

RESEARCH

Open Access



The sputum microbiome is distinct between COPD and health, independent of smoking history

Koironi Haldar^{1†}, Leena George^{1†}, Zhang Wang², Vijay Mistry¹, Mohammadali Yavari Ramsheh¹, Robert C. Free¹, Catherine John³, Nicola F. Reeve³, Bruce E. Miller⁴, Ruth Tal-Singer⁴, Adam J. Webb⁵, Anthony J. Brookes⁵, Martin D. Tobin³, Dave Singh⁶, Gavin C. Donaldson⁷, Jadwiga A. Wedzicha⁷, James R. Brown⁸, Michael R. Barer¹ and Christopher E. Brightling^{1,9*} 

Abstract

Background: Airway bacterial dysbiosis is a feature of chronic obstructive pulmonary disease (COPD). However, there is limited comparative data of the lung microbiome between healthy smokers, non-smokers and COPD.

Methods: We compared the 16S rRNA gene-based sputum microbiome generated from pair-ended Illumina sequencing of 124 healthy subjects (28 smokers and 96 non-smokers with normal lung function), with single stable samples from 218 COPD subjects collected from three UK clinical centres as part of the COPDMAP consortium.

Results: In healthy subjects *Firmicutes*, *Bacteroidetes* and *Actinobacteria* were the major phyla constituting 88% of the total reads, and *Streptococcus*, *Veillonella*, *Prevotella*, *Actinomyces* and *Rothia* were the dominant genera. *Haemophilus* formed only 3% of the healthy microbiome. In contrast, *Proteobacteria* was the most dominant phylum accounting for 50% of the microbiome in COPD subjects, with *Haemophilus* and *Moraxella* at genus level contributing 25 and 3% respectively. There were no differences in the microbiome profile within healthy and COPD subgroups when stratified based on smoking history. Principal coordinate analysis on operational taxonomic units showed two distinct clusters, representative of healthy and COPD subjects (PERMANOVA, $p = 0.001$).

Conclusion: The healthy and COPD sputum microbiomes are distinct and independent of smoking history. Our results underline the important role for *Gammaproteobacteria* in COPD.

Keywords: COPD, Healthy airway, Microbiome, *Haemophilus*, *Proteobacteria*

Background

Chronic obstructive pulmonary disease (COPD) is characterised by inflammation and irreversible airflow obstruction. Before the advent of culture-independent DNA profiling methods, the healthy lung was deemed a

sterile niche while COPD samples would frequently culture *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Moraxella catarrhalis* and *Pseudomonas aeruginosa*. It is considered that pathogenic bacteria gain a foothold in damaged airways contributing to further lung pathology through release of noxious bacterial products and provocation of host inflammation [1, 2].

In recent years, use of high-throughput 16S rRNA gene based sequencing has demonstrated that rich, complex bacterial communities exist in the airways of both health and COPD, with overlapping bacterial

* Correspondence: ceb17@le.ac.uk

[†]Koironi Haldar and Leena George contributed equally to this work.

¹Institute for Lung Health, NIHR, BRC, Department of Respiratory Sciences, College of Life Sciences, University of Leicester, Leicester LE1 7RH, UK

⁹Institute for Lung Health, University Hospitals of Leicester, Groby Road, Leicester LE3 9QP, UK

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

composition observed [3, 4]. In COPD aerobic, facultative and anaerobic organisms colonise the airways [3, 5], with *Proteobacteria* and *Firmicutes* being the two major phyla reported in the microbiome and *Haemophilus* and *Streptococcus*, the respective dominant genera [3, 4, 6, 7]. At exacerbation, shifts in bacterial composition, characterised by a relative increase in *Proteobacteria* that falls in response to antibiotics has been observed, suggesting an association with the aetiology of COPD exacerbations [8, 9]. Furthermore, the ratio of *Gammaproteobacteria* to *Firmicutes* identifies the subgroup with dynamic changes in their microbiome during exacerbation, suggesting a potential use of this ratio as a biomarker for targeting antimicrobial treatment [10].

While *Proteobacteria* have been associated with COPD exacerbation events, the role of the microbiome in the stable state and important differences in composition with health are unclear [3, 4, 6]. Furthermore, changes in the microbiome that may associate with development of COPD in smokers are unclear. Differences have been shown in the nasal and oropharyngeal microbiome between smokers and non-smokers [11] but microbiome data from healthy smokers and non-smokers is limited making the findings inconclusive to contextualize the pathological basis of the observations in COPD.

To address this, we have used sputum collected from a substantial number of well characterised healthy volunteers to investigate the impact of smoking on healthy lower airway microbiome; to explore if there are differences in microbiome between health and COPD and its association with smoking in health.

Methods

Subjects and study samples

Healthy volunteers ($n = 251$) (excluded participants with asthma, COPD or bronchiectasis) were selected from the Extended Cohort for E-health, Environment and DNA (EXCEED) cohort and assessed at a single centre, Glenfield Hospital, Leicester. Participants with ≥ 10 pack year (PY) smoking were grouped as healthy smokers and remaining as occasional / never smokers (< 10 PY) [12]. Participant demographics and clinical characteristics including lung function, blood and sputum cell differentials were recorded. The comparator COPD group comprised of a single stable (non-exacerbation) visit sputum collected from 218 subjects (included ex- and current smokers) at three centres, Leicester, London and Manchester, as part of the COPDMap consortium (www.copdmap.org; NCT01620645) [13]. All the participants in both studies had a minimum of 6 weeks antibiotic free period before their sample collection. Sputum induction was undertaken if an adequate spontaneous sample was not produced. In most of the healthy participants induced sputum was collected. Both cohort studies

had ethical approval and all subjects gave written informed consent before the performance of any study-related assessments.

All patients provided written informed consent using protocols approved by the local Ethics Committees at each site (London- 11/L0/1630; Manchester- 10/H/1003/108; Leicester- 07/H0406/157).

16S rRNA gene sequencing

As a part of standard routine, for both studies, sputum plugs were separated from the salivary contents to minimise the oral bacterial contamination in samples. Samples with only salivary contents were not processed for microbial work. Sputum plugs, were stored (-80°C) and processed for high-throughput sequencing similar to the COPD cohort samples [13]. Briefly, bacterial genomic DNA was extracted from the homogenised (0.1% dithiothreitol) plugs using the lysozyme-based lysis procedure from Qiagen DNA Mini kit (Qiagen, CA, USA) as per manufacturer's protocol. Out of 251 samples only 137 had adequate DNA concentration for microbiome analysis. Amplicon library was generated utilizing 28 PCR cycles and targeting the V4 hypervariable region of the 16S rRNA gene with 515F: 5' GTGCCAGCMGCCGCGGTAA3', 806R: 5'GGACTACHVGGGTWCTAAT3' primers, including Illumina sequencing adapters and a 12 bp Golay barcode sequence attached to forward primer. Pair-ended sequencing was performed using multiplex libraries on the Illumina MiSeq platform. Sequencing run included a commercial mock community DNA (ZymoBIOMICS microbial DNA standard) as a positive control and DNA extraction negative control (each batch of healthy sample DNA extraction included a DNA extraction negative control and a single pooled aliquot was prepared from all of these controls for sequencing) and PCR negative control for reagent contamination check. PCR negative controls didn't produce any reads and DNA extraction negative control only produced 8 raw reads and therefore were not processed further.

COPDMap single stable samples, utilized here for comparative analysis with healthy samples, had their sequence data generated as part of that study along with appropriate sequencing controls [13] and the sequence data are deposited at the National Centre for Biotechnology Information Sequence Read Archive (SRP102480).

Microbiome analysis

Reads were processed using QIIME pipeline version 1.9.1 [14] after adaptor trimming and removing low quality reads with Trimmomatic 0.36 [14, 15]. Pair-ended sequences were joined using fastq-join with a minimum 10 bp overlap [16]. The joined sequences were filtered with a Phred score ≥ 20 and processed to remove

contaminating sequences and chimeras using UCHIME [17]. Sequence reads are deposited at the National Centre for Biotechnology Information (SRA accession: PRJNA491861) <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA491861/>.

Based on rarefaction curves, healthy and COPD MAP samples, were normalised to 11,000 reads sequencing depth leaving 124 samples in healthy and 218 samples in COPD MAP for microbiome analysis. Operational taxonomic units were generated at 97% sequence identity using close reference (OTU) method and Greengenes database (version 13_8) and assigned taxonomic identities with the RDP classifier [18]. Alpha diversity index was generated based on the number of OTUs (observed_OTUs and Chao1-richness measure) and their distribution (Shannon index- diversity measure) within a sample. Beta diversity index was based on weighted UniFrac distance measure [19] (phylogenetic distance-based) between OTUs present in each sample and visualized using PCoA plots.

Microbial function prediction through 16S rRNA gene sequences

Predictive microbial functional profile was generated using the PICRUSt software (v1.0.0) which infers the pathway content of the microbiome by assigning bacterial functional genes for the OTUs, normalised for 16S rRNA gene copy number, using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database [20]. Statistical analysis and visualisation plots were generated using Statistical Analysis of Metagenomic Profiles (STAMP) software [21].

Statistical analysis

Univariate statistical analyses were performed using GraphPad Prism (Version 7, San Diego, CA). Parametric and nonparametric data are presented as mean (SEM) and median (interquartile range) respectively. Most of the dataset did not meet the normal distribution criteria (Shapiro-wilk test). Therefore, non-parametric Mann-Whitney test and Kruskal-Wallis (KW) test were performed for between-group comparison of two or greater than two groups respectively. For the KW test, Dunn's multiple comparison test was used to determine statistical significance of pair-wise comparisons. For parametric data, equivalent parametric statistical tests were performed. For categorical data, Chi-square test was performed.

PERMANOVA [22] was performed to test if the overall healthy microbiome was different to COPD using weighted UniFrac distance measure as input. The individual OTUs contributing to differences in health and COPD were identified by performing a non-parametric group-significance test and a corrected p -value of ≤ 0.05

for multiple testing conditions was considered statistically significant. To rank the discriminating taxonomic groups between health and COPD, linear discriminant analysis (LDA) effect size (LEfSe) analysis was performed [23], which detects taxonomies with differential abundance ($p \leq 0.05$) using the KW test and then assigns a LDA based effect size score. A threshold of ≥ 3.6 LDA score was used instead of default value of 2 to feature the most discriminant bacterial groups [24]. To determine if any individual microbiome constituents were associated with any subject metadata, MaAsLin was performed [25] and Benjamini-Hochberg (FDR) corrected $q < 0.05$ (FDR-adjusted P value) for multiple comparisons was considered statistically significant.

Results

Table 1 summarizes the demographics and clinical features of the healthy and COPD cohorts (refer to 'Methods' section for more details on both the cohort selection criteria).

Healthy sputum microbiome

A total of 1424 OTUs at 97% sequence identity were observed in 124 healthy samples after rarefaction. Most OTUs belonged to *Firmicutes* (55% \pm 13%) followed by *Bacteroidetes* (21% \pm 11%) and *Actinobacteria* (12% \pm 6%) phyla (Fig. 1a). *Streptococcus* (30% \pm 13%) was the most abundant genus followed by *Veillonella* (17% \pm 9%), *Prevotella* (16% \pm 10%), *Actinomyces* (6% \pm 5%), *Rothia* (5% \pm 4%) and *Granulicatella* (3% \pm 3%) (Fig. 1b). The *Proteobacteria* phylum constituted 7% (\pm 7%) of the bacterial community with *Haemophilus* (3% \pm 5%) as its dominant genus.

Compared to participants with a smoking history of < 10 pack year (PY), the subgroup with ≥ 10 PY showed a higher proportion of *Firmicutes* and a lower proportion of *Bacteroidetes* at phylum level (Fig. 1a); followed by higher *Streptococcus* and lower *Prevotella* at the genus level but these differences did not reach statistical significance (Fig. 1b). Beta diversity based principal coordinate analysis (PCoA) plots did not reveal distinct microbiome clusters for the two smoking subgroups (Fig. 1c). Shannon index was higher ($p < 0.01$) in healthy subjects with < 10 PY smoking history compared with ≥ 10 PY, suggesting a more diverse microbiome in the former (Fig. 1d) but there was no significant difference in observed_OTUs or Chao1 index.

COPD MAP stable sputum microbiome

A total of 2329 OTUs at 97% sequence identity were observed in 124 stable COPD samples after rarefaction. The *Proteobacteria* (51% \pm 12%) phylum constituted half of the COPD microbiome with *Haemophilus* (25% \pm 8%) as its most abundant genus followed by *Erwinia* (7% \pm

Table 1 Healthy and COPD subject characteristics

	Healthy control (n = 124)	Healthy control PY ≥10 (n = 28)	Healthy control PY < 10 (n = 96)	COPD (n = 218)	p-value*
Gender female (n)	73	15	58	60	< 0.0001
Age (years)	61 (54, 67)	65 (58, 69)	59 (53, 67)	69 (64,74)	< 0.0001
BMI	27.7 (24.3, 31.2)	29.3 (25.8, 34.4)	27.3 (24.2, 29.9)	26.7 (23.32,30.06)	0.0325
Smoking history (Pack years)	0 (0, 8)	29 (17.3, 45.8)	0 (0,0)	46 (34,64)	< 0.0001
MRC Dyspnoea scale	0 (0, 1)	0 (0, 1)	0 (0, 1)	2 (1,2)	< 0.0001
VAS dyspnoea (mm scale)	0 (0, 2)	0 (0, 3.5)	0 (0, 2)	30 (15,49)	< 0.0001
Post BD FEV1 (L)	2.87 (2.43, 3.37)	2.82 (2.33, 3.24)	2.88 (2.46, 3.40)	1.45 (1.03,1.86)	< 0.0001
Post BD FEV1% Predicted	109 (100, 118)	105 (92, 118)	110 (100, 118)	57 (42, 69)	< 0.0001
Post BD FEV1/FVC %	80 (77, 83)	78 (75,82)	80 (78, 83)	51 (41,59)	< 0.0001
Blood White cell count (×10 ⁹ /L)	5.8 (4.9, 7.3)	6.4 (5.1, 7.5)	5.7 (4.8, 7.25)	7.2 (6.3,8.8)	< 0.0001
Blood Neutrophil count (× 10 ⁹ /L)	3.3 (2.7, 4.2)	3.8 (2.8, 4.3)	3.2 (2.6, 4.1)	4.7 (3.9, 5.7)	< 0.0001
Blood Eosinophil count (× 10 ⁹ /L)	0.14 (0.09, 0.23)	0.18 (0.11, 0.25)	0.14 (0.09, 0.22)	0.21 (0.14, 0.27)	< 0.0001
Sputum TCC (×10 ⁶ /g)	2.4 (1.1, 4.2)	3.19 (1.6, 5.9)	2.2 (1, 3.6)	10.5 (4.7,13.0)	< 0.0001
Sputum eosinophils %	0.25 (0, 0.75)	0.25 (0.18, 2.38)	0.25 (0, 0.75)	0.75 (0.2)	0.0029
Sputum neutrophil (%)	60 (43, 74)	65 (42, 75)	56 (43, 73)	75 (40, 89)	0.0011

Data is expressed as median (IQR). BMI body mass index, Sputum TCC sputum total cell count, FEV1 forced expiratory volume in 1 s, Post BD FEV1 spirometry recording post bronchodilator; FVC forced vital capacity. * = represents Kruskal-Wallis test comparing between healthy (PY ≥10), healthy (PY < 10) and COPD

3%), *Cronobacter* (6% ± 2%) and *Moraxella* (3% ± 7%) (Fig. 2a and b). The two other abundant phyla were *Firmicutes* (29% ± 9%) and *Bacteroidetes* (16% ± 5%) dominated respectively by *Veillonella* (16% ± 9%), *Granulicatella* (7% ± 3%) and *Streptococcus* (5% ± 2%) and by *Prevotella* (14% ± 5%) at the genus level (Fig. 2a and b).

COPD ex-smokers and current smokers showed similar bacterial composition at phylum and genus level (Fig. 2a and b) and no significant difference in their microbial diversity measurements (Fig. 2c and d). There was no significant difference in the microbiome profile of the COPD samples from the three centres (efigure 1).

Healthy vs COPD microbiome

In the healthy microbiome *Streptococcus* (28%) from the *Firmicutes* (55%) phylum was the predominant constituent, while *Haemophilus* (3%) from *Proteobacteria* (7%) was present at low levels. In contrast, for the COPD cohort *Haemophilus* (25%) was the most dominant genus with a low proportion of *Streptococcus* (5%) observed (Fig. 3a and b). PCoA analysis showed distinct clusters of healthy and COPD subjects with significant difference in their microbiome by permutation multivariate analysis of variance (PERMANOVA $p = 0.01$) (Fig. 2c). COPD samples had more OTUs identified compared to healthy individuals, with significantly higher alpha diversity indices (Fig. 3d). Linear discriminant effect size (LEfSe) analysis revealed a higher abundance of *Gammaproteobacteria* species and lower proportion of *Firmicutes*, *Bacteroidetes* and *Actinobacteria* taxa to be

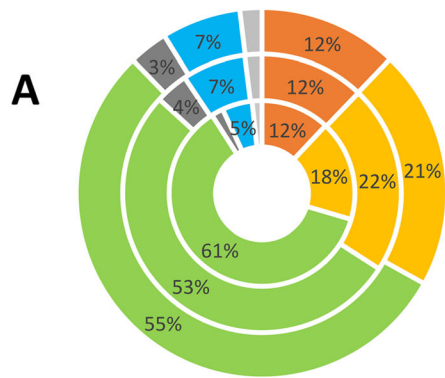
the major contributors in differentiating COPD from health (Fig. 4). OTUs differentiating the two groups are presented in Table 2.

Multivariate analysis by linear models (MaAsLin) analysis detected no significant association either in COPD or in health between the bacterial groups and clinical characteristics related to smoking, lung function and symptom score.

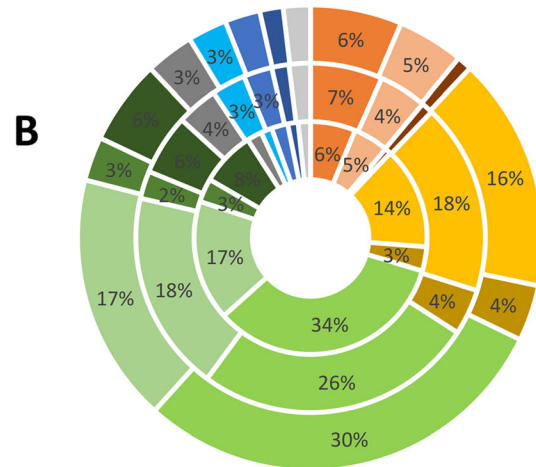
Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) analysis was performed to predict functional gene content from the 16S rRNA gene content. PCoA plots based on this analysis showed distinct clusters of COPD and health (Fig. 5a). Functional genes associated with Bacterial motility proteins, lipopolysaccharide biosynthesis, ABC transporters and secretion systems were in higher proportion in COPD while metabolic pathways were more abundant in healthy subjects (Fig. 5b).

Discussion

Our study is the largest to compare the sputum microbiome between health and COPD and clear differences between these groups were identified. *Firmicutes*, *Bacteroidetes* and *Actinobacteria* comprised 88% of the sputum microbiome in healthy participants, with *Streptococcus*, *Prevotella* and *Veillonella* as the dominant genera. *Haemophilus*, the dominant genus in COPD was present in health at a significantly lower proportion. Healthy smokers with ≥10 PY smoking history showed a trend towards a higher ratio of *Streptococcus* to *Prevotella*.



Actinobacteria Bacteroidetes
 Firmicutes Fusobacteria
 Proteobacteria Other_phyla



Actinomyces Rothia
 Actinobacteria_others Prevotella
 Bacteroidetes_others Streptococcus
 Veillonella Granulicatella
 Firmicutes_others Fusobacteria_others
 Haemophilus Unclassified_Neisseriaceae
 Proteobacteria_others Others

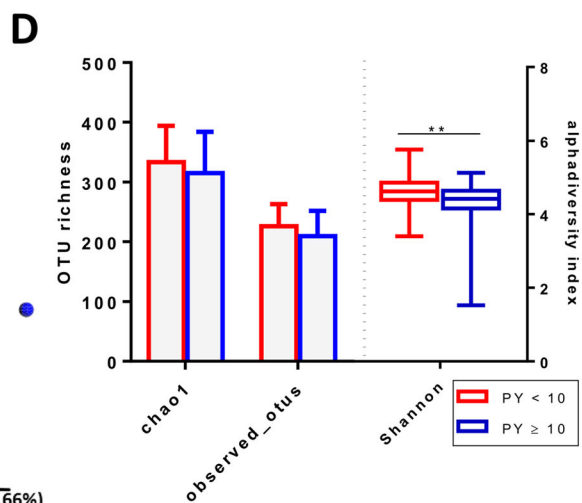
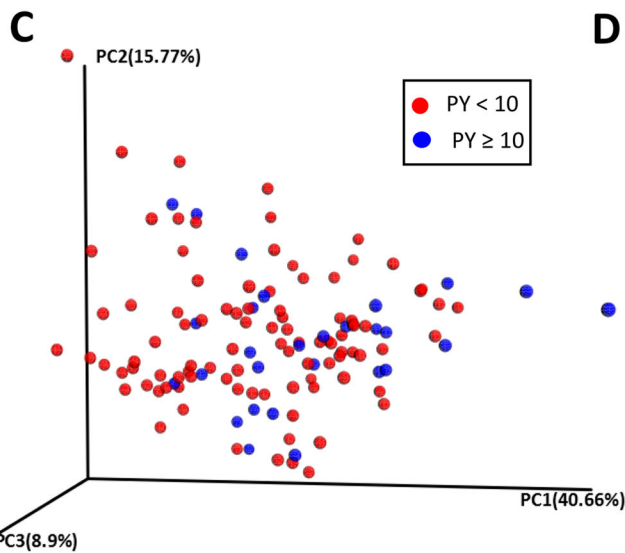
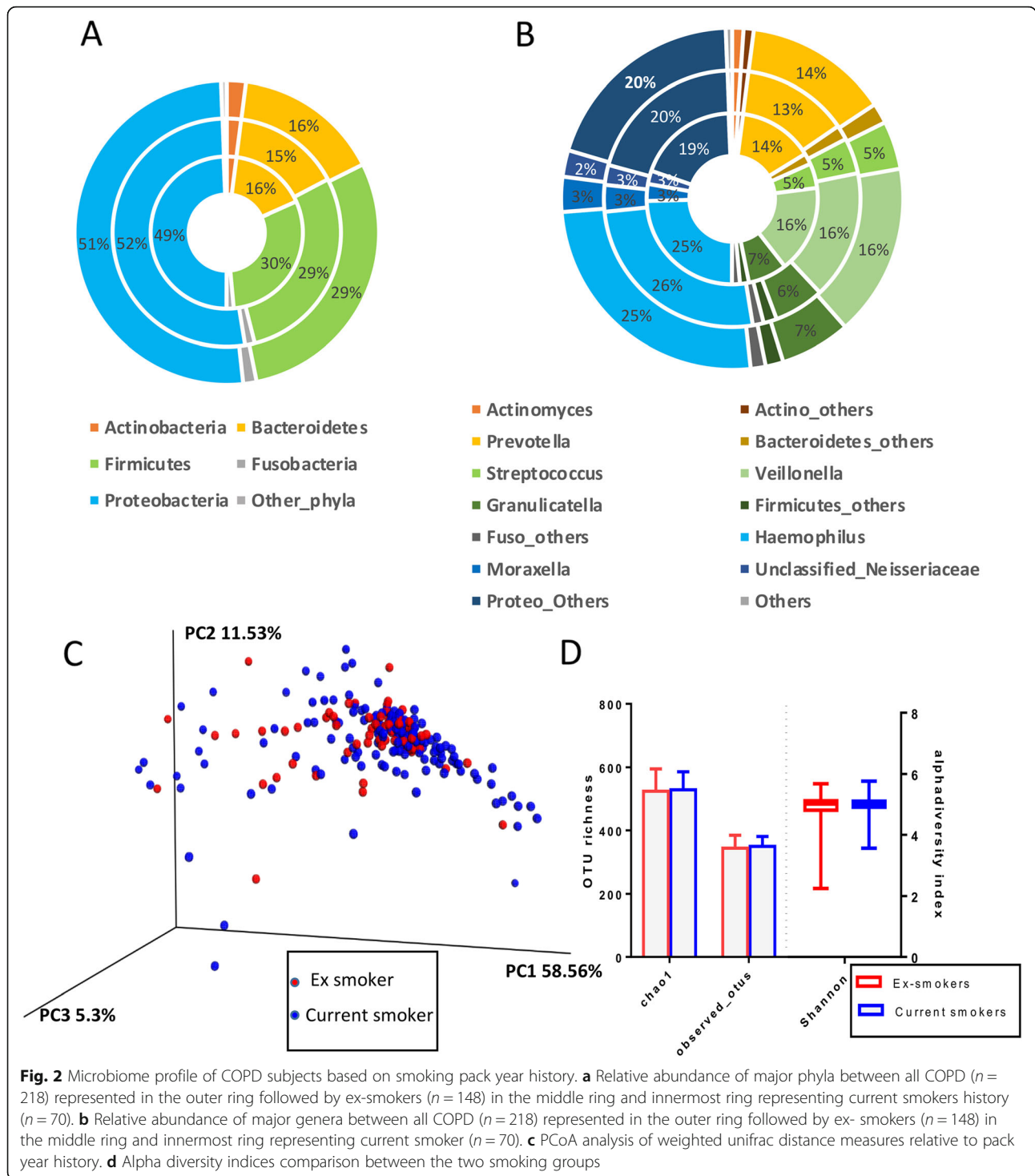
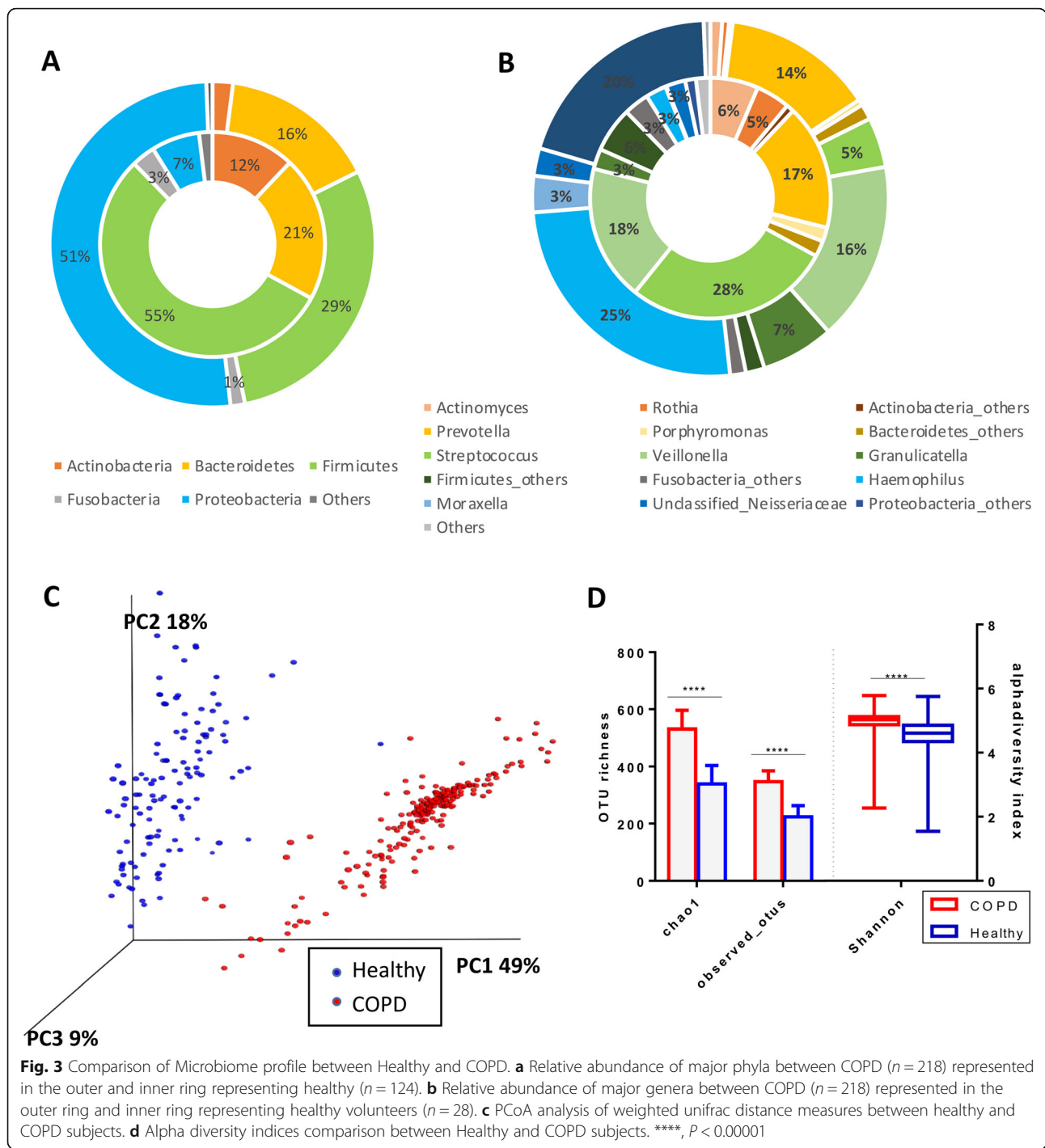


Fig. 1 Microbiome profile of Healthy volunteers based on smoking pack year history. **a** Relative abundance of major phyla between all healthy ($n = 124$) represented in the outer ring followed by healthy < 10 PY smoking history subgroup ($n = 96$) in the middle ring and innermost ring representing healthy ≥ 10 PY history subgroup ($n = 28$). **b** Relative abundance of major genera between all healthy ($n = 124$) represented in the outer ring followed by healthy < 10 PY ($n = 96$) in the middle ring and innermost ring representing healthy ≥ 10 PY history ($n = 28$). **c** Principal coordinate analysis (PCoA) analysis of weighted unifrac distance measures relative to pack year history. **d** Alpha diversity indices comparison between < 10 PY and ≥ 10 PY smoking sub-groups. Chao1 and observed_otus are represented as bar chart as mean and standard deviation; Shannon index is represented by box whisker plot showing median, interquartile range and minimum and maximum. **: $P < 0.01$



Existing comparative respiratory microbiome data are divided, with some studies reporting an overlapping microbial composition between health and COPD [3, 4, 26] while others have shown the COPD microbiome to be distinct [5, 6, 27]. These discordant outcomes likely reflect the underlying heterogeneity in COPD groups and small sample sizes of healthy individuals (< 20),

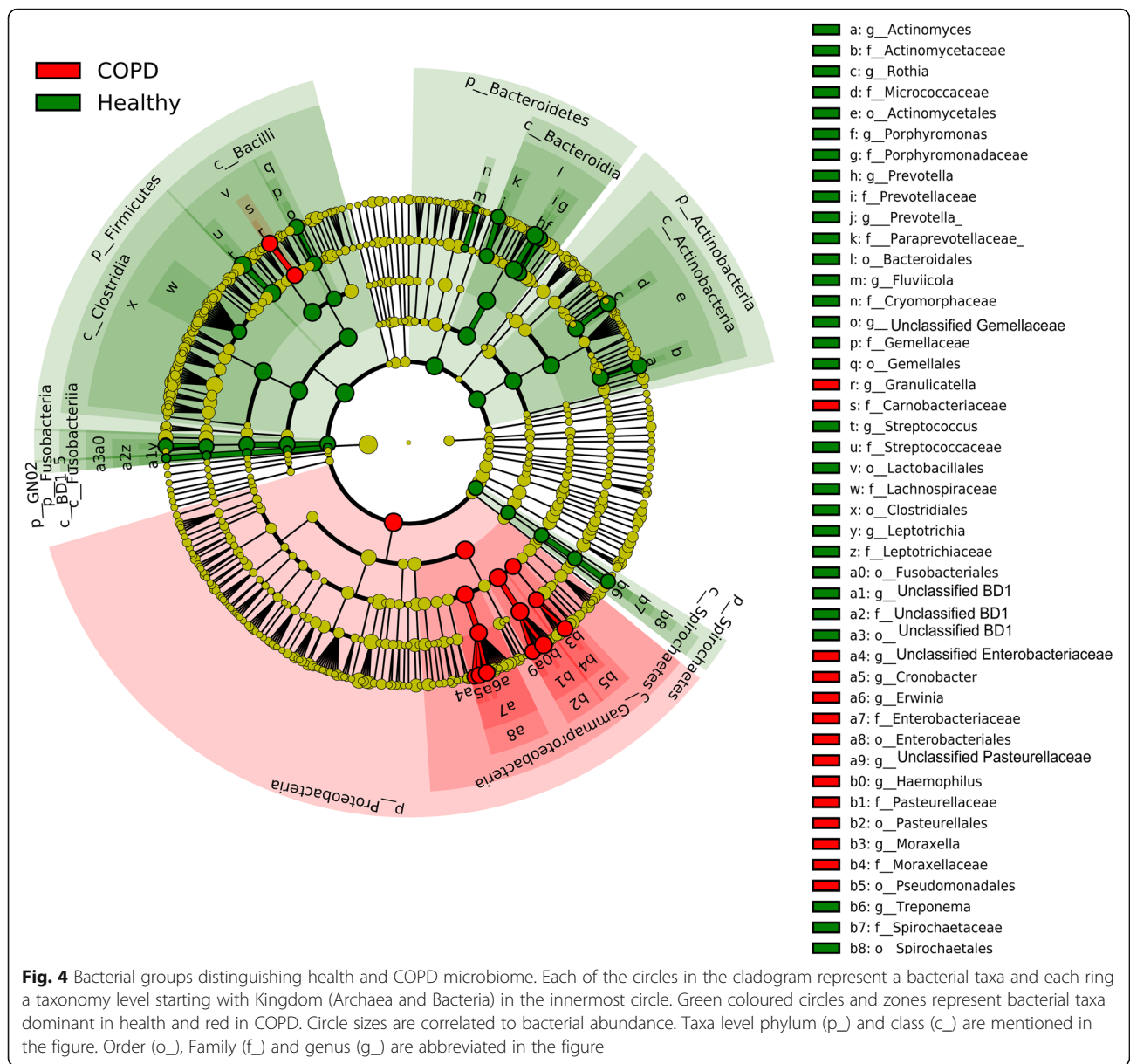
undermining the strength of these studies [3–6, 26, 27]. However, similar to our observation, higher levels of *Proteobacteria*, especially *Haemophilus*, in COPD [3–5] and relatively higher proportion of *Prevotella*, *Veillonella* and *Actinomyces species* in health have been observed [5, 6] but differences did not reach significance. Contrary to our observations, most studies have reported similar or a



higher abundance of *Firmicutes* and especially *Streptococcus* in COPD compared with health [4, 6]. One reason for this might be that composition varies between samples depending upon the type of treatment received, disease severity and inflammation. Previous studies have shown association between very severe COPD and eosinophilic phenotypes with dominance of *Firmicutes*, while *Proteobacteria* are predominant in moderate

COPD and the bacterial related phenotype [9, 28]. The COPD cohort analysed here was mainly of moderate-to-severe severity with high neutrophil counts.

Contrary to most studies, we found a higher alpha diversity in COPD compared to health [3, 6]. Although the COPD sample reads were reanalysed with the healthy at a normalised sequence depth, they were sequenced as part of COPDMAP study which involved a much larger



sample size [13], including different disease stages, and this may have contributed higher COPD diversity. Moreover, our COPD cohort was older than the healthy group and from moderate to severe GOLD stage. Higher diversity has been associated with both increasing disease severity and age in COPD [4, 27, 29].

Although a strong association exists between smoking and both airway inflammation and COPD, the determinants for developing COPD in smokers are not clear. We hypothesised that smoking associated microbiome changes in health would help in understanding the role of microbes in transition from health towards COPD. Consistent with previous studies, we found no significant difference between the microbiome of smokers with ≥10PY history and the < 10 PY group [3, 4, 6, 30].

However, similar to Morris and colleagues [30], a trend towards lower proportions of both *Bacteroidetes* and *Proteobacteria* in smokers with ≥10PY history was observed, suggesting subtle effects of smoking on the airway microbiome. Other pathological factors may therefore be important in shaping the microbiome in COPD. Hypoxia and chronic systemic inflammation related factors, which are features of COPD, have been reported to be associated with the airway microbiome [31] and may be relevant to the differences observed in our COPD cohort.

PICRUSt analysis showed relatively higher lipopolysaccharide biosynthesis products in COPD. Lipopolysaccharides are present in the outer membrane of *Proteobacteria* and together with pathogen-associated

Table 2 OTU groups that distinguish the Healthy and the COPD subjects

OTU	Taxonomy	Healthy_ mean	COPD_ mean	FDR_ P	Bonferroni_ P
579,608	Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	658	88	0.001	0-020
787,709	Actinobacteria Actinobacteria Actinomycetales Actinomycetaceae Actinomyces	363	64	0.001	0-020
1,078, 207	Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	7	184	0.001	0-020
865,469	Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Haemophilus	297	2150	0.001	0-020
585,419	Firmicutes Clostridia Clostridiales Veillonellaceae Veillonella	1611	1214	0.001	0-020
1,083, 037	Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae Moraxella	12	345	0.001	0-020
968,954	Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	906	2	0.001	0-020
932,696	Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Erwinia	0	782	0.001	0-020
579,924	Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae	2	210	0.001	0-020
1,083, 194	Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	1290	5	0.001	0-020
1,027, 587	Firmicutes Bacilli Lactobacillales Carnobacteriaceae Granulicatella	57	702	0.001	0-020
1,017, 181	Actinobacteria Actinobacteria Actinomycetales Micrococcaceae Rothia	488	63	0.001	0-020
667,570	Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Cronobacter	0	702	0.001	0-020
935,742	Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae Prevotella	298	101	0.001	0-020
342,427	Firmicutes Clostridia Clostridiales Veillonellaceae Veillonella	44	284	0.001	0-020

Non-parametric t-test was performed on rarefied OTU table at 11,000 reads with OTUs that were observed in minimum 25% of the total number of samples and had a minimum contribution of 1% of the total reads

molecular patterns, induce strong and damaging pro-inflammatory responses. In keeping with this, our previous study showed that sputum chemokine interleukin-8, known to play a key role in COPD inflammation, is positively correlated with *Haemophilus* and *Moraxella*, suggesting these bacteria trigger the excessive production of this chemokine [9]. Moreover, *Haemophilus* has been implicated in a dysbiotic role by co-inclusion of its related phylotypes and depletion of *Firmicutes*, *Bacteroidetes* and *Actinobacteria* that are involved in pathways for production of anti-inflammatory compounds [8, 9].

Although antibiotic treatment has been associated with suppression of *Proteobacteria* in COPD [8, 9], it is not true for all cases [10]. With the increasing urgency for effective antibiotic stewardship, research is needed to better understand the impact of both acute and long term antimicrobial therapy on the COPD microbiome. In this respect, alternate therapeutic strategies such as *H. influenzae* vaccination, or highly selective antimicrobial approaches such as phage therapy may effectively reverse some dysbiotic with prognostic benefit.

A limitation of this study is that the lung microbiome has been analysed from sputum samples which can be contaminated with the microbiome of the oropharynx. However, we emphasise that this effect will have been limited by sputum plug selection for the analysis. We did not perform longitudinal sampling to demonstrate reproducibility of the

sputum microbiome over time in healthy participants. For COPD we have previously demonstrated that the sputum microbiome is comparable between time-points when sampling at their stable state [32]. The effects on the microbiome of using sputum induction as the predominant sampling technique in the healthy control group are also not known, but it is noteworthy that the predominant bacterial constituents of our healthy microbiome are consistent with the respiratory microbiome detected by investigating BAL and bronchial samples reported in previous studies [3, 6]. This suggests that our observations are robust and representative of the bacterial composition of the lung microbiome. A major incentive to work with sputum is its compatibility with routine clinical practice as any findings are therefore more readily translated into established care pathways. In this study we have not characterized the viral and fungal communities, and this will be important to understand their role in health and disease.

Conclusions

In summary, clear and significant differences exist between the lung microbiome in health and COPD, with dysbiosis in COPD characterised by increased abundance of *Proteobacteria* especially *Haemophilus*. The changes observed in COPD are distinct from the microbiome in smokers without COPD, suggesting an association between airway damage and dysbiosis. The

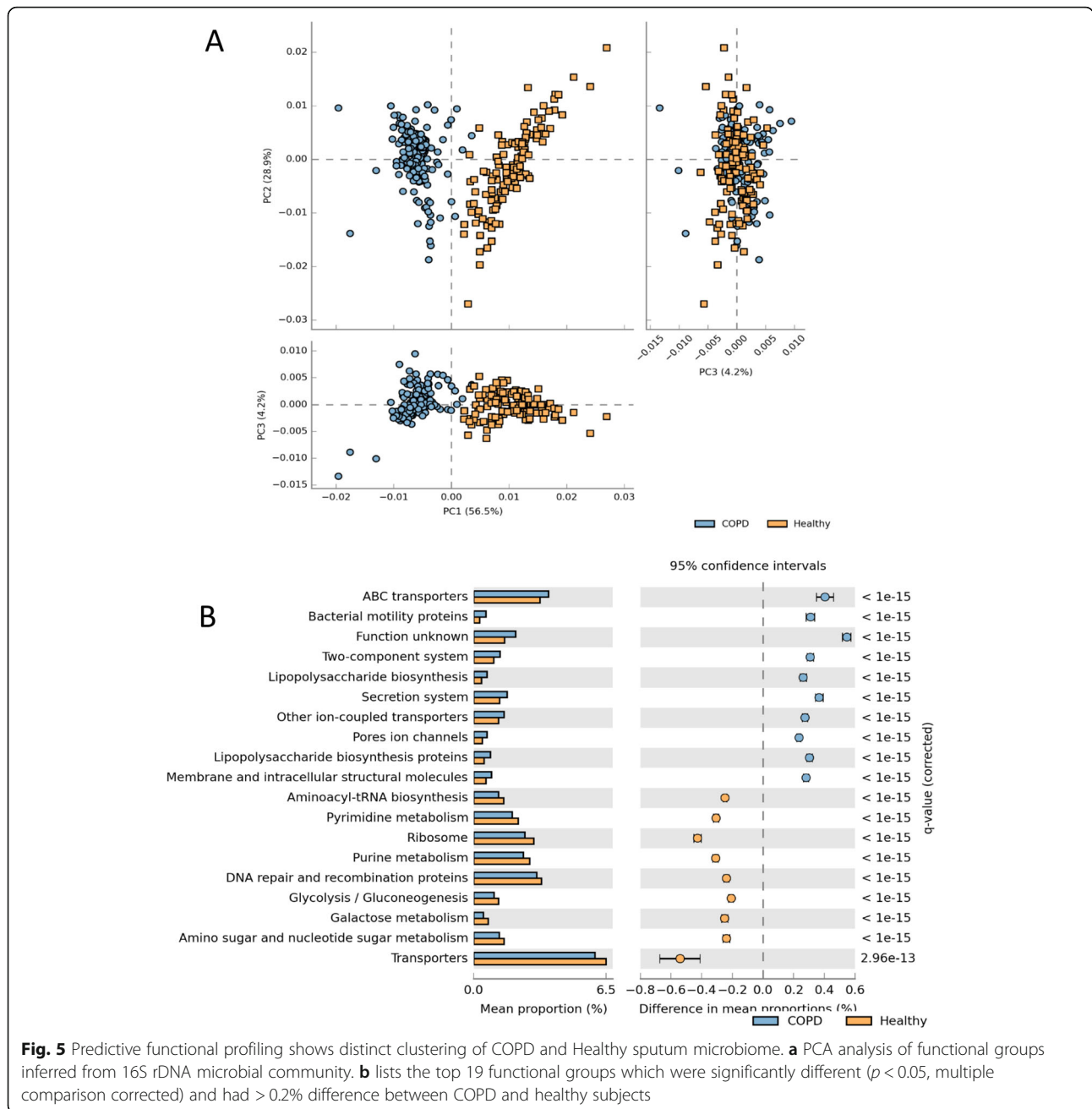


Fig. 5 Predictive functional profiling shows distinct clustering of COPD and Healthy sputum microbiome. **a** PCA analysis of functional groups inferred from 16S rDNA microbial community. **b** lists the top 19 functional groups which were significantly different ($p < 0.05$, multiple comparison corrected) and had $> 0.2\%$ difference between COPD and healthy subjects

pathogenesis and pathological significance of dysbiosis in COPD remains unclear. Longitudinal studies are needed to determine whether, and to what extent, the onset and progression of COPD are attributable to an altered lung microbiome.

Abbreviations

COPD: Chronic obstructive pulmonary Disease; EXCEED: Extended Cohort for E-health, Environment and DNA; KW: Kruksal-Wallis; MaAsLin: Multivariate analysis by linear models; OTU: Operational taxonomic unit; PY: Pack year smoking history; PCoA: Principal coordinate analysis; PERM ANOVA: Permutation multivariate analysis of variance; SEM: Standard error mean; LefSe: Linear discriminant effect size

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12931-020-01448-3>.

Additional file 1.

Acknowledgements

The EXCEED study gratefully acknowledges the support of all participants and staff who have contributed to the study. The study has been supported by the University of Leicester, Leicester City Council, the NIHR Leicester Biomedical Research Centre, the NIHR Clinical Research Network East Midlands, and the Medical Research Council (grant G0902313 to MDT) and the Wellcome Trust (grant 202849 to MDT).

Authors' contributions

KH undertook the data analysis and statistical analysis. LG and VM undertook patient recruitment, data collection, and were involved in data analysis. KH and MRB led the microbiological assessments. Samples and analyses were contributed by LG, CJ, NFR and MT (EXCEED), and ZW, JRB, DS, JAW, GCD, BEM and RTS (COPDMAP) study. RF and AJW were responsible for data management. CEB led the design of the study, data collection, data interpretation, data analysis and had full access to the data and was responsible for the integrity of the data and final decision to submit. All the authors contributed to drafting the manuscript or revising it critically for important intellectual content; all authors gave final approval of the version to be published; and all authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This study was part supported by the MRC (COPDMAP), National Institute for Health Research Leicester Biomedical Research Centre, UK, AirPROM (FP7–270194). MDT is in receipt of a Wellcome Trust Investigator Award (202849). CJ holds a MRC Clinical Research Training Fellowship (MR/P00167X/1). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Availability of data and materials

Exceed healthy sample 16 rRNA gene sequence reads are deposited at the National Centre for Biotechnology Information (SRA accession: PRJNA491861) <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA491861>. COPDMAP 16S rRNA gene sequence data are deposited at the National Centre for Biotechnology Information Sequence Read Archive (SRP102480) and associated data available from (<https://www.ncbi.nlm.nih.gov/search/all/?term=SRP102480>).

Ethics approval and consent to participate

All patients provided written informed consent using protocols approved by the local Ethics Committees at each site (London- 11/L0/1630; Manchester-10/H/1003/108; Leicester- 07/H0406/157) for COPDMAP study and the EXCEED study ethics number is: 13/EM/0226.

Consent for publication

Not applicable.

Competing interests

KH, LG, ZW, VM, MYR, RCF, NFR, AJW, AJB, MRB have nothing to declare. JRB, BEM and RTS are employees and shareholders of GSK; GCD reports grants and personal fees from Astrazeneca and grants from Micom Ltd and American Thoracic Society; MDT reports grants from GSK and other from Orion; JAW reports grants from GSK, grants from Johnson and Johnson, other from Novartis, other from Boehringer Ingelheim, other from Astra Zeneca, other from GSK, grants from GSK, grants from Astra Zeneca, grants from Boehringer Ingelheim, grants from Novartis; DS reports grants and personal fees from GlaxoSmithKline, grants and personal fees from AstraZeneca, grants and personal fees from Boehringer Ingelheim, grants and personal fees from Chiesi, personal fees from Cipla, personal fees from Genentech, grants and personal fees from Glenmark, grants and personal fees from Menarini, grants and personal fees from Mundipharma, grants and personal fees from Novartis, personal fees from Peptinnovate, grants and personal fees from Pfizer, grants and personal fees from Pulmatrix, grants and personal fees from Therevance, grants and personal fees from Verona; CEB reports grants and personal fees from GSK, grants and personal fees from Novartis, grants and personal fees from Genentech, grants and personal fees from Chiesi, personal fees from Sanofi/Regeneron, grants and personal fees from 4DPharma, grants and personal fees from BI, grants and personal fees from Mologics, grants and personal fees from Gossamer, grants and personal fees from AZ/MedImmune, grants and personal fees from TEVA, outside the submitted work. CEB reports grants from MRC COPDMAP, grants from AirPROM (FP7–270194), during the conduct of the study.

Author details

¹Institute for Lung Health, NIHR, BRC, Department of Respiratory Sciences, College of Life Sciences, University of Leicester, Leicester LE1 7RH, UK. ²Institute of Ecological Science, School of Life Science, South China Normal University, Guangzhou 510631, China. ³Department of Health Sciences, NIHR,

BRC, University of Leicester, Leicester LE1 7RH, UK. ⁴Global Medical, GSK, Collegeville, PA 19426, USA. ⁵Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. ⁶University of Manchester and University Hospital of South Manchester, Manchester M23 9QZ, UK. ⁷National Heart and Lung Institute, Imperial College London, London SW3 6NP, UK. ⁸Computational Biology, Human Genetics, Research and Development (R&D), GlaxoSmithKline (GSK), Collegeville, PA 19426, USA. ⁹Institute for Lung Health, University Hospitals of Leicester, Groby Road, Leicester LE3 9QP, UK.

Received: 9 April 2020 Accepted: 8 July 2020

Published online: 14 July 2020

References

- Sethi S, Murphy TF. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med*. 2008;359(22):2355–65.
- Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet*. 2014;384(9944):691–702.
- Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One*. 2011;6(2):e16384.
- Pragman AA, Kim HB, Reilly CS, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS One*. 2012;7(10):e47305.
- Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. *PLoS One*. 2010;5(1):e8578.
- Einarsson GG, Comer DM, McIlreavey L, Parkhill J, Ennis M, Tunney MM, et al. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax*. 2016;71(9):795–803.
- Cabrera-Rubio R, Garcia-Núñez M, Setó L, Antó JM, Moya A, Monsó E, et al. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. *J Clin Microbiol*. 2012;50(11):3562–8.
- Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *J Clin Microbiol*. 2014;52:2813–23.
- Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. *Eur Respir J*. 2016;47(4):1082–92.
- Haldar K, Bafadhel M, Lau K, Berg A, Kwambana B, Kebabdzé T, et al. Microbiome balance in sputum determined by PCR stratifies COPD exacerbations and shows potential for selective use of antibiotics. *PLoS One*. 2017;12(8):e0182833.
- Charlson ES, Chen J, Custers-Allen R, Bittering K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One*. 2010;5(12):e15216.
- Johannessen A, Omenaas ER, Bakke PS, Gulsvik A. Implications of reversibility testing on prevalence and risk factors for chronic obstructive pulmonary disease: a community study. *Thorax*. 2005;60(10):842–7.
- Wang Z, Singh R, Miller BE, Tal-Singer R, Van Horn S, Tomsho L, et al. Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study. *Thorax*. 2018;73(4):331–8.
- Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr Protoc Microbiol*. 2012;27:1E–5.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
- Aronesty E. Command-line tools for processing biological sequencing data. 2011. Available from: <https://github.com/ExpressionAnalysis/ea-utils>.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27(16):2194–200.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7.
- Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71(12):8228–35.
- Langille MGJ, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotech*. 2013;31(9):814–21.

21. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014;30(21):3123–4.
22. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Aust Ecol*. 2001;26(1):32–46.
23. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011; 12(6):R60.
24. Bergen B, Endres S, Engel A, Zark M, Dittmar T, Sommer U, et al. Acidification and warming affect prominent bacteria in two seasonal phytoplankton bloom mesocosms. *Environ Microbiol*. 2016;18(12):4579–95.
25. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13(9):R79.
26. Diao W, Shen N, Du Y, Erb-Downward JR, Sun X, Guo C, et al. Symptom-related sputum microbiota in stable chronic obstructive pulmonary disease. *Int J Chronic Obstruct Pulmon Dis*. 2018;13:2289–99.
27. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;185(10):1073–80.
28. Pragman AA, Lyu T, Baller JA, Gould TJ, Kelly RF, Reilly CS, et al. The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. *Microbiome*. 2018;6:7.
29. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. Chronic obstructive pulmonary disease lung microbiota diversity may be mediated by age or inhaled corticosteroid use. *J Clin Microbiol*. 2015;53:1050.
30. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*. 2013;187(10):1067–75.
31. Shukla SD, Walters EH, Simpson JL, Keely S, Wark PAB, O'Toole RF, et al. Hypoxia-inducible factor and bacterial infections in chronic obstructive pulmonary disease. *Respirology*. 2020;25(1):53–63.
32. Bafadhel M, Haldar K, Barker B, Patel H, Mistry V, Barer MR, et al. Airway bacteria measured by quantitative polymerase chain reaction and culture in patients with stable COPD: relationship with neutrophilic airway inflammation, exacerbation frequency, and lung function. *Int J Chron Obstruct Pulmon Dis*. 2015;10:1075–83.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

