

# Promising Intestinal Microbiota Associated with Clinical Characteristics of COPD Through Integrated Bioinformatics Analysis

Tianwen Lai<sup>1,\*</sup>, Chaole Luo<sup>1,\*</sup>, Yalian Yuan<sup>2,\*</sup>, Jia Fang<sup>1,\*</sup>, Yun Wang<sup>1</sup>, Xiantong Tang<sup>1</sup>, Lihuan Ouyang<sup>1</sup>, Keyan Lin<sup>1</sup>, Bin Wu<sup>2</sup>, Weimin Yao<sup>1</sup>, Ruina Huang<sup>1</sup>

<sup>1</sup>Department of Respiratory and Critical Care Medicine, the First Dongguan Affiliated Hospital, Guangdong Medical University, Dongguan, 523121, People's Republic of China; <sup>2</sup>Respiratory Diseases Research Institute, Affiliated Hospital of Guangdong Medical University, Zhanjiang, 524001, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Ruina Huang; Weimin Yao, Email wolidedd@163.com; yaoweimin@gdmu.edu.cn

**Introduction:** Chronic obstructive pulmonary disease (COPD), an incurable chronic respiratory disease, has become a major public health problem. The relationship between the composition of intestinal microbiota and the important clinical factors affecting COPD remains unclear. This study aimed to identify specific intestinal microbiota with high clinical diagnostic value for COPD.

**Methods:** The fecal microbiota of patients with COPD and healthy individuals were analyzed by 16S rDNA sequencing. Random forest classification was performed to analyze the different intestinal microbiota. Spearman correlation was conducted to analyze the correlation between different intestinal microbiota and clinical characteristics. A microbiota-disease network diagram was constructed using the gut MDisorder database to identify the possible pathogenesis of intestinal microorganisms affecting COPD, screen for potential treatment, and guide future research.

**Results:** No significant difference in biodiversity was shown between the two groups but significant differences in microbial community structure. Fifteen genera of bacteria with large abundance differences were identified, including *Bacteroides*, *Prevotella*, *Lachnospira*, and *Parabacteroides*. Among them, the relative abundance of *Lachnospira* and *Coprococcus* was negatively related to the smoking index and positively related to lung function results. By contrast, the relative abundance of *Parabacteroides* was positively correlated with the smoking index and negatively correlated with lung function findings. Random forest classification showed that *Lachnospira* was the genus most capable of distinguishing between patients with COPD and healthy individuals suggesting it may be a potential biomarker of COPD. A *Lachnospira* disease network diagram suggested that *Lachnospira* decreased in some diseases, such as asthma, diabetes mellitus, and coronavirus disease 2019 (COVID-19), and increased in other diseases, such as irritable bowel syndrome, hypertension, and bovine lichen.

**Conclusion:** The dominant intestinal microbiota with significant differences is related to the clinical characteristics of COPD, and the *Lachnospira* has the potential value to identify COPD.

**Keywords:** chronic obstructive pulmonary disease, intestinal microbiota, 16S rDNA sequencing, *Lachnospira*, gut-lung axis

## Introduction

Chronic obstructive pulmonary disease (COPD) is an incurable chronic respiratory disease, characterized by persistent respiratory symptoms and airflow restriction due to airway or alveolar abnormalities.<sup>1</sup> The morbidity and mortality of COPD are rising every year worldwide and the World Health Organization estimates that COPD will be the third leading cause of death in the world by 2030.<sup>2</sup> Although our understanding of COPD is improving, the pathological mechanism of COPD has not been fully elucidated. Existing evidence shows that in addition to smoking, which is the main risk factor for COPD,<sup>3,4</sup> other environmental factors, age, and infection are associated with low lung function.<sup>5-7</sup> In recent years, the fecal microbiome is thought to differ between patients with COPD and healthy individuals,<sup>8</sup> and intestinal microbiota

composition is closely interacted with lung immunity.<sup>9</sup> However, the relationship between the composition of intestinal microbiota and important clinical factors related to COPD remains to be further studied.

The intestinal microbiota is the largest and most complex microecosystem of the human body and participates in many physiological and pathological activities, including an important role in lung function.<sup>10</sup> In recent years, intestinal microbiota imbalance has been recognized as a risk factor for COPD.<sup>11</sup> In COPD patients, an increase in gram-negative bacteria can aggravate the release of intestinal endotoxin, cause indirect lung injury, and accelerate the progress of the disease,<sup>12</sup> whereas intestinal mucosal hypoxia and ischemia caused by metabolic abnormalities in patients with COPD lead to the loss of intestinal integrity and ultimately an imbalance of intestinal microbiota.<sup>13</sup> These studies reveal that COPD and intestinal microbiota bidirectionally influence each other. Moreover, the structure of intestinal microbiota can be changed by COPD risk factors (smoking and gender) in normal/pathological conditions.<sup>14–16</sup> It can be inferred that there must be a causal relationship between intestinal microbiota imbalance and the development of COPD. Herein, understanding the correlation between intestinal microbiota and these clinical features is conducive to the early identification of patients requiring intervention and improving their prognosis. Here, we used high-throughput 16S rDNA sequencing to analyze the intestinal microbiota and correlate these data with the clinical characteristics of patients with COPD to help identify specific intestinal microbiota with high clinical diagnostic value. This may help to develop new and advanced methods to control intestinal microbiota to prevent the deterioration and progression of COPD.

## Methods

### Cohort Design and Sample Collection

This study was carried out in the Department of Respiratory Medicine, Affiliated Hospital of Guangdong Medical University from October 2019 to May 2021. The study was approved by the Ethics Committee of the Affiliated Hospital of Guangdong Medical University (approval No. PJ2015101KT). We recruited 35 healthy control participants and 37 patients with COPD treated in the outpatient department. Patients who met the inclusion criteria were enrolled after signing the informed consent form. All procedures were implemented in accordance with the Declaration of Helsinki. The inclusion criteria were: (1) findings in line with the diagnostic criteria of the Global Initiative on Chronic Obstructive Pulmonary Disease (GOLD; revised in 2019) based on pulmonary function tests (after inhalation of bronchodilators), eg, first forced expiratory volume (FEV<sub>1</sub>)/forced vital capacity (FVC)<0.7 and (2) stable condition. The exclusion criteria were: (1) existence of primary intestinal diseases; (2) pregnancy; (3) presence of active pulmonary tuberculosis and other diseases; (4) use of antibiotics, systemic intravenous hormones, or probiotics in the last 4 weeks; and (5) incomplete data.

The names, gender, age, medical history, and medication history of all participants were collected. Vital capacity was measured after the administration of bronchodilators to assess airway obstruction. We collected blood samples for clinical chemical analysis after fasting for at least 8 h and measured the proportion of white blood cells and neutrophils with an automatic analyzer (Sysmex XN3000). We collected fresh fecal samples (1 g of each sample) from the middle and inside parts of each participant's feces. During the sampling process, we avoided touching the inner surface of the container and ensured to avoid contamination (such as from urine). Fecal samples were stored in a freezer at  $-80^{\circ}\text{C}$  until further processing.

### Specimen Collection and 16S rDNA Sequencing

The total bacterial DNA was isolated from fecal samples using the HiPure stool DNA kit (No. D3141; Guangzhou, China). In brief, 150–200 mg fecal sample was diluted in 1.2 mL Buffer SSL and bathed in  $70^{\circ}\text{C}$  water for 10 minutes. After centrifugation, the supernatant was collected and incubated with 20  $\mu\text{L}$  Proteinase K and 250  $\mu\text{L}$  Buffer AL at  $70^{\circ}\text{C}$  for 10 minutes. DNA of the mixture was isolated by the bead-beating method and stored at  $-20^{\circ}\text{C}$  before 16S rDNA sequencing.

The V3+V4 region of 16S rDNA was amplified with Q5<sup>®</sup> High-Fidelity DNA Polymerase (New England Biolabs, USA) using specific primers with barcodes. 16S rDNA underwent two rounds of amplification procedures. The first round of amplification procedures was initial denaturation at  $95^{\circ}\text{C}$  for 5 minute, followed by 30 cycles at  $95^{\circ}\text{C}$  for 1 minute,  $65^{\circ}\text{C}$  for 1 minute, and  $72^{\circ}\text{C}$  for 1 minute and extension at  $72^{\circ}\text{C}$  for 7 minute. The second round of amplification procedures was the same as in the first round except that the cycle number was changed to 12. The primer sequence was 341F: CCTACGGGNGGCWGCAG, 806R: GGACTACHVGGGTATCTAAT. The PCR amplification

products were cut, recycled, and quantified with the ABI StepOnePlus Real-Time PCR System (Life Technologies). We mixed equal amounts of purified amplification products, connected the sequencing connector, constructed the sequencing library, and sequenced the samples using a HiSeq2500 PE250.

## Analysis of 16sDNA Sequencing Data

Raw reads were filtered using FASTP (version 0.18.0)<sup>17</sup> according to the following rules: reads containing more than 10% of unknown nucleotides (N) and reads containing less than 50% of bases with quality (Q-value) $>20$  were removed. Paired-end clean reads were merged as raw tags using FLASH (version 1.2.11)<sup>18</sup> with a minimum overlap of 10 bp and mismatch error rates of 2%. Noisy sequences of raw tags were filtered under specific filtering conditions (raw tags with the default quality threshold is  $\leq 3$  were broken and tags with continuous high-quality base length are less than 75% were filtered) to obtain the high-quality clean tags. The obtained clean tags clustered the SILVA database (version 138.1) into operational taxonomic units (OTUs) of  $\geq 97\%$  similarity using the UPARSE (version 9.2.64)<sup>19</sup> pipeline. All chimeric tags were removed using the UCHIME algorithm<sup>20</sup> and finally obtained effective tags for further analysis. The tag sequence with the highest abundance was selected as a representative sequence within each cluster. The Mothur software was used to remove redundant tag sequences and select unique tag sequences (each unique sequence represented a group of identical tag sequences, and the number of tags represented by each unique tag was defined as the abundance of this tag). To study the intestinal microbiota composition and diversity of samples, UPARSE software was used to cluster all effective tags of all samples. By default, it clustered the sequences into OTUs with 97% identity and calculated the absolute abundance and relative information of each OTU in each sample. In the process of building OTUs, UPARSE selected representative sequences (the tag sequence with the highest abundance in OTUs) and used the naïve Bayesian assignment algorithm of the RDP Classifier to annotate intestinal microbiota using the Greengenes database with a confidence threshold of 0.8–1).

After obtaining the intestinal microbiota annotation of each OTU, to study the systematic relationship among OTUs, KRONA (version 2.6) was used to visually display the intestinal microbiota annotation results, and the change in intestinal microbiota abundance of different samples at each classification level was displayed in the form of a stacked bar chart. For clarity and readability, we only showed the top 10 intestinal microbiota whose expression abundance reaches 2% in at least one sample. The remaining intestinal microbiota were uniformly classified as “Other”, and the tags that could not be annotated at this level were classified as “Unclassified”. We used QIIME software (version 1.9.1)<sup>21</sup> to calculate the alpha diversity index and draw the corresponding curves. When comparing the two study groups, we used the Wilcoxon rank sum test. We used the R package *vegan* (version 2.5.3) (<https://cran.r-project.org>) to analyze the beta diversity, performed a nonmetric multidimensional scale (NMDS) analysis based on the unweighted\_unifrac distance between samples to show differences in microbial community structure, and used the analysis of similarities (ANOSIM) test to determine significant differences in distances between and within groups. Intestinal microbiota comparison between control and COPD groups was calculated by the Wilcoxon rank test in R Project Vegan package (version 2.5.3) (<https://cran.r-project.org>). Bonferroni correction was used to obtain the q-value of the p value after comparison between the two groups. The significance was set as q-value  $\leq 0.05$ , and the differential intestinal microbiota was screened according to q-value  $\leq 0.05$ .

## Random Forest Analysis of the Different Intestinal Microbiota

Based on the R package *random Forest* (version 4.6.12),<sup>22</sup> 15 differential intestinal microbiota were trained using a random forest classifier. The training and test sets were randomly assigned at 7:3. Cross-validation grid search was used for hyperparameter tuning and a random forest model was built. After parameter debugging, when the number of trees included in the random forest was set to 500 (ntree=500), the error decreased significantly in the early stage with the growth of each tree and stabilized in the later stage, indicating that the ntree setting during model fitting was reasonable. When the number of variables included in each decision tree was 6 (mtry=6), and the accuracy of cross validation was the highest. The confidence interval (CI) of each area under the receiver operating characteristic (ROC) curve (AUC) was calculated using the R package *pROC* (1.10.0)<sup>23</sup> to analyze ROC curves, and *ggplot2* (version 2.2.1)<sup>24</sup> was used to plot

the graphs. Based on the Gini index and mean decrease accuracy, the intestinal microbiota was ranked according to the weight of their contributions to the random forest model.

## Correlation Analysis Between Different Intestinal Microbiota and Clinical Characteristics

The R package *psych* (version 1.8.4) was used for Spearman correlation analyses. The correlation coefficient was calculated based on the relative abundance of intestinal microbiota and clinical characteristics, and correlations were displayed using a heat map.  $p < 0.05$  was considered statistically significant. A correlation coefficient  $r$  of 0–0.2 was considered as uncorrelated or very weak, 0.2–0.4 as weak, 0.4–0.6 as medium, 0.6–0.8 as strong, and 0.8–1 as very strong.

## Disease-Related Network Analysis of Key Intestinal Microbiota

gutMDisorder<sup>25</sup> is a database related to intestinal microorganisms, which provides information on the correlation of intestinal microorganisms with human/murine diseases, drugs, food, and other intervention factors. This is of great convenience for the research on the pathogenesis of host diseases, as well as their diagnosis and treatment, based on the regulation of intestinal microorganisms. We used this database to understand the network of associations between key intestinal microbiota and disease phenotypes, explore the possible pathogenesis of intestinal microbes regarding their impact on COPD, screen potential therapeutic measures, and guide future research.

## Results

### Patients' Baseline Characteristics

37 patients with COPD and 35 healthy control participants were enrolled. Patients were all assessed during periods of stable disease. The smoking status, medication history, CAT scores, vital capacity measurements (FEV1% and FEV1/FVC% predicted value), and blood cell counts were recorded and compared between COPD patients and healthy controls. Compared with healthy control, the proportion of smokers in patients with COPD was significantly higher ( $p < 0.01$ ), the proportions of total leukocytes and neutrophils in patients with COPD were significantly increased ( $p < 0.01$ ) and FEV1% and FEV1/FVC% predicted value were significantly decreased (both  $p < 0.01$ ; Table 1).

**Table 1** Characteristics of the Study Participants

Parameters	Healthy Controls (n=35)	Patients with COPD (n=37)	P-value
Female/Male	17/18	15/22	0.636
Age, years	64.6 ± 1.16	66.4 ± 1.25	0.289
Smoking, pack-years	5.0 ± 1.6	30.1 ± 3.9	<0.01
FEV1, %predicted	94.9 ± 2.38	41.1 ± 2.44	<0.01
FEV1/FVC, %	84.1 ± 1.92	46.4 ± 1.68	<0.01
WBC (×10 <sup>9</sup> )	6.8 ± 0.23	8.5 ± 0.50	<0.01
Neutrophils (%)	59.3 ± 1.39	68.9 ± 2.06	<0.01
CAT scores	NA	15.8 ± 1.0	NA
Treatment regimens, n (%)			
ICS/LABA	NA	21 (56.7)	NA
LAMA/LABA	NA	16 (43.2)	NA
SABA	NA	20 (54.1)	NA
Theophylline	NA	31 (83.7)	NA

**Notes:** Data are presented as mean ± SEM unless stated otherwise.

**Abbreviations:** CAT, COPD assessment test; WBC, white blood cells; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; COPD, chronic obstructive pulmonary disease; ICS, inhaled corticosteroid; LABA, long-acting  $\beta$  agonist; SABA, short-acting  $\beta$  agonist; LAMA, long-acting muscarinic agonist; NA, not applicable.

## 16S rDNA Sequencing Data Quality Assessment

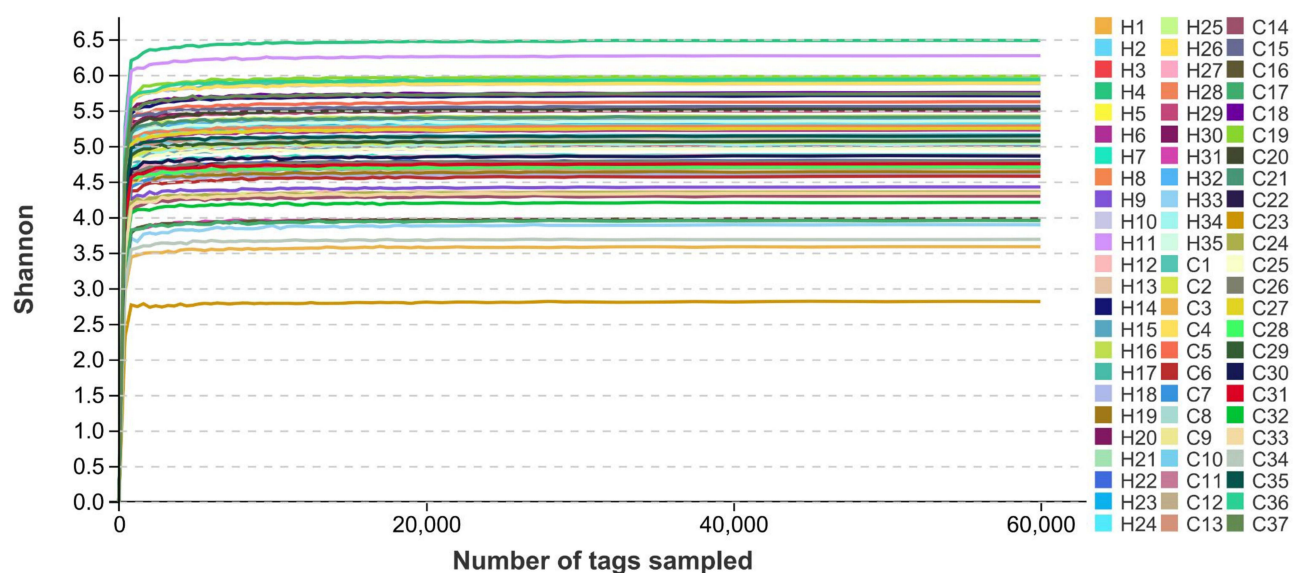
To compare the compositions of intestinal bacterial communities between patients with COPD and healthy individuals, we sequenced 16S rDNA of 72 fecal samples. After sequencing to obtain raw readings, we filtered low-quality readings, merged them, and filtered them to ensure that they were aggregated into OTUs using the most valid data. A total of 7,143,756 available raw tags were obtained from all 72 fecal samples, and the tag sequences were checked for redundancy using the software package Mothur to generate 7,109,296 unique tags. Finally, 3301 OTUs were obtained at a similar level of 97%. To verify whether the sequenced data of this study were large enough to reflect the diversity of the original intestinal microorganisms, the Shannon sparse curve analysis was conducted on the sequencing results. The Shannon sparse curve of all samples showed a plateau, indicating that the 16S rDNA sequence sampling was sufficient and that the samples in this study covered most of the intestinal microbiota information (Figure 1).

## Alpha Diversity Analysis

The alpha diversity of the samples was represented by several indexes in Table 2. Sobs (characterize the actual number of OTU), Chao1 richness estimator, and ACE (an index that estimates the number of OUT in a community) indexes were used to measure the intestinal microbiota diversity of the sample, and the Pielou index was used to assess intestinal microbiota evenness. Simpson index and Shannon index comprehensively reflected the intestinal microbiota diversity and evenness. The results showed no significant difference in Sobs, Chao1, ACE, Shannon, Simpson, and Pielou indexes between the two study groups. We concluded that the two groups were not significantly different regarding intestinal microbiota diversity and evenness between the two groups.

## Beta Diversity Analysis

NMDS analysis is based on the fact that the gut microbial species information is reflected in the multidimensional space in the form of points and the degree of difference between different samples is reflected by the distance between points. Stress value (range from 0 to 1) is used to represent NMDS and to indicate a better representation of the true distances at lower values. We used NMDS to test the clustering model of sample OTUs of different samples according to the unweighted\_unifrac distance metric. The results showed a difference in the overall community composition between COPD and healthy intestinal microbiota (Figure 2A), and the NMDS intensity index stress (stress value=0.158) was able to explain the significance of the model. The differences between the two groups were compared using the ANOSIM test based on the unweighted\_unifrac distance. The ANOSIM results illustrated that the difference between the two groups



**Figure 1** Shannon sparse curve of each sample OTU operational taxon unit.

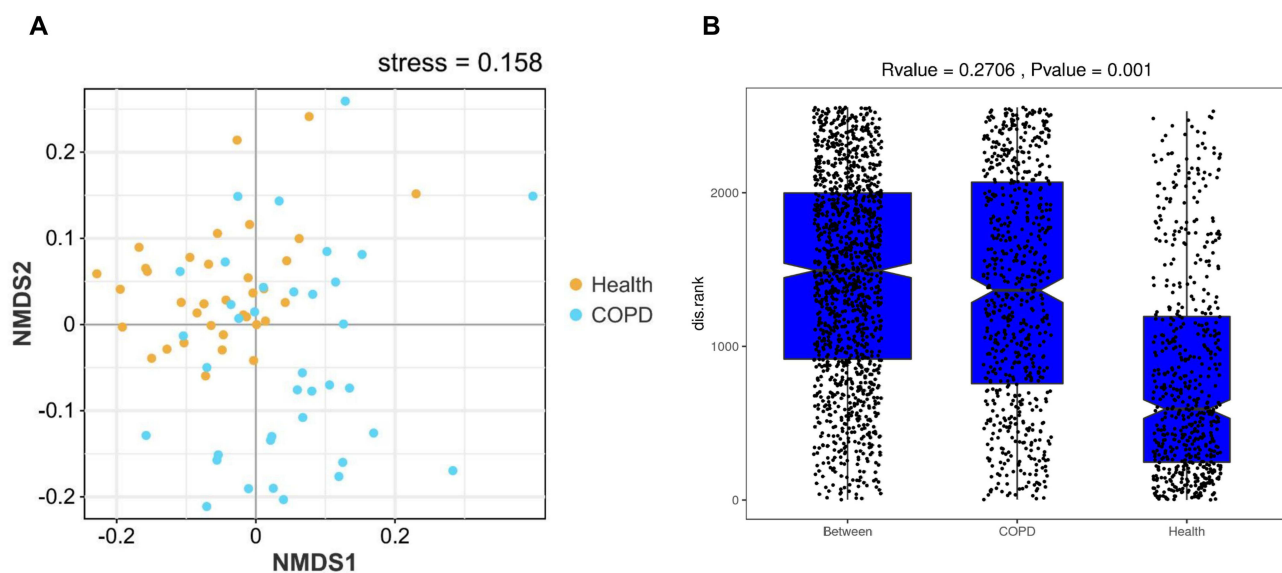
**Table 2** Description of methods and Results of Microbial Diversity and Evenness Assessment

	Sobs	Shannon	Simpson	Chao I	ACE	Pielou
COPD	614.4864865	4.876082167	0.900774647	892.8842121	871.4961703	0.526382406
Healthy control	603.8571429	5.137543605	0.922601047	876.6779727	870.4104571	0.556407099
P-value (Wilcoxon)	0.946097	0.125986	0.156596	0.796765	0.875444	0.093333

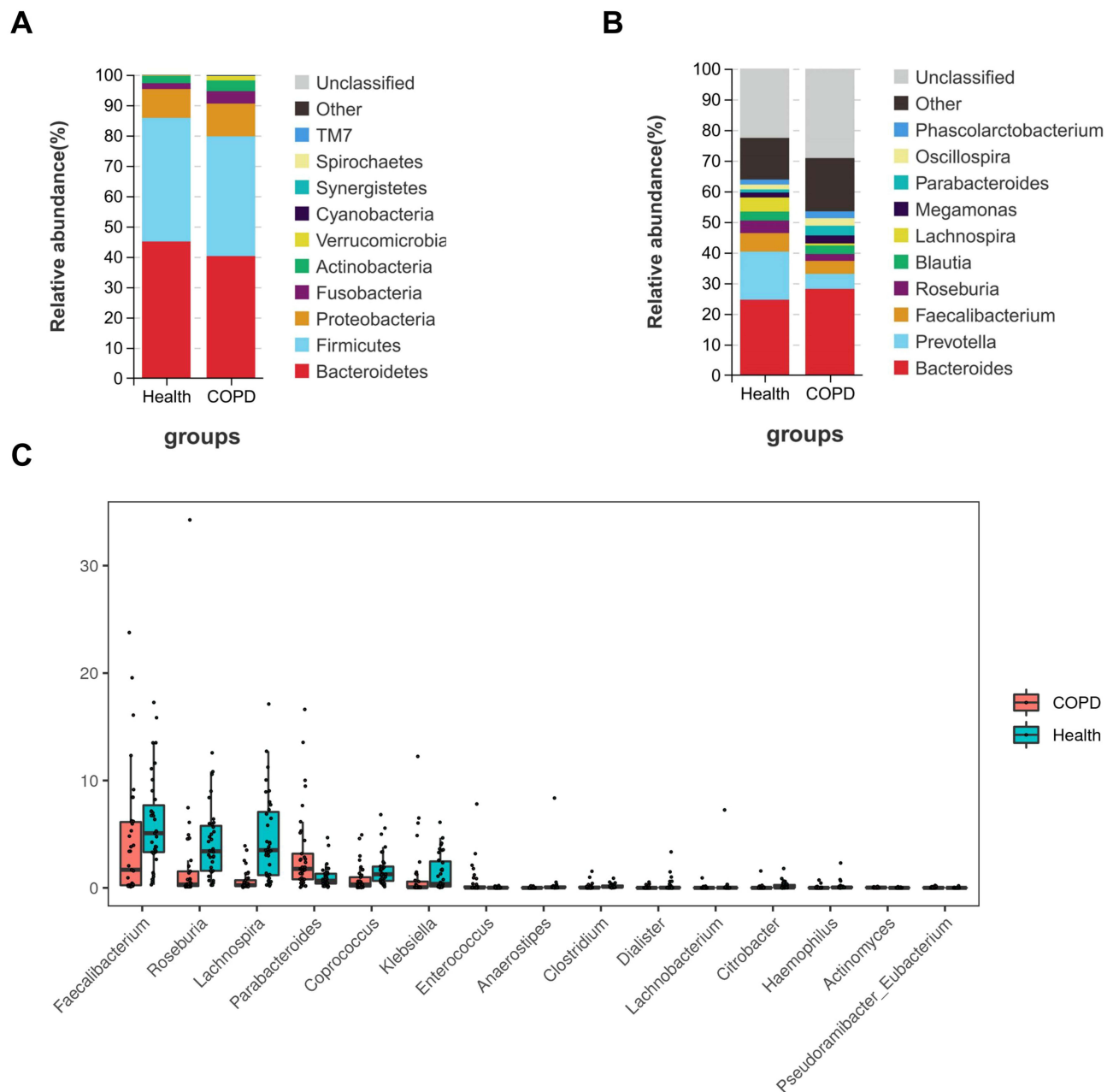
was greater than that within each group, and the difference between the two groups was significant ( $r=0.2706$ ,  $P=0.001$ ; [Figure 2B](#)). These findings suggested that the bacterial communities differed from each other.

## Differences in Intestinal Microbiota

The distributions of the bacterial populations were analyzed using a stacked bar chart. The microbial communities of the two groups of samples at different classification levels showed some changes. At the phylum level, the dominant bacterial communities of the two groups were Bacteroidetes, Firmicutes, Proteobacteria, Actinobacteria, and Fusobacteria ([Figure 3A](#)). For intestinal microbiota with highly abundant OTUs in the top 1000 and a relative abundance (number of intestinal microbiota tags in the sample/total number of tags in the sample) of at least one sample reaching more than 0.1% in the group, we used the R package *vegan* to carry out Wilcoxon testing. P-values for comparison were corrected using Bonferroni to obtain q-values, and the significance level was set to  $q\text{-value} \leq 0.05$  to compare the differences between the two groups of samples. The results showed no significant difference in the mean OTU abundance of these dominant bacterial communities at the phylum level ([Supplementary Table 1](#)). The distribution of dominant bacterial communities at the genus level of the two groups was listed in [Figure 3B](#) and included *Bacteroides*, *Prevotella*, *Faecalibacterium*, *Roseburia*, *Lachnospira*, *Parabacteroides*, etc. By Wilcoxon testing, we found 15 bacterial communities at the genus level with significant mean abundance changes in the COPD group compared with the healthy controls, among which *Faecalibacterium*, *Roseburia*, *Lachnospira*, and *Coprococcus*, etc. were significantly reduced but *Parabacteroides*, etc. were significantly increased ([Figure 3C](#), [Supplementary Table 2](#)). Given that gender is a factor affecting gut microbiota, we analyzed the difference in gut microbiota between males and females in COPD populations. The results showed no significant differences in intestinal microbiota between men and women ([Supplementary Tables 3 and 4](#)).



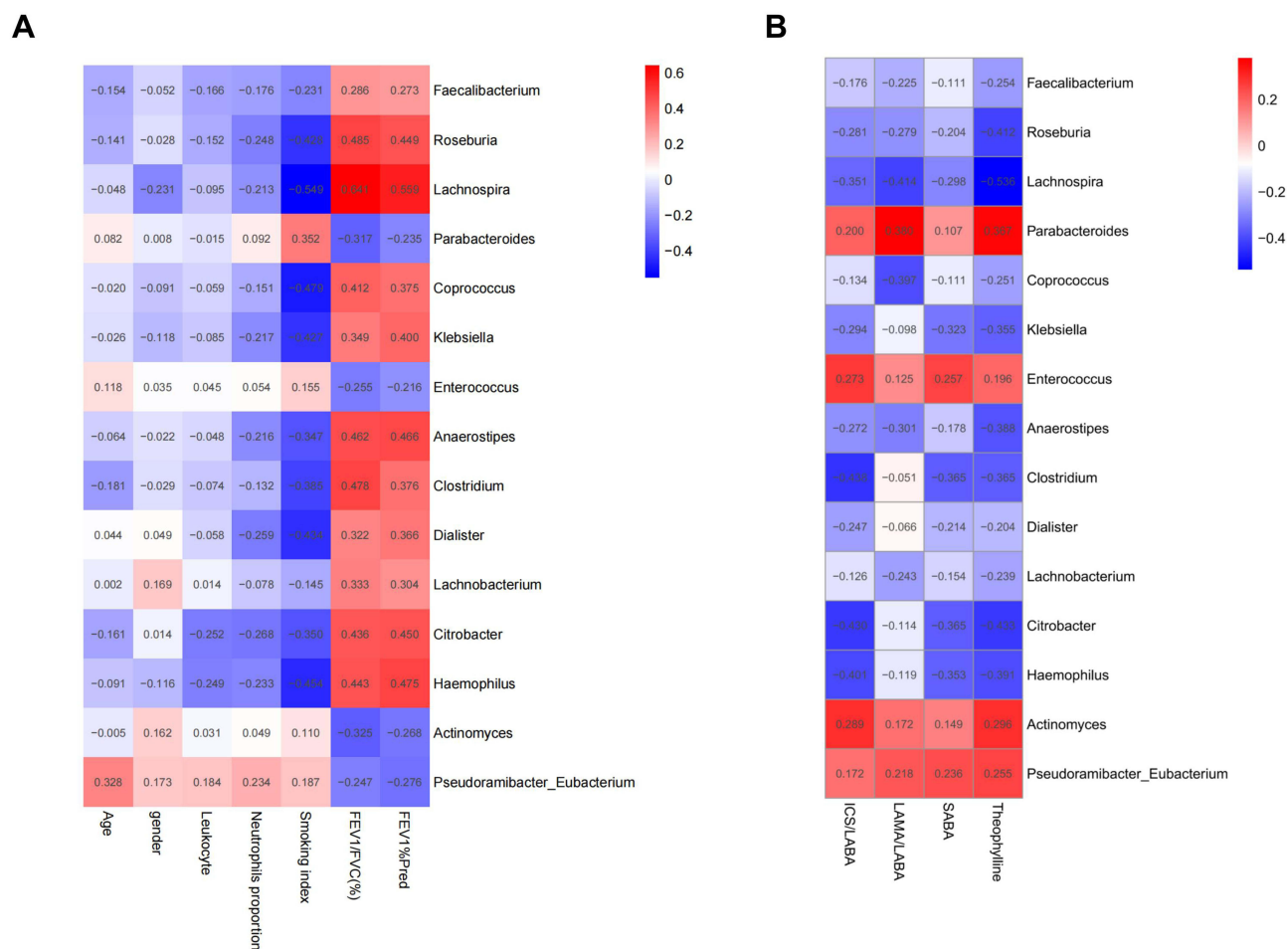
**Figure 2** Beta diversity analysis. **(A)** NMDS of the unweighted\_unifrac distance metric. **(B)** ANOSIM of the unweighted\_unifrac distance. Between represents the distribution of two-sample distances between health and COPD. Each scatter point represents the distance between two samples.



**Figure 3** Differences in intestinal microbiota. **(A)** A stacked bar chart showed the distribution of dominant intestinal microbiota at the phylum level. **(B)** A stacked bar chart showed the distribution of dominant intestinal microbiota at the genus level. **(C)** Wilcoxon test results showed the relative OTU abundance of 15 genus intestinal microbiotas in COPD patients was significantly different from healthy controls. Each scatter represents the relative abundance of this bacterium in each sample.

## Associations Between Differential Intestinal Microbiota and Clinical Characteristics

To determine the relationship between the two groups of different intestinal microbiota and clinical characteristics, we conducted Spearman correlation analyses. As shown in [Figure 4A](#), the relative abundance of several kinds of intestinal microbiota (*Roseburia*, *Lachnospira*, *Coprococcus*, *Klebsiella*, *Anaerostipes*, *Clostridium*, *Dialister*, *Citrobacter*, and *Haemophilus*) were negatively correlated with the smoking index and positively correlated with FEV1%predicted and FEV1/FVC (%). The relative abundance of *Parabacteroides* was positively correlated with the smoking index (*Parabacteroides*  $r=0.352$ ,  $p<0.01$ ) and negatively correlated with FEV1%predicted (*Parabacteroides*  $r=-0.235$ ,  $p<0.05$ ) and FEV1/FVC (%) (*Parabacteroides*  $r=-0.317$ ,  $p<0.01$ ). The relative abundance of *Pseudoramibacter\_Eubacterium* was negatively correlated with several base characteristics (age, Leukocyte, Neutrophils proportion, and smoking) of COPD



**Figure 4** Heat map of Spearman correlation analysis regarding differential intestinal microbiota and clinical characteristics (A) and treatment regimens (B). The correlation index in COPD or health samples is indicated by the color bar on the right side of the heat map.

patients. No significant correlation was observed between gender and different intestinal microbiota ( $p > 0.05$ ). We also investigated the correlation between different intestinal microbiota and treatment regimens and found a positive or negative correlation between them. Of particular note was that *Lachnospira* and *Coprococcus* were negatively correlated with the LAMA/LABA (*Lachnospira*  $r = -0.414$ ,  $p < 0.01$ ; *Coprococcus*  $r = -0.397$ ,  $p < 0.01$ ), and negatively correlated with the Theophylline (*Lachnospira*  $r = -0.536$ ,  $p < 0.01$ ; *Coprococcus*  $r = -0.251$ ,  $p < 0.05$ ) (Figure 4B). *Parabacteroides* positively correlated with LAMA/LABA and Theophylline ( $r = 0.380$ ,  $p < 0.01$ ;  $r = 0.367$ ,  $p < 0.01$ ) (Figure 4B). *Actinomyces* and *Pseudoramibacter\_Eubacterium* positively correlated with four COPD indexes (ICS/LABA, LAMA/LABA, SABA and Theophylline) (Figure 4B). The p value of correlation analysis was shown in the [Supplementary Tables 5 and 6](#).

## Identification of Intestinal Microbiota in Distinguishing the COPD Group from the Control Group

To determine whether 15 differentiated intestinal microbiota compositions can distinguish between patients with COPD and healthy participants and to identify key intestinal microbiota with the ability to predict COPD, we used the machine learning random forest classifier to analyze the intestinal microbiome in relation to clinical characteristics of the two groups, as well as ROC curve analyses to evaluate the diagnostic value of candidate bacteria in distinguishing between healthy individuals and patients with COPD. After parameter debugging, when the number of decision trees included in the random forest was set to 500, with the growth of each tree, the error decreased significantly in the early stage and tended to be stable in the later stage, suggesting that the setting of tree in the model fitting process is reasonable



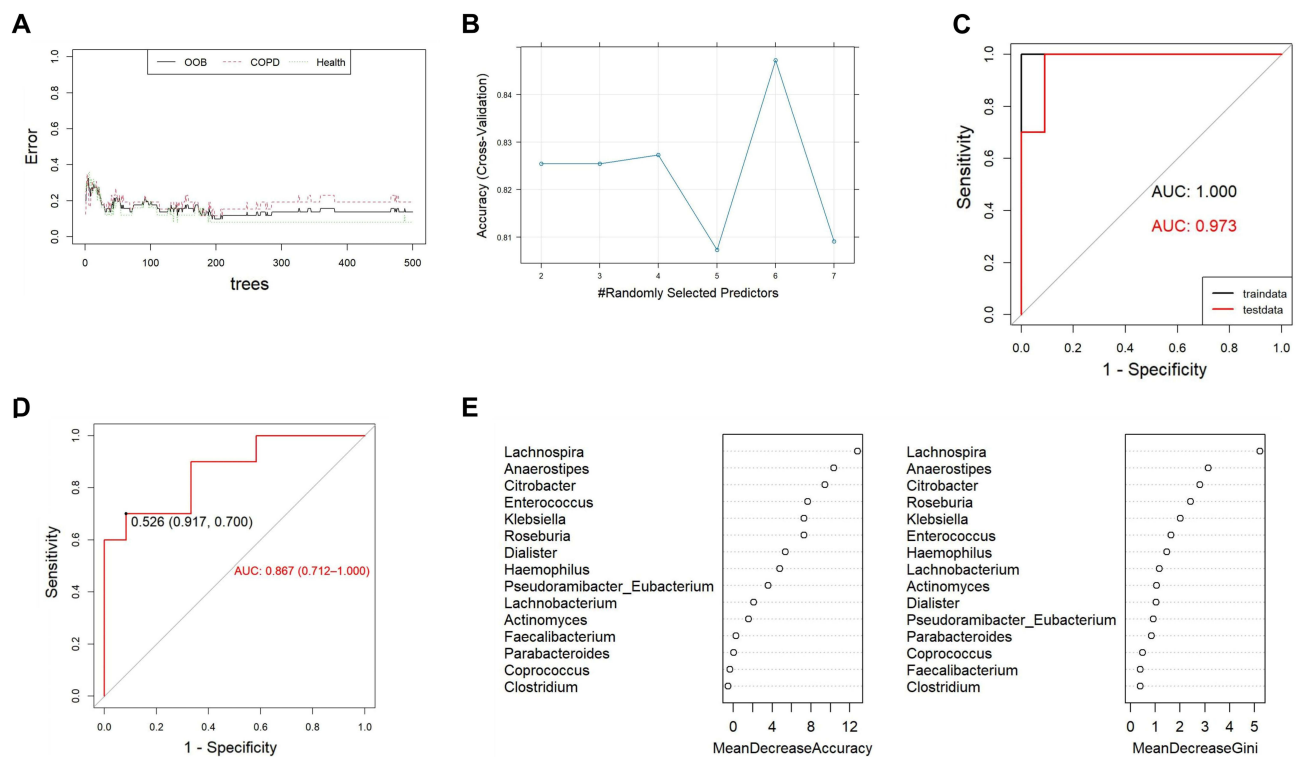
(Figure 5A). When the number of variables included in each tree was equal to 5, the accuracy of cross-validation was the highest (Figure 5B). The AUC of ROC curve evaluation in the training set and testing set were 1.000 and 0.973, respectively (Figure 5C). The AUC of the testing set was not significantly reduced relative to the training set suggesting no overfitting was considered. The ROC curve evaluation AUC of the random forest model was 0.867 suggesting good prediction efficiency (Figure 5D). The mean decrease accuracy plot showed that among those 15 differentiated intestinal microbiotas, *Lachnospira* has the most important ability to distinguish COPD and the mean decrease in Gini coefficient showed that *Lachnospira* contributed the most to this random forest (Figure 5E).

## Disease-Related Network of Key Intestinal Microbiota Based on the gutMDisorder Database

The microbial disease network diagram based on the gutMDisorder database showed the association between *Lachnospira* and various diseases. *Lachnospira* decreases in some diseases, such as asthma, diabetes mellitus, and coronavirus disease 2019 (COVID-19), and increases in other diseases, such as irritable bowel syndrome, hypertension, and bovine lichen. *Lachnospira* changes in anorexia nervosa and ulcerative colitis are controversial; however, no information on the relationship between *Lachnospira* and COPD was found (Figure 6A and B). Plantain seed, exercise, dietary fiber, and other measures can increase the abundance of *Lachnospira*, while cefprozil, short-chain galactooligosaccharides, obesity, proton pump inhibitors, and other measures can reduce the abundance of *Lachnospira* (Figure 6C and D).

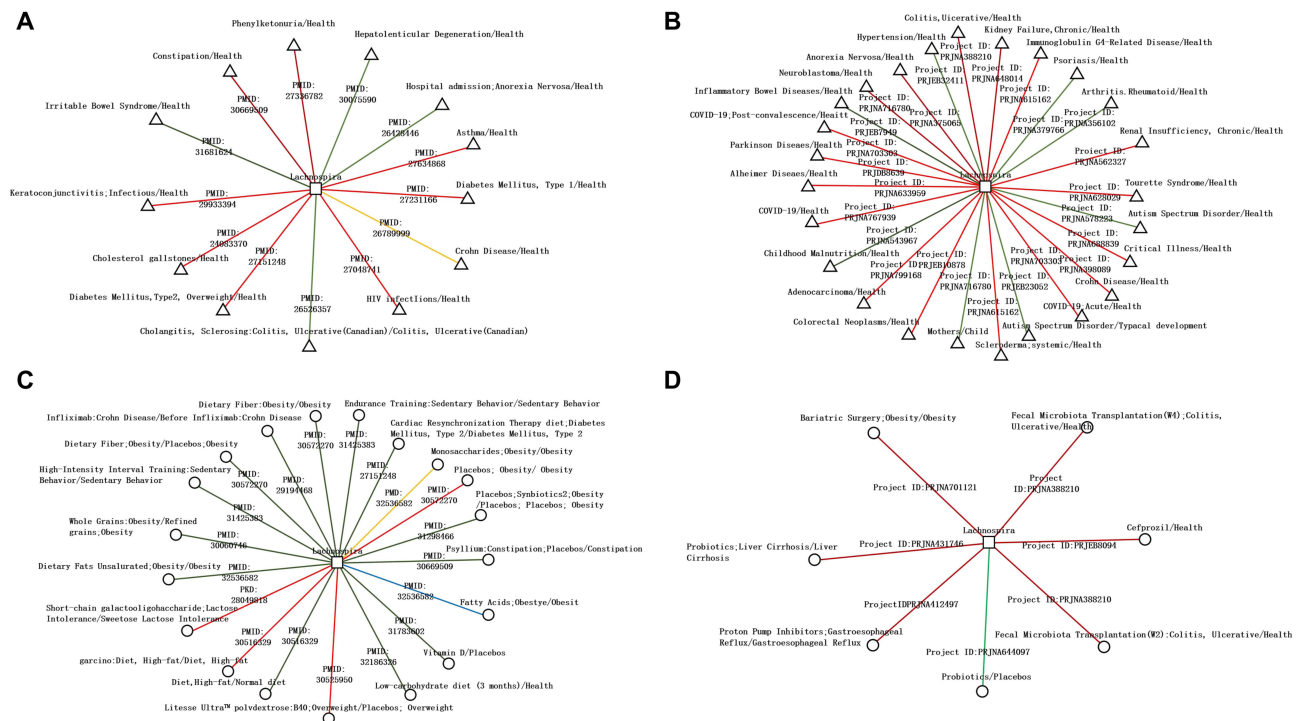
## Discussion

The regulation of lung immune responses by the intestinal microbiota is called the “gut-lung axis”, which may be involved in the pathogenesis of COPD<sup>26</sup> and contributes to the frequency of acute COPD attacks.<sup>8</sup> However, the gut



**Figure 5** The ability of intestinal microbiota to identify patients with COPD. **(A)** The illustration of the error changed with the number of trees. **(B)** Illustration of the tuning of the mtry hyperparameters. **(C)** ROC curve of the training set and testing set of the Random Forest Model. **(D)** ROC curve of the Random Forest Model of differential intestinal microbiota. **(E)** The mean decreases in the Gini coefficient and accuracy plot.

**Abbreviations:** AUC, the area under the curve; ROC, receiver operating curve.



**Figure 6** Disease-related network of key intestinal microbiota. **(A and B)** Diseases types associated with *Lachnospira* based on Literature-based associations **(A)** and Raw data-based associations **(B)**. **(C and D)** Interventions change the abundance of *Lachnospira* based on Literature-based associations **(C)** and Raw data-based associations **(D)**. **Notes:** A circle node represents an intervention-related comparison, a square a gut microbe, and a triangle a phenotype-related comparison. A red link indicates an increase in gut microbes, a green link indicates a decrease in gut microbes, a blue link indicates the presence of a gut microbe, and an orange link indicates the absence of a gut Supplementary microbe.

microbial populations associated with the pathogenesis of COPD and their pathways of action are not well understood. In the present study, we found that the fifteen genera intestinal microbiota of intestinal microbiota in COPD patients were significantly different from healthy controls and showed negative and positive correlations with COPD clinical characteristics indicating the potential importance of intestinal microbiota in COPD pathology. Consistent with a previous study,<sup>8</sup> our findings suggest that the difference in intestinal microbiota between COPD patients and healthy individuals is mainly reflected in the dominant microbiota.

The dominant population in the intestinal microbiota of COPD patients and healthy people was very similar at both phylum and genus levels. But at the genus level, compared with healthy people, the abundance of dominant intestinal microbiota in COPD patients was increased or decreased (*Bacteroides*, *Prevotella*, *Faecalibacterium*, *Roseburia*, *Lachnospira*, *Parabacteroides*, etc. were reduced, and *Parabacteroides*, etc. were increased). The phenomenon of trade-off at the genus level makes the difference between intestinal microbiota unable to be reflected at the phyla level (at the phylum level, no significant difference was in the relative abundance of dominant bacteria.). COPD-related abundance changes of the genus of intestinal microbiota in our study are both consistent and inconsistent between our study and other studies. For example, in a study of 16S rRNA gene sequencing of COPD patients' faeces,<sup>8</sup> similarly, they also found that *Roseburia* and *Lachnospira* reduced in abundance in COPD; while the genera species with increased abundance identified by them were not found in our study. Chiu et al found that Bacteroidetes were more abundant in controls and Firmicutes were more abundant in COPD patients with reduced lung function.<sup>27</sup> Although we observed changes in the relative abundance of *Bacteroidetes* and *Firmicutes* in COPD patients (both with mild decreases), no significant change was observed between-group comparisons. The differences between the results of these studies and our results may be related to the different parameters of included COPD patients such as disease status, dietary factors, medication use, and geographical factors of the included patients. Therefore, the specific intestinal microbiota associated with COPD remains to be discovered.

Interestingly, we found that the differential gut microbiota between COPD and healthy controls had significant positive and negative correlations with risk factors in COPD patients. Age and gender are risk factors for COPD, and the gut microbiota is different across age or gender.<sup>28,29</sup> However, our correlation analysis showed that age and gender were not significantly correlated with gut microbiota, suggesting that the effects of age and gender on COPD progression were independent of gut microbiota. We focused on the relationship between smoking and gut microbiota. So far, whether smoking affects the progress of COPD through an influence on intestinal microbiota remains controversial. In a study with healthy subjects as the study target, smokers had changes in the gut microbiota compared with nonsmokers, predictive of changes in the gut microbiota with smoking status.<sup>30</sup> The research of Bowerman et al showed that no significant changes in intestinal microbiota composition were observed in COPD patients with and without smoking.<sup>8</sup> A previous study observed that smoking cessation altered the composition of the microbiota, such as increased abundance of *Actinomycetia* and *Chlamydia* and decreased abundance of *Bacteroidetes* and *Proteobacteria*.<sup>31</sup> We, therefore, analyzed the correlation between smoking and the abundance of different intestinal microbiota to explore the specific intestinal microbiota affected by smoking, and several kinds of intestinal microbiota (*Rosebunia*, *Lachnospira*, *Coprococcus*, *Klebsiella*, *Anserostipesc*, etc.) were found to be negatively correlated with the smoking index. We paid particular attention to *Lachnospira*. In a study in mice,<sup>32</sup> smoke exposure induced dramatic changes in bacterial activity and microbiota composition, in particular, a significant increase in *Lachnospiraceae* activity was observed. In a cigarette smoking-induced COPD mouse model, an increased abundance of *Lachnospiraceae* was positively correlated with the severity of COPD.<sup>33</sup> Contrary to animal experiments, the abundance of *Lachnospiraceae* seems to be reduced in COPD patients and smokers.<sup>8,34</sup> Consistent with these human experiments, we found that *Lachnospira* was negatively correlated with the smoking index and moderately or positively correlated with pulmonary function indexes in COPD patients. Up to now, the correlation network among smoking, *Lachnospira*, and COPD has not been identified and needs further research. In addition to gut microbiota, a previous study found that oral and nasal microbiota, such as *Actinomyces*, *Megasphaera* and *Corynebacterium*, etc. were significantly elevated by smoking in COPD subjects.<sup>16</sup> Overall, as the most important risk factor for COPD, smoking affects the activity of microbiota, but the elaboration of the relationship between this effect and the development of COPD needs more research.

Machine learning is often used as a data analysis method to evaluate the correlation between biological parameters, especially random forest models, which are used to predict target variables according to various explanatory variables.<sup>35</sup> In recent years, some cohort studies have used random forest classifications to illustrate the predictive contribution and discriminatory ability of explanatory variables (such as microbial metabolites) to target variables (such as disease-related factors).<sup>36,37</sup> Based on random forest machine learning, we constructed a model to distinguish COPD patients from healthy individuals by combining data from fifteen intestinal microbiotas at the genus level. According to the ROC results, our classification model achieved good results in discriminating patients with COPD from healthy participants, suggesting that intestinal microbiota can effectively distinguish COPD patients from the general population. Based on its model contribution, we found that *Lachnospira* played the most important role in identifying patients with COPD. Considering the close correlation between *Lachnospira* and the clinical phenotype of COPD, we believe that *Lachnospira* has a strong diagnostic potential in identifying COPD.

We obtained the association network between *Lachnospira* and diseases from a gut MDisorder database.<sup>25</sup> *Lachnospira* belongs to the family *Lachnospiraceae* and is considered to be involved in a variety of inflammatory diseases, such as ulcerative colitis,<sup>38</sup> irritable bowel syndrome,<sup>39</sup> and asthma.<sup>40</sup> Like other intestinal microbiota, *Lachnospira* actively participates in the metabolism of a variety of carbohydrates and produces short-chain fatty acids such as acetic acid and butyric acid from carbohydrate fermentation,<sup>41–44</sup> which affects the activity and function of immune cells. Several studies have demonstrated that butyrate affects immune cell function. For example, Yip et al elaborated on the extensive effects of butyrate on immune cells in allergic asthma.<sup>43</sup> In COPD, butyrate was proven to mediate pulmonary inflammation via the gut-lung axis.<sup>45</sup> This evidence indicates that *Lachnospira* and its metabolic short-chain fatty acids are closely related to immune regulation. Although we did not find an association between *Lachnospira* and COPD based on the database gutMDisorder analysis, our results showed that *Lachnospira* was associated with clinical features of COPD, and its related information had a good ability to identify COPD patients. The reason for this discrepancy does not exclude the fact that the failure of the gutMDisorder database to fully cover

recent advances. In any case, there is a potential link between *Lachnospira* and its metabolism of short-chain fatty acids and the development of COPD. Our findings provide new evidence for the role of the gut-lung axis in COPD, which needs to be verified by further studies.

This study has several limitations. First, the number of participants was small, so further research with a larger sample size is needed to confirm our results. Second, because most patients in the COPD group were smokers, we cannot confirm differences in intestinal microbiota between smokers and non-smokers among COPD patients, so the causal relationship between the three factors remains to be further studied. Finally, since this is not an intervention study, we cannot fully understand some factors that may affect intestinal microbiota and COPD disease status. Therefore, it is necessary to conduct further intervention studies to determine the relationship between COPD and the intestinal microbiome. In conclusion, more research is needed to clarify the causal relationship between gut microbiota and COPD.

## Data Sharing Statement

The raw data can be gained from the corresponding author.

## Ethics Statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the Affiliated Hospital of Guangdong Medical University, case number PJ2015101KT. All participants provided their written informed consent to participate in this study.

## Author Contributions

Tianwen Lai, Ruina Huang, Lihuan Ouyang, Bin Wu, and Yalian Yuan were involved in the conception and design of the study. Jia Fang, Yun Wang, Lihuan Ouyang, Xiantong Tang, Keyan Lin, and Chaole Luo were responsible for data acquisition. Ruina Huang was responsible for data analysis and manuscript writing. Weimin Yao participated study design and also provided effective scientific suggestions for this study. Tianwen Lai provided science supervise and critically revise manuscripts. All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

## Funding

This research was supported by the National Natural Science Foundation of China (82170030, 82370038), the Young Innovative Talents Project of Guangdong Normal Universities (No. 2018KQNCX099).

## Disclosure

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Ferri S, Paoletti G, Pelaia C, Heffler E, Canonica GW, Puggioni F. COPD and biologic treatment: state of the art. *Curr Opin Allergy Clin Immunol*. 2023;23(4):309–318. doi:10.1097/ACI.0000000000000920
2. Pollak J, van Agteren JE, Esterman AJ, Carson-Chahhoud KV. Psychological therapies for the treatment of depression in chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2019;3(3):CD012347. doi:10.1002/14651858.CD012347.pub2
3. Rao W, Wang S, Duleba M, et al. Regenerative Metaplastic Clones in COPD Lung Drive Inflammation and Fibrosis. *Cell*. 2020;181(4):848–864. e818. doi:10.1016/j.cell.2020.03.047
4. Wang C, Zhou J, Wang J, et al. Progress in the mechanism and targeted drug therapy for COPD. *Signal Transduction and Targeted Therapy*. 2020;5(1):546.
5. Alter P, Kahnert K, Trudzinski FC, et al. Disease Progression and Age as Factors Underlying Multimorbidity in Patients with COPD: results from COSYCONET. *Int J Chron Obstruct Pulmon Dis*. 2022;17:1703–1713. doi:10.2147/COPD.S364812
6. Linden D, Guo-Parke H, Coyle PV, et al. Respiratory viral infection: a potential “missing link” in the pathogenesis of COPD. *Eur Respir Rev*. 2019;28(151):180063. doi:10.1183/16000617.0063-2018
7. Agusti A, Vogelmeier C, Faner R. COPD 2020: changes and challenges. *Am J Physiol Lung Cell Mol Physiol*. 2020;319(5):L879–L883. doi:10.1152/ajplung.00429.2020

8. Bowerman KL, Rehman SF, Vaughan A, et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nat Commun.* 2020;11(1):5886. doi:10.1038/s41467-020-19701-0
9. Yadava K, Pattaroni C, Sichelstiel AK, et al. Microbiota Promotes Chronic Pulmonary Inflammation by Enhancing IL-17A and Autoantibodies. *Am J Respir Crit Care Med.* 2016;193(9):975–987. doi:10.1164/rccm.201504-0779OC
10. Espirito Santo C, Caseiro C, Martins MJ, Monteiro R, Brandao I. Gut Microbiota, in the Halfway between Nutrition and Lung Function. *Nutrients.* 2021;13(5):1716. doi:10.3390/nu13051716
11. Zeng X, Yang H, Yang Y, Gu X, Ma X, Zhu T. Associations of Clinical Characteristics and Intestinal Flora Imbalance in Stable Chronic Obstructive Pulmonary Disease (COPD) Patients and the Construction of an Early Warning Model. *Int J Chron Obstruct Pulmon Dis.* 2021;16:3417–3428. doi:10.2147/COPD.S330976
12. Mayhew D, Devos N, Lambert C, et al. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. *Thorax.* 2018;73(5):422–430. doi:10.1136/thoraxjnl-2017-210408
13. Sun Z, Zhu QL, Shen Y, Yan T, Zhou X. Dynamic changes of gut and lung microorganisms during chronic obstructive pulmonary disease exacerbations. *Kaohsiung J Med Sci.* 2020;36(2):107–113. doi:10.1002/kjm2.12147
14. Silveyra P, Fuentes N, Rodriguez Bauza DE. Sex and Gender Differences in Lung Disease. *Adv Exp Med Biol.* 2021;1304:227–258. doi:10.1007/978-3-030-68748-9\_14
15. Koliada A, Moseiko V, Romanenko M, et al. Sex differences in the phylum-level human gut microbiota composition. *BMC Microbiol.* 2021;21(1):131. doi:10.1186/s12866-021-02198-y
16. Agarwal DM, Dhotre DP, Kumbhare SV, et al. Disruptions in oral and nasal microbiota in biomass and tobacco smoke associated chronic obstructive pulmonary disease. *Arch Microbiol.* 2021;203(5):2087–2099. doi:10.1007/s00203-020-02155-9
17. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–i890. doi:10.1093/bioinformatics/bty560
18. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27(21):2957–2963. doi:10.1093/bioinformatics/btr507
19. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10(10):996–998. doi:10.1038/nmeth.2604
20. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011;27(16):2194–2200. doi:10.1093/bioinformatics/btr381
21. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335–336.
22. Liaw A, M W. Classification and Regression by randomForest. *R News.* 2002;23(23):18–22.
23. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *Int J Med.* 2011;12(1):77. doi:10.1186/1471-2105-12-77
24. Wickham H. ggplot2. *ggplo 2.* 2011;3(2):180–185.
25. Cheng L, Qi C, Zhuang H, Fu T, Zhang X. gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* 2020;48(D1):D554–D560. doi:10.1093/nar/gkz843
26. Zhang D, Li S, Wang N, Tan H-Y, Zhang Z, Feng Y. The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases. *Front Microbiol.* 2020;11.
27. Chiu YC, Lee SW, Liu CW, Lan TY, Wu LS. Relationship between gut microbiota and lung function decline in patients with chronic obstructive pulmonary disease: a 1-year follow-up study. *Respir Res.* 2022;23(1):10. doi:10.1186/s12931-022-01928-8
28. Coman V, Vodnar DC. Gut microbiota and old age: modulating factors and interventions for healthy longevity. *Exp Gerontol.* 2020;141:111095. doi:10.1016/j.exger.2020.111095
29. Kim YS, Unno T, Kim BY, Park MS. Sex Differences in Gut Microbiota. *World J Mens Health.* 2020;38(1):48–60. doi:10.5534/wjmh.190009
30. Lee SH, Yun Y, Kim SJ, et al. Association between Cigarette Smoking Status and Composition of Gut Microbiota: population-Based Cross-Sectional Study. *J Clin Med.* 2018;7(9):282. doi:10.3390/jcm7090282
31. Biedermann L, Zeitz J, Mwyni J, et al. Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. *PLoS One.* 2013;8(3):e59260. doi:10.1371/journal.pone.0059260
32. Allais L, Kerckhof FM, Verschuere S, et al. Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ Microbiol.* 2016;18(5):1352–1363. doi:10.1111/1462-2920.12934
33. Lai HC, Lin TL, Chen TW, et al. Gut microbiota modulates COPD pathogenesis: role of anti-inflammatory Parabacteroides goldsteinii lipopolysaccharide. *Gut.* 2022;71(2):309–321. doi:10.1136/gutjnl-2020-322599
34. Prakash A, Peters BA, Cobbs E, et al. Tobacco Smoking and the Fecal Microbiome in a Large, Multi-ethnic Cohort. *Cancer Epidemiol Biomarkers Prev.* 2021;30(7):1328–1335. doi:10.1158/1055-9965.EPI-20-1417
35. Goecks J, Jalili V, Heiser LM, Gray JW. How Machine Learning Will Transform Biomedicine. *Cell.* 2020;181(1):92–101. doi:10.1016/j.cell.2020.03.022
36. Sawane K, Hosomi K, Park J, et al. Identification of Human Gut Microbiome Associated with Enterolignan Production. *Microorganisms.* 2022;10(11):2169. doi:10.3390/microorganisms10112169
37. Seethaler B, Nguyen NK, Basrai M, et al. Short-chain fatty acids are key mediators of the favorable effects of the Mediterranean diet on intestinal barrier integrity: data from the randomized controlled LIBRE trial. *Am J Clin Nutr.* 2022;116(4):928–942. doi:10.1093/ajcn/nqac175
38. Kevans D, Tyler AD, Holm K, et al. Characterization of Intestinal Microbiota in Ulcerative Colitis Patients with and without Primary Sclerosing Cholangitis. *J Crohns Colitis.* 2016;10(3):330–337. doi:10.1093/ecco-jcc/jjv204
39. Zhu S, Liu S, Li H, et al. Identification of Gut Microbiota and Metabolites Signature in Patients With Irritable Bowel Syndrome. *Front Cell Infect Microbiol.* 2019;9:346. doi:10.3389/fcimb.2019.00346
40. Stiemsma LT, Arrieta MC, Dimitriu PA, et al. Shifts in Lachnospira and Clostridium sp. in the 3-month stool microbiome are associated with preschool age asthma. *Clin Sci.* 2016;130(23):2199–2207. doi:10.1042/CS20160349
41. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio.* 2014;5(2):e00889. doi:10.1128/mBio.00889-14

42. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol*. 2016;13(12):691–706. doi:10.1038/nrgastro.2016.165
43. Luo H, Yang R, Zhao Y, et al. Recent advances and strategies in process and strain engineering for the production of butyric acid by microbial fermentation. *Bioresour Technol*. 2018;253:343–354. doi:10.1016/j.biortech.2018.01.007
44. Guo P, Zhang K, Ma X, He P. Clostridium species as probiotics: potentials and challenges. *J Anim Sci Biotechnol*. 2020;11:24. doi:10.1186/s40104-019-0402-1
45. Jiang M, Li Z, Zhang F, et al. Butyrate inhibits iLLC2-mediated lung inflammation via lung-gut axis in chronic obstructive pulmonary disease (COPD). *BMC Pulm Med*. 2023;23(1):163. doi:10.1186/s12890-023-02438-z

International Journal of Chronic Obstructive Pulmonary Disease

Dovepress

### Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>