Analysis of sputum microbial diversity in patients with different histological types of lung cancer

Zhuonan Ran

Southwest Medical University

Jiexing Liu

Southwest Medical University

Caiyan Xin

Southwest Medical University

Fen Wang

Southwest Medical University

Shan Zeng

Southwest Medical University

Hua Xiao

Southwest Medical University

Xiang Shen

Southwest Medical University

Bin Xiong

Southwest Medical University

Zhangyong Song (✉ szy83529@163.com)

Southwest Medical University   https://orcid.org/0000-0003-1190-485X

Abstract

Background: The respiratory microbiome of lung cancer patients is different from healthy people. Finding the characteristics of microorganisms is conducive to exploring new directions for the assist in distinguishing different histological types of lung cancer.

Methods: 16S rRNA sequencing and internal transcribed spacer sequencing were performed respectively on sputum samples from 8 adenocarcinomas, 10 squamous cell carcinomas, 5 small cell lung cancers, and 3 combined-small cell lung cancer patients. And then, bioinformatics analysis of their bacterial microbiome and fungal mycobiome were conducted. And the correlation between respiratory microorganisms and clinical characteristics was further analyzed.

Results: The difference in species diversity of the respiratory tract microorganisms of different histological types of lung cancer is statistically significant. Moreover, smoking can increase the diversity of respiratory bacteria. Nevertheless, quitting smoking is beneficial to the transformation of the respiratory tract flora to a non-smoking state. Spearman's correlation further showed that *Fusarium* was positively related to *Streptococcus*, *Neisseria*, *Prevotella-7*, and *Veillonella*, and *Penicillium* was positively related to *Neisseria*.

Conclusions: Based on the differences of microorganisms among different histological types of lung cancer, deep composition characterization of lung cancer-related microorganisms and their clinical characteristics is an important advancement in understanding the role of microorganisms in lung cancer development.

Introduction

With the development of sequencing technology, researchers have demonstrated that microbes are widespread in the human body and play different roles in diseases progression and treatment. On the one hand, microbe infection can promote the development of cancer [1–3]. Microorganisms may promote the development of lung cancer (LC) by inflammation, immune response, genotoxicity, and other mechanisms [4]. Inflammatory factors, such as interleukin (IL), tumor necrosis factor (TNF)-α, and cyclooxygenase (COX-2), are closely related to the occurrence of tumor [5]. The positive rate of *Helicobacter pylori* (*HP*) in patients with gastric cancer is positively correlated with the expression of migration inhibition factor (MIF) [6]. And the expression of MIF can promote cell transformation, tumor cell proliferation, and metastasis. MIF is also up-regulated expression in non-small cell lung cancer (NSCLC) [7]. Meanwhile, microbial fermentation may also produce potentially toxic and cancer-promoting metabolites, such as ammonia, amine, phenols, sulfides, and nitrosamines [8]. On the other hand, investigations had found that microbes might be useful in the treatment of tumors. For example, transplantation of faecal bacteria can enhance the efficacy of EGFR-TKI (epidermal growth factor receptor tyrosine kinase inhibitors) and anti-PD (Programmed cell death protein)-1 in LC mice [9, 10]. In addition, intestinal flora may affect the pulmonary immune response through blood circulation and

lymphatic circulation, thus elevating the anti-pulmonary-tumor effect [11]. In a world, further studies are needed to carry out, which will be helpful in the joint analysis of the microbial effect and cancer development.

The lung is more likely to colonize with microbes because of communicating with the outside world through the trachea [12]. The composition of bacteria in oropharynx, nasopharynx and lung of healthy people is different. *Prevotella*, *Streptococcus*, *Veilonella*, *Haemophilus*, and *Neisseria* were the main bacteria genera in respiratory tract of healthy people [13, 14]. And *Cladosporium*, *Eurotium*, *Penicillium*, and *Aspergillus* were the most common fungi taxa in respiratory tract of healthy people [15]. LC as a highly malignant tumor with high prevalence rate [16], its pulmonary microbial composition has been confirmed to be different from that of tumor free people by several studies. *Streptococcus*, *Granulicatella*, *Veillonella*, and *Neisseria* were the major abundance changed in LC patients compared to the tumor free population [17−20]. To confirm the relationship microbial effect and LC development, many investigations were carrying out. Both *Mycobacterium tuberculosis* and HP infection had been confirmed to associate with the development of LC [21, 22]. Recent investigation also had confirmed that chronic infections in the lung can increase the risk of LC [4]. Moreover, the differences in microbial composition between adenocarcinoma (AD) and squamous cell carcinomas (SCC) patients were also found [23]. However, up to now, there are few researches about lung microbial diversity in patients with different histological types of LC.

In the present, the characteristics of sputum bacterial microbiome and fungal mycobiome distribution in patients with different histological types of LC were focused. By employing the bacterial 16S rRNA and fungal internal transcribed spacer (ITS) high-throughput sequencing, the distinct composition characterization of LC-related microorganisms on the sputum were showed. Our finding broadens the new concept that there are differences of microorganisms among different histological types of LC. The deep composition and functional characterization may serve to open new avenues in understanding the relationship mechanism of microorganisms and LC.

Materials And Methods

Subjects and Sample Collection: A total of 26 sputum specimens were collected from clinical patients with advanced LC at the Affiliated Hospital of Southwest Medical University, Luzhou, China. All these patients were histologically diagnosed as LC. Patient information was anonymised and their identity was removed before analysis. Written informed consent was obtained from all participants. None of the patients had a history of long-term immunosuppressive use, thoracic surgery, autoimmune diseases, endotracheal intubation, invasive/noninvasive mechanical ventilation, or oral diseases. After gargling, sputum samples were collected in a sterile cup. Samples were transferred rapidly on dry ice to laboratory, and stored at -80°C. Patients' clinical indicators had been recorded: basic patient information including gender and age, smoking history, pathological types and clinical stages of lung cancer, inflammatory

cytokine levels including white blood cells (WBC) count, neutrophil rate, lymphocyte rate and C-reactive protein.

Sample pretreatment and genomic DNA isolation

The samples were thawed on ice. And then, samples were treated with 5 ml of 30 % aqueous methanol and 500 µl of a methanol-dithiothreitol (DTT) solution, made up by adding 0.24 g DTT to 1 mL of 30 % aqueous methanol, mixed for 15min, and centrifuged at 20000rpm for 10 min [24]. The supernatant was discarded and the remaining substances were used for DNA isolation. The FastDNA SPIN Kit for soil (MP Biomedicals Inc, United States) was used to the genomic DNA isolation. The procedure is according to the instructions in the kit. Genomic DNA was eluted in to 80 µl of ultrapure water. The extracted DNA was then electrophoreted, which produces bright bands that lead to successful DNA extraction. The extracted DNA was stored in -20°C refrigerator.

Library construction and sequencing: Based on 16S rRNA and ITS conserved regions, primers were designed. 16S rRNA sequencing forward primer was: 5 '-CCGTCAATTCMTTTRAGTTT-3', reverse primer was: 5 '-GTGCCAGCMGCCGCGG-3'. The ITS1 sequencing forward primer was: 5 '-GCTGCGTTCTTCATCGATGC-3', reverse primer was: 5 '-GGAAGTAAAAGTCGTAACAAGG-3'. And sequencing adapters at the ends of primers were added. PCR amplification and purify, quantify and homogenize the products to form a sequencing library was performed. After that, the qualified library was sequenced with Illumina HiSeq 2500 profiling by BioMarker Technologies (Beijing, China). And the results were stored in the FASTQ file format, which contains the sequence information of the reads and their corresponding the sequencing were quality information. To filter the raw reads obtained by sequencing Trimmomatic v0.33 software was used [25]. Then, to identify and remove primer sequences to obtain high-quality reads, cut adapt 1.9.1 software was explored [26]. To splice the high-quality reads of each sample through overlap, FLASH v1.2.7 software [27] was used. And the resulting splicing sequence is clean reads. To identify and remove the chimera sequence to get the effective reads, UCHIME v4.2 software [28] was used.

Statistical analysis

USEARCH software [29] was used to cluster/denoise high-quality sequences at 97.0 % similarity level to obtain OTUs (operational taxonomic units) on the basis of Silva (Bacteria) and UNITE (fungi) taxonomic database. And their species classification based on the sequence composition of OTUs was getted. Based on the OTU results, taxonomic analysis on the samples at each classification level to obtain a community structure diagram was performed. Metastats software [30] was used to carry out T test on species abundance data between groups.

Perform alpha and beta diversity analysis on different LC types, smoking and non-smoking, and infection indicators. The alpha diversity index was evaluated by QIIME2 software [31]. QIIME software [32] was used for beta diversity. The T test was used to evaluate the differences in alpha diversity index among different

treatments. The Unweightde Unifrac algorithm was used for bacterial beta diversity analysis. The binary-Jaccard algorithm was used for fungal beta diversity analysis. Based on the appeal of the two distance matrices, Non-MetricMulti-Dimensional Scaling (NMDS) [33], Heatmap [34], PERMANOVA [35] and Unweighted Pair-group Method with Arithmetic Mean (UPGMA) [36] are used for multi-dimensional beta diversity analysis.

Results

Patient cohort and sequencing quality

There were 8 samples of AD, 10 samples of SCC, 5 samples of small cell lung cancer (SCLC), and 3 samples of combined-small cell lung cancer (C-SCLC). Among all the patients, 19 had smoking history and 6 had chemotherapy history. Patients' basic information is presented in Table 1. By using 16S rRNA sequencing, a total of 2,388,223 original sequences were obtained from 26 samples, and 1,764,852 high-quality sequences were obtained after splice and filtration. The average sequence length of all samples was 415–427 bp, and bases with mass value greater than or equal to 30 accounted for 94.61–95.53 % of the total base number. Through ITS sequencing, a total of 2,058,555 original sequences were obtained from 26 samples, and 1,747,974 high-quality sequences were obtained after splicing and filtering. The average sequence length of all samples was 215–263 bp. And bases with mass value were greater than or equal to 30 accounted for 94.31–98.34 % of the total base number.

Table 1
Clinical and pathologic characters of enrolled patients

Feature	AD	SCC	SCLC	C-SCLC
number	8	10	5	3
Gender				
Male	5	10	5	3
Female	3	0	0	0
Drinking history				
Yes	3	8	2	3
No	5	2	3	0
Smoking history				
Yes	5	8	3	3
No	3	2	2	0
Chemotherapy history				
Yes	2	3	2	0
No	6	7	3	3
AD: adenocarcinoma; SCLC: small cell lung cancer; SCC: squamous cell carcinoma; C-SCLC: combined-SCLC.				

Bacterial microbial composition is basically similar, but there are some species have statistical difference

In total, 13 kinds of bacteria phylum, 108 kinds of bacteria genus, and 52 kinds of bacteria species were found. The pulmonary bacterial flora of LC patients mainly belongs to *Firmicutes* (34.48 %), *Proteobacteria* (22.88 %), *Bacteroidetes* (22.60 %), and *Actinobacteria* (9.28 %). The main genera are *Streptococcus* (21.35 %), *Neisseria* (15.57 %), *Prevotella* (9.73 %) and *Veillonella* (5.37 %). They have different proportions in each group (Figs. 1A and 1B). Metastats software was used to carry out T test on species abundance data between groups. *Firmicutes* was more abundant in SCC than in AD and SCLC. It's interesting to note that the abundance of *Firmicutes* in SCC (38.7 %) was significantly higher than that in AD (29.6 %) ($P = 0.017$). However, most of the family level species with significant difference in abundance between AD and SCC belonged to *Bacteroidetes*. It suggests that the difference between SCC and AD in *Firmicutes* is not caused by a few species, however, by a wide range of species abundance changes. *Synergistetes* and *Spirochaetes* were more abundant in AD than in SCC and SCLC. And the difference of their abundance between AD and SCC was statistically significant ($P < 0.05$). The species with statistically differences were summarized in Table 2. AD and SCC have the largest number of different species.

Table 2
The significant difference analysis between groups

Phylum	AD			SCC			P value
	Mean	Variance	Std.err	Mean	Variance	Std.err	
Synergistetes	1.60E-03	2.62E-06	5.73E-04	1.69E-04	7.98E-08	8.93E-05	0.004
Firmicutes	2.96E-01	4.27E-03	2.31E-02	3.87E-01	5.69E-03	2.39E-02	0.017
Spirochaetes	1.54E-02	2.45E-04	5.54E-03	3.70E-03	2.52E-05	1.59E-03	0.047
Family	Mean	Variance	Std.err	Mean	Variance	Std.err	P value
Desulfovibrionaceae	4.19E-04	6.85E-07	2.93E-04	2.78E-06	7.75E-11	2.78E-06	0.001
Marinifilaceae	3.03E-04	2.47E-07	1.76E-04	5.42E-06	1.31E-10	3.62E-06	0.001
Sphingobacteriaceae	0.00E+00	0.00E+00	0.00E+00	2.56E-04	6.57E-07	2.56E-04	0.001
Lentimicrobiaceae	6.60E-04	4.90E-07	2.47E-04	3.89E-05	5.04E-09	2.25E-05	0.003
Synergistaceae	1.60E-03	2.62E-06	5.73E-04	1.69E-04	7.98E-08	8.93E-05	0.004
Propionibacteriaceae	3.07E-04	1.55E-07	1.39E-04	4.19E-05	2.92E-09	1.71E-05	0.011
Defluviitaleaceae	2.82E-04	3.61E-08	6.72E-05	5.93E-05	1.55E-08	3.93E-05	0.014
Bacteroidaceae	3.33E-03	4.60E-05	2.40E-03	9.74E-05	3.34E-08	5.78E-05	0.028
Rikenellaceae	6.36E-04	3.61E-07	2.12E-04	1.24E-04	6.26E-08	7.91E-05	0.042
Family	AD			SCLC			P value
	Mean	Variance	Std.err	Mean	Variance	Std.err	
Tannerellaceae	1.25E-03	3.66E-07	2.14E-04	3.99E-04	1.73E-07	1.86E-04	0.004
Lentimicrobiaceae	6.60E-04	4.90E-07	2.47E-04	1.26E-05	7.94E-10	1.26E-05	0.008
Leptotrichiaceae	4.35E-02	7.68E-04	9.80E-03	1.73E-02	4.20E-05	2.90E-03	0.009

Phylum	AD			SCC			*P* value
	Mean	Variance	Std.err	Mean	Variance	Std.err	
Phylum	SCC			SCLC			P value
	Mean	Variance	Std.err	Mean	Variance	Std.err	
Tenericutes	5.95E-04	4.16E-07	2.04E-04	9.18E-05	3.17E-08	7.96E-05	0.041
Family	Mean	Variance	Std.err	Mean	Variance	Std.err	P value
Leptotrichiaceae	3.44E-02	4.08E-04	6.39E-03	1.73E-02	4.20E-05	2.90E-03	0.017
Ruminococcaceae	4.55E-03	1.65E-05	1.28E-03	1.40E-03	1.30E-06	5.11E-04	0.024
Flavobacteriaceae	1.47E-02	2.27E-04	4.76E-03	3.67E-03	6.05E-06	1.10E-03	0.026
Mycoplasmataceae	4.61E-04	3.09E-07	1.76E-04	5.82E-05	1.59E-08	5.63E-05	0.031
Actinomycetaceae	2.65E-02	2.39E-04	4.89E-03	1.50E-02	4.33E-05	2.94E-03	0.045

Alpha diversity showed differences among groups (Figs. 1C and 1D). Chao1 assesses species abundance. And Shannon assesses species diversity which is affected by species abundance and uniformity in sample communities. AD has a higher Chao1 value than SCC (P = 0.0088) and SCLC (P = 0.0042). AD also has a higher Shannon value than SCC (P = 0.0324). Both species abundance and species diversity of AD is the highest. And the alpha diversity of SCC is the lowest. There was no statistical difference between SCC and SCLC in alpha diversity.

The similarity of bacterial species diversity was higher in AD than in SCC

According to the results of NMDS, the distribution of each sample in SCC and SCLC is dispersed, while in AD is concentrated (Fig. 2A). According to the Heatmap results, the color gradient from blue to red indicates the distance between samples from near too far. It was further observed that AD, SCC, and SCLC formed separate clusters except for individual samples. The within-group difference AD was the smallest and SCC was the largest. And SCLC is between AD and SCC. AD and SCC showed great between-group difference (Fig. 2B). PERMANOVA (permutational multivariate analysis of variance) was used to further verify the results of NMDS and Heatmap. AD has the smallest beta distance and SCC has the largest beta distance. The beta distance between groups was slightly larger than within groups. *P* value is less than 0.05 meaning the high reliability of the test (Fig. 2C). In conclusion, among the three groups, the AD group had the highest bacterial species diversity similarity, however, the SCC group had the lowest species bacterial diversity similarity.

C-SCLC differs from SCLC in bacterial diversity

Although C-SCLC has the same histological components as SCLC. Its bacterial species diversity is different from SCLC. *Proteobacteria* abundance in SCLC is greater than C-SCLS ($P = 0.0238$). In genus level, the abundance of *Neisseria*, *Aggregatibacter*, *Haemophilus*, *Peptostreptococcus*, *Gemella*, *Stomatobaculum*, and *Parvimonas* in the SCLC was significantly greater than that in the C-SCLC ($P < 0.05$), however, the abundance of *Tannerella* in the SCLC was significantly less than that in the C-SCLC ($P < 0.005$). The sample hierarchy cluster tree is obtained through UPGMA. And C-SCLC and SCLC samples are clustered into two separate clusters (Fig. 3A). PERMANOVA results showed that the beta distance of C-SCLC was greater than that of SCLC. The beta distance between two groups was significantly larger than within groups (Fig. 3B). These data suggested that there was a difference in the composition of bacterial species between C-SCLC and SCLC.

Smoking affect the sputum bacteria composition: In AD group, the ratio of smoking to non-smoking was 5:3. In phylum level, the abundance of *Spirochaetes* and *Bacteroidetes* in the smoking group was significantly greater than that in the non-smoking group ($P < 0.05$). And the abundance of *Actinobacteria* in the smoking group was significantly less than that in the non-smoking group ($P = 0.0106$). In genus level, the abundance of *Mogibacterium*, *Prevotella-2*, *Treponema-2*, *Pseudomonas*, *Alloprevotella*, *Fretibacterium*, *Leptotrichia*, and *Ralstonia* in the smoking group was significantly greater than that in the non-smoking group ($P < 0.05$), however, the abundance of *Corynebacterium* and *Rothia* in the smoking group was significantly less than that in the non-smoking group ($P < 0.05$). The abundance of many species has increased in the smoking group, which suggesting that smoking may increase the risk of bacterial invasion. The results of the Heatmap showed that the bacteria species diversity was more similar within groups rather than between groups (Fig. 4A). Although the A11 sample was in the smoking group, its species diversity was more similar with non-smoking group. Tracking the clinical information of A11 samples found that it's from a patient who had quit smoking for more than 10 years. This suggests that quitting smoking may lead to a gradual shift in respiratory bacteria to non-smoking status. PERMANOVA showed that the beta distance was significantly greater in the smoking group than in the non-smoking group (Fig. 4B). These phenomena were not observed in the fungal sequencing results (data not shown). Altogether, these data suggests that smoking can reduce diversity similarity of bacteria.

Fungal composition of sputum in lung cancer patients

Fungal sequencing reads were also clustered. The OTU value of each sample is 79–349. A total of 1324 OTUs were obtained from the 26 samples. The number of OTU shared by each groups was small (Fig. 5A). In total, 10 kinds of fungal phylum, 328 kinds of fungal genus, and 359 kinds of fungal species were found. Unlike bacteria, there are exist many unclassified species. The main fungal composition of sputum of LC patients at phylum level is *Ascomycota* (74.77 %), *Basidiomycota* (11.89 %), *Mucoromycota* (1.39 %), *Rozellomycota* (1.22 %), and Unclassified (10.50 %) (Fig. 5B). And the composition at genus level is *Candida* (18.52 %), *Cladosporium* (8.32 %), *Fusarium* (5.42 %), *Leptobacillium* (2.98 %), *Aspergillus* (2.64 %), and Unclassified (23.63 %).

The diversity of fungi varies greatly among individuals

The Chao1 value of AD group is greater than that of SCC group ($P = 0.0328$). And the Chao1 value of C-SCLC is less than the other three groups ($P < 0.05$). These suggest that AD fungus species abundance was higher than SCC. And C-SCLC fungus species abundance was lower than the other three groups (Fig. 6A). The beta distance between each sample is far (Fig. 6B). It indicates that unlike bacteria, the fungal composition of the sputum of LC patients has great individual differences. And there was no obvious clustering of fungal diversity among the groups.

Correlations between bacteria and fungi

Spearman's rank correlation coefficient [37] was used to analyze the correlation between bacterial and fungal species abundance. The correlation between the top three abundance species at the phylum level and the top five abundance species at the genus level in bacteria and fungi was analyzed, respectively. Scatter diagrams showing the correlations between bacteria and fungi were plotted in Fig. 7. There is negative correlation between *Proteobacteria* and *Ascomycota* ($P = 0.05$). *Fusarium* was positively correlated with *Streptococcus*, *Neisseria*, *Prevotella-7*, and *Veillonella* ($P < 0.05$). And *Penicillium* was positively correlated with *Neisseria* ($P < 0.05$).

Inflammatory cytokines and respiratory microbiome

According to the median of WBC, neutrophil ratio, lymphocyte ratio, and C-reactive protein (CRP), the AD group was divided into low-median inflammatory factor group and high-median inflammatory factor group. Alpha diversity and beta diversity of bacteria and fungi were analyzed. There was no statistical difference in the Chao1 and Shannon values between the high and low median inflammatory factor groups ($P > 0.05$) (suppl-table 1, suppl-table 2). In terms of beta diversity, there was no obvious similarity in the species diversity of bacteria or fungi in the two groups. In the SCC group, also did not observe the correlation between inflammatory factors and microbiome diversity.

Discussion

In this study, high-throughput sequencing was used to sequence the sputum microbiome and mycobiome of LC patients. This is a thorough study of respiratory microbiome in different histological types of LC. In addition to bacteria, the human respiratory tract is colonized with fungal mycobiome. In fact, perez-Cobas's study confirmed that human airways are also colonized with archaea [38]. This suggests that the respiratory tract is a microenvironment with a variety of microbial communities. Our results also confirmed the effects of smoking on the respiratory microorganisms. It provided preliminary data for the related institutional research on the LC and microorganisms.

Using different tissue samples, it has been confirmed that the lung microbiome of LC patients is different from that of tumor free people. In the bronchial brush samples, the abundance of *Streptococcus* in LC patients was significantly higher than that in healthy groups by 16S rRNA sequencing. *Neisseria* also was

confirmed as an increasing trend in the cancerous sites of LC patients. And the alpha diversity of LC patients decreased significantly [39]. Moreover, the abundance of *Streptococcus*, *Prevotella*, and *Bifidobacterium* in LC tumor tissue was significantly higher than that in emphysema lung tissue [40]. In addition, the abundance of *Streptococcus*, *Abiotrophia*, and *Granulicatella* in sputum samples of LC patients was significantly higher than that of healthy people[41]. However, there are few studies on the differences of the microbiome between different histological types of LC.

AD and SCC are the two most common types of LC. Previous research found there were 9 genera to be significantly correlated with the occurrence of SCC as compared to AD (*Acidovorax*, *Brevundimonas*, *Comamonas*, *Tepidimonas*, *Rhodoferax*, *Klebsiella*, *Leptothrix*, *Polaromonas*, *Anaerococcus*) [42]. And similar to our results, the phylogenetic diversity of AD tumor tissues was significantly greater than that of SCC [23]. Moreover, in saliva samples, the abundances of *Streptococcus* and *Porphyromonas* in AD were higher than in SCC [17]. In addition, compared with SCC patients, AD patients have the larger bacterial species abundance, species diversity, and inter-group species diversity similarity (Figs. 1 and 2). However, there is no precise research to confirm the relationship between these different composition microorganisms and AD and SCC development. Additional experiments are needed to further elucidate these mechanisms.

When SCLC is combined with any other NSCLC histologic type, the symptom is referred to as C-SCLC [43]. At present, few studies on respiratory microbiome of SCLC and C-SCLC have been carried out. Our results showed that the bacterial species abundance, species diversity, and species diversity similarity of SCLC are between AD and SCC (Figs. 1 and 2). However, compared to AD and SCC, there is a significantly different species in SCLC. Although SCLC and C-SCLC have the same histological components, the bacterial species diversity of C-SCLC is less similar. And there is a difference in the composition of bacterial species between C-SCLC and SCLC (Fig. 3). Whether the combination of the two histological types increased individual differences in C-SCLC requires further confirmation by exploring larger sample size investigations. In addition, the detection of biological components of exhaled breath can be used to diagnose cancer, and even distinguish LC of different tissue types [44]. In this cue, the detection of the respiratory microbiome may also be a way to identify different histological types of LC.

It is known that smoking influence the lung microorganism's composition. Previous studies had confirmed that the oral bacterial microbiome diversity and bacterial species of mice exposed to cigarette smoke was significantly lower than that of the control group. And the oral microbiome diversity of mice gradually recovered after the cigarette smoke exposure was stopped for 3 months [45]. Similar to previous researches, our results showed there was a significant difference in the respiratory bacterial microbiome between smokers and non-smokers (Fig. 4). Besides, the microbiological function analysis of smokers and non-smokers is significantly different. Compared with non-smokers, smokers showed higher tricarboxylate utilization and lactic acid racemization [46]. Smoking was also associated with an increased abundance of potentially pathogenic bacteria, including *Streptococcus*, *Fusobacterium*, *Prevotella*, *Haemophilus*, and *Treponema*[47]. And tobacco impairs innate immune defenses, leading to

changes in the abundance, taxonomic composition, and phylogenetic diversity of respiratory microbiome. This, in turn, leads to an inflammatory response that further damages lung defenses and further dysregulation of the respiratory microbiome [48]. In conclusion, smoking can increase the chance of microbial colonization and increase the diversity of respiratory bacterial microbiome. Interestingly, our results also confirmed that long-term cessation can restore the respiratory bacterial microbiome to the status of non-smokers.

As known, *Candida* and *Aspergillus*, common opportunistic pathogens, tend to cause disease when the body's immune system is weakened [49]. LC patients are more likely to be complicated with fungal infection, resulting in prolonged hospital stay, increased treatment costs, poor prognosis, and even life-threatening [50]. Combination of these reasons, these colonizing opportunistic pathogens may cause disease and clinical symptoms. It had confirmed that *Candida* infection may be associated with oral and esophageal cancers [51]. And fungi may accelerate the pancreatic ductal adenocarcinoma [52]. Unlike bacteria, the species diversity of fungi varies greatly among individuals (Fig. 6). And this conclusion is consistent with previous studies [15].

However, it is known that bacteria and fungi co-exist in the human respiratory tract. *Candida albicans* and *Escherichia coli* have a synergistic effect, which can increase the mortality of mice infected with abdominal cavity [53]. And the β-1,3-glucan of *C. albicans* can also enhance the resistance of *Staphylococcus* aureus in a multi-microbial biofilm environment [54]. It also found that the interactions between *Aspergillus fumigatus*, *Cryptococcus* and bacteria, could be a new direction for antifungal drugs [55]. In this study, it was found that there was a correlation between the abundance of most bacteria and fungi. And *Fusarium*, in particular, was positive associated with four species of bacteria. Further experiments are needed to elucidate the positive association in the LC.

Microorganisms that exist in the human body can be divided into symbiotic bacteria and pathogens. Symbiotic bacteria generally do not cause infection and inflammation, and pathogens are the cause of infection, which is also the basis of the infection dichotomy [56]. In our study, no relationship between inflammatory factors and the diversity of the respiratory tract microbiome was found. This suggests that the respiratory tract microorganisms in patients with lung cancer were mostly symbiotic bacteria. However, there were many opportunistic pathogens in the respiratory tract of patients with lung cancer, such as *Streptococcus*, *Staphylococcus*, *Candida*, and *Aspergillus*. It interacted with microorganisms to cause infections, which in turn caused changes in inflammatory factors. Another speculation is that the inflammatory response caused by the respiratory microbiome only caused changes in the tumor microenvironment, and did not cause significant changes in blood inflammatory factors.

There exist several limitations to the study. Since the incidence of lung cancer in men was higher than that in women, the proportion of men and women in this study was different. Because the incidence of SCLC and C-SCLC was lower than that of AD and SCC, the sample size of SCLC and C-SCLC was smaller. The sputum samples may have oral flora contamination compared to the bronchoalveolar lavage (BAL)

samples, however, we believe that the sputum sample contains a larger collection of microorganisms. The sputum sample contains the collection of microorganisms in the bronchi and alveoli of the left and right lungs, while BAL only contains the collection of microorganisms in a certain segment of the bronchi and alveoli. And it provides a non-invasive method of obtaining upper bronchial tract samples and also involves minimal patient discomfort.

Conclusion

In the present study, the differences in the respiratory microbiome among patients with different histological types of LC were dissected. In addition, the effect of smoking on the respiratory microbiome was found. Moreover, it provided the cue that there is a reaction between the pulmonary bacteria and pulmonary fungi in lung cancer patients. Such systematic and high throughput-wide characterization can be effectively used to understand the interaction between pulmonary microorganisms and LC development.

Declarations

Ethics approval and consent

The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of Affiliated Hospital of Southwest Medical University (KY2020082). All patients were informed about the study, informed about the whereabouts of the samples, and signed an informed consent form. Before analysis, patient information is anonymized.

Consent for publication

All patients were informed about the study. And get their consent to publish their data.

Availability of data and material

The datasets generated during and/or analyses during the current study are available in the Sequence Read Archive (PRJNA722474, PRJNA722576).

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

Author contributions

Zhuonan R, Jiexing L, Zhangyong S, Bin X conceived and designed the study.

Zhuonan R, Jiexing L, Fen W, Caiyan X, Zhangyong S performed the experiments.

Zhuonan R, Jiexing L, Zhangyong S,Hua Xiao, Bin X analyzed the data.

Fen W, Caiyan X, Xiang S, Shan Z contributed analysis tools.

Zhuonan R, Jiexing L, Zhangyong S, Bin X provided critical inputs on design, analysis, and interpretation of the study.

All the authors had access to the data.

All authors read and approved the final manuscript as submitted.

References

1. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393:169–82.
2. Helicobacter C, Collaborative G. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut. 2001;49:347–53.
3. Thompson MP, Kurzrock R. Epstein-Barr virus and cancer. Clin Cancer Res. 2004;10:803–21.
4. Ran Z, Liu J, Wang F, Xin C, Xiong B, Song Z. Pulmonary micro-ecological changes and potential microbial markers in lung Cancer patients. Front Oncol. 2021;10:576855.
5. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? Lancet. 2001;357:539–45.
6. Yoon K, Kim N, Park Y, Kim BK, Park JH, Shin CM, et al. Correlation between macrophage migration inhibitory factor and autophagy in Helicobacter pylori-associated gastric carcinogenesis. PLoS One. 2019;14:e0211736.
7. Wang WM, Liu JC. Effect and molecular mechanism of mir-146a on proliferation of lung cancer cells by targeting and regulating MIF gene. Asian Pac J Trop Med. 2016;9:806–11.
8. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13:800–12.
9. Heshiki Y, Vazquez-Uribe R, Li J, Ni Y, Quainoo S, Imamovic L, et al. Predictable modulation of cancer treatment outcomes by the gut microbiota. Microbiome. 2020;8:28.
10. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359:97–103.

11. Bingula R, Filaire M, Radosevic-Robin N, Bey M, Berthon JY, Bernalier-Donadille A, et al. Desired turbulence? gut-lung axis, immunity, and lung cancer. J Oncol. 2017;2017:5035371.

12. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial topography of the healthy human lower respiratory tract. mBio. 2017;8:e02287-16.

13. 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.

14. Moffatt MF, Cookson WO. The lung microbiome in health and disease. Clin Med (Lond). 2017;17:525–9.

15. Tipton L, Ghedin E, Morris A. The lung mycobiome in the next-generation sequencing era. Virulence. 2017;8:334–41.

16. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol. 2016;893:1–19.

17. Yan X, Yang M, Liu J, Gao R, Hu J, Li J, et al. Discovery and validation of potential bacterial biomarkers for lung cancer. Am J Cancer Res. 2015;5:3111–22.

18. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, et al. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer. 2016;102:89–95.

19. Cameron SJ, Lewis KE, Beckmann M, Allison GG, Ghosal R, Lewis PD, et al. The metabolomic detection of lung cancer biomarkers in sputum. Lung Cancer. 2016;94:88–95.

20. Hosgood HD 3rd, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen. 2014;55:643–51.

21. Pilaniya V, Gera K, Kunal S, Shah A. Pulmonary tuberculosis masquerading as metastatic lung disease. Eur Respir Rev. 2016;25:97–8.

22. Deng B, Li Y, Zhang Y, Bai L, Yang P. Helicobacter pylori infection and lung cancer: a review of an emerging hypothesis. Carcinogenesis. 2013;34:1189–95.

23. Yu G, Gail MH, Consonni D, Carugno M, Humphrys M, Pesatori AC, et al. Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. Genome Biol. 2016;17:163.

24. Cameron SJS, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, et al. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. PLoS One. 2017;12:e0177062.

25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.

26. Martin M. Cut adapt removes adapter sequences from high-throughput sequencing reads. Embnet Journal. 2011;17:10–2.

27. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957–63.

28. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194–200.

29. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10:996–8.

30. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. PLoS Comput Biol. 2009;5:e1000352.

31. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:1091.

32. Lawley B, Tannock GW. Analysis of 16S rRNA Gene amplicon sequences using the QIIME software package. Methods Mol Biol. 2017;1537:153–63.

33. Looft T, Johnson TA, Allen HK, Bayles DO, Alt DP, Stedtfeld RD, et al. In-feed antibiotic effects on the swine intestinal microbiome. Proc Natl Acad Sci USA. 2012;109:1691–6.

34. Qu Z, Lau CW, Nguyen QV, Zhou Y, Catchpoole DR. Visual Analytics of Genomic and Cancer Data: A Systematic Review. Cancer Inform. 2019;18:1176935119835546.

35. Kelly BJ, Robert G, Kyle B, Scott SM, Lewis JD, Collman RG, et al. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. Bioinformatics. 2015;15:2461–8.

36. Hua GJ, Hung CL, Lin CY, Wu FC, Chan YW, Tang CY, et al. MGUPGMA: A Fast UPGMA Algorithm With Multiple Graphics Processing Units Using NCCL. Evol Bioinform Online. 2017;13:1176934317773422.

37. Hazra A, Gogtay N. Biostatistics series module 6: correlation and linear regression. Indian J Dermatol. 2016;61:593–601.

38. Perez-Cobas AE, Ginevra C, Rusniok C, Jarraud S, Buchrieser C. Persistent legionnaires' disease and associated antibiotic treatment engender a highly disturbed pulmonary microbiome enriched in opportunistic microorganisms. mBio. 2020;11:e00889-20.

39. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, et al. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer. 2018;142:769–78.

40. Liu Y, O'Brien JL, Ajami NJ, Scheurer ME, Amirian ES, Armstrong G, et al. Lung tissue microbial profile in lung cancer is distinct from emphysema. Am J Cancer Res. 2018;8:1775–87.

41. Hosgood HD 3rd, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen. 2014;55:643–51.

42. Leigh GK, White JR, Vargas AJ, Bliskovsky VV, Beck JA, Natalia VM, et al. Interaction between the microbiome and tp53 in human lung cancer. Genome Biol. 2018;19:123–38.

43. Lei Y, Feng H, Qiang H, Shang Z, Chang Q, Qian J, et al. Clinical characteristics and prognostic factors of surgically resected combined small cell lung cancer: a retrospective study. Lung Cancer.

2020;146:244–51.

44. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. Br J Cancer. 2010;103:542–51.

45. Hilty M, Wuthrich TM, Godel A, Adelfio R, Aebi S, Burgener SS, et al. Chronic cigarette smoke exposure and pneumococcal infection induce oropharyngeal microbiota dysbiosis and contribute to long-lasting lung damage in mice. Microb Genom. 2020;6:1–14.

46. Al Bataineh MT, Dash NR, Elkhazendar M, Alnusairat DMH, Darwish IMI, Al-Hajjaj MS, et al. Revealing oral microbiota composition and functionality associated with heavy cigarette smoking. J Transl Med. 2020;18:421.

47. Panzer AR, Lynch SV, Langelier C, Christie JD, McCauley K, Nelson M, et al. Lung microbiota is related to smoking status and to development of acute respiratory distress syndrome in critically Ill trauma patients. American Journal of Respiratory Critical Care Medicine. 2018;197:621–31.

48. Mammen MJ, Sethi S. COPD and the microbiome. Respirology. 2016;21:590–9.

49. Köhler JR, Hube B, Puccia R, Casadevall A, Perfect JR. Fungi that Infect Humans. Microbiol Spectr. 2017;5:1–29.

50. Bereket W, Hemalatha K, Getenet B, et al. Update on bacterial nosocomial infections. Eur Rev Med Pharmacol Sci. 2012;16:1039–44.

51. Gholizadeh P, Eslami H, Yousefi M, Asgharzadeh M, Aghazadeh M, Kafil HS. Role of oral microbiome on oral cancers, a review. Biomed Pharmacother. 2016;84:552–8.

52. Dambuza IM, Brown GD. Fungi accelerate pancreatic cancer. Nature. 2019;574:184–5.

53. Klaerner HG, Uknis ME, Acton RD, Dahlberg PS, Carlone-Jambor C, Dunn DL. *Candida albicans* and *Escherichia coli* are synergistic pathogens during experimental microbial peritonitis. J Surg Res. 1997;70:161–5.

54. Kong EF, Tsui C, Kucharikova S, Andes D, Van Dijck P, Jabra-Rizk MA. Commensal protection of Staphylococcus aureus against antimicrobials by *Candida albicans* biofilm matrix. mBio. 2016;7:e01316-65.

55. Wang F, Xin C, Liu J, Ran Z, Zhao C, Song Z. Interactions between invasive fungi and symbiotic bacteria. World J Microbiol Biotechnol. 2020;36:137.

56. Wu BG, Segal LN. The Lung Microbiome and Its Role in Pneumonia. Clin Chest Med. 2018;39:677–89.
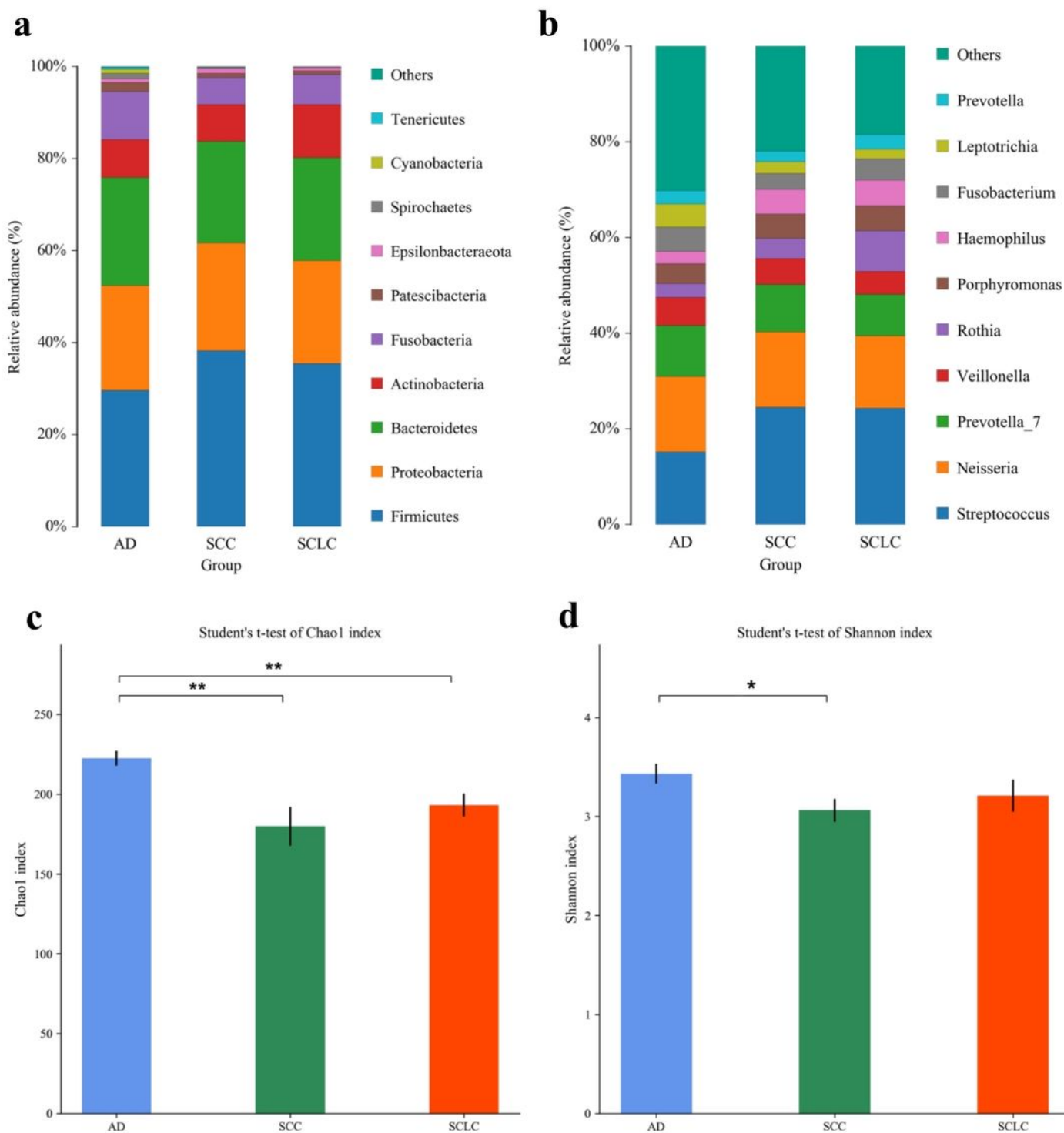
Figures

Figure 1

Bacterial species distribution histogram and alpha diversity. (a) Species distribution histogram of different groups in phylum level. One color represents one phylum, the length of the color block present the relative richness proportion of the phylum. (b) Bacterial species distribution histogram of different groups in genus level. (c) Histogram of alpha diversity index differences between groups. Chao1 index measure species richness, i.e. the number of species. (d) Shannon indexes was used to measure species

diversity, it was affected by species richness and community evenness in the sample community. Pairwise comparisons were made between each group, * P< 0.05, ** P< 0.01, indicated a significant difference between the two groups.
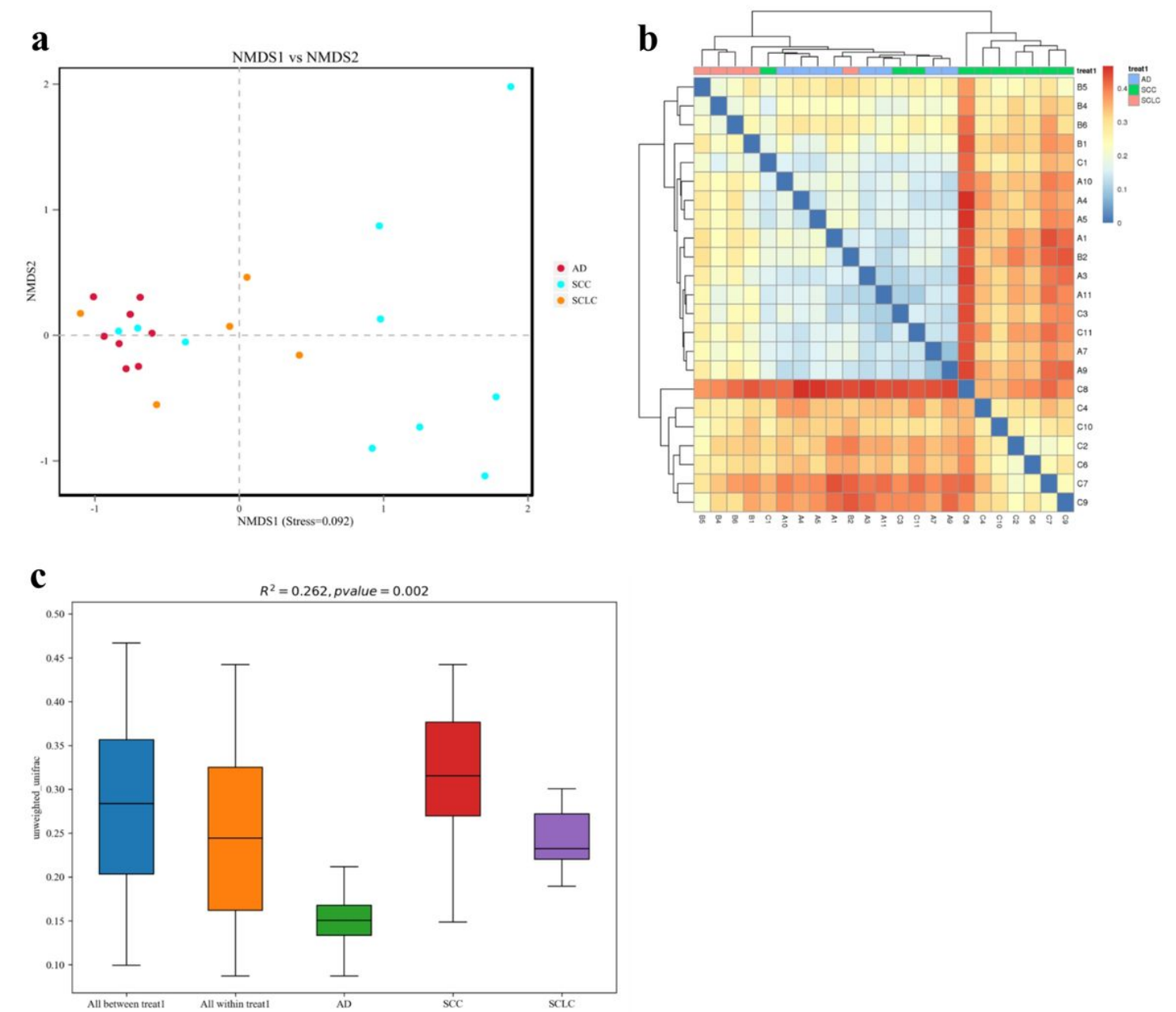
Figure 2

AD, SCC, and SCLC bacterial beta diversity analysis. (a) Non-MetricMulti-Dimensional Scaling analysis. Dots represent all samples; different colors represent different groups, distance between dots presents the difference, when Stress is less than 0.2, it indicates that NMDS analysis has certain reliability. The closer the samples are on the coordinate diagram, the higher the similarity is. (b) Sample clustering heatmap. Color gradient from blue to red represents the distance between samples from close to far. Treat1 represent different groups. X-axis and Y-axis represent different samples. (c) Permutational multivariate analysis of variance. R2 represents the interpretation degree of sample difference between different

groups. A larger R2 indicates that grouping has a higher interpretation degree to the difference, the group difference is higher, and P value less than 0.05 indicate a high reliability of the test. Y-axis represents beta distance; The box above "All between" represents the beta distance data of samples between all groups, while the box above "All within" represents the beta distance data of samples within all groups. The box below represents the beta distance data of samples within different groups.
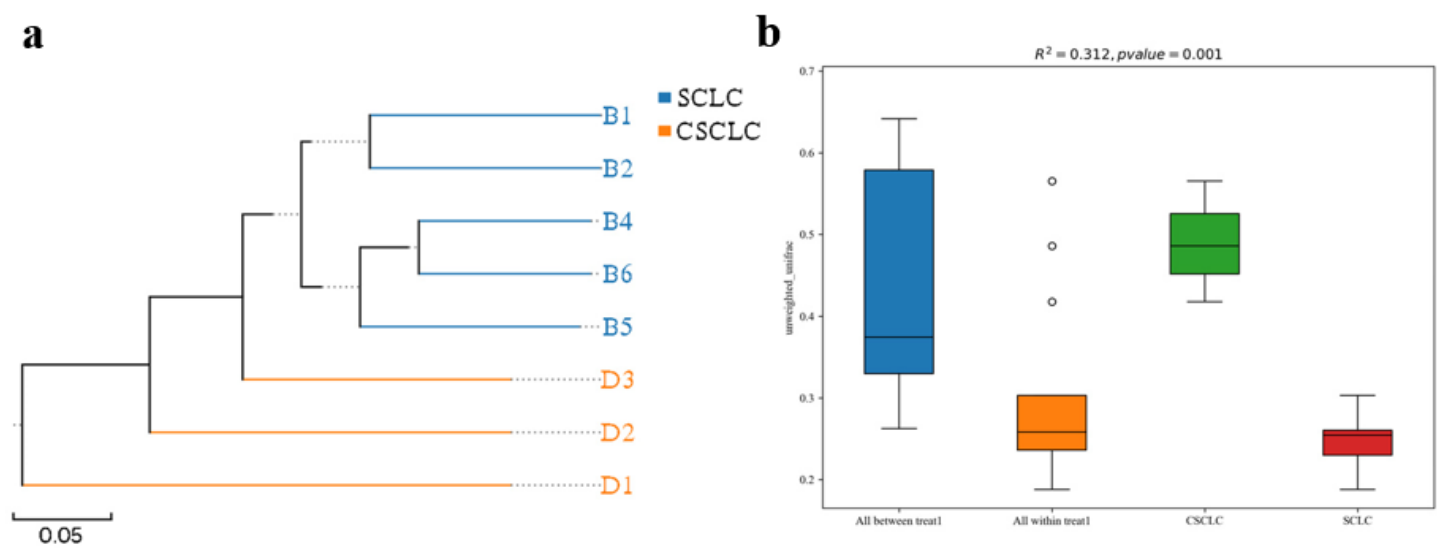
Figure 3

SCLC and C-SCLC bacterial beta diversity analysis. (a) Unweighted pair-group method with arithmetic mean analysis. The closer the samples are, the shorter the branch length is, the more similar the species compositions of the two samples are. Different colors represent different groups. (b) Permutational multivariate analysis of variance. R2 represents the interpretation degree of sample difference between different groups. A larger R2 indicates that grouping has a higher interpretation degree to the difference, the group difference is higher, and P value less than 0.05 indicate a high reliability of the test. Y-axis represents beta distance. The box above "All between" represents the beta distance data of samples between all groups, while the box above "All within" represents the beta distance data of samples within all groups. The box below represents the beta distance data of samples within different groups.
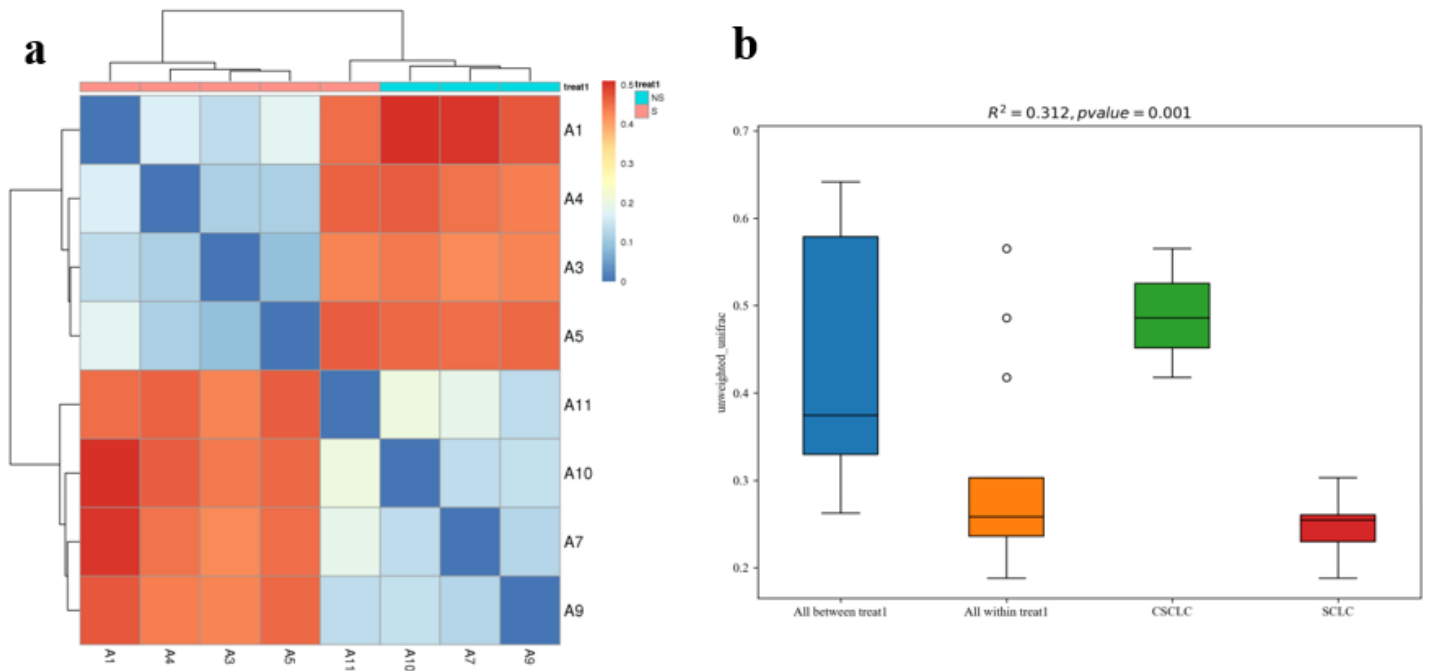
Figure 4

Smoker and non-smoker bacterial beta diversity analysis. (a) Sample clustering heatmap. Color gradient from blue to red represents the distance between samples from close to far. Treat1 represent different groups. X-axis and Y-axis represent different samples. (b) Permutational multivariate analysis of variance. R2 represents the interpretation degree of sample difference between different groups. A larger R2 indicates that grouping has a higher interpretation degree to the difference, the group difference is higher, and P value less than 0.05 indicate a high reliability of the test. Y-axis represents beta distance. The box above "All between" represents the beta distance data of samples between all groups, while the box above "All within" represents the beta distance data of samples within all groups. The box below represents the beta distance data of samples within different groups.
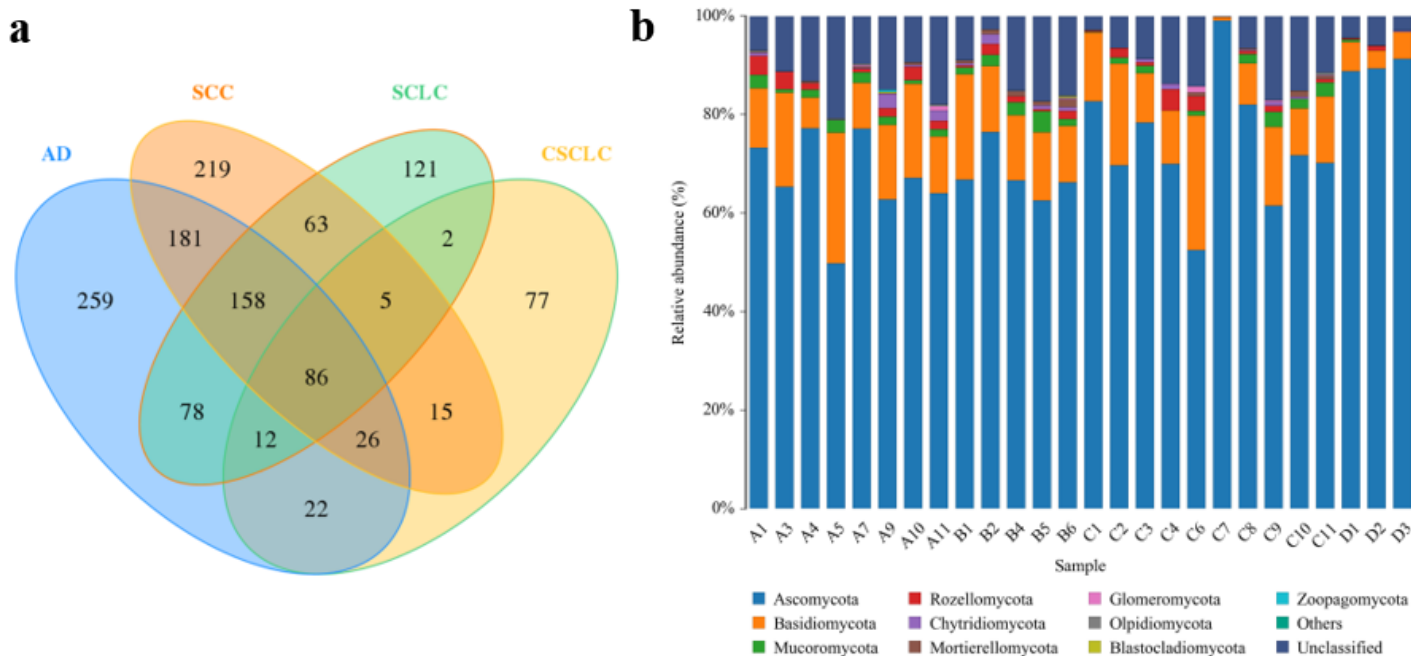
Figure 5

Fungal OTU analysis and histogram of species distribution. (a) OTU-Venn diagram. Different colors are presented by different samples. The number of overlapping area among graphs with different colors is the number of OTU shared by multi-samples. (b) Fungal species distribution histogram of different groups in phylum level. One color represents one phylum, the length of the color block present the relative richness proportion of the phylum.
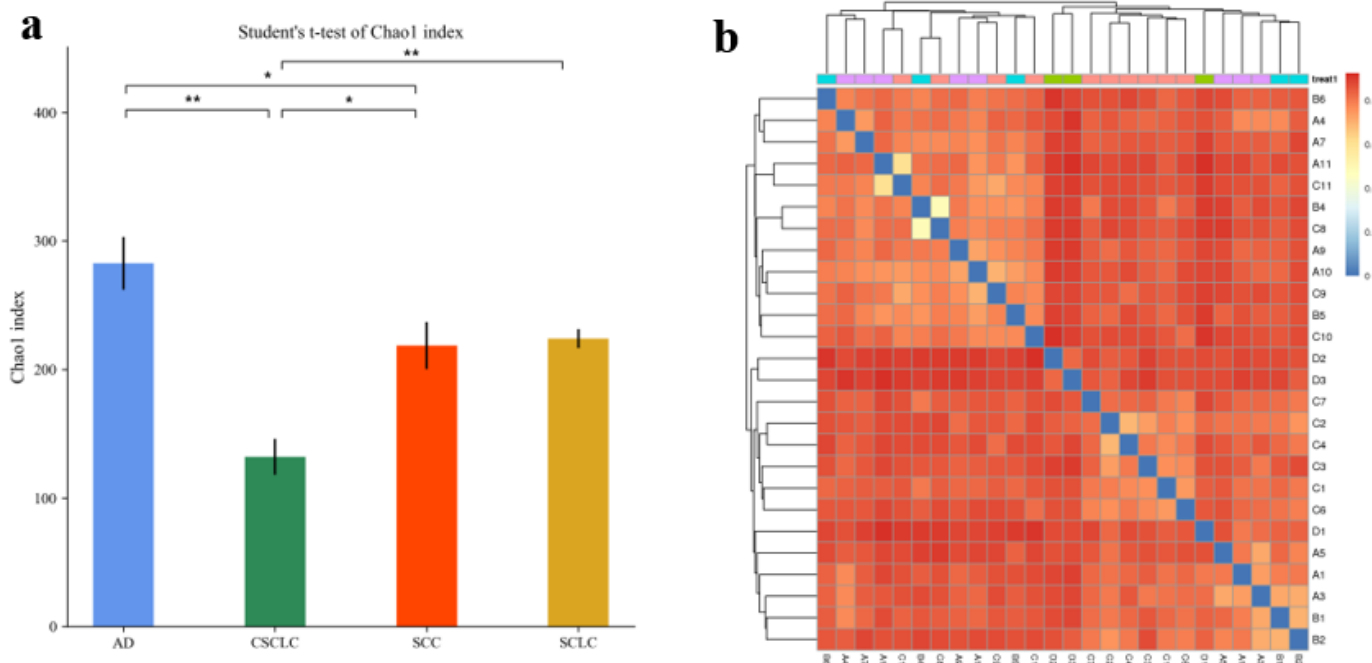

Figure 6

Fungal alpha diversity and beta diversity. (a) Histogram of alpha diversity index differences between groups. Chao1 index measure species richness, i.e. the number of species. Pairwise comparisons were made between each group, * P< 0.05, ** P< 0.01, indicated a significant difference between the two groups. (b) Sample clustering heatmap. Color gradient from blue to red represents the distance between samples from close to far. Treat1 represent different groups. X-axis and Y-axis represent different samples.
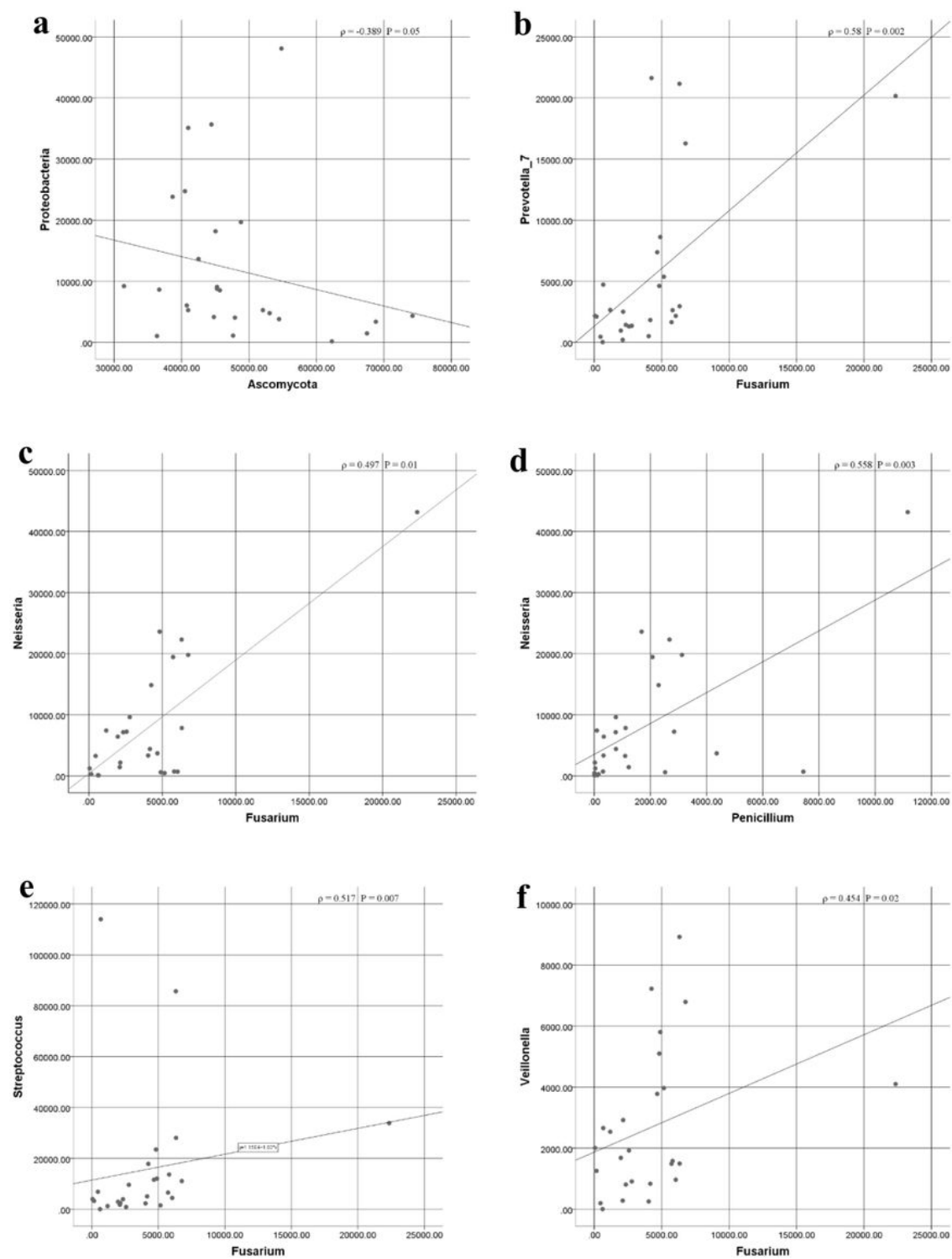

Figure 7

The correlation between bacteria and fungi. The X axis represents the abundance of fungal species, and the Y axis represents the abundance of bacterial species. ρ is Spearman correlation coefficient, and it has statistical significance when P value < 0.05. (a) Correlation analysis of Proteobacteria and Ascomycota. (b) Correlation analysis of Prevotella-7 and Fusarium. (c) Correlation analysis of Neisseria and Fusarium. (d) Correlation analysis of Neisseria and Penicillium. (e) Correlation analysis of Streptococcus and Fusarium. (f) Correlation analysis of Veillonella and Fusarium.