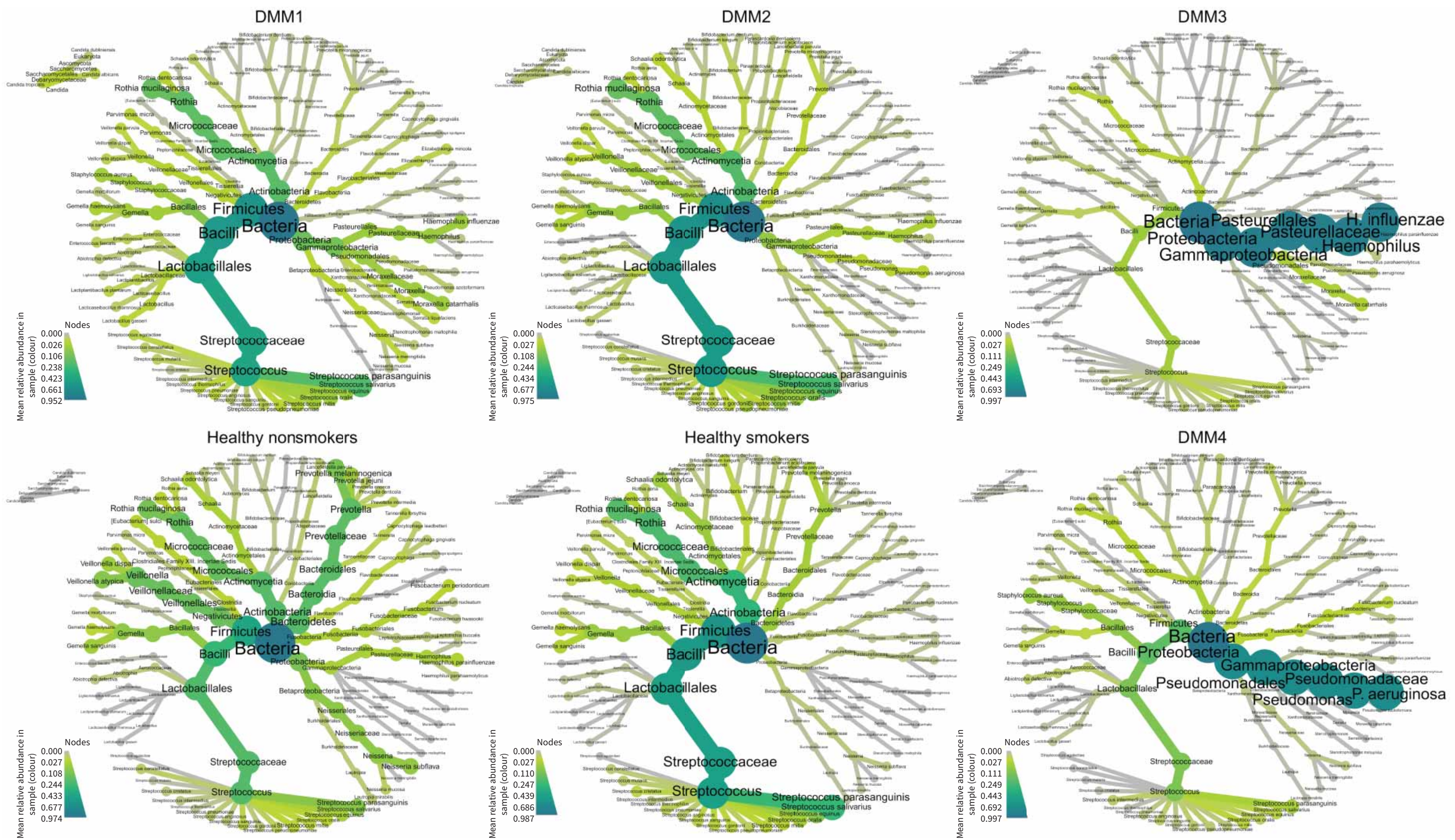


FIGURE 4 Taxonomic representation of the airway metagenome in a heat tree format. The bronchiectasis cohort is separated into four Dirichlet multinomial mixture (DMM) clusters (DMM1 n=40, DMM2 n=35, DMM3 n=16, DMM4 n=10). For comparison, the airway metagenomes of healthy nonsmokers (n=88, Hannover, Germany) and healthy smokers (n=8, Dundee, UK) are shown. Heat trees were generated by the metacoder package [19] included in the Wochenende pipeline and represent mean relative abundances per taxonomic rank from phylum to species calculated from RPMM values (reads per million base pairs) (node colour). Node size is comparable between all groups, whereas the grey background illustrates all species present in the study.

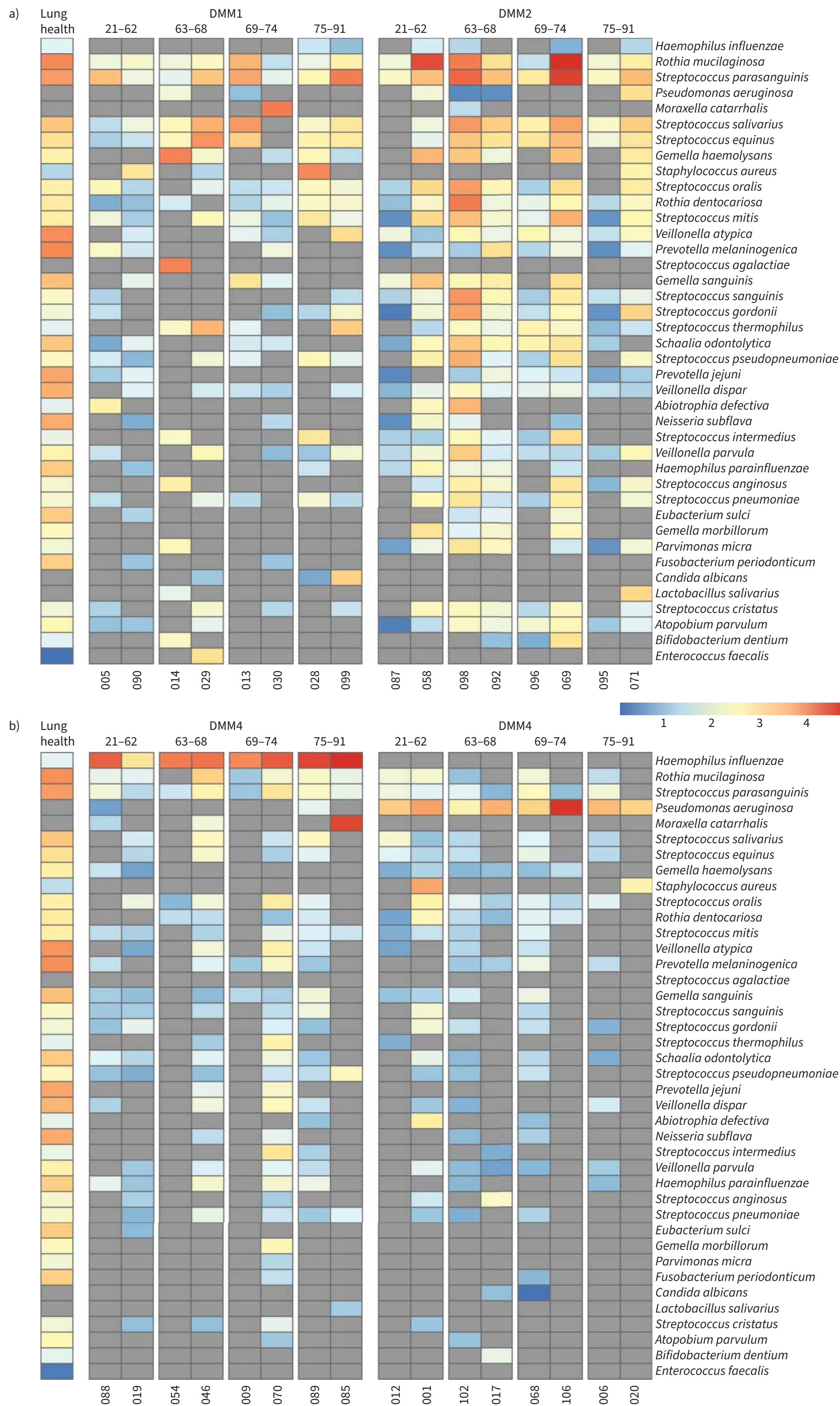


FIGURE 5 Heatmap of the top 40 microbial species in bronchiectasis and a cumulative metagenome from healthy nonsmokers. Two patient samples were randomly selected per age quartile (21–62, 63–68, 69–74 and

75–91 years) and Dirichlet multinomial mixture (DMM) cluster. The species were compared with a cumulative metagenome of 88 induced sputa from healthy nonsmokers aged 20–59 years. A grey square indicates an absent species. Colour indicates logarithmic RPMM values (reads per million base pairs) from low (blue) to high abundance (red). a) The first panel depicts the most variable cluster DMM1 and the more diverse cluster DMM2. b) The second panel illustrates the less diverse pathogen-driven clusters DMM3 and DMM4.

The composition of the third group varies between the published reports, which probably does not reflect the spatiotemporal divergence of BE microbiota around the globe but more the differential range of taxa that are resolved by the respective diagnostic microbiology laboratory and the inherent limitations of 16S rDNA amplicon analysis. The 16S rDNA sequence has its merits as being the universal molecular marker of the bacterial tree of life [27], but the amplicon sequencing of a variable region often falls short to discern bacteria at the species level. This limitation is particularly relevant for the diagnosis of the microbiota in the respiratory tract that consist of many Firmicutes such as *Streptococcus* species that are not differentiated by sequencing of a variable region of the 16S rDNA gene.

Shotgun sequencing of the metagenome overcomes these limitations. Consistent with two previous reports on the metagenome of 55 sputa from people with BE [8, 9], our metagenome datasets segregated into clusters characterised by a predominance of *Pseudomonas*, *Streptococcus* and *Haemophilus*. Thanks to algorithms that minimise false-positive and false-negative taxonomic assignments of reads [14], we could identify all bacterial species in the samples represented by at least 100 reads. Thus, the non-*P. aeruginosa*/non-*H. influenzae* group could be partitioned into two clusters comprising a divergent set of commensals and opportunistic pathogens. In future longitudinal studies, it will be interesting to assess whether a patient's assignment to a cluster will be stable or could switch over time.

Since this study implemented the sputum metagenome of healthy smokers and healthy nonsmokers, which has not been considered in all previous studies on the BE sputum microbiome, the dysbiosis of the BE airway metagenome in the absence of typical pathogens could be clarified. The members of the genera *Rothia* and *Streptococcus* can thrive in the BE lung habitat, whereas numerous other species of the healthy airway microbiome are partly or completely depleted. The strongest depletion was seen for the phylogenetically unrelated species *F. periodonticum*, *E. sulci* and *N. subflava*. We have observed the same pattern of depletion in sputa collected from individuals with cystic fibrosis [12]. *F. periodonticum* and *E. sulci* have so far not been investigated in the context of respiratory health and disease. However, low levels of *N. subflava* have been identified as a biomarker of COPD disease severity in Chinese patients [28]. Interestingly, based on murine infection models *N. subflava* has been qualified as a pathobiont in Southeast Asians with BE [29]. The differential abundance of *Neisseria* species and the association with BE disease in European and Asian patient populations may reflect a varying impact of ethnicity, lifestyle and polymicrobial interactome in the airways. In addition, the different spectrum of aetiologies already seen as a gradient among European populations [20] may contribute to the opposite roles attributed to *N. subflava*.

We have tested the association of the alpha diversity measures of the patient's metagenome with several clinical outcome measures, such as lung function, anthropometry, treatment and frequency of pulmonary exacerbations (see table 1), which all turned out to be nonsignificant with the exception of the BSI [21]. In other words, the high heterogeneity of aetiologies, in accordance with BE being a heterogeneous disease and especially the low frequencies of certain aetiologies, did not allow for multivariate statistical analysis taking aetiology as an outcome variable of metagenome data. The BSI, which is based on the evaluation of data from a prospective BE cohort [21], was superior to any other standard measure to link the airway microbiome with disease.

Our study has limitations. We did not perform any genome-based functional analysis. Any comparative metatranscriptomics, metabolomics and metaproteomics study would be beyond the scope of this work because of the technical challenges involved in handling the low microbial biomass samples. Moreover, we did not assess the geographic diversity of the BE sputum metagenomes, which has already been addressed by one of us by comparing samples from Scotland and the Southeast Asian Malacca peninsula [8, 29]. In addition, the sputa from patients and smokers were collected from visitors to a Dundee clinic and those from healthy nonsmokers were retrieved from people living in the Hannover region. Third, induced sputa were collected from the healthy controls because most healthy people cannot spontaneously mobilise respiratory secretions and we wanted a consistent sampling method for all controls. Conversely, spontaneous sputum was sampled from the people with BE. Thus, the same mode of sampling and storage was not guaranteed. Moreover, the healthy control group was on average younger than the people with BE.

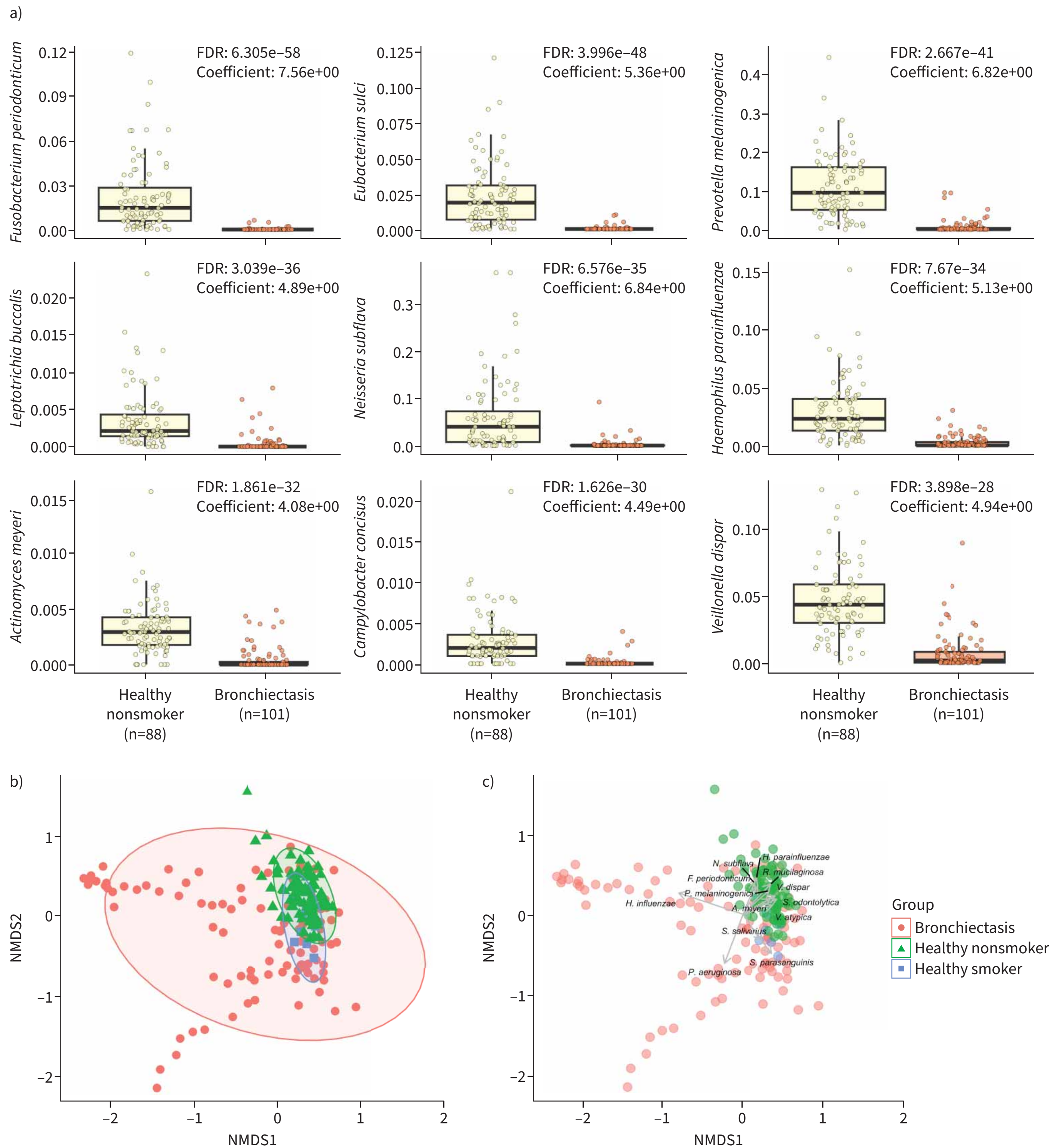


FIGURE 6 Characterisation of common and differential members in the bronchiectasis (BE) cohort, healthy nonsmokers and healthy smokers. a) Representative taxa of the nine genera that are most strongly depleted in BE sputa compared to healthy nonsmoker controls (MaAsLin2 analysis [17]). The coefficient value (effect size) indicates the contrast between health and BE sputa. b) Nonmetric multidimensional scaling (NMDS) of Bray–Curtis dissimilarity indices to assess beta diversity among the BE cohort, healthy nonsmokers and healthy smokers (stress=0.14, k=3). c) Key species that drive the NMDS. The vectors indicate the fit of these major species that significantly contribute to the clustering of samples (p=0.001, R>0.15, envfit analysis).

However, as the gradient of depletion noted in patients with BE matches with that observed in patients with cystic fibrosis seen at the Hannover clinic [12], the divergent geographic origin of samples from Scotland and Northern Germany should not have substantially influenced the outcome of our study.

In summary, the sputum metagenome of our study cohort could be segregated into four clusters. Apart from the samples dominated by singular pathogens such as *H. influenzae*, *P. aeruginosa* or *S. aureus*, many specimens contained no *prima facie* pathogens. The polymicrobial communities were mainly an assembly of *Rothia* and *Streptococci* species, but other common members of a healthy airway microbiome were underrepresented or even lacking. Even in the absence of classical pathogens, all samples from individuals with BE showed this mode of dysbiosis. These species that differentiate BE from lung health could be exploited in the future as biomarkers to monitor the longitudinal course of a patient, to trace the efficacy of disease management or even to function as an end-point for clinical trials.

Provenance: Submitted article, peer reviewed.

Acknowledgments: The authors are indebted to Marie-Madlen Pust (Department of Paediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany) for her intellectual input at the beginning of the study. We thank Marie Dorda (RCUG, Hannover Medical School, Hannover, Germany) for her excellent technical support.

Author contributions: L. Wiehlmann and B. Tümmler acquired funding. I. Rosenboom, L. Wiehlmann, J.D. Chalmers and B. Tümmler conceived the study and developed the experimental design. H. Richardson, M.B. Long, A. Shoemark and J.D. Chalmers recruited patients and provided clinical samples. C. Davenport and L. Wiehlmann provided technical support and wet-lab and *in silico* resources. I. Rosenboom and A. Thavarasa processed samples. I. Rosenboom performed DNA sequencing and data analysis. I. Rosenboom, J.D. Chalmers and B. Tümmler drafted the manuscript. All authors critically revised the manuscript and approved its final version.

Conflict of interest: A. Shoemark reports consulting fees from Spirovant and Translate Bio. J.D. Chalmers reports grants or contracts from AstraZeneca, Boehringer Ingelheim, Insmed, Gilead Sciences, Novartis, Genentech, GlaxoSmithKline and Trudell; fees from AstraZeneca, Boehringer Ingelheim, Insmed, Gilead Sciences, Novartis, Genentech, GlaxoSmithKline, Trudell, Zambon, CSL Behring, Janssen and Antabio; and is an associate editor of this journal. B. Tümmler reports grants from Vertex Pharmaceuticals. All other authors have nothing to disclose.

Support statement: This work was supported by grants from the Volkswagenstiftung and Niedersächsisches Ministerium für Wissenschaft und Kultur (Big Data initiative, ZN3432) to B. Tümmler and L. Wiehlmann, and Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC 2155 "RESIST" to B. Tümmler. Funding information for this article has been deposited with the Crossref Funder Registry.

Ethics statement: Sputum sampling from bronchiectasis patients and healthy smokers in Dundee was approved by the research ethics committee (approval numbers 16/NW/0101 and 17/LO/1961). The study regarding healthy nonsmokers was approved by the ethics committee of Hannover Medical School (MHH, Nr.9299_BO_K_2020). Written informed consent was obtained from all participants.

References

1 Polverino E, Goeminne PC, McDonnell MJ, *et al.* European Respiratory Society guidelines for the management of adult bronchiectasis. *Eur Respir J* 2017; 50: 1700629.
2 Richardson H, Dicker AJ, Barclay H, *et al.* The microbiome in bronchiectasis. *Eur Respir Rev* 2019; 28: 190048.
3 Chalmers JD, Chang AB, Chotirmall SH, *et al.* Bronchiectasis. *Nat Rev Dis Primers* 2018; 4: 45.
4 Woo TE, Lim R, Heirali AA, *et al.* A longitudinal characterization of the non-cystic fibrosis bronchiectasis airway microbiome. *Sci Rep* 2019; 9: 6871.
5 Tiew PY, Jaggi TK, Chan LLY, *et al.* The airway microbiome in COPD, bronchiectasis and bronchiectasis–COPD overlap. *Clin Respir J* 2021; 15: 123–133.
6 Dicker AJ, Lonergan M, Keir HR, *et al.* The sputum microbiome and clinical outcomes in patients with bronchiectasis: a prospective observational study. *Lancet Respir Med* 2021; 9: 885–896.
7 Beghini F, McIver LJ, Blanco-Míguez A, *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 2021; 10: e65088.
8 Mac Aogáin M, Narayana JK, Tiew PY, *et al.* Integrative microbiomics in bronchiectasis exacerbations. *Nat Med* 2021; 27: 688–699.
9 Narayana JK, Aliberti S, Mac Aogáin M, *et al.* Microbial dysregulation of the gut–lung axis in bronchiectasis. *Am J Respir Crit Care Med* 2023; 207: 908–920.

10 Rigauts C, Aizawa J, Taylor SL, *et al. Rothia mucilaginosa* is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. *Eur Respir J* 2022; 59: 2101293.
11 Rosenboom I, Scheithauer T, Friedrich FC, *et al.* Wochenende – modular and flexible alignment-based shotgun metagenome analysis. *BMC Genomics* 2022; 23: 748.
12 Pienkowska K, Pust M-M, Gessner M, *et al.* The cystic fibrosis upper and lower airway metagenome. *Microbiol Spectr* 2023; 11: e0363322.
13 Pallenberg ST, Pust M-M, Rosenboom I, *et al.* Impact of elexacaftor/tezacaftor/ivacaftor therapy on the cystic fibrosis airway microbial metagenome. *Microbiol Spectr* 2022; 10: e0145422.
14 Pust M-M, Tümmler B. Identification of core and rare species in metagenome samples based on shotgun metagenomic sequencing, Fourier transforms and spectral comparisons. *ISME Commun* 2021; 1: 2.
15 Breitwieser FP, Baker DN, Salzberg SL. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol* 2018; 19: 198.
16 Rodriguez-R LM, Gunturu S, Tiedje JM, *et al.* Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 2018; 3: e00039-18.
17 Mallick H, Rahnavard A, McIver LJ, *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 2021; 17: e1009442.
18 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012; 7: e30126.
19 Foster ZSL, Sharpton TJ, Grünwald NJ. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput Biol* 2017; 13: e1005404.
20 Chalmers JD, Polverino E, Crichton ML, *et al.* Bronchiectasis in Europe: data on disease characteristics from the European Bronchiectasis Registry (EMBARC). *Lancet Respir Med* 2023; 11: 637–649.
21 Chalmers JD, Goeminne P, Aliberti S, *et al.* The bronchiectasis severity index. An international derivation and validation study. *Am J Respir Crit Care Med* 2014; 189: 576–585.
22 Cuthbertson L, Felton I, James P, *et al.* The fungal airway microbiome in cystic fibrosis and non-cystic fibrosis bronchiectasis. *J Cyst Fibros* 2021; 20: 295–302–.
23 Metersky ML, Aksamit TR, Barker A, *et al.* The prevalence and significance of *Staphylococcus aureus* in patients with non-cystic fibrosis bronchiectasis. *Ann Am Thorac Soc* 2018; 15: 365–370.
24 Purcell P, Jary H, Perry A, *et al.* Polymicrobial airway bacterial communities in adult bronchiectasis patients. *BMC Microbiol* 2014; 14: 130.
25 Rogers GB, Zain NMM, Bruce KD, *et al.* A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc* 2014; 11: 496–503.
26 Woo TE, Lim R, Surette MG, *et al.* Epidemiology and natural history of airway infections in non-cystic fibrosis bronchiectasis. *ERJ Open Res* 2018; 4: 00162-2017.
27 Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; 87: 4576–4579.
28 Diao W, Shen N, Du Y, *et al.* Symptom-related sputum microbiota in stable chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 2289–2299.
29 Li L, Mac Aogáin M, Xu T, *et al. Neisseria* species as pathobionts in bronchiectasis. *Cell Host Microbe* 2022; 30: 1311–1327.e8.