

RESEARCH

Open Access



# Analysis of sputum microbial metagenome in COPD based on exacerbation frequency and lung function: a case control study

Wei Li<sup>1†</sup>, Bingbing Wang<sup>2†</sup>, Min Tan<sup>2</sup>, Xiaolian Song<sup>2</sup>, Shuanshuan Xie<sup>2</sup> and Changhui Wang<sup>2\*</sup>

## Abstract

**Background:** The role of the sputum microbiome in chronic obstructive pulmonary disease (COPD) progression remains elusive. As the advent of the new culture-independent microbial sequencing technique makes it possible to disclose the complex microbiome community of the respiratory tract. The aim of this study was to use metagenomic next-generation sequencing (mNGS) to confirm whether there are differences in sputum microbiome of COPD between different exacerbation frequencies and lung function.

**Methods:** Thirty-nine COPD patients were divided into a frequent exacerbators (FE) group (n = 20) and a non-frequent exacerbators (NFE) (n = 19) group according to their exacerbation history, or a mild group (FEV<sub>1</sub>/pre ≥ 50%, n = 20) and a severe group (FEV<sub>1</sub>/pre < 50%, n = 19) according to the lung function. Sputum was collected during their stable phase, followed by DNA extraction, untargeted metagenomic next-generation sequencing (mNGS) and bioinformatic analysis.

**Results:** mNGS identified 3355 bacteria, 71 viruses and 22 fungi at the specie level. It was found that Shannon index and Simpson index in FE group was lower than that in NFE group (p = 0.005, 0.008, respectively) but similar between mild and severe groups. Out of top 10 bacteria taxa, Veillonella, Fusobacterium and Prevotella jejuni had a higher abundance in NFE group, Rothia had a higher abundance in mild group. Linear discriminant analysis revealed that many bacterial taxa were more abundant in NFE group, and they mostly belonged to Actinobacteria, Bacteroidetes and Fusobacteria phyla. Frequency of exacerbations was also found to be negatively correlated with alpha diversity (with Shannon index, r = - 0.423, p = 0.009; with Simpson index, r = - 0.482, p = 0.002). No significant correlation was observed between alpha diversity and FEV<sub>1</sub>/pre.

**Conclusions:** Microbiome diversity in FE group was lower than that in NFE group. There was a significant difference in microbiome taxa abundance between FE and NFE groups, or mild and severe groups. These findings demonstrated that sputum microbiome community dysbiosis was associated with different exacerbation frequencies and lung function in stable COPD.

**Keywords:** Chronic obstructive pulmonary disease (COPD), Metagenomic next-generation sequencing, Sputum, Microbiome, Exacerbation frequency, Lung function

<sup>†</sup>Wei Li and Bingbing Wang are contributed equally to this work.

\*Correspondence: wang-chang-hui@hotmail.com

<sup>2</sup> Department of Respiratory Medicine, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai 200072, China  
Full list of author information is available at the end of the article

## Background

Chronic obstructive pulmonary disease (COPD) is a common chronic disease characterized by persistent respiratory symptoms and airflow limitation. COPD is the third leading cause of death in the world, responsible for approximately 6% of total deaths [1].



The impact of COPD on mortality and morbidity will persist due to the aging population and urbanization. Mounting evidence has shown that dysbiosis of pulmonary microbiota participates in COPD progress [2]. Exacerbation frequency and lung function are two major indexes of COPD. The former is related to airflow limitation, symptoms and health-related quality of life impairment [3, 4], and the latter is an index revealing severity of airflow limitation. Till now, the difference in sputum microbiome of COPD between different exacerbation frequencies or lung function is still poorly understood.

The primary aim of the present study was using Metagenomic Next-Generation Sequencing (mNGS), also known as shotgun metagenomic sequencing, to confirm whether there are differences in sputum microbiome of COPD between different exacerbation frequencies and lung function.

## Methods

### Patient selection and grouping

Patients who used to seek hospital treatment for COPD exacerbation in the Department of Respiratory Medicine of Shanghai Tenth People's Hospital (Shanghai, China) were interviewed by telephone, and those who met with the inclusion criteria and agreed to participate in the research were invited to the said hospital for further medical checkups from April to May 2021. The inclusion criteria for stable COPD patients were: (a) symptoms of dyspnea, chronic cough or expectoration, and/or a history of exposure to risk factors for the disease; (b) the presence of a post-bronchodilator  $FEV_1/FVC < 0.7$ ; and (c) no symptomatic deterioration in recent 30 days [5]. Patients with malignancies, heart failure, cognitive disorders, and severe kidney or liver dysfunction were excluded. Basic demographic information, smoking history, exacerbation frequency in the last 1 year, modified Medical Research Council (mMRC) dyspnea scale and the COPD Assessment Test (CAT) were obtained. In addition, the participating patients were required to receive spirometry. Written informed consent was obtained from all participating patients before initiation of the investigation. The study protocol was approved by the Institutional Review Boards of the said hospital.

Patients with at least 2 treated exacerbation events in last 1 year were divided into a FE group, and patients with 1 or less treated exacerbation events in last 1 year were divided into a NFE group. In addition, the patients were also reclassified into a mild group ( $FEV_1/pre \geq 50\%$ ) and a severe group ( $FEV_1/pre < 50\%$ ) according to the lung function.

### Sampling

Samples were the first expectorated sputum obtained in the morning. The patients were required to clean their mouth before expectorating the sputum spontaneously into a DNA-free tube.

### Nucleic acid extraction, library preparation and sequencing

DNA was extracted from all samples using a QIAamp® UCP Pathogen DNA Kit (Qiagen) following the manufacturer's instructions. Human DNA was removed using Benzonase (Qiagen) and Tween20 (Sigma) [6]. 10 nanograms DNA samples were used for library construction through Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA) [7]. Library was qualitatively assessed by Qubit dsDNA HS Assay kit, followed by High Sensitivity DNA kit (Agilent) on an Agilent 2100 Bioanalyzer. Library pools were then loaded onto an Illumina Nextseq 550Dx sequencer for 75 cycles of single-end sequencing to generate approximately 20 million reads for each library. For negative controls, we also prepared peripheral blood mononuclear cell (PBMC) samples with  $10^5$  cells/mL from healthy donors in parallel with each batch, using the same protocol, and sterile deionized water was extracted alongside the specimens to serve as non-template controls (NTC) [7, 8]. DNA-free water went through DNA extraction and mNGS analysis as a blank control group to assess the degree of background contamination associated with DNA extraction kit and sequencing reagents together.

### Bioinformatics analysis

Trimmomatic [9] was used to remove low quality reads, adapter contamination, and duplicate reads, as well as those shorter than 50 bp. Low complexity reads were removed by Kcomplexity with default parameters [10]. Human sequence data were identified and excluded by mapping to a human reference genome (hg38) using Burrows-Wheeler Aligner software [11].

After alignment, multiple indicators were comprehensively evaluated to have the list of suspected microorganisms. Then, microbiota composition profiles were inferred from quality filtered forward reads using Kraken V.2.1.2 and Bracken V.2.6.2 with the `k2_pluspf_20210517` database. The site by species counts and relative abundance tables were input into R-base V.4.1.0 for statistical analysis. Alpha diversity of the microbiota profile for each subject was assessed by group at the different level data using the Vegan package in R (version 2.5.7). Principal component analysis (PCA) ordinations were used to visualize the clustering of the samples based on the compositional profiles at genus or specie level. Different groups

were assessed using permutational multi-variate analysis of variance (PERMANOVA) and PCA stat with the Vegan package in R (version 2.5.7). The result of PCOA was stated by the function of `dudi.pco` in `ade4` package in R (version 1.7.18). Associations of specific microbial species or genus with patient parameters were identified using the linear discriminant analysis effect size (LEfSe) [12].

### Statistical analysis

SPSS Statistics 26 was used to perform statistical analysis of clinical variables comparison between groups. Results for categorical variables were expressed as absolute and relative frequencies while results for continuous variables were expressed as means and standard deviations (SD) if normally distributed or as medians and interquartile ranges (IQR) when the distribution was not normal. Chi squared test was conducted for all categorical variables. As for continuous variables, two independent-sample t-test was conducted for normally distributed continuous variables and Mann–Whitney U test was conducted for not normally distributed. Correlation analysis were performed through Spearman's correlation analysis.  $p$  value  $< 0.05$  was considered statistically significant.

## Results

### Patient characteristics

Altogether 39 stable COPD patients were enrolled in our study (Table 1). The range of the post-bronchodilator

FEV<sub>1</sub>/FVC was 25.67–68.76%. According to their exacerbation history in the last 1 year, 19 fell in NFE group and 20 in FE group; according to their post-bronchodilator FEV<sub>1</sub>/pre value, 19 fell in severe COPD group and 20 in mild COPD group. Gender, age, smoking history and FEV<sub>1</sub>/FVC were comparable between FE and NFE groups, while BMI and FEV<sub>1</sub>/pre in NFE group were significantly higher than those in FE group ( $p = 0.038, 0.014$ , respectively). Gender, age and smoking history were also comparable between mild and severe groups, except that BMI in mild group was significantly higher than that in severe group ( $p = 0.041$ ).

### Bacterial, viral and fungal composition

Deep metagenome sequencing totally identified 1012 bacteria, 27 viruses and 18 fungi at genus level, and 3355 bacteria, 71 viruses and 22 fungi at specie level.

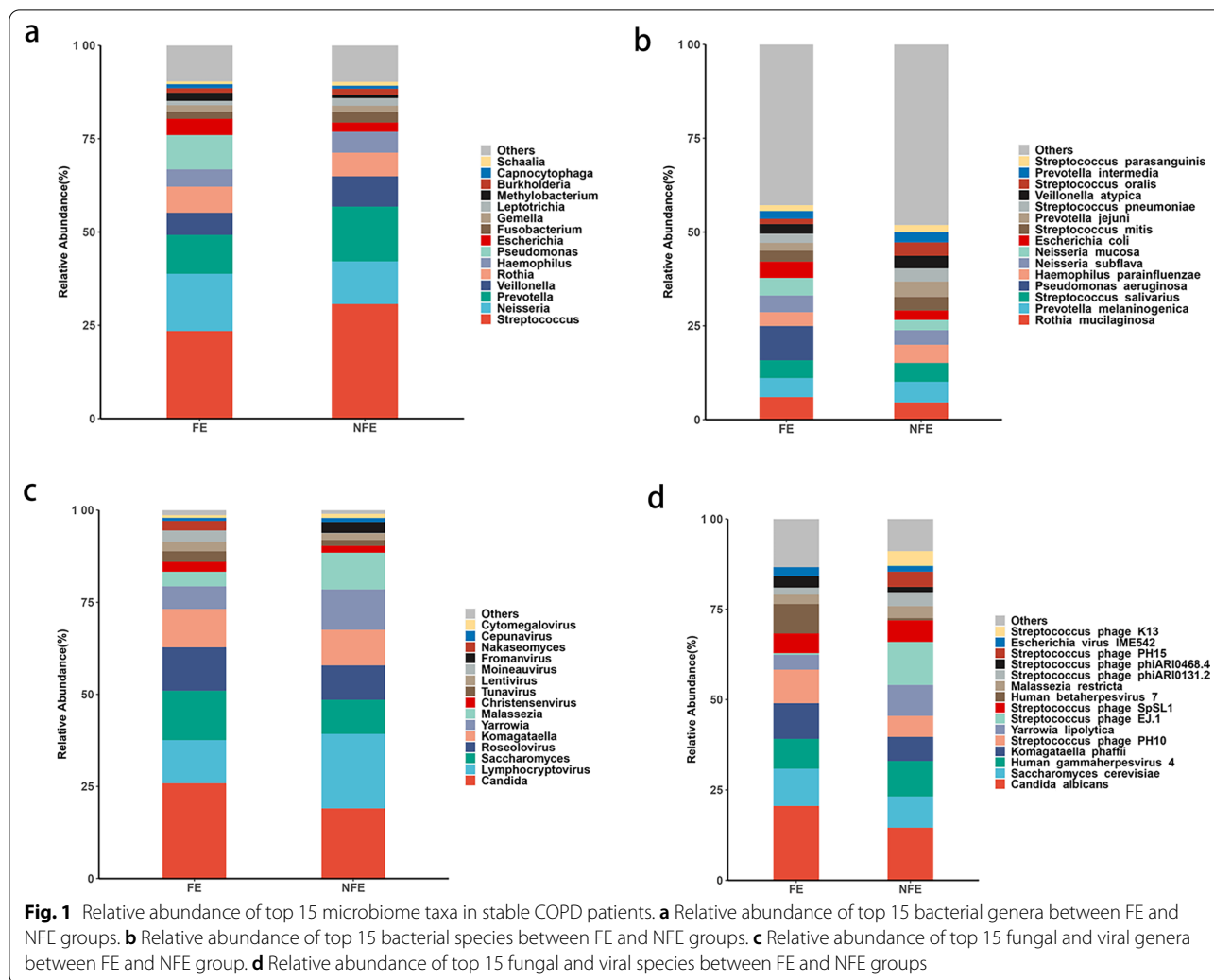
The microbiome composition comparison of the blank control group and sample group was displayed in Additional file 1: Fig. S1. As shown in the Additional file 1: Fig. S1, the major microbiome components of blank control group were completely different from the sample group, which means that the degree of background contamination was low.

The compositions of the top 15 bacteria, viruses and fungi in the sputum of FE and NFE groups at both genus and specie level are shown in Fig. 1. The top 15 bacterial genera were *Streptococcus*, *Neisseria*, *Prevotella*, etc. *Candida* was the most dominant fungal genus in stable

**Table 1** Clinical characteristics of the study population

|                           | All               | NFE               | FE                | p value | Severe            | Mild             | p value |
|---------------------------|-------------------|-------------------|-------------------|---------|-------------------|------------------|---------|
| N                         | 39                | 19                | 20                |         | 19                | 20               |         |
| Gender, Male (%)          | 29 (70.7%)        | 13 (68.4%)        | 16 (80%)          | 0.408   | 15 (78.9%)        | 14 (70%)         | 0.716   |
| Age (years)               |                   |                   |                   |         |                   |                  |         |
| Range                     | 56–88             | 60–84             | 56–88             |         | 56–88             | 60–84            |         |
| Mean $\pm$ sd             | 69.18 $\pm$ 7.25  | 68.68 $\pm$ 6.84  | 69.65 $\pm$ 7.77  | 0.683   | 66 (61, 72)       | 70 (67, 74.75)   | 0.118   |
| BMI                       |                   |                   |                   |         |                   |                  |         |
| Range                     | 15.43–32.28       | 16.60–32.28       | 15.43–29.38       |         | 15.43–30.19       | 16.60–32.28      |         |
| Mean $\pm$ sd             | 23.74 $\pm$ 4.13  | 25.14 $\pm$ 4.03  | 22.41 $\pm$ 3.86  | 0.038   | 22.36 $\pm$ 4.23  | 25.05 $\pm$ 3.66 | 0.041   |
| Smoking history           |                   |                   |                   |         |                   |                  |         |
| Ever-smoker               | 27(69.2%)         | 12(63.2%)         | 15(75%)           |         | 14(73.7%)         | 13(65%)          |         |
| Never-smoker              | 12(30.8%)         | 7(36.8%)          | 5(25%)            | 0.423   | 5(26.3%)          | 7(35%)           | 0.557   |
| FEV <sub>1</sub> /pre (%) |                   |                   |                   |         |                   |                  |         |
| Range                     | 20.3–97.7         | 21.8–97.7         | 20.3–93.7         |         | 20.3–47.6         | 50.4–97.7        |         |
| Mean $\pm$ sd             | 49.62 $\pm$ 20.91 | 58.6(39.2, 69.4)  | 37.55(25.4, 54.7) | 0.009   | 32.1(23.1, 39.2)  | 64(56.13, 70.88) | <0.001  |
| FEV <sub>1</sub> /FVC (%) |                   |                   |                   |         |                   |                  |         |
| Range                     | 25.67–68.76       | 25.67–68.76       | 27.03–68.24       |         | 25.67–61.48       | 46.71–68.76      |         |
| Mean $\pm$ sd             | 48.36 $\pm$ 12.25 | 51.28 $\pm$ 11.94 | 45.59 $\pm$ 12.17 | 0.149   | 39.60 $\pm$ 10.08 | 56.69 $\pm$ 7.35 | <0.001  |

FE frequent exacerbators, NFE non-frequent exacerbators, sd standard deviation, N number, BMI body mass index, FEV<sub>1</sub> forced expiratory volume in 1 s, FEV<sub>1</sub>/pre FEV<sub>1</sub> percent predicted, FVC forced vital capacity. Non-normal distributed data were expressed as medians and interquartile ranges (IQR)



COPD. Lymphocryptovirus was the most dominant virus in stable COPD. In terms of species, the top 15 dominant bacteria were *Rothia mucilaginosa*, *Prevotella melaninogenica*, *Streptococcus salivarius*, etc. *Candida albicans* was the most abundant fungus while human gammaherpesvirus 4, also named Epstein-Barr virus, was the most abundant virus in stable COPD.

#### Diversity comparison between groups

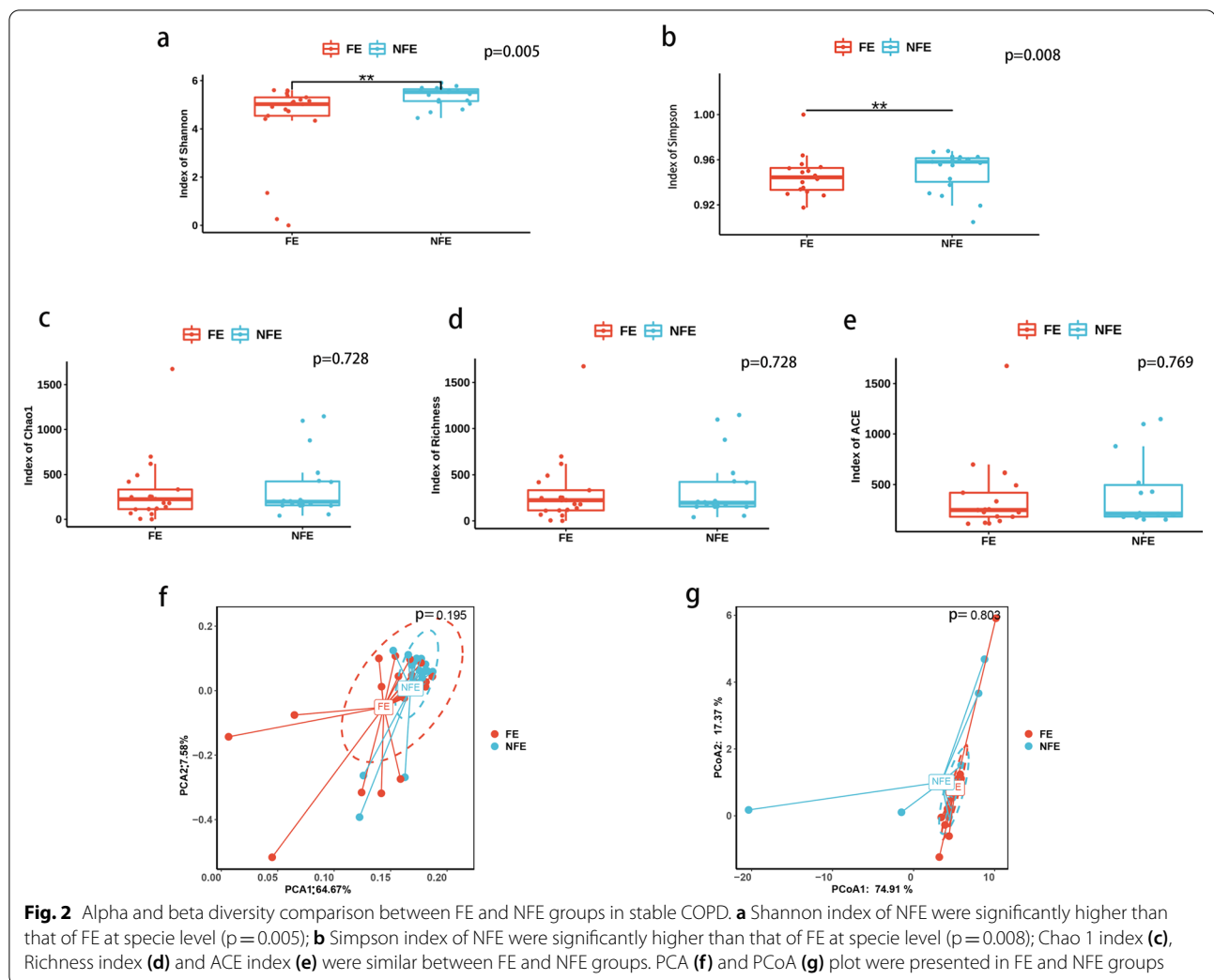
We then compared alpha and beta diversity of the sputum microbiome between groups. Both Simpson index ( $p=0.008$ ) and Shannon index ( $p=0.005$ ) of NFE were significantly higher than those of FE at specie level (Fig. 2a, b). Chao1 index, Richness index and ACE index were similar between FE and NFE groups at specie level (Fig. 2c–e). Besides, PCA plot and PCoA showed no significant difference between NFE and FE groups (Fig. 2f, g). At genus level, no significant difference of diversity was detected between FE and NFE group. No significant

difference in alpha and beta diversity was detected between mild and severe groups at genus or specie level.

#### Differential taxa between groups

The relative abundance top 10 bacterial taxa was compared between FE and NFE, mild and severe group (Fig. 3). Mann–Whitney U test identified *Veillonella*, *Fusobacterium* and *Prevotella jejuni* were significantly more abundant in NFE group compared with FE group. *Rothia* were more abundant in mild group compared with severe group.

A LefSe analysis was also performed to identify differences in microbiome distribution between COPD FE and NFE subgroups. The significantly enriched genera and species ( $LDA > 2$ ) in each group are displayed in Fig. 4. The relative abundance of *Fusobacterium*, *Leptotrichia*, *Porphyromonas*, etc. and *Prevotella jejuni*, *Prevotella intermedia*, *Leptotrichia wadei*, etc. was significantly increased in NFE group (Fig. 4a, b). While *Moraxella*



catarrhalis was significantly increased in FE group (Fig. 4b). As for fungi and virus, Streptococcus phage EJ 1 was increased in NFE group (Fig. 4c). Differences in microbiome distribution between mild and severe groups were also identified. The relative abundance of Rothia, Prevotella oris, Rothia dentocariosa, Tannerella forsythia, etc. was significantly increased and the abundance of Bifidobacterium, Clostridioides difficile, Enterobacter hormaechei and Brucella melitensis was significantly decreased in mild COPD group (Fig. 4d, e). In terms of virus and fungi, the relative abundance of Equine infectious anemia virus and Escherichia virus BIFF was both increased in severe group (Fig. 4f).

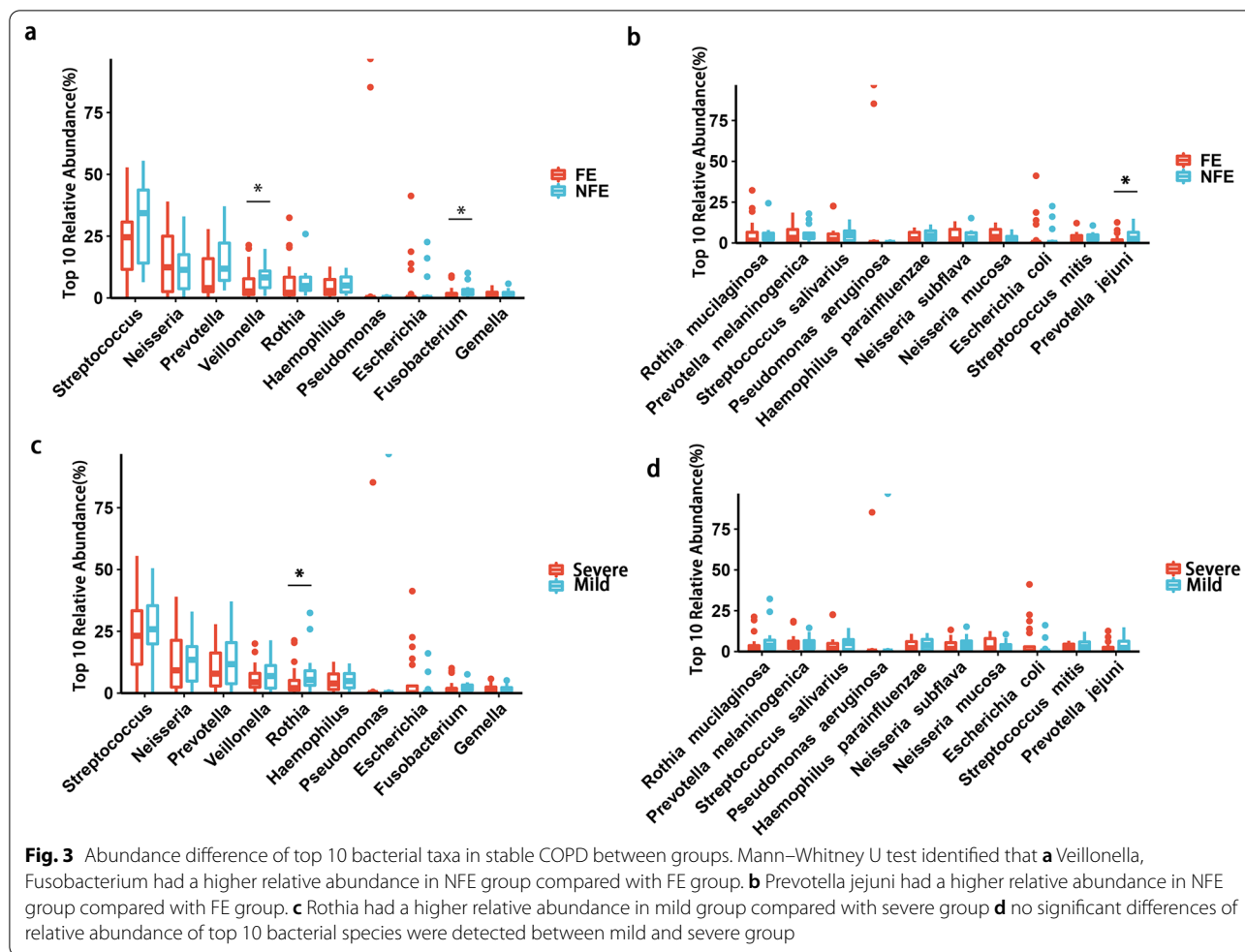
#### Correlation between the microbiome and clinical indexes

Correlation analysis was performed between clinical variables (frequency of exacerbation, age, smoking index, BMI, CAT, mMRC, FEV<sub>1</sub>, FEV<sub>1</sub>/pre (%) and FEV<sub>1</sub>/FVC) and microbiome diversity index (Shannon and Simpson

index). Spearman's correlation analysis showed that the frequency of exacerbation had a negative correlation with Shannon index ( $r = -0.423$ ,  $p = 0.009$ ) and Simpson index ( $r = -0.482$ ,  $p = 0.002$ ), and FEV<sub>1</sub>/pre showed no significant correlation with either Shannon or Simpson index (Fig. 5). In addition, age was also found to be negatively correlated with both Shannon and Simpson indexes (Additional file 1: Fig. S2).

An additional correlation analysis was performed between clinical variables (FEV<sub>1</sub>/pre and frequency of exacerbations) and the abundance of the microbiome genera and species that were identified by LDA analysis. Bifidobacterium and Bifidobacterium dentium proved to have a positive correlation with FEV<sub>1</sub>/pre ( $r = 0.325$  or  $0.331$ ) (Fig. 6a). A total of 29 bacterial genera and species showed a negative correlation with the frequency of exacerbation ( $r$  ranged from  $-0.463$  to  $-0.334$ ), 79.3% of which belonged to phylum Actinobacteria, Bacteroidetes and Fusobacteria (Fig. 6b).



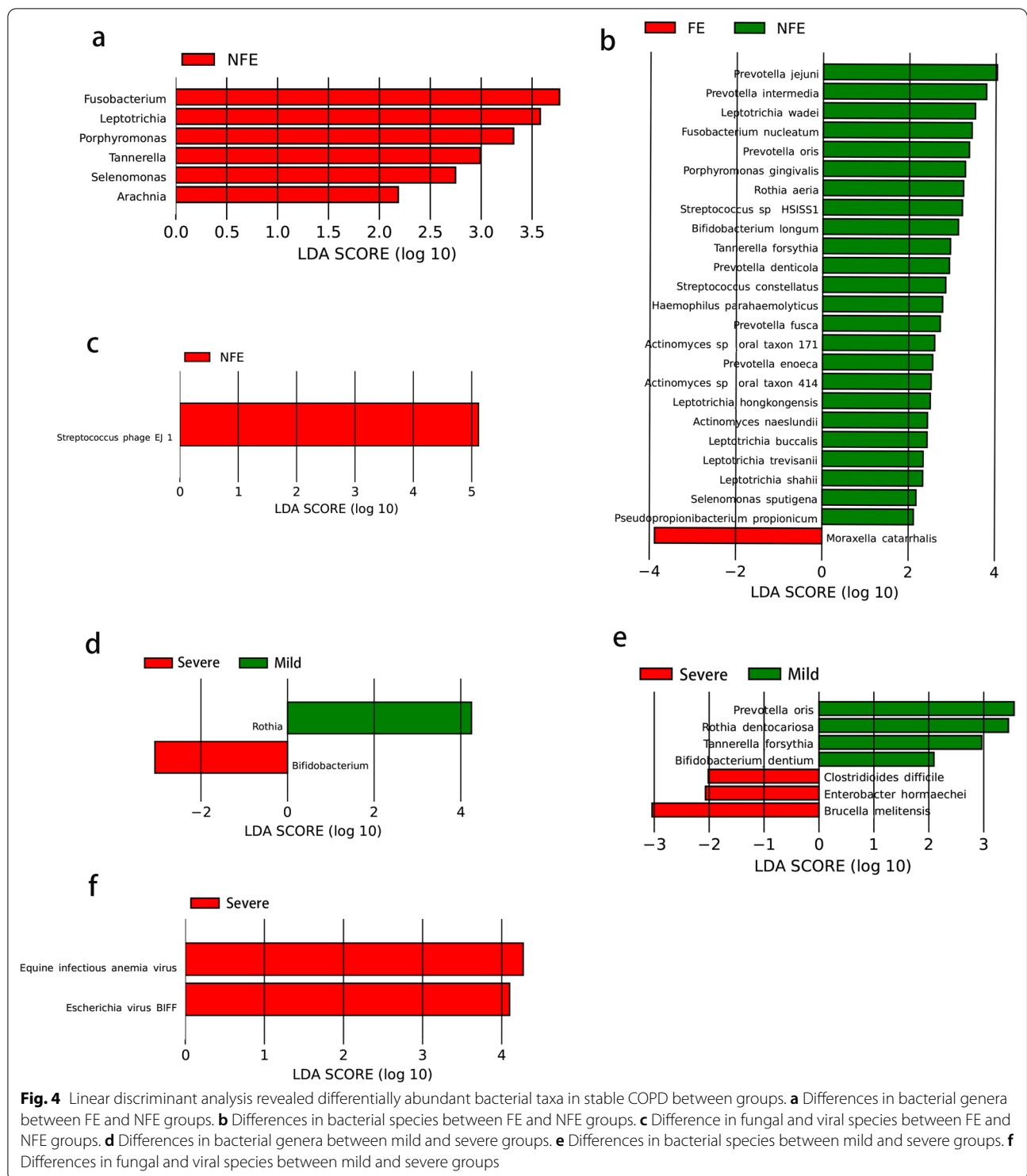


The Spearman correlations between the top 30 bacteria genus and species are shown in Fig. 7 and used to explore potential co-existence and co-exclusion relationships.

### Discussion

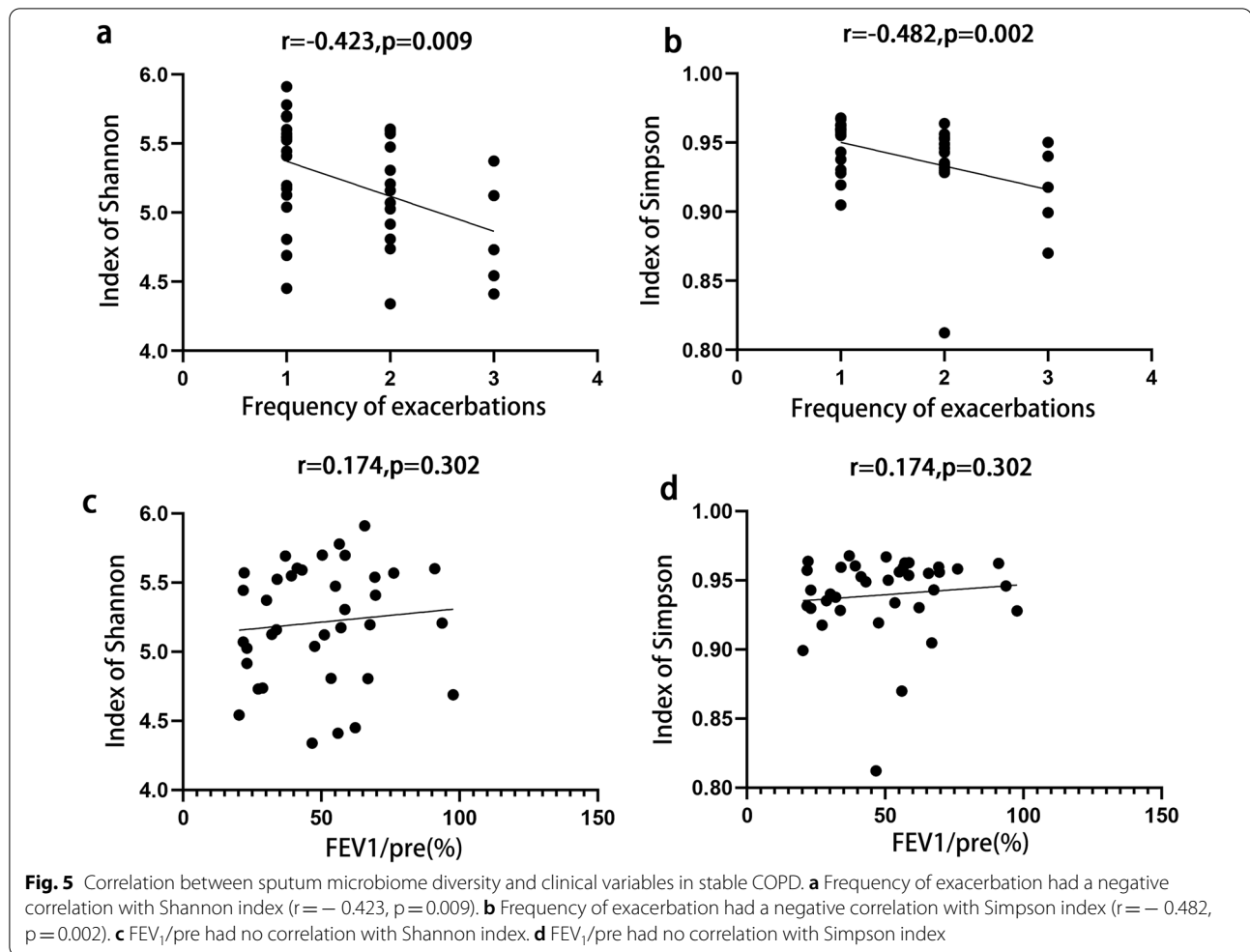
To the best of our knowledge, this is the first study reporting the use of mNGS to investigate the sputum microbiome in COPD of different exacerbation frequencies or lung function. The key findings of our study are as follows. First, the microbiome alpha diversity in FE group was significantly decreased as compared with NFE group, which was verified in the subsequent analysis, showing that the frequency of exacerbation was negatively correlated with microbiome diversity; microbiome diversity was similar between mild and severe groups. Second, the microbiome composition underwent significant changes. More specific bacterial taxa were enriched in NFE group and they mostly belonged to phylum Actinobacteria, Bacteroidetes and Fusobacteria. More specific bacterial taxa were enriched in mild group and they mostly belonged to phylum Actinobacteria and Bacteroidetes.

It was found in our study that microbiome alpha diversity in FE group was significantly decreased at species level, and there was a negative correlation between the frequency of exacerbation and alpha diversity, which is consistent with the finding of Pragman et al. [13, 14], who analyzed the oral wash and sputum samples from COPD patients with 16S rRNA, and found that NFE samples were more diverse than FE samples [14]. Yang et al. [15] applied 16S rRNA to investigate the sputum microbiome of COPD outpatients, and discovered that Chao1 index and observed OTUs in high-risk exacerbator (HRE) group were significantly lower than those in low-risk exacerbator (LRE) group, while Shannon index was similar between the two group, knowing that Shannon index addresses the richness and evenness of the community, while Chao1 and Observed OTUs emphasizes the number of species. The reason for the difference may also come from different grouping criteria. In the study by Yang et al., cases of COPD with <2 moderate exacerbation and no severe exacerbation per year were assigned to LRE



group, and otherwise to HRE group. In addition, different sequencing techniques (16S rRNA vs. mNGS) may also be the possible reason. It was found in our study that  $\alpha$  diversity was similar between mild and severe groups, which is consistent with Yang's finding [16].

Out of top 10 bacteria taxa, Veillonella, Fusobacterium and Prevotella jejuni were found to be more enriched in NFE group compared with FE group. Veillonellae are anaerobic Gram-negative cocci, usually found in the mouth, gastrointestinal tract and the vagina. Veillonella



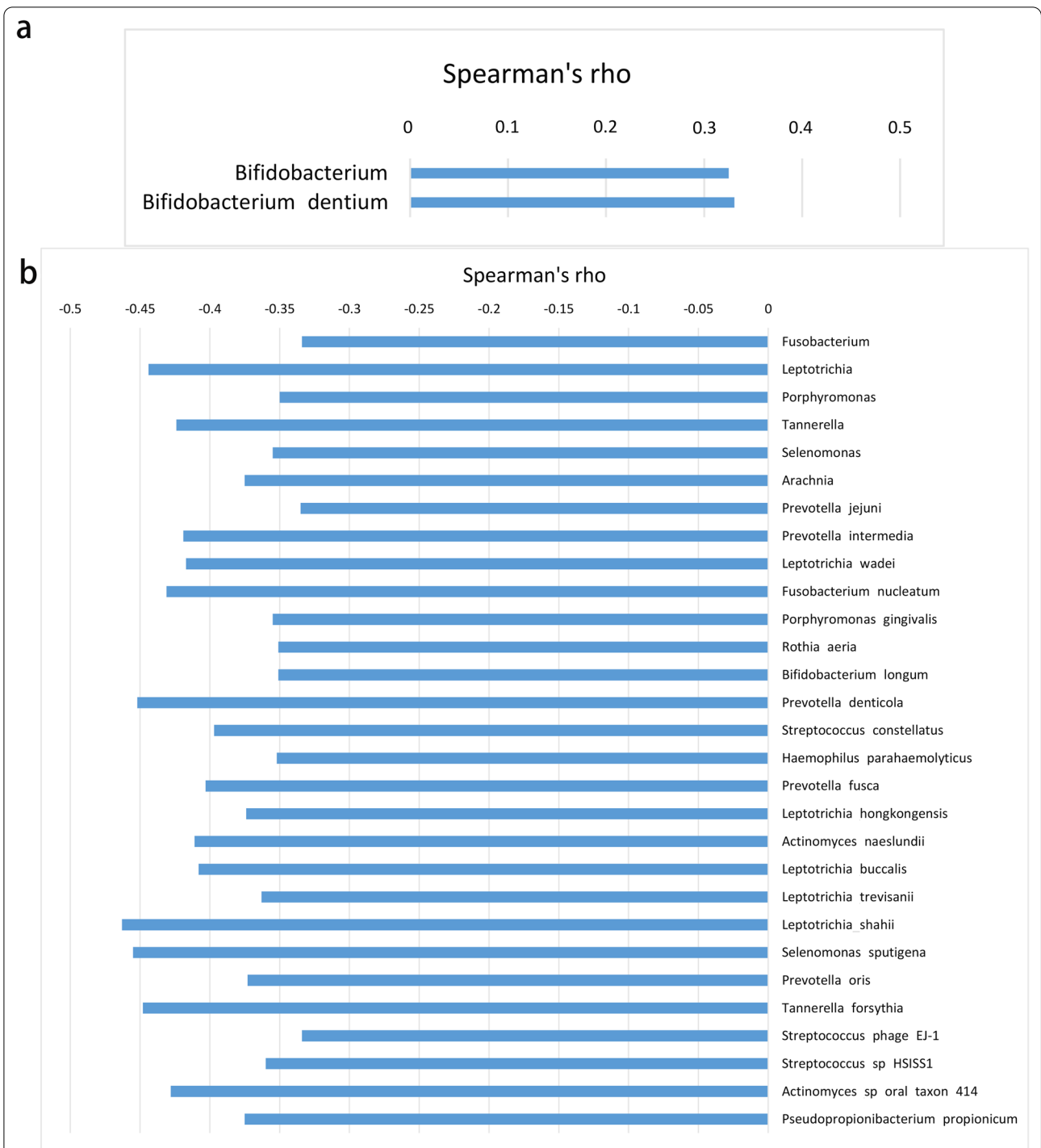
species were found to have a strong relationship with bronchoalveolar lavage (BAL) lymphocyte counts, BAL neutrophil counts and exhaled nitric oxide of early COPD patients [17]. In another study, Veillonella were found strongly associated with a Th17 lung inflammation phenotype in participants without known pulmonary disease [18, 19]. These findings seemed contrary to ours, which may due to different study population. Little was known about the association of Fusobacterium, Prevotella jejuni and COPD yet. Rothia were more abundant in mild group compared with severe group. The genus Rothia are Gram-positive, and found as colonizers of the human oral cavity. In a cohort of adults with bronchiectasis, the abundance of Rothia species was negatively correlated with pro-inflammatory markers (interleukin (IL)-8 and IL-1 $\beta$ ) and matrix metalloproteinase (MMP)-1, MMP-8 and MMP-9 in sputum. Rothia species were also found inhibit NF- $\kappa$ B activation in epithelial cells [20].

Our LDA analysis revealed that many bacterial taxa were more abundant in NFE group, and they mostly

belonged to phylum Actinobacteria, Bacteroidetes and Fusobacteria. Meanwhile, some bacterial taxa were more enriched in mild group, and they mostly belonged to phylum Actinobacteria and Bacteroidetes. Pragman et al. [14] also found that Actinomyces was more common in the sputum of NFE group as compared with FE group. Guo et al. [20] reported that large numbers of bacteria that colonized on the surface of the respiratory mucosa of healthy people were mainly Actinobacteria and Bacteroidete. He et al. [21] found that community-acquired pneumonia presented a depletion of some genera of the phylum Bacteroidetes in bronchoalveolar lavage samples. All these studies agree with our research in some sense. We found that species more enriched in FE or severe group mostly belonged to phylum Proteobacteria, which agrees with the finding of the over-representation of Proteobacteria phylum in severe COPD in other studies [13, 21, 22].

Bifidobacterium and Bifidobacterium dentium showed a positive correlation with  $FEV_1/pre$ . Some studies

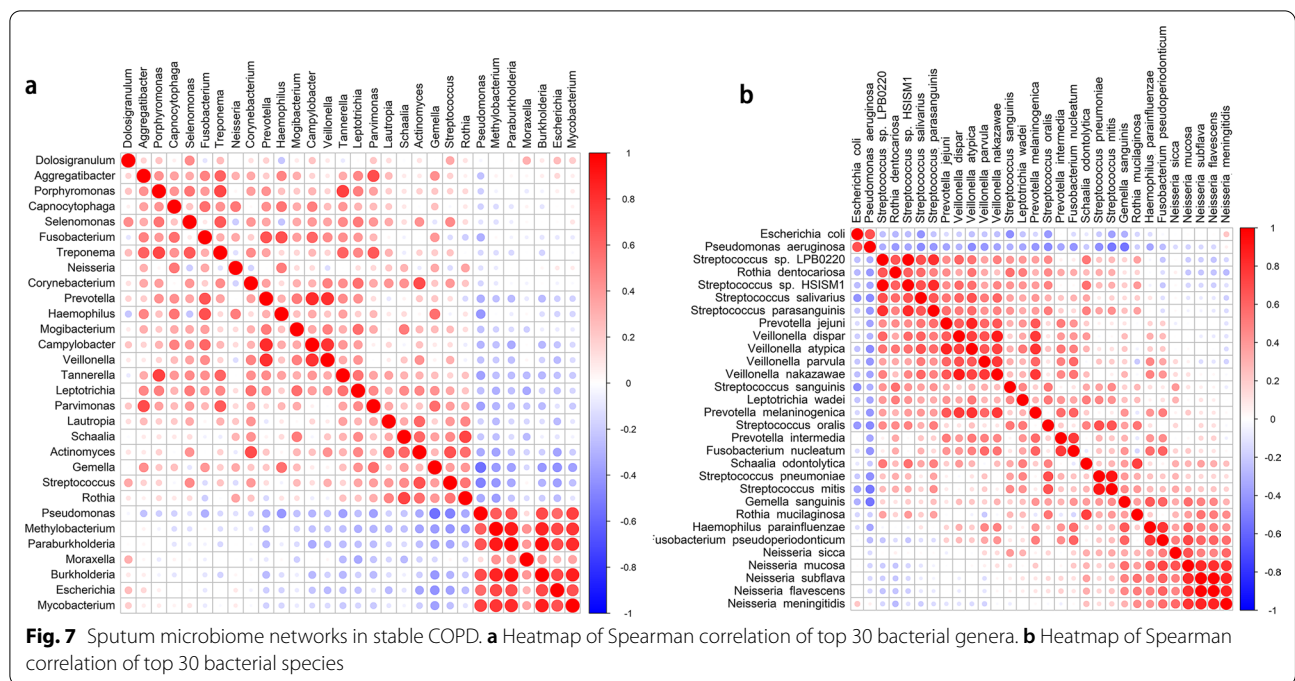




**Fig. 6** Correlation between relative abundance of the microbiome and clinical variables. **a** Bifidobacterium and Bifidobacterium dentium had a positive correlation with FEV<sub>1</sub>/pre ( $r = 0.325$  or  $0.331$ ); **b** A total of 29 bacterial genera and species showed a negative correlation with the frequency of exacerbation ( $r$  ranged from  $-0.463$  to  $-0.334$ )

reported that Bifidobacterium was good to airway inflammation in asthma [23, 24], which seemingly agrees with our finding. In addition, we identified a lot of bacterial genera or species mainly belong to phylum Fusobacteria,

Bacteroidetes and Actinobacteria. They were found to be negatively correlated with the frequency of exacerbation, which is first reported in our study.



To the best of knowledge, this is the first study reporting the use of mNGS to compare the sputum microbiome in COPD of different exacerbation frequencies or lung function. Compared with 16S rRNA, mNGS has the advantages of (1) covering not only bacteria but also eukaryotes and DNA viruses; (2) higher taxonomic resolution at species level instead of genus level; (3) lower potential of bias due to the untargeted nature of the methodology; (4) enabling to discover new or unexpected organism [25]. As the sputum was collected during stable phase of COPD patients, the effect of antibiotics should have been avoided. Nowadays, mNGS may play a role in identifying pathogen in COPD exacerbation. Our research applied mNGS to investigate the correlation of sputum microbiome and disease severity during COPD stable phase, which means in the future mNGS may be a strong tool to predict COPD progress and direct COPD treatment.

There are some limitations in this study. Firstly, the sample size was not large enough, and therefore larger COPD cohort studies are required to verify our findings and conclusions. In addition, our study is a cross-sectional study which only allowed for correlation analysis and did not allow for causal relationship exploration. To form a longitudinal cohort, the patients enrolled in this study will be followed up continuously. Finally, the baseline FEV<sub>1</sub>/pre was not balanced between NFE and FE groups. Although we analyzed differences between groups with different lung functions to minimize the limitation, some bias may not be avoidable.

### Conclusion

Microbiome diversity was lower in frequent COPD exacerbators, but similar between different lung function groups. The microbiome composition changed between different groups as well. It is still necessary to gain a better understanding about the underlying mechanism of the microbiome and exacerbation frequency or lung function for the sake of improving the clinical outcome of COPD treatment.

### Abbreviations

COPD: Chronic obstructive pulmonary disease; mNGS: Metagenomic next-generation sequencing; FE: Frequent exacerbators; NFE: Non-frequent exacerbators; FEV<sub>1</sub>: Forced expiratory volume in 1 s; FEV<sub>1</sub>/pre: FEV<sub>1</sub> percent predicted; n: Number; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; 16S rRNA: 16S ribosomal RNA; mMRC: Modified Medical Research Council dyspnea scale; CAT: COPD Assessment Test; NTC: Non-template controls; PERMANOVA: Permutational multi-variate analysis of variance; PCA: Principal component analysis; PCoA: Principal co-ordinates analysis; LEfSe: Linear discriminant analysis effect size; BMI: Body mass index; FVC: Forced vital capacity; HRE: High-risk exacerbator; LRE: Low-risk exacerbator; OTU: Operational taxonomic units; LDA: Linear discriminant analysis; BAL: Bronchoalveolar lavage.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12931-022-02246-9>.

**Additional file 1.** Supplemental Figures.

### Acknowledgements

The authors thank Wei Liu at Vision Medicals Center for Infection Diseases for guidance with data processing and Shunxing Zhang for English polishing.

### Author contributions

Conception and design, CHW, MT, SSX; Sample acquisition, WL, XLS; Sample processing, WL; Data analysis, WL, MT; Drafting of manuscript, BBW; Data interpretation, manuscript revision, final approval, all. All authors read and approved the final manuscript.

### Funding

This work was supported by Shanghai Science and Technology (18410720900). The funder has no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

Raw sequence data have been deposited in the NCBI "Sequence Read Archive" (SRA) under the project accession PRJNA858967. Other data generated or analysed during this study are included in this published article.

### Declarations

#### Ethics approval and consent to participate

The institutional review board for human studies at Shanghai Tenth People's Hospital approved the protocol (21K85) and written informed consent was obtained from the subjects. All methods were carried out in accordance with the Declaration of Helsinki.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Geriatrics, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai 200072, China. <sup>2</sup>Department of Respiratory Medicine, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai 200072, China.

Received: 10 May 2022 Accepted: 9 November 2022

Published online: 19 November 2022

### References

- The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Accessed 22 April 2022.
- Liu J, Ran Z, Wang F, Xin C, Xiong B, Song Z. Role of pulmonary microorganisms in the development of chronic obstructive pulmonary disease. *Crit Rev Microbiol*. 2021;47:1–12.
- Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax*. 2002;57:847–52.
- Le Rouzic O, Roche N, Cortot AB, Tillie-Leblond I, Masure F, Perez T, Boucot I, Hamouti L, Ostinelli J, Pribil C, et al. Defining the "Frequent Exacerbator" phenotype in COPD: a hypothesis-free approach. *Chest*. 2018;153:1106–15.
- Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2022.
- Amar Y, Lagkouravdos I, Silva RL, Ishola OA, Foessel BU, Kublik S, Schöler A, Niedermeier S, Bleuel R, Zink A, et al. Pre-digest of unprotected DNA by Benzonase improves the representation of living skin bacteria and efficiently depletes host DNA. *Microbiome*. 2021;9:123.
- Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, Stryke D, Pham E, Fung B, Bolosky WJ, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res*. 2019;29:831–42.
- Li H, Gao H, Meng H, Wang Q, Li S, Chen H, Li Y, Wang H. Detection of pulmonary infectious pathogens from lung biopsy tissues by metagenomic next-generation sequencing. *Front Cell Infect Microbiol*. 2018;8:205.
- Xu Y, Kang L, Shen Z, Li X, Wu W, Ma W, Fang C, Yang F, Jiang X, Gong S, et al. Dynamics of severe acute respiratory syndrome coronavirus 2 genome variants in the feces during convalescence. *J Genet Genomics*. 2020;47:610–7.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.
- Galiana A, Aguirre E, Rodriguez JC, Mira A, Santibañez M, Candela I, Llavero J, Garcinuño P, López F, Ruiz M, et al. Sputum microbiota in moderate versus severe patients with COPD. *Eur Respir J*. 2014;43:1787–90.
- Pragman AA, Knutson KA, Gould TJ, Isaacson RE, Reilly CS, Wendt CH. Chronic obstructive pulmonary disease upper airway microbiota alpha diversity is associated with exacerbation phenotype: a case-control observational study. *Respir Res*. 2019;20:114.
- Shannon CE. The mathematical theory of communication 1963. *MD Comput*. 1997;14:306–17.
- Yang CY, Li SW, Chin CY, Hsu CW, Lee CC, Yeh YM, Wu KA. Association of exacerbation phenotype with the sputum microbiome in chronic obstructive pulmonary disease patients during the clinically stable state. *J Transl Med*. 2021;19:121.
- Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, Chen H, Berger KI, Goldring RM, Rom WN, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome*. 2013;1:19.
- Caverly LJ, Huang YJ, Sze MA. Past, present, and future research on the lung microbiome in inflammatory airway disease. *Chest*. 2019;156:376–82.
- Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol*. 2016;1:16031.
- Rigauts C, Aizawa J, Taylor S, Rogers GB, Govaerts M, Cos P, Ostyn L, Sims S, Vandeplassche E, Sze M, et al. *Rothia mucilaginosa* is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. *Eur Respir J* 2021.
- Cameron SJ, Lewis KE, Huws SA, Lin W, Hegarty MJ, Lewis PD, Mur LA, Pachebat JA. Metagenomic sequencing of the chronic obstructive pulmonary disease upper bronchial tract microbiome reveals functional changes associated with disease severity. *PLoS ONE*. 2016;11: e0149095.
- Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS ONE*. 2012;7: e47305.
- Sagar S, Morgan ME, Chen S, Vos AP, Garssen J, van Bergenhenegouwen J, Boon L, Georgiou NA, Kraneveld AD, Folkerts G. Bifidobacterium breve and Lactobacillus rhamnosus treatment is as effective as budesonide at reducing inflammation in a murine model for chronic asthma. *Respir Res*. 2014;15:46.
- van de Pol MA, Lutter R, Smids BS, Weersink EJ, van der Zee JS. Synbiotics reduce allergen-induced T-helper 2 response and improve peak expiratory flow in allergic asthmatics. *Allergy*. 2011;66:39–47.
- Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J Clin Invest*. 2022;132:e154944.

### Publisher's Note

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