Characteristics of Esophagus Flora in Patients with Esophageal Squamous Cell Carcinoma in Central China

Shuixiang He (✉ dyyyjxk@mail.xjtu.edu.cn)

Xi'an Jiaotong University

Xiao-bo Liu

Xi'an Jiaotong University

Zi-ye Gao

Taihe Hospital

Shu Jin

Taihe Hospital

Sandeep Pandey

Taihe Hospital

Bo Gao

Taihe Hospital

Qiang Tong

Taihe Hospital

Abstract

Microecology may be involved in tumorigenesis and development through the introduction of chronic inflammation and may be related to esophageal squamous cell carcinoma (ESCC). This study was to observe the characteristics of ESCC flora and to preliminarily analyze the key genera in ESCC. A total of 72 ESCC patients and 20 healthy individuals esophageal tissue samples, genomic DNA was extracted, PCR was performed on the bacterial 16SrRNA gene sequence V4 hypervariable region, and bacterial characteristics were analyzed by IlluminMiSeq sequencing. The esophageal flora abundance of ESCC patients (Alpha diversity: Shannon index, P = 0.5088; Simpson index, P = 0.5894; Chao1 index, P = 0.0029) and variability (Beta diversity: PC1 16.62%, P_{PC1} = 0.0034, P_{PC2} =8.5e-08) were significantly higher than HC tissue. The five most abundant phylum in ESCC group were *Firmicutes, Bacteroidietes, Proteobacteria, Fusobacteria* and *Actinobacteria,* while the HC group were *Proteobacteria, Firmicutes, Bacteroidietes, Fusobacteria* and *Actinobacteria*. The top five genus in the ESCC group were *Streptococcus, Prevotella, Haemophilus, Neisseria* and *Veillonella*; while the top five genus in the HC group were *Streptococcus, Neisseria, Veillonella, Prevotella* and *Haemophilus*. The ESCC group has advantages over HC in *Aerobic、Anaerobic、Contains-Mobile-Elements、Facultatively-Anaerobic、Forms-Biofilms、Potentially-pathogenic、Stress-Tolerant*, and the difference was statistically significant (P <0.05). The esophageal microbial community structure of ESCC patients is different from that of HC, suggesting that the corresponding changes in esophageal microflora have a certain correlation with ESCC. It indicates that the detection of esophageal microorganisms has potential as a means of early diagnosis or screening of future ESCC.

Introduction

Esophageal carcinoma (EC) ranks seventh in cancer incidence rate and sixth in mortality in the world[1]. EC has two pathological types, namely, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) [2]. China's EC incidence rate ranks in the top five globally [3], with ESCC accounting for more than 90%[4]. The exact cause of EC is unclear [5] the Scientific research project of Hubei Provincial Department of Education. Common risk factors include tobacco, alcohol, low social status, chronic atrophic gastritis, poor oral hygiene, tooth loss, charcoal and wood cooking, high temperature diet, family history of esophageal cancer, etc. [6, 7]. Microecology may be involved in tumorigenesis and development through the introduction of chronic inflammation [8, 9]. Microecology may be related to ESCC [3]. Known risk factors, such as tobacco, may be involved in the pathogenesis of ESCC by inducing changes in microecology [10].

Human microbiota (HM) includes trillions of bacteria, archaea, fungi and viruses that interact with the human body[11]. Normal HM plays a role in human nutrition and drug metabolism, maintaining the integrity of the intestinal mucosal barrier, immune regulation, and protection against pathogens [12], and its disorders are related to a variety of diseases, including tumors[13].The number of bacteria in the digestive tract is 10 times the total number of human cells[11]. Most of the bacteria have a specific

spatial distribution and cannot be cultured. The oral cavity, esophagus, and to rectum are diverse and different in number [14].

Research on microecology and ESCC is still in the early stage. Knowledge on the functions of microecology and its host is limited [15]. To explore the microecology status of ESCC patients can help further studies on ESCC. In this paper, illuminaHiseq technology was used to observe the characteristics of ESCC flora and to preliminarily analyze the key genera in ESCC. It provides theoretical basis for the study on ESCC flora.

1. Materials And Methods

1.1 Sample source

Seventy-two ESCC patients who were hospitalized in the digestive endoscopy center and thoracic surgery of taihe hospital from July 2018 to July 2019 were regarded as group B and were all confirmed by pathology. Twenty healthy individuals in the center during the same period were regarded as healthy control (HC, group A) by endoscope examination. There was no significant difference in gender and age between the two groups (Supplementary Table S1).

The study protocol was reviewed and approved by the Taihe Hospital Ethics Committee, and all patients written consent were obtained before participating in this research. Moreover, this study was conducted following the provisions of the Helsinki Declaration.

1.2. Inclusion and exclusion criteria

Inclusion criteria: (1) Age ≥ 18 years, patients in group B histological diagnosis of ESCC; (2) Patients in good condition with no metabolic diseases, such as diabetes, hyperlipidemia, and infectious diseases; (3) Did not take antibiotics, acid inhibitors, and probiotics that affect the esophagus flora in the past 2 months; (4) No special eating habits; (5) No serious liver and kidney disease and immune deficiency; and (6) Informed consent for this study.

Exclusion criteria: (1) Received treatments that affect microecology in the past 2 months; (2) Combined with autoimmune diseases;(3)Past or present tumors other than esophageal cancer; (4) Incomplete information; and (5) Researchers consider them inappropriate for the study.

1.3 Standard Collection

Endoscopy center specimens: 6−8 h after fasting, the patients were enrolled for gastroscopy. They gargled with warm water before the examination, and 4−8 specimens were taken after esophageal lesions were found. Two specimens were immediately placed in liquid nitrogen after being labeled, and the remaining tissues were routinely stored in formalin and sent to the pathology department. The researchers in group A collected 2 esophageal mucosa, and the treatment was performed as previously

described. Surgical specimens were frozen and sent to the pathology department after ex vivo, and appropriate samples were selected for subsequent research in accordance with the inclusion criteria.

1. 1.4 Detection platform
2. 1.4.1 DNA extraction

Based on SDS Lysis Buffer freeze-thaw method, the genome DNA of all samples was extracted by PowerMax extraction kit (MoBio Laboratories, Carlsbad, CA, USA) and frozen at -20 °C for later use. The number and quality of DNA were measured by NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

1.4.2 16 SrDNA Amplicon Pyrosequencing

Polymerase chain reaction (PCR) amplification of 16SrRNA gene in V4 region was conducted by Barcode using a 7-bp specific sequence. Primer 515F (5'- GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3'). The PCR reaction system included 50µl:25 µl phusion high-fidelity PCR master mix with HF buffer, F/R primers at 3 µl each (10 uM), 10 µl DNA template, and 6 µl ddH$_2$O. Loop parameters were as follows: predegenerate at 98 °C for 30 s; and 25 cycles of degeneration at 98 °C for 15 s, 58 °C annealing for 15 s, 72 °C extension for 15 s; 72 °C final extension for 1 min. PCR product was purified by AMPure XP Beads (Beckman Coulter, Indianapolis, IN) and quantified with PicoGreen dsDNA. Assay Kit (Invitrogen, Carlsbad, CA, USA). Sequencing was performed on the Illumina HiSeq4000 pair-end 2 × 150 bp platform (Illimina Inc., USA).

1.4.3 OUT clustering and species annotation

Vsearch version 2.4.4 was used to analyze operational taxonomic units (OTUs). OTUs was clustered by 97% similarity. The species of representative sequence was annotated by SILVA128 database, and the composition of flora of all samples at Phylum, Class, Order, Family, Genus level were counted. The abundance and classification of OTUs was recorded.

1.4.4 Bioinformatics analysis and statistical analyses

Sequence data analysis mainly done by using QIIME (Quantitative Insights Into Microbial Ecology, version 1.8.0, http://qiime.org/) and R package (www.r-project.org, version 3.2.0). The Alpha diversity index at OTU level was calculated by using QIIME software, and the samples' OTU abundance and uniformity were compared. Beta diversity examines the similarity of the colony structure of different samples, and clusters are better in the sample set.

UniFrac distance measure was calculated by using QIIME software [16], and a principal component analysis (PCA), principal co-ordinates analysis (PCoA), and nonmetric multidimensional scaling (NMDS) maps were generated to analyze the Beta diversity of different samples in microecology.

The t-test and the Monte Carlo permutation examination were used to draw a box plot to compare the difference in Unifrac distance between groups. The differentiation markers of microecology flora structure between groups were evaluated by PERMANOVA (Permutational multivariate analysis of variance) [17]. Vegan from R package was used to visualize classified groups and abundance based on the MEGAN software [18, 19] and Graphical Phylogenetic Analysis (GraPhlAn) [20]. A Venn graph was generated based on R package "Venn Diagram". The common OTUs between samples or groups were visualized. The difference in the classification level within or between groups was compared by R package Kruskal method.

Linear discriminant analysis effect size (LEfSe) analysis [21] combines linear discriminant analysis with Kruskal–Wallis and Wilcoxon rank sum tests, to screen out intergroup biomarkers, such as species, with significant difference. By Random Forest Analysis, R package "random Forest" default setting was used to compare the differences between groups. Microbial function was predicted on the basis of PICRUSt [22]. Metagenomic Profiles (STAMP) software package version 2.1.3 [23] was used for further analysis of the output file. Parallel-META 3 (version 3.3.2) was used to complete the Beta-diversity analysis in species and functions based on Meta-Storms distance. Statistical analysis was performed using SPSS 21.0 version (SPSS Inc., Chicago, USA). When $P < 0.05$, the difference was considered statistically significant.

2 Results

2.1 Sample sequencing data

The offline data met the test requirements after using sequence tags, clean-tags, and OUT. After clustering with 97% similarity, 4003 OUTs were obtained, of which 1290 were in the HC group (group A), 3763 in the ESCC group (group B), and 1050 OUTs in both groups (Fig. 1).

2.2 Alpha diversity analysis

Shannon index and Simpson index can predict species diversity. By comparison, the Shannon and Simpson indexes of the ESCC group were higher than those of the HC group, but the difference was not statistically significant ($P > 0.05$). Chao1 index is used to estimate the total number of species in the community. The chao1 index of the two groups was statistically significant ($P < 0.05$), indicating that the types of microorganisms in the two groups are not different, but the number of microorganisms in ESCC group was higher than that in HC group (Fig. 2).

2.3 Beta diversity analysis

Beta diversity examines the similarity of the colony structure of different samples, and clusters are better in the sample set. PCoA can find the most important coordinates in the distance matrix and observe the differences between individuals or groups. PC1 and PC2 represent suspected influencing factors for the shift in the microbial composition of the two groups. PC1 in the two groups was 16.62%, $P_{PC1}=0.0034$, suggesting a difference in the composition of microecological flora between the two groups (Fig. 3A).

However, NMDS analysis is more stable than PCoA for complex structured data ranking results. The two groups in this study can be distinguished (Fig. 3B).

2.4 Differential LEfSe analysis

Phascolarctobacterium, Dialister, Clostridiales, S24_7, Rikenellaceae, Odoribacter and other bacterial genus in ESCC patients were higher than HC, while *Neisseria、Neisseriales、Bradyrhizobiaceae、F16、Helicobacteraceae、Helicobacter* bacterial genus were higher in HC patients than in ESCC patients, the difference was statistically significant ($P < 0.05$, Fig. 4A and B).

2.5 Structural analysis of esophagus flora in two groups

Because the esophageal mucosa flora did not conform to the normal distribution. Thus, Wilcox test was used to analyze the relative abundance of the two groups of bacteria. Significant analysis was performed at the above levels. This study only shows the results at the phylum and genus levels, and the analysis of the microbial flora composition at other levels is not listed (Supplementary Table S 2–4).

2.5.1 Analysis of microbial flora composition at the phyla level

The two samples consisted of 12 phyla. The five most abundant phyla in ESCC group are *Firmicutes, Bacteroidietes, Proteobacteria, Fusobacteria* and *Actinobacteria.* The five most abundant phyla in the HC group are *Proteobacteria, Firmicutes, Bacteroidietes, Fusobacteria* and *Actinobacteria*. The significant different phylum between two groups were *Cyanobacteria, Proteobacteria, SR1* and *TM7* ($P < 0.05$). Specifically, compared with the HC group, the abundance of *Cyanobacteria* in the ESCC group is increased, while the abundance of *Proteobacteria, SR1* and *TM7* is decreased ($P < 0.05$, Table 1)). This indicates that the colonization flora changes dynamically during the occurrence and development of ESCC, which lays a foundation for further research on the role of colonization flora in the occurrence and development of ESCC..

Table 1
Significant differences in phylum levels between the two groups (mean ± sd)

name	HC	ESCC	*P*-value*
Cyanobacteria	0.000 ± 0.001	0.003 + 0.008	0.041
Proteobacteria	0.332 ± 0.157	0.217 ± 0.181	0.022
SR1	0.003 ± 0.003	0.001 ± 0.004	0.023
TM7	0.005 ± 0.004	0.002 ± 0.004	0.001
Notes: HC: healthy controls, ESCC: esophageal squamous cell carcinoma; *:Wilcoxon.test			

2.5.2 Analysis of microbial flora composition at genus level

The top five genera in the ESCC group are *Streptococcus, Prevotella, Haemophilus, Neisseria* and *Veillonella*; while the top five genera in the HC group were *Streptococcus, Neisseria, Veillonella, Prevotella* and *Haemophilus* ($P < 0.05$).The comparison shows that the abundance of *Bifidobacterium, Collinsella, Parabacteroides, Paraprevotella, Coprococcus, Lachnospira, Roseburia, Faecalibacterium, Ruminococcus, Dialister, Megamonas*, and *Megamonas* were higher in ESCC group, and the abundance of *Prevotella, Lysinibacillus, Streptococcus, Megasphaera, Veillonella, Leptorichia, Ralstonia, Neisseria, Helicobacter, Haemophilus* and *Acinetobacter* were lower than the abundance of HC group ($P < 0.05$, Table 2).

Table 2
Significant differences in genus levels between the two groups (mean ± sd)

name	HC	ESCC	*P*-value*
Actinomyces	0.015 ± 0.012	0.003 ± 0.004	0.000
Rothia	0.011 ± 0.014	0.007 ± 0.025	0.001
Bifidobacterium	0.001 ± 0.003	0.013 ± 0.041	0.000
Collinsella	0.000 ± 0.001	0.002 ± 0.004	0.046
Parabacteroides	0.003 ± 0.008	0.007 ± 0.010	0.001
Porphyromonas	0.015 ± 0.012	0.015 ± 0.027	0.005
Paraprevotella	0.000 ± 0.000	0.001 ± 0.002	0.000
Prevotella	0.048 ± 0.038	0.034 ± 0.066	0.002
Lysinibacillus	0.015 ± 0.017	0.006 ± 0.014	0.013
Streptococcus	0.120 ± 0.126	0.071 ± 0.081	0.005
Coprococcus	0.000 ± 0.001	0.001 ± 0.003	0.001
Lachnospira	0.002 ± 0.005	0.005 ± 0.007	0.044
Roseburia	0.001 ± 0.002	0.004 ± 0.006	0.013
Faecalibacterium	0.004 ± 0.011	0.020 ± 0.033	0.012
Ruminococcus	0.001 ± 0.004	0.005 ± 0.008	0.001
Dialister	0.001 ± 0.009	0.006 ± 0.013	0.004
Megamonas	0.004 ± 0.012	0.008 ± 0.020	0.027
Megasphaera	0.004 ± 0.006	0.002 ± 0.005	0.034
Phascolarctobacterium	0.002 ± 0.003	0.008 ± 0.011	0.025
Veillonella	0.051 ± 0.051	0.024 ± 0.032	0.001
Leptotrichia	0.009 ± 0.010	0.006 ± 0.017	0.002
Sutterella	0.001 ± 0.004	0.003 ± 0.005	0.005
Ralstonia	0.011 ± 0.016	0.001 ± 0.004	0.000
Neisseria	0.071 ± 0.096	0.025 ± 0.057	0.000
Helicobacter	0.009 ± 0.014	0.002 ± 0.016	0.000

Notes: HC: healthy controls,ESCC: esophageal squamous cell carcinoma; *:Wilcoxon.test

name	HC	ESCC	*P*-value*
Haemophilus	0.040 ± 0.035	0.032 ± 0.047	0.018
Acinetobacter	0.003 ± 0.004	0.002 ± 0.009	0.030
Notes: HC: healthy controls,ESCC: esophageal squamous cell carcinoma; *:Wilcoxon.test			

Table 3
Comparison of Phenotype Classification Based on BugBase

BugBase	KS_pvalue
Aerobic	0.000826148
Anaerobic	0.000439153
Contains_Mobile_Elements	7.84E-05
Facultatively_Anaerobic	0.03818119
Forms_Biofilms	0.0180091
Gram_Negative	0.08919101
Gram_Positive	0.08919101
Potentially_Pathogenic	0.009671794
Stress_Tolerant	0.006169115

2.6 Random forest classification Tree genus classification effect and ROC curve (genus level)

The random forest method was used to select the top60 species to establish a model (Fig. 5), and then the ROC curve was used to verify that our model is reliable and can effectively distinguish two groups of samples (AUC = 0.90, Fig. 6).

2.7 Comparison of Phenotypic Classification Based on BugBase to Predict the Function of Microbial Metabolism

Bugbase mainly performs phenotypic prediction, including Gram-positive, Gram-negative, biofilm formation, pathogenicity, mobile components, oxygen demand, including anaerobic, aerobic, facultative bacteria) and oxidative stress tolerance[20, 24]. Our tudies show that the ESCC group has advantages over HC in *Aerobic, Anaerobic, Contains-Mobile-Elements, Facultatively-Anaerobic, Forms-Biofilms, Potentially-pathogenic, Stress-Tolerant*, and the difference was statistically significant (*P* < 0.05, Table 4).

3 Discussion

The esophagus is located between the flora-rich oropharynx and stomach,and has less species than oral populations [24]. In 1990s, culture studies have suggested that aseptic esophagus populations or a small

amount of bacteria that are swallowed or gastric reflux [25]. Pei *et al* [26] confirmed that the esophagus flora mainly includes six phyla: *Firmicutes, Bacteroides, Actinobacteria, Proteobacteria, Fusobacteria* and *TM7*. Yang *et al* [27] divided the esophagus flora into two types: I and II. The former is composed of Gram-positive bacteria mainly distributed in the esophagus of normal people, and the latter is composed of Gram-negative bacteria mainly existing in Barrett's esophagus (BE) or esophagitis patients.

Li [28] and their colleagues found that the normal esophagus flora is similar to the oral cavity flora. Firmicutes is the most abundant phyla, and no specific flora exists in different parts of the esophagus. Recent studies have shown that [15, 29] the abundance of normal esophageal archaea, phages, and other flora is low, mainly including *Streptococcus, Prevotella, Veillonella, Clostridium, Haemophilus, Neisseria* and *Porphyromonas*. In the present study, the HC group includes 12 phyla. The most abundant phytobacteria are *Bacteroidia, Bacilli, Gammaproteobacteria, Clostridia* and *Betaproteobacteria*, which are basically consistent with the report except for some differences with regard the impact of different regions on dietary structure and living environment.

Understanding the microecology of ESCC is highly important. Peters Ab *et al* [30] found that oral *Porphyromonas gingivalis* increases the risk of ESCC. Chen *et al* reported that the oral flora diversity of ESCC patients is reduced and that the *Lautropia, Bulleidia, Catonella, Corynebacterium, Peptococcus* and *Cardiobacterium* abundance in ESCC patients are less than that in non-ESCC. Wang *et al*[8] confirmed that *Actinomyces* and *Atopobium* were associated with high risk in saliva of genus level ESCC patients, while *Fusobacterium* and *Porphyromonas* were associated with healthy people. Esophageal squamous dysplasia (ESD) is a precancerous lesion of ESCC. Yu *et al* [32] the Scientific research project of Hubei Provincial Department of Education found that the low microbial abundance and composition of the upper digestive tract microbiota in the Chinese population are related to ESD. Yamamura *et al* [33] confirmed that *Fusobacterium nucleus* is significantly higher in 23% of EC (EAC, ESCC, etc.) than in non-cancerous non-tumor tissue (P = 0.021) and is related to EC severity and prognosis. Liu *et al* [34]reported that the abundance of esophageal phyla *Bacteroidetes, Firmicutes*, and *Spirochaetes* is increased in ESCC patients, *proteobacteria* is decreased in patients with positive lymphatic metastasis. *Prevotella* and *Treponema* were more abundant in positive patients than in negative patients, which were related to the poor prognosis of the patient and may be indicators of prognosis in ESCC. Gao *et al* [35] found that *Porphyromonas gingivalis* is often located in ESCC and ESD, and its infection is related to disease severity and prognosis, which can be used as a marker for the diagnosis and outcome of ESCC. Passing the graduate student flora may be beneficial to the early diagnosis, disease evaluation, and prognosis of ESCC.

In the present study, the abundance of *Cyanobacteria* in the ESCC group increased, while the abundance of *Proteobacteria, SR1* and *TM7* decreased. It is suggested that *Proteobacteria, SR1* and *TM7* may have a protective effect on normal esophagus. During the development of ESCC, the flora balance was broken, and *Cyanobacteria* may be involved in the development of ESCC, however, the detailed mechanism needs to be further studied

We found that the abundance of *Bifidobacterium, Collinsella, Parabacteroides, Paraprevotella, Coprococcus, Lachnospira, Roseburia, Faecalibacterium, Ruminococcus, Dialister, Megamonas*, and *Megamonas* were higher in ESCC group, and the abundance of *Prevotella, Lysinibacillus, Streptococcus, Megasphaera, Veillonella, Leptorichia, Ralstonia, Neisseria, Helicobacter, Haemophilus* and *Acinetobacter* were lower than that of HC group. Shao *et al* [36] the Scientific research project of Hubei Provincial Department of Education reported that the ESCC flora is mainly composed of *Firmicutes, Bacteroidetes* and *Proteobacteria. Clostridium* and *Streptococcus* are less abundant in ESCC tissues than in non-tumor tissues. The abundance of *Bifidobacteria, Clostridium prasium*, and *Ruminococcus aureus* in the feces of EC patients was significantly lower than that of HC population (*P*< 0.05) [36]. A study by the University of Chicago [38] found that the intestine contains high abundance Malignant melanoma mice of Bifidobacterium have stronger CD8[+] T cell activity and better efficacy of receiving immune checkpoint inhibitors. Recent studies [39] have shown that *Bifidobacterium* can effectively fight cancer cells and is associated with substantial improvement of gastrointestinal cancer. Liu *et al* [34] found that *Prevotella* and *Treponema* are more abundant in patients with positive lymphatic metastasis in ESCC and that *Streptococcus* and *Prevotella are* associated with poor prognosis in ESCC patients, suggesting that *Streptococcus* and *Prevotella* are prognostic indicators of ESCC. In future research, cell culture systems and animal models should be added to examine the pathogenic role of microorganisms in ESCC. Research on the above-mentioned flora may help to formulate new strategies for ESCC prevention, diagnosis, early intervention and treatment.

Alpha diversity analysis of the two groups of bacteria showed that the diversity of the ESCC group flora was higher than that of the HC group. This result may be related to factors, such as the change in flora during the occurrence and development of ESCC and the relative specificity of different types of tumors. In Beta diversity analysis, the PCoA study found that the red dots representing the EC specimens and the blue dots representing the HC specimens overlap on the coordinate axis, and the factor that caused the two groups of PC1 to separate was 16.62%. NMDS analysis showed that the esophageal bacterial variation in the ESCC group was greater than that in the control group. It was suggested that the ESCC group had high abundance and large variation.

LEfSe analysis found that the high risk of ESCC may be related to *Phascolarctobacterium, Dialister, Clostridiales, S24_7, Rikenellaceae*, and so on. Currently, *Helicobacter pylori* (*H.pylori*) infection is generally associated with EA [40], but whether or not ESCC is related to this infection remains controversial [41, 42]. The presents tudy found that the abundance of *Helicobacter* flora in the EC group was reduced, but this genus has a large number of strains, and the correlation between *H. pylori* and ESCC needs to be studied.

We use the random forest method and select the TOP60 species to establish a model. The ROC curve is used to verify that the model is reliable and can effectively distinguish two groups of samples. Further comparison of the phenotypic types of BugBase found that the ESCC group has advantages over the HC group in metabolism, such as *Aerobic, Anaerobic, Containers-Mobile-Elements, Facultatively-Anaerobic*, and *Forms-Biofilms*, suggesting that the above related flora is rich and can be expanded from bacterial

functions in the later stage. Although the pathogenesis of ESCC is not caused solely by bacterial flora activity, bacterial flora homeostasis may play an important role in the occurrence and development of EC and provide important clues and references for ESCC.

Colletively, data from the present study reveales that the esophageal microbial community structure of ESCC patients is different from that of HC, and the corresponding changes in esophageal microflora maybe a certain correlation with ESCC. It is indicated that the detection of esophageal microorganisms has potential as a means of early diagnosis or screening of future ESCC. However, this study also has shortcomings. First, the ESCC patients included in the study were mostly male, and the number of ESCC was small. Second, the analysis of the flora of early esophageal cancer and precancerous lesions was lacking. Monitoring the changes in the composition of the flora during different stages of tumorigenes is and development will help clarify the mechanism of ESCC. The sample size and pathological type will be increased in our future studies to explore further the role of esophageal flora in ESCC.

Contribution

L-XB, TQ, and H-SX contributed conceptualization, funding acquisition, writing–review, and editing. L-XB and G-ZY wrote the manuscript. JS, PS and GB contributed methodology and software .L-XB TQ collected and detected clinical samples. All authors revised the manuscript and approved the final manuscript.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Data submission

The raw data used in this study were uploaded to Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under Bioproject PRJNA 615067.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin.

2. Baba Y, Iwatsuki M, Yoshida N, Watanabe M, Baba H. Review of the gut microbiome and esophageal cancer: Pathogenesis and potential clinical implications. Ann Gastroenterol Surg. 2017;1:99–104.

3. Abnet CC, Arnold M, Wei WQ. (2018) Epidemiology of Esophageal Squamous Cell Carcinoma. GASTROENTEROLOGY.**154**, 360–373.

4. Wen X, Wen D, Yang Y, Chen Y, Wang G, Shan B. Urban-Rural Disparity in Helicobacter Pylori Infection-Related Upper Gastrointestinal Cancer in China and the Decreasing Trend in Parallel with Socioeconomic Development and Urbanization in an Endemic Area. Annals of Global Health. 2017;83:444–62.

5. Neto AG, Whitaker A, Pei Z. Microbiome and potential targets for chemoprevention of esophageal adenocarcinoma. SEMINARS IN ONCOLOGY. 2016;43:86–96.

6. Liang H, Fan JH, Qiao YL. Epidemiology, etiology, and prevention of esophageal squamous cell carcinoma in China. Cancer Biology Medicine. 2017;14:33–41.

7. Patel K, Wakhisi J, Mining S, Mwangi A, Patel R. (2013) Esophageal Cancer, the Topmost Cancer at MTRH in the Rift Valley, Kenya, and Its Potential Risk Factors. ISRN Oncol.**2013**, 503249.

8. Wang Q, Rao Y, Guo X, Liu N, Liu S, Wen P, Li S, Li Y. Oral Microbiome in Patients with Oesophageal Squamous Cell Carcinoma. Sci Rep. 2019;9:19055.

9. Liu XB, Gao ZY, Sun CT, Wen H, Gao B, Li SB, Tong Q. The potential role of P.gingivalis in gastrointestinal cancer: a mini review. Infect Agent Cancer. 2019;14:23.

10. Vogtmann E, Flores R, Yu G, Freedman ND, Shi J, Gail MH, Dye BA, Wang GQ, Klepac-Ceraj V, Paster BJ, Wei WQ, Guo HQ, Dawsey SM, Qiao YL. Abnet CC. (2015) Association between tobacco use and the upper gastrointestinal microbiome among Chinese men. Cancer Causes Control.26, 581–8.

11. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J. A human gut microbial gene catalogue established by metagenomic sequencing. NATURE. 2010;464:59–65.

12. (2012) Structure, function and diversity of the healthy human microbiome. NATURE.486, 207–214.

13. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8:51.

14. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Nageshwar, RD. Role of the normal gut microbiota. World J Gastroenterol. 2015;21:8787–803.

15. Deshpande NP, Riordan SM, Castano-Rodriguez N, Wilkins MR, Kaakoush NO. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. Microbiome. 2018;6:227.

16. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. NATURE METHODS. 2010;7:335–6.

17. Anderson MJ, Ellingsen KE, McArdle BH. Multivariate dispersion as a measure of beta diversity. ECOLOGY LETTERS. 2006;9:683–93.

18. Huson DH, Weber N. Microbial community analysis using MEGAN. Methods Enzymol. 2013;531:465–85.

19. Mitra S, Stark M, Huson DH. Analysis of 16S rRNA environmental sequences using MEGAN. BMC GENOMICS. 2011;12(Suppl 3):17.

20. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;3:e1029.

21. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. GENOME BIOLOGY. 2011;12:R60.

22. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega TR, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. NATURE BIOTECHNOLOGY. 2013;31:814–21.

23. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. BIOINFORMATICS. 2014;30:3123–4.

24. Yang L, Chaudhary N, Baghdadi J, Pei Z. Microbiome in reflux disorders and esophageal adenocarcinoma. CANCER JOURNAL. 2014;20:207–10.

25. Gagliardi D, Makihara S, Corsi PR, Viana AT, Wiczer MV, Nakakubo S, Mimica LM. Microbial flora of the normal esophagus. DISEASES OF THE ESOPHAGUS. 1998;11:248–50.

26. Pei Z, Yang L, Peek RM Jr, Pride LS DT and Blaser MJ. Bacterial biota in reflux esophagitis and Barrett's esophagus. World J Gastroenterol. 2005;11:7277–83.

27. Yang L, Lu X, Nossa CW, Francois F, Peek RM, Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. GASTROENTEROLOGY. 2009;137:588–97.

28. Dong L, Yin J, Zhao J, Ma SR, Wang HR, Wang M, Chen W, Wei WQ. Microbial Similarity and Preference for Specific Sites in Healthy Oral Cavity and Esophagus. Front Microbiol. 2018;9:1603.

29. El-Zimaity H, Di Pilato V, Novella RM, Brcic I, Rajendra S, Langer R, Dislich B, Tripathi M, Guindi M, Riddell R. (2018) Risk factors for esophageal cancer: emphasis on infectious agents. Ann N Y Acad Sci.

30. Peters BA, Wu J, Pei Z, Yang L, Purdue MP, Freedman ND, Jacobs EJ, Gapstur SM, Hayes RB, Ahn J. Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers. CANCER RESEARCH. 2017;.**77**:6777–87.

31. Chen X, Winckler B, Lu M, Cheng H, Yuan Z, Yang Y, Jin L, Ye W. Oral Microbiota and Risk for Esophageal Squamous Cell Carcinoma in a High-Risk Area of China. PLoS One. 2015;10:e143603.

32. Yu G, Gail MH, Shi J, Klepac-Ceraj V, Paster BJ, Dye BA, Wang GQ, Wei WQ, Fan JH, Qiao YL, Dawsey SM, Freedman ND, Abnet CC. Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. Cancer Epidemiol Biomarkers Prev. 2014;23:735–41.

33. Yamamura K, Baba Y, Nakagawa S, Mima K, Miyake K, Nakamura K, Sawayama H, Kinoshita K, Ishimoto T, Iwatsuki M, Sakamoto Y, Yamashita Y, Yoshida N, Watanabe M, Baba H. Human Microbiome Fusobacterium Nucleatum in Esophageal Cancer Tissue Is Associated with Prognosis. CLINICAL CANCER RESEARCH. 2016;22:5574–81.

34. Liu Y, Lin Z, Lin Y, Chen Y, Peng XE, He F, Liu S, Yan S, Huang L, Lu W, Xiang Z, Hu Z. Streptococcus and Prevotella are associated with the prognosis of oesophageal squamous cell carcinoma. JOURNAL OF MEDICAL MICROBIOLOGY. 2018;67:1058–68.

35. Gao S, Li S, Ma Z, Liang S, Shan T, Zhang M, Zhu X, Zhang P, Liu G, Zhou F, Yuan X, Jia R, Potempa J, Scott DA, Lamont RJ, Wang H, Feng X. Presence of Porphyromonas gingivalis in esophagus and its association with the clinicopathological characteristics and survival in patients with esophageal cancer. Infect Agent Cancer. 2016;11:3.

36. Shao D, Vogtmann E, Liu A, Qin J, Chen W, Abnet CC, Wei W. Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. CANCER. 2019;125:3993–4002.

37. Li NN, Bai CM, Zhao L, Ge YP. [Gut Microbiome Differences between Gastrointestinal Cancer Patients and Healthy People]. Zhongguo Yi Xue Ke Xue Yuan Xue Bao. 2019;41:636–45.

38. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, Benyamin FW, Lei YM, Jabri B, Alegre ML, Chang EB, Gajewski TF. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. SCIENCE. 2015;350:1084–9.

39. Bahmani S, Azarpira N, Moazamian E. Anti-colon cancer activity of Bifidobacterium metabolites on colon cancer cell line SW742. Turkish Journal of Gastroenterology. 2019;30:835–42.

40. Tomasello G, Giordano F, Mazzola M, Jurjus R, Jurjus A, Damiani P, Nobile S, Carini F, Leone A. Helicobacter pylori and Barretts esophagus: a protective factor or a real cause? J Biol Regul Homeost Agents. 2017;31:9–15.

41. Vohlonen IJ, Hakama M, Harkonen M, Malila N, Pukkala E, Koistinen V, Sipponen P. Oesophageal cancer incidence in 20-year follow-up in a population-based sample of 12 000 middle-age men with or without Helicobacter pylori infection in Finland. GUT. 2018;67:1201–2.

42. Poyrazoglu OB, Dulger AC, Gultepe BS. Helicobacter Pylory infection in patients with esophageal squamous cell carcinoma. Clinics. 2017;72:150–3.
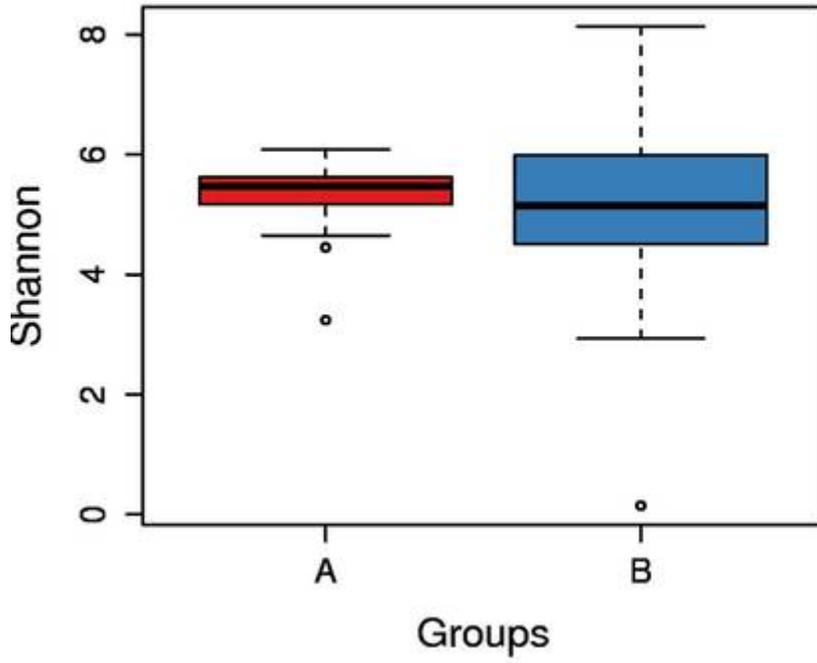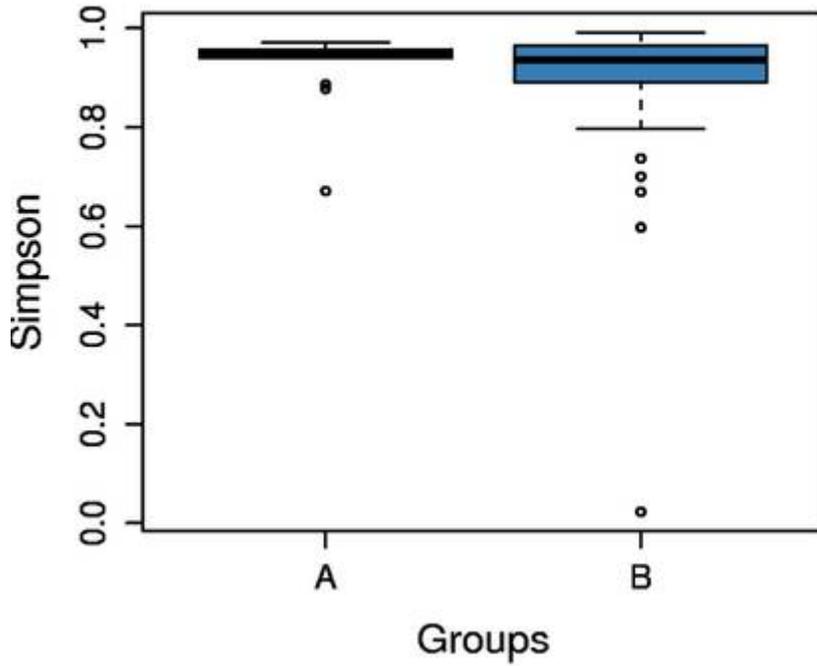
Figures
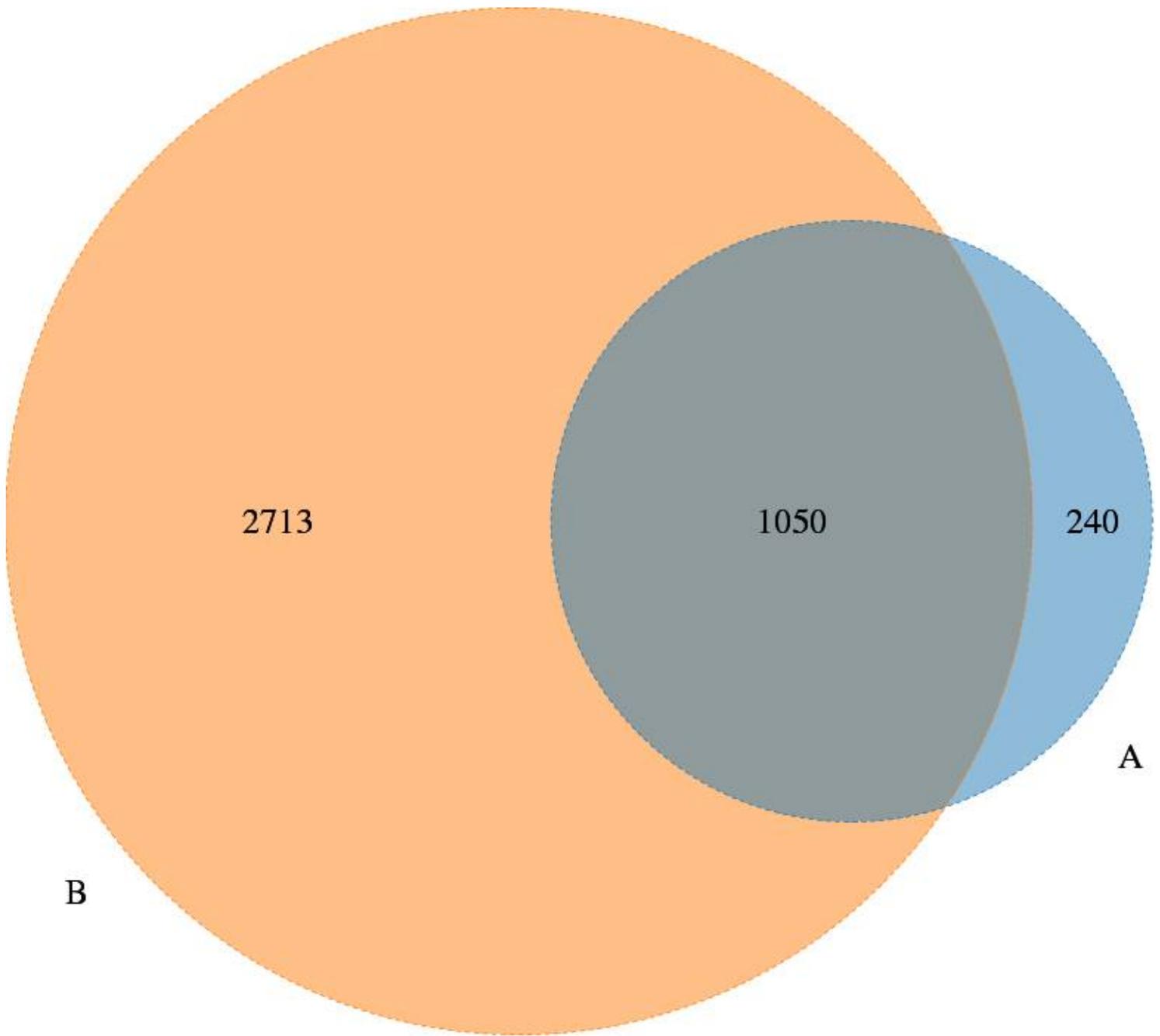
Figure 1

Venn diagram showing the distribution of operational taxonomic units (OTUs) detected in the two groups
Notes: blue:HC group, yellow: ESCC group.

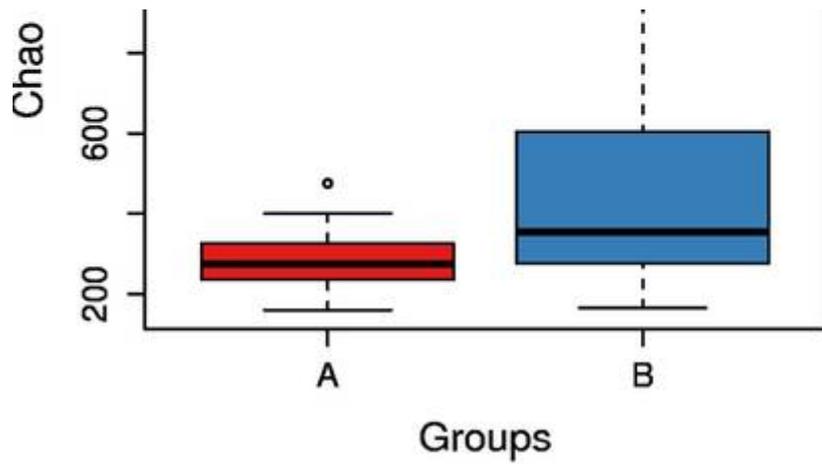**P= 0.5088**

P= 0.5894

P= 0.0029

Figure 2

The alpha microbial diversity box diagrams Notres: From top to bottom are shannon, simpson and chao1, with the pvalue of the statistical test marked at the top. The abscissa is the name of the group, and the ordinate is the value of the alpha diversity index under different groups. The box chart can display 5 statistics (minimum value, first quartile, median, third quartile, and maximum value, that is, 5 lines from bottom to top). Outliers are marked with "o". P-value is the value of the rank sum test corresponding to the Alpha diversity index
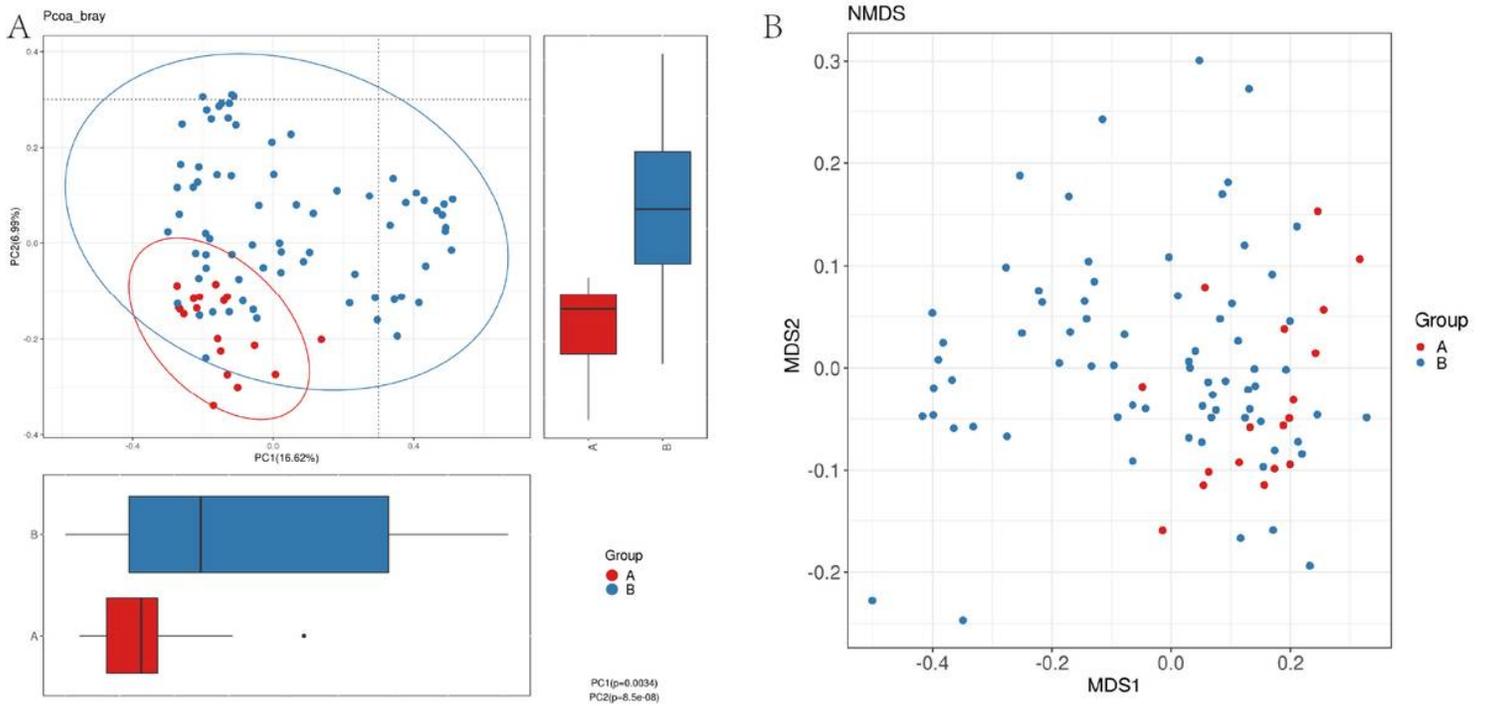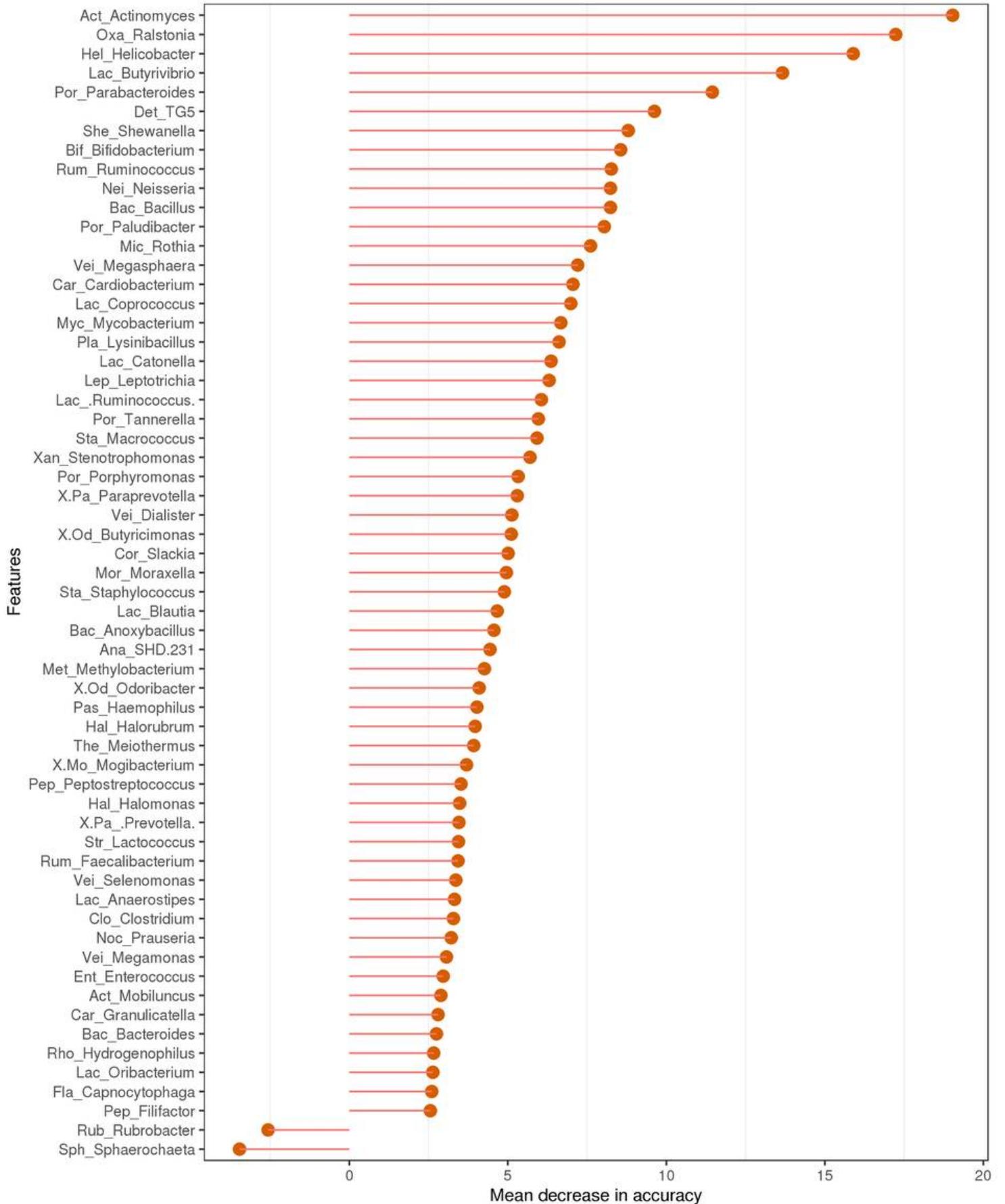
Figure 3

The beta microbial diversity box diagrams (A).PCoA plots for unweighted unifrac distances of the variation in microbiota composition detected in ESCC and HC groups (remove the sample name). Notes: The horizontal box diagram is the distribution of the values grouped on the first column coordinates, the vertical box diagram is the distribution of the values grouped on the second principal coordinates, and the values of the two vectors are marked in the lower right corner. Each point in the figure represents a sample, and the points of the same color come from the same group, and the distance reflects the similarity of the samples. (B) .Scatter plot of two sets of NMDS analysis results. Notes: Each point in the figure represents a sample,. The distance from the response sample is similar. The greater the distance between the two points, the greater the difference in community structure between the two groups. The blue represents the ESCC group, and the red represents the HC group.
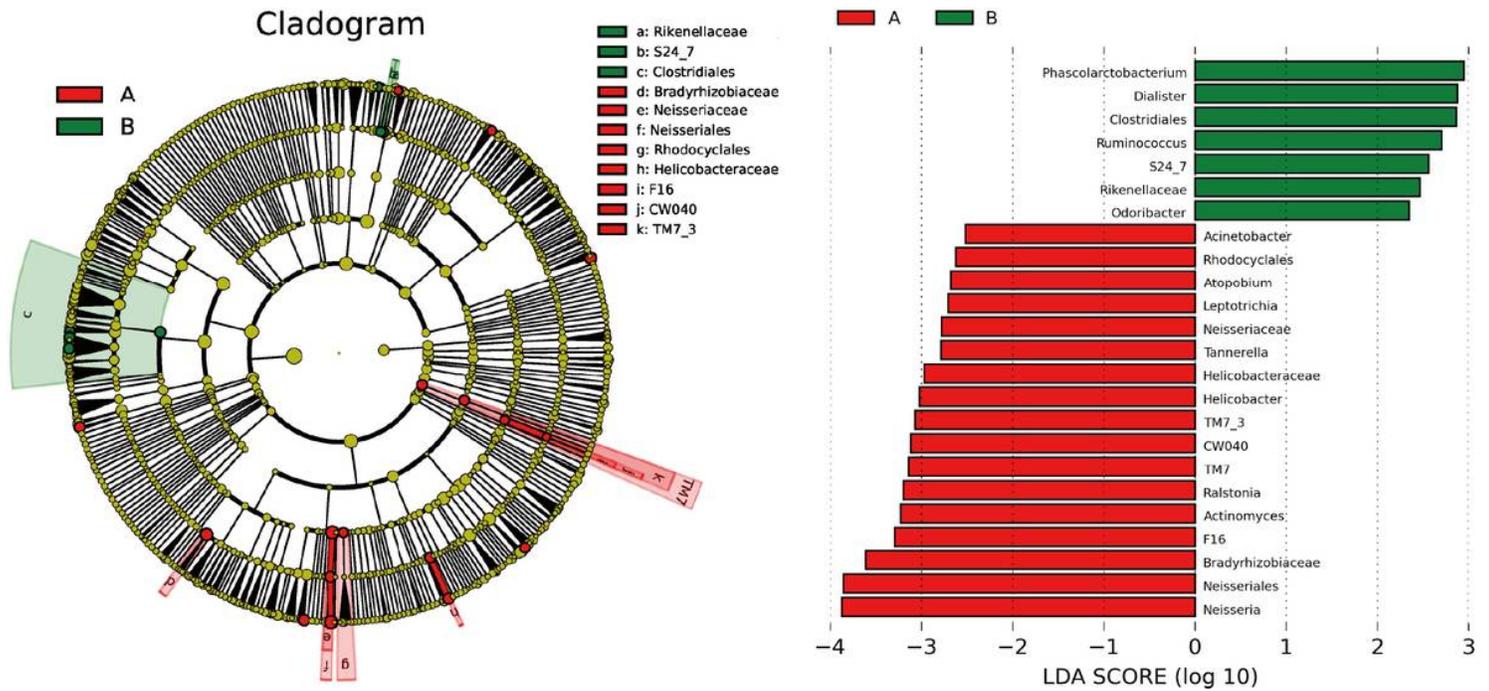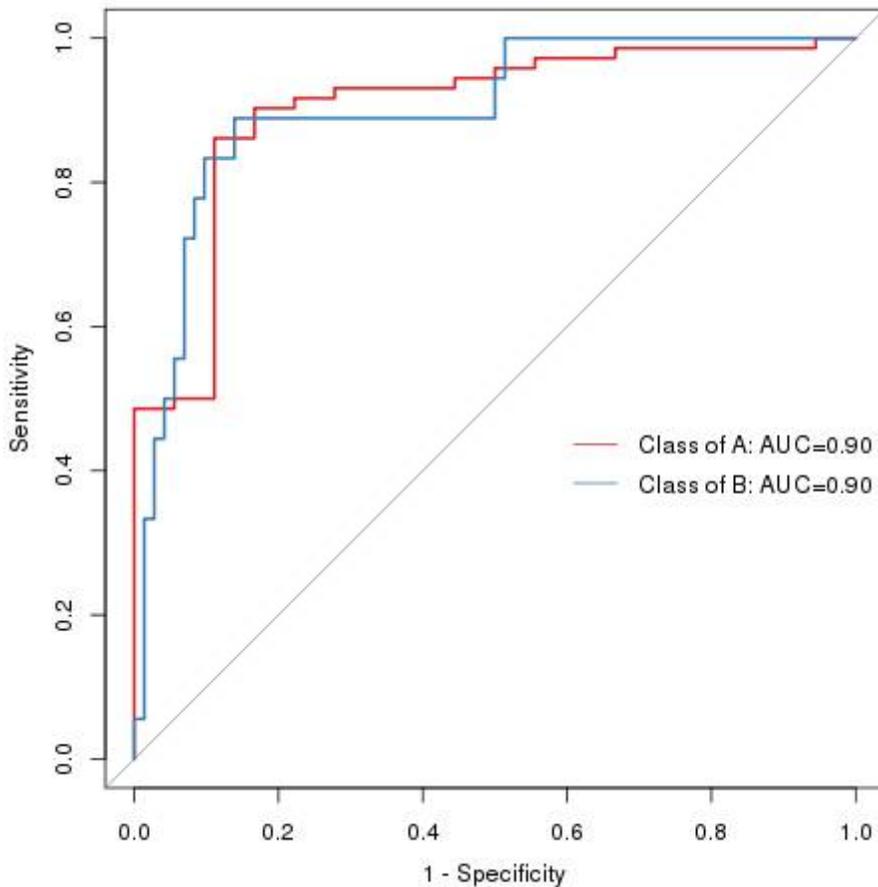
Figure 4

Species composition analysis. (A). Cladogram view of the representative microbial structure among the two groups, (B). Differential bacteria between the two groups by LEfSe analysis Notes: Dominant microbial classes for the specific groups are represented by different colors. The diameter of each dot is proportional to the OTU abundance. Red represents the HC group and green represents the ESCC group. Significantly different logarithmic LDA score was set to 2. Only the groups with significant differences are shown in the figure.

Random-Forest Score (error rate = 45.14%)

Figure 5

Species importance map Notes:The abscissa is the importance level, and the ordinate is the species name sorted according to importance. The figure reflects the genus of bacteria in the classifier that plays a major role in the classification effect, arranged from largest to smallest. Error rate indicates the error rate of the random forest method for predicting classification using the features below. Higher indicates that the classification accuracy based on the genus characteristics is not high, and the genus characteristics may not be obvious between the groups. Taking all levels as an example, the top 60 species are used for the drawing.

Figure 6

The ROC curve used to analyze the clinical accuracy of using differential bacteria obtained from the ESCC group and HC group for the diagnosis of ESCC Notes: On the ROC curve, the point closest to the upper left of the graph is the critical value with higher sensitivity and specificity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS4.docx

- TableS3.docx
- TableS2.docx
- TableS1.docx