

Comparative Study on the Esophageal Microflora of Different Tissue Types in Patients With Esophageal Squamous Cell Carcinoma

Xiaobo liu

First Affiliated Hospital of Xi'an Jiaotong University

Ziye Gao

Taihe Hospital

Wen Xu

Taihe Hospital

JianChao Meng

Taihe Hospital

Chuantao Sun

Taihe Hospital

Hui Wen

Taihe Hospital

Qiang Tong (✉ xinl2012@163.com)

Taihe Hospital

Shuixiang He

First Affiliated Hospital of Xi'an Jiaotong University

Research Article

Keywords: Esophageal squamous cell carcinoma, Postoperative tissue, Esophageal mucosa, 16sRNA, High-throughput sequencing

Posted Date: February 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-153745/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Comparative study on the esophageal microflora of different tissue types in patients with esophageal squamous cell carcinoma

Xiao-bo Liu^{1,2,3}, *Zi-ye Gao*⁴, *Wen Xu*², *Jian-Chao Meng*⁴, *Chuan-tao Sun*², *Hui Wen*²,

Qiang Tong^{2*}, *Shui-xiang He*^{1*}

1.Department of Gastroenterology, the First Affiliated Hospital of Medical College, Xi'an Jiaotong University, Xi'an, 710061, Shanxi, China

2.Department of Gastroenterology, Taihe Hospital, Hubei University of Medicine, Shiyan, 442000, Hubei, China;

3.Hubei Key Laboratory of Embryonic Stem Cell Research, Taihe Hospital, Hubei University of Medicine, Shiyan, 442000, Hubei, China;

4.Department of Oncology, Taihe Hospital, Hubei University of Medicine, Shiyan, 442000, Hubei, China;

**Corresponding author:*

Professor Shui-xiang He, Department of Gastroenterology, the First Affiliated Hospital of Xi'an Jiaotong University, 76 Yanta West Road, Xi'an, Shanxi 710061, P.R. China. Tel/ Fax:

+86-7198801431; E-mail: dyyyjxk@mail.xjtu.edu.cn;

Professor Qiang Tong, Department of Gastroenterology, Taihe hospital, Hubei University of Medicine, 32 south renmin road, Shiyan, Hubei 442000, People's republic of china, Tel:

+86-7198801421; Email: xinl2012@163.com.

Running title: Microflora of different ESCC tissues.

Abstract:

Objective: This study aimed to compare the differences in microbiota between the postoperative tissues and esophageal mucosa tissues of patients with esophageal squamous cell carcinoma (ESCC).

Methods: Seventy-two patients who had ESCC and diagnosed in Taihe Hospital were selected from July 2018 to July 2019 to participate in this work. Then, 27 postoperative tissues and 45 mucosa samples of ESCC were collected. The sequence V₄ hypervariable region was amplified, and Illumina MiSeq sequencing was performed to analyze the differences between the two groups.

Results: Results revealed that the Shannon and Chao1 indices of the postoperative esophageal cancer tissue group (Group A) were significantly higher ($P < 0.05$) than those of the esophageal mucosa group (Group B). The Simpson index of Group A was higher than that of Group B, but the difference was not significant ($P > 0.05$). The beta diversity of the two groups was also not significantly different ($P > 0.05$). LEfSe analysis showed that the abundance of *Megasphaera*, *Actinobacteria*, *Enterobacteriaceae*, and *Enterobacteriales* in Group A was significantly higher than that in Group B, but the abundance of *Mogibacteriaceae* in Group B was significantly higher than that in Group A ($P < 0.05$). At the phylum level, *Actinobacteria* and *Verrucomicrobiae* were more abundant in the postoperative tissue group than in the esophageal mucosa group. The abundance of *Fusobacteriia*, *SRI*, and *Spirochaetes* in the postoperative tissue group was significantly lower than that of the esophageal mucosa group ($P < 0.05$).

Conclusion: The source of the sample should be considered in studies on the esophageal flora. More relevant findings were obtained from postoperative tissues than from the normal mucosa.

Keywords: Esophageal squamous cell carcinoma; Postoperative tissue; Esophageal mucosa;

Introduction

Esophageal carcinoma (EC) ranks seventh in terms of the global incidence of malignant tumors and sixth in terms of mortality[1]. This type of cancer mainly includes two pathological types, namely, esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC)[2]. The incidence of EC in China ranks among the top five in the world [3], the sixth in terms of the incidence of malignant tumors in the country, and fourth in terms of death [4]. ESCC accounts for over than 90% of all EC cases [5].

As part of the tumor microenvironment, microorganisms may participate in tumor development by inducing chronic or persistent inflammation [6]. However, the microecological composition of each part is not uniform, and different parts of the gastrointestinal tract may have specific microecological communities [7]. Gender, obese body type, age, food, host genetic background, environment, antibacterial drugs, and other factors affect microbial structures [8-12]. Moreover, different methods of material extraction may affect research results on the digestive tract flora.

The microflora in the digestive tract is related to the occurrence and development of ESCC [13]. Changes in the esophageal flora should be studied, or specific bacterial changes should be detected; such studies may be beneficial to the early diagnosis, evaluation, and favorable prognosis judgment of ESCC [14-16]. Sampling methods for research on the flora causing esophageal diseases include saliva collection, oropharyngeal swabs, esophageal mucosal swabs, endoscopic biopsies, endoscopic mucosal resection specimens, surgical biopsies after esophageal

surgeries, esophageal string tests (ESTs), and new Cytosponge devices [17-21]. Studies on the esophageal flora of patients with ESCC remain in infancy, and the most suitable type of samples for this disease is unknown. This study aimed to identify an ideal method for studies on the esophageal flora of patients with ESCC by comparing differences in the bacterial flora between surgical tissues and endoscopic esophageal mucosa tissues.

1. Materials and methods

1.1 Sample source

A total of 72 patients who had ESCC and diagnosed via digestive endoscopy and thoracic surgery at Taihe Hospital from July 2018 to July 2019 were selected for this research and divided into two groups. Group A was composed of 27 esophageal cancer postoperative tissue samples, and Group B was assigned 45 esophageal mucosa samples. Gender and age were not significantly different between the two groups (Table 1).

The following inclusion criteria of ESCC were applied : age ≥ 18 years; pathological diagnosis of ESCC; without metabolic diseases (such as diabetes), hyperlipidemia, or other infectious diseases; good general condition; no intake of antibiotics, acid suppressants, or probiotics within the past 2 months; balanced diet and no special dietary habits; and no serious liver, kidney, and immunodeficiency diseases. The exclusion criteria were as follows: use of drugs affecting the microecology of the esophagus in the past 2 months; complications of metabolic or infectious diseases; presence of tumors other than ESCC; incomplete data; and not considered suitable for inclusion by the researchers.

The inclusion criteria of the healthy control group were established with reference to the literature [22,23]: the enrolled persons were ≥ 18 years old and were generally in good condition;

endoscopy showed no abnormal esophagus lesions and pathologically proved normal squamous epithelium, the exclusion criteria were the same as ESCC.

The study protocol was reviewed and approved by the Taihe Hospital Ethics Committee, and all patients' written consent were obtained before they were allowed to participate in this research. Moreover, this study was conducted in accordance with the provisions of the Helsinki Declaration.

1.2 Method

1.2.1 Sample collection

Endoscopy center specimens were obtained, gastroscopy was performed 6–8 hours after fasting, and warm water was used for gargling before examination. After the esophageal tumor lesions were found, four to eight specimens were collected with sterile biopsy forceps for examination. Two specimens were marked, placed in sterile cryopreservation tubes, and frozen in liquid nitrogen. The remaining tissues were routinely stored in formalin and sent to the pathology department. Surgical specimens were cryopreserved and sent to the pathology department, and appropriate samples were chosen for follow-up studies in accordance with the inclusion criteria.

1.2.2 DNA extraction

The DNA of the sample was extracted with a Power Maxkit (MoBio Laboratories, Carlsbad, CA, USA) by using sodium dodecyl sulfate lysate freeze–thaw method. The purity and quantity of the DNA were determined with NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) , and the sample was frozen at $-20\text{ }^{\circ}\text{C}$ for later use.

1.2.3 16srDNA sequencing

The V_4 region of the 16S rRNA gene was amplified through a polymerase chain reaction

(PCR). The primers included 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). The PCR system (50 µl) was composed of the following: 25 µl of buffer, 3 µl each of F/R primers (10 µM), 10 µl of DNA template, and 6 µl of ddH₂O. The loop parameters were as follows: pre-denaturation at 98 °C for 30 s, 25 cycles of denaturation at 98 °C for 15 s, annealing at 58 °C for 15 s, and extension at 72 °C for 15 s, and final extension at 72 °C for 1 min. The PCR products were purified with AMPure (Beckman Coulter, Indianapolis, IN), quantified with PicoGreen dsDNA, assay kit (Invitrogen, Carlsbad, CA, USA), and sequenced on a 2×150 bp pair-end IlluminaHiSeq4000platform (Illumina Inc., USA) and completed at Shanghai Biotechnical Medical Laboratory Center, Shanghai, China.

1.2.4 OTUs clustering and species annotation

Operational taxonomic units (OTUs) were analyzed with V search version 2.4.4 and clustered with a similarity of 97%. Representative sequences were annotated on the basis of the SILVA128 database. The abundance and classification of the OTUs were recorded.

1.2.5 Bioinformatic analysis and statistical analysis

Quantitative insights into microbial ecology (QIIME version 1.8.0) and R package (www.r-project.org, version 3.2.0) were used to analyze the data. Alpha diversity indices, including Chao1, Shannon, Simpson, and abundance-based coverage estimator, were calculated. The abundance and uniformity of OTUs were compared, and UniFrac distance was calculated[24].

Principal component analysis (PCA) and nonmetric multidimensional scaling (NMDS) plots were drawn for the beta analysis of the sample flora structure. The *t*-test and Monte Carlo permutation test were used to draw box plots and compare differences in the UniFrac distance between groups. Permutational multivariate analysis of variance (PERMANOVA) was performed

to evaluate the markers of community structure differentiation among groups[25].

Vegan from R package, MEGAN[26, 27], and Graphical Phylogenetic Analysis (GraPhlAn) were used to visualize the groups and abundances [28]. Venn diagrams were generated using the Venn Diagram module of the R package to visualize common and unique OTUs between samples or groups. The Kruskal method in the R-package was adopted to compare the differences in classification levels within or between groups.

Linear discriminant analysis (LDA) was combined with a Kruskal–Wallis test through linear discriminant analysis effect size (LEfSe) analysis [29] to screen out biomarkers, i.e., species with remarkable differences between groups. Random forest analysis was performed using the default settings of the random forest module in R to compare differences between groups. Microbial functions were predicted in accordance with PICRUSt [30]. Statistical Analysis of Metagenomic Profiles (STAMP) software package version 2.1.3 (31) was used to further analyze the output file. Parallel-META 3 (version 3.3.2) was utilized to analyze β diversity in species and function based on Meta-Storms distance. Statistical analysis was performed using SPSS 21.0 version (SPSS Inc., Chicago, USA). Differences were considered statistically significant at $P < 0.05$.

2 Results

2.1 Sample sequencing data

After clustering was performed with 97% similarity, 3,656 OTUs, including 2,926 in the esophageal cancer postoperative tissue group (Group A) only, 2,772 in the esophageal mucosa group (Group B) only, and 2,042 OTUs in both groups, were obtained. A total of 884 OTUs were

unique to Group A, and 730 OTUs were unique OTUs to Group B (Fig. S1).

2.2 Alpha diversity analysis

The Shannon and Chao indices of postoperative tissue (Group A) samples were significantly higher than those of esophageal mucosa group tissue (Group B) samples ($P < 0.05$). The Simpson index of Group A was higher than that of Group B, but the difference was not significant ($P > 0.05$). These findings indicated that the diversity of microbial flora in postoperative tissues was higher than that in the esophageal mucosa group (Fig. 1 A 1–3).

2.3 Beta diversity analysis

PC1 and PC2 represented the potential factors influencing the deviation of the microbial composition of the two groups. For the two groups, PC1=19.14%, $P_{PC1}=0.67$, and $P_{PC2}=0.42$, suggesting that the bacterial compositions in the two groups had no significant difference (Fig. 1B). NMDS analysis was more stable for data with a complex structure than for data with a simple structure, and the overall flora of the two groups could not be clearly distinguished. This result demonstrated that the overall composition of the flora of the two groups exhibited no significant difference (Fig. 1C).

2.4 Differential LEfSe analysis

The abundance of *Megasphaera*, *Actinobacteria*, *Enterobacteriaceae*, and *Enterobacteriales* in the esophageal postoperative tissue (Group A) was higher than that in the esophageal mucosal tissue (Group B), but the abundance of *Mogibacteriaceae* in the esophageal mucosa tissue group (Group B) was higher than that of the esophageal tissue (Group A). The difference was statistically significant ($P < 0.05$, Fig. 2 A and B).

2.5 Structural analysis of the two groups of the esophageal flora

Our study only showed the results at phylum and genus levels and did not include an analysis of the microbial flora composition at other levels.

2.5.1 Analysis of microbial flora composition at the phylum level

The two groups of the samples were considerably different at the phylum level, and five major phyla were identified. *Actinobacteria* and *Verrucomicrobiae* were more abundant in the postoperative tissue group than in the esophageal mucosa group. The abundance of *Fusobacteriia*, *SRI*, and *Spirochaetes* was significantly lower in the postoperative tissue group than in the esophageal mucosa group ($P < 0.05$, Table 2).

2.5.2 Microbial flora composition analysis at a genus level

At the genus level, *Collinsella*, *Bacteroides*, *Parabacteroides*, *Butyricimonas*, *Paraprevotella*, *Gemella*, *Enterococcus*, *Blautia*, *Coprococcus*, *Lachnospira*, *Roseburia*, *Faecalibacterium*, *Oscillospira*, *Ruminococcus*, *Megamonas*, *Megasphaera*, *Phascolarctobacterium*, *Sutterella*, and *Akkermansia* were more abundant in the postoperative tissue group than in the esophageal mucosa group, whereas the abundance of *Bifidobacterium*, *Porphyromonas*, *Prevotella*, *Prevotella*, *Catonella*, *Oribacterium*, *Ruminococcus*, *Peptostreptococcus*, *Selenomonas*, *Parvimonas*, *Fusobacterium*, *Leptotrichia*, *Ralstonia*, *Campylobacter*, *Actinobacillus*, and *Treponema* in the former was significantly lower than that in the latter ($P < 0.05$, Table 3).

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

The top 60 species were selected for the random forest method to establish a model. The error rate was 22.59% (Fig.3A). The receiver operating characteristic (ROC) curve confirmed that our model was reliable and could effectively distinguish between the two groups of samples

(AUC=0.86, Fig. 3B).

2.7 Comparison of phenotype classification based on BugBase

The Gram-positive bacteria in the postoperative tissue group exhibited an advantage over the Gram-positive bacteria in the mucosal tissue group. By contrast, the Gram-negative bacteria in the postoperative tissue group were significantly weaker than those in the mucosal tissue group ($P < 0.05$, Figure S2). The two groups were similar under the following conditions: aerobic, anaerobic, presence of mobile elements, facultatively anaerobic, forms-biofilms, potentially pathogenic, and stress-tolerant conditions, and the difference was not significant ($P > 0.05$, Table 4).

3 Discussion

The human microbiota (HM) includes trillions of bacteria, archaea, fungi, and viruses that interact with the human body [32]. Normal HM plays a role in human nutrition, drug metabolism, maintenance of the integrity of the intestinal mucosal barrier, immunomodulation, and protection against pathogens [33]. Changes in the microbial community composition are related to many diseases, including tumors [34,35]. Bacteria were first found in tumors over a century ago. Different tumor types have unique flora, but the characterization of tumor microbiomes is often challenging because of their low biomass [36]. The microbiota, as a part of the tumor microenvironment, plays an important role in tumorigenesis and metastasis. However, the microecology of the different body parts has not been unified.

The number of bacteria in the digestive tract is 10 times the total number of human cells [30]. Most of the bacteria have a specific spatial distribution and are not cultivable. The flora in

the oral cavity, esophagus, and rectum are diverse in terms of type and quantity [37]. The composition of the gastrointestinal flora in the same body undergoes remarkable changes [35]. The abundance of the flora in the upper digestive tract is higher than that in the lower digestive tract [38], and the flora of the left and right colons show remarkable differences [39,40]. Thus, the specificity of the tissue flora should be considered in other related studies.

Different sampling methods may affect the results of studies on the flora. Some scholars have explored approaches to obtain a reasonable sampling method. The most commonly used sample types on intestinal flora studies are fecal and mucosal biopsy specimens, but these two methods have limitations [41]. Araujo-Perez F *et al.* [42] found that the bacterial diversity of an anal swab is higher than that of rectal biopsy. Kotar T *et al.* [43] and Fair K *et al.* [44] obtained similar conclusions and confirmed that the sampling time of the anal swab influences sequencing results. An oral swab is more practical and cost efficient than tissue biopsy in studies on oral flora in mice. Thus, swabbing maybe the preferred sampling method [45]. Budding A E *et al.* [41] believed that the anal swab results of flora can be reproduced whether sample collection is performed at home or in outpatient clinics; thus, anal swab is suitable for clinical diagnosis.

The esophagus contains many types of bacteria, and an abundant flora can be found between the oropharynx and the stomach. Some esophageal flora in the stomach are similar to those in the oral cavity. Hence, the different parts of the esophagus have no specific flora [18, 46]. The abundance of *Archaea* and *phages* in the normal esophagus is low; the normal esophagus also contains *Streptococcus*, *Prevotella*, *Veillonella*, *Clostridium*, *Haemophilus*, *Neisseria*, *Porphyromonas*, and other bacteria [47, 48]. Shao *et al.* [49] found that the microbial environment of ESCC is composed of *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*. The abundance of

Clostridium in tumor tissues increased (3.2% vs.1.3%), whereas the abundance of *Streptococcus* decreased (12.0% vs. 30.2%) compared with that in nontumor tissues.

Studies have been performed to enhance the sampling methods of esophageal flora. Liu *et al.*[50] believed that the swabs and biopsies of patients with ESCC have similar microbial profiles. However, Gall *et al.*[18] and colleagues believed that the amount of DNA recovered from amucosal chip brush is greater than that from mucosal samples in esophageal adenocarcinoma. Okereke I *et al.*[51] studied Barrett's esophagus and confirmed that swabs obtained from the oropharynx or an endoscope cannot replace the biopsies of esophageal mucosa. Further research [19] also verified that mucosal biopsy should be used for analysis on the esophageal flora.

Alpha diversity can reflect the diversity of a microbial community. The Chao1 index describes the richness of a community and reflects the number of microbial members, such as OTUs, in a community. The Shannon and Simpson indices reflect the uniformity of a community and the abundance among members. We found that the Chao1 and Shannon indices of the postoperative tissue group increased compared with those of the mucosal tissue group. Although the Simpson index of the postoperative tissue group was higher than that of the mucosal tissue group, but the difference between the two groups was not significant, suggesting that the postoperative tissue flora was richer than the mucosal tissue flora. Increased and good uniformity was noted, i.e., the abundance of the bacteria between the groups was slightly different. Beta diversity was used to examine the similarity of the colony structure of different samples, and good clustering was found in the sample set. The two groups of bacteria, i.e., P_{PC1} and P_{PC2} , had >0.05 , indicating that the diversity of the two groups exhibited no significant difference.

LEfSe analysis revealed that the flora of the two groups included considerably different

species. *Megasphaera*, *Actinobacteria*, *Enterobacteriaceae*, and *Enterobacteriales* were more abundant in the postoperative tissues of the esophagus than in the mucosal tissues. *Mogibacteriaceae* was more abundant in the mucosal tissue group than in the postoperative tissue group. The bacterial species of the two groups were compared at the phylum and genus levels. The predominant phyla in the postoperative tissue group were *Actinobacteria* and *Verrucomicrobiae*. The dominant phyla in the mucosal tissue group were *Fusobacteriia*, *SRI*, and *Spirochaetes*.

Genus-level analysis showed different dominant bacteria in the two groups of flora. The different distributions of the flora in the esophageal tissues could be explained as follows. The flora may participate in the occurrence and development of ESCC, and the abundance of bacteria in changes with the tumor progression and invasion of ESCC. Differences between the two groups might be caused by variations in pH and pressure in the esophagus, intragastric food, gastric acid, bile reflux, and other undetermined factors.

The random forest method was adopted, and the top 60 species were selected to establish the model, and the reliability of the model was verified with the ROC curve, which could effectively distinguish the two groups of samples. BugBase mainly conducts phenotypic predictions, including Gram staining, biofilm formation, pathogenicity, mobile components, oxygen demand (e.g., anaerobic bacteria, aerobic bacteria, and facultative bacteria), and oxidative stress tolerance [52]. The comparison of the BugBase phenotypes of the two groups showed differences in Gram-negative and Gram-positive bacteria, and this finding might be related to the abovementioned variation in the distribution of bacterial groups. Future search on the function of bacteria should be carried out.

Although the activity of flora is not the only factor in the pathogenesis of ESCC, dys

bacteriosis may play an important role in the occurrence and development of ESCC. This finding provided important clues and directions for related studies. Tissue sources should be considered in studies on the flora, and postoperative tissues may offer more meaningful results than mucosal samples in studies on esophageal flora. This study had some limitations that could affect the interpretation of our results. First, the flora of different parts of the esophagus and postoperative tissues were not compared. Second, other sampling methods, such as endoscopic smear, were not performed. Third, this study did not explore the flora of early esophageal cancer and precancerous lesions. Such analysis could monitor changes in bacterial flora at different stages of tumor development. Lastly, future studies should involve a larger sample type and size to determine the best approach for assessing esophageal samples. In this study, specific bacteria were detected, and the role of these species in ESCC should be studied further in future studies.

References:

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, 2018.
2. Baba, Y., Iwatsuki, M., Yoshida, N., Watanabe, M., and Baba, H. Review of the gut microbiome and esophageal cancer: Pathogenesis and potential clinical implications. *Ann Gastroenterol Surg*, *1*: 99-104, 2017.
3. Abnet, C. C., Arnold, M., and Wei, W. Q. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology*, *154*: 360-373, 2018.
4. Chen, R., Zheng, R. S., Zhang, S. W., Zeng, H. M., Wang, S. M., Sun, K. X., Gu, X. Y., Wei, W. W., and He, J. [Analysis of incidence and mortality of esophageal cancer in China, 2015]. *Zhonghua Yu Fang Yi Xue Za Zhi*, *53*: 1094-1097, 2019.
5. Wen, X., Wen, D., Yang, Y., Chen, Y., Wang, G., and 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. *Ann Glob Health*, *83*: 444-462, 2017.
6. Wang, Q., Rao, Y., Guo, X., Liu, N., Liu, S., Wen, P., Li, S., and Li, Y. Oral Microbiome in Patients with Oesophageal Squamous Cell Carcinoma. *Sci Rep*, *9*: 19055, 2019.
7. Zoetendal, E. G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., Bork, P., Wels, M., de Vos, W. M., and Kleerebezem, M. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J*, *6*: 1415-26, 2012.
8. Cuevas-Sierra, A., Riezu-Boj, J. I., Guruceaga, E., Milagro, F. I., and Martinez, J. A. Sex-Specific Associations between Gut Prevotellaceae and Host Genetics on Adiposity. *Microorganisms*, *8*, 2020.
9. La-Ongkham, O., Nakphaichit, M., Nakayama, J., Keawsompong, S., and Nitisinprasert, S. Age-related changes in the gut microbiota and the core gut microbiome of healthy Thai humans. *3 Biotech*, *10*: 276, 2020.
10. Eck, A., Rutten, N., Singendonk, M., Rijkers, G. T., Savelkoul, P., Meijssen, C. B., Crijns, C. E., Oudshoorn, J. H., Budding, A. E., and Vlieger, A. M. Neonatal microbiota development and the effect of early life antibiotics are determined by two distinct settler types. *PLoS One*, *15*: e0228133, 2020.
11. Neckovic, A., van Oorschot, R., Szkuta, B., and Durdle, A. Investigation of direct and indirect transfer of microbiomes between individuals. *Forensic Sci Int Genet*, *45*: 102212, 2020.
12. Pan, H. W., Du LT, Li, W., Yang, Y. M., Zhang, Y., and Wang, C. X. Biodiversity and richness shifts of mucosa-associated gut microbiota with progression of colorectal cancer. *Res Microbiol*, *171*: 107-114, 2020.

13. Peters, B. A., Wu, J., Pei, Z., Yang, L., Purdue, M. P., Freedman, N. D., Jacobs, E. J., Gapstur, S. M., Hayes, R. B., and Ahn, J. Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers. *Cancer Res*, 77: 6777-6787, 2017.
14. Deshpande, N. P., Riordan, S. M., Castano-Rodriguez, N., Wilkins, M. R., and Kaakoush, N. O. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome*, 6: 227, 2018.
15. Elliott, D., Walker, A. W., O'Donovan, M., Parkhill, J., and Fitzgerald, R. C. A non-endoscopic device to sample the oesophageal microbiota: a case-control study. *Lancet Gastroenterol Hepatol*, 2: 32-42, 2017.
16. Yu, G., Gail, M. H., Shi, J., Klepac-Ceraj, V., Paster, B. J., Dye, B. A., Wang, G. Q., Wei, W. Q., Fan, J. H., Qiao, Y. L., Dawsey, S. M., Freedman, N. D., and Abnet, C. C. Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. *Cancer Epidemiol Biomarkers Prev*, 23: 735-41, 2014.
17. Elliott, D., Walker, A. W., O'Donovan, M., Parkhill, J., and Fitzgerald, R. C. A non-endoscopic device to sample the oesophageal microbiota: a case-control study. *Lancet Gastroenterol Hepatol*, 2: 32-42, 2017.
18. Gall, A., Fero, J., McCoy, C., Claywell, B. C., Sanchez, C. A., Blount, P. L., Li, X., Vaughan, T. L., Matsen, F. A., Reid, B. J., and Salama, N. R. Bacterial Composition of the Human Upper Gastrointestinal Tract Microbiome Is Dynamic and Associated with Genomic Instability in a Barrett's Esophagus Cohort. *PLoS One*, 10: e0129055, 2015.
19. Okereke, I. C., Miller, A. L., Hamilton, C. F., Booth, A. L., Reep, G. L., Andersen, C. L., Reynolds, S. T., and Pyles, R. B. Microbiota of the Oropharynx and Endoscope Compared to the Esophagus. *Sci Rep*, 9: 10201, 2019.
20. Fillon, S. A., Harris, J. K., Wagner, B. D., Kelly, C. J., Stevens, M. J., Moore, W., Fang, R., Schroeder, S., Masterson, J. C., Robertson, C. E., Pace, N. R., Ackerman, S. J., and Furuta, G. T. Novel device to sample the esophageal microbiome--the esophageal string test. *PLoS One*, 7: e42938, 2012.
21. Kageyama, S., Takeshita, T., Takeuchi, K., Asakawa, M., Matsumi, R., Furuta, M., Shibata, Y., Nagai, K., Ikebe, M., Morita, M., Masuda, M., Toh, Y., Kiyohara, Y., Ninomiya, T., and Yamashita, Y. Characteristics of the Salivary Microbiota in Patients With Various Digestive Tract Cancers. *Front Microbiol*, 10: 1780, 2019.
22. Liu N, Ando T, Ishiguro K, et al. Characterization of bacterial biota in the distal esophagus of Japanese patients with reflux esophagitis and Barrett's esophagus[J]. *BMC Infect Dis*, 2013,13:130. DOI:10.1186/1471-2334-13-130.

23. Nasrollahzadeh D, Malekzadeh R, Ploner A, et al. Variations of gastric corpus microbiota are associated with early esophageal squamous cell carcinoma and squamous dysplasia[J]. *Sci Rep*, 2015,5:8820. DOI:10.1038/srep08820.
24. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7: 335-6, 2010.
25. Anderson, M. J., Ellingsen, K. E., and McArdle, B. H. Multivariate dispersion as a measure of beta diversity. *Ecol Lett*, 9: 683-93, 2006.
26. Huson, D. H., and Weber, N. Microbial community analysis using MEGAN. *Methods Enzymol*, 531: 465-85, 2013.
27. Mitra, S., Stark, M., and Huson, D. H. Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics*, 12 Suppl 3: S17, 2011.
28. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., and Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ*, 3: e1029, 2015.
29. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., and Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol*, 12: R60, 2011.
30. Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkpile, D. E., Vega, T. R., Knight, R., Beiko, R. G., and Huttenhower, C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*, 31: 814-21, 2013.
31. Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30: 3123-4, 2014.
32. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., 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, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., 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, S. D., and Wang, J. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464: 59-65, 2010.
33. Structure, function and diversity of the healthy human microbiome. *Nature*, 486: 207-14, 2012.

34. Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. The healthy human microbiome. *Genome Med*, 8: 51, 2016.
35. Vuik, F., Dicksved, J., Lam, S. Y., Fuhler, G. M., van der Laan, L., van de Winkel, A., Konstantinov, S. R., Spaander, M., Peppelenbosch, M. P., Engstrand, L., and Kuipers, E. J. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United European Gastroenterol J*, 7: 897-907, 2019.
36. Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., Arora, R., Cogdill, A. P., Khan, M., Ologun, G., Bussi, Y., Weinberger, A., Lotan-Pompan, M., Golani, O., Perry, G., Rokah, M., Bahar-Shany, K., Rozeman, E. A., Blank, C. U., Ronai, A., Shaoul, R., Amit, A., Dorfman, T., Kremer, R., Cohen, Z. R., Harnof, S., Siegal, T., Yehuda-Shnaidman, E., Gal-Yam, E. N., Shapira, H., Baldini, N., Langille, M., Ben-Nun, A., Kaufman, B., Nissan, A., Golan, T., Dadiani, M., Levanon, K., Bar, J., Yust-Katz, S., Barshack, I., Peeper, D. S., Raz, D. J., Segal, E., Wargo, J. A., Sandbank, J., Shental, N., and Straussman, R. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*, 368: 973-980, 2020.
37. Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., and Nageshwar, R. D. Role of the normal gut microbiota. *World J Gastroenterol*, 21: 8787-803, 2015.
38. Vasapolli, R., Schutte, K., Schulz, C., Vital, M., Schomburg, D., Pieper, D. H., Vilchez-Vargas, R., and Malfertheiner, P. Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals. *Gastroenterology*, 157: 1081-1092.e3, 2019.
39. Saffarian, A., Mulet, C., Regnault, B., Amiot, A., Tran-Van-Nhieu, J., Ravel, J., Sobhani, I., Sansonetti, P. J., and Pedron, T. Crypt- and Mucosa-Associated Core Microbiotas in Humans and Their Alteration in Colon Cancer Patients. *mBio*, 10, 2019.
40. Kashiwagi, S., Naito, Y., Inoue, R., Takagi, T., Nakano, T., Inada, Y., Fukui, A., Katada, K., Mizushima, K., Kamada, K., Uchiyama, K., Handa, O., Ishikawa, T., and Itoh, Y. Mucosa-Associated Microbiota in the Gastrointestinal Tract of Healthy Japanese Subjects. *Digestion*, 101: 107-120, 2020.
41. Budding, A. E., Grasman, M. E., Eck, A., Bogaards, J. A., Vandenbroucke-Grauls, C. M., van Bodegraven, A. A., and Savelkoul, P. H. Rectal swabs for analysis of the intestinal microbiota. *PLoS One*, 9: e101344, 2014.
42. Araujo-Perez, F., McCoy, A. N., Okechukwu, C., Carroll, I. M., Smith, K. M., Jeremiah, K., Sandler, R. S., Asher, G. N., and Keku, T. O. Differences in microbial signatures between rectal mucosal biopsies and rectal swabs. *Gut Microbes*, 3: 530-5, 2012.
43. Kotar, T., Pirs, M., Steyer, A., Cerar, T., Soba, B., Skvarc, M., Poljsak, P. M., and Lejko, Z. T. Evaluation of rectal swab use for the determination of enteric pathogens: a prospective study of

diarrhoea in adults. *Clin Microbiol Infect*, 25: 733-738, 2019.

44. Fair, K., Dunlap, D. G., Fitch, A., Bogdanovich, T., Methe, B., Morris, A., McVerry, B. J., and Kitsios, G. D. Rectal Swabs from Critically Ill Patients Provide Discordant Representations of the Gut Microbiome Compared to Stool Samples. *mSphere*, 4, 2019.

45. Hernandez-Arriaga, A., Baumann, A., Witte, O. W., Frahm, C., Bergheim, I., and Camarinha-Silva, A. Changes in Oral Microbial Ecology of C57BL/6 Mice at Different Ages Associated with Sampling Methodology. *Microorganisms*, 7, 2019.

46. Dong, L., Yin, J., Zhao, J., Ma, S. R., Wang, H. R., Wang, M., Chen, W., and Wei, W. Q. Microbial Similarity and Preference for Specific Sites in Healthy Oral Cavity and Esophagus. *Front Microbiol*, 9: 1603, 2018.

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

48. Deshpande, N. P., Riordan, S. M., Castano-Rodriguez, N., Wilkins, M. R., and Kaakoush, N. O. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome*, 6: 227, 2018.

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

50. Liu, A. Q., Vogtmann, E., Shao, D. T., Abnet, C. C., Dou, H. Y., Qin, Y., Su, Z., Wei, W. Q., and Chen, W. A Comparison of Biopsy and Mucosal Swab Specimens for Examining the Microbiota of Upper Gastrointestinal Carcinoma. *Cancer Epidemiol Biomarkers Prev*, 28: 2030-2037, 2019.

51. Okereke, I., Hamilton, C., Reep, G., Krill, T., Booth, A., Ghouri, Y., Jala, V., Andersen, C., and Pyles, R. Microflora composition in the gastrointestinal tract in patients with Barrett's esophagus. *J Thorac Dis*, 11: S1581-S1587, 2019.

52. Thomas, A. M., Jesus, E. C., Lopes, A., Aguiar, S. J., Begnami, M. D., Rocha, R. M., Carpinetti, P. A., Camargo, A. A., Hoffmann, C., Freitas, H. C., Silva, I. T., Nunes, D. N., Setubal, J. C., and Dias-Neto, E. Tissue-Associated Bacterial Alterations in Rectal Carcinoma Patients Revealed by 16S rRNA Community Profiling. *Front Cell Infect Microbiol*, 6: 179, 2016.

Abbreviations

ESCC: esophageal squamous cell carcinoma; EC: esophageal carcinoma; EAC :esophageal adenocarcinoma; ESTs: esophageal string tests; OTUs: operational taxonomic units;PCA: principal component analysis; NMDS: nonmetric multidimensional scaling ; HM : human microbiota.

Authors' contributions

L-XB, G-ZY and H-SX contributed conceptualization, writing–review, and editing. L-XB, M-JC and G-ZY wrote the manuscript. G-ZY and M-JC contributed methodology and software. S-CT, XW and WH collected and detected clinical samples. TQ and H-SX contributed funding acquisition, and editing. All authors revised the manuscript and approved the final manuscript.

Acknowledgment

We are grateful to all the participants in the study whose tissues donations made this study possible. We sincerely thank Dr. Zi-wei Fan and Jiang-man Zhao from Biotechnology Co., Ltd. (Shanghai, China) for their selfless help.

Funding

This study was funded by the Scientific research project of Hubei Provincial Department of Education (B2013107), the 2016 Joint Diagnostic Medicine Research Project of Taihe Hospital (2016JD02), and the Foundation of Taihe Hospital (2019JJXM032, 2020JJXM032).

Availability of data and materials

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

Ethics approval and consent to participate

The study protocol was reviewed and approved by the Taihe Hospital Ethics Committee (No.2018KS020), and all patients received information concerning their participation in the study and gave written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Figure legends:

Fig.1 Comparison of alpha and beta diversity of the esophageal flora after esophageal surgery (Group A) and in the esophageal mucosa group (Group B)

Notes: A: A-1 to A-3 represent the Shannon, Simpson, and Chao1 indices, respectively, *P*-value is marked at the top of each figure. The abscissa contains the name of the group, and the ordinate shows the alpha diversity index of the different groups. The box chart can display five statistics (i.e., minimum value, first quartile, median, third quartile, and maximum value) as five lines from bottom to top. Outliers are marked "o". *P*-value corresponds to the value of the rank sum test

corresponding to the alpha diversity index.

B. PCoA plots of the unweighted Unifrac distances of the variation in microbiota composition detected in the ESCC and HC groups (sample name was removed). The horizontal box diagram shows the distribution of the values grouped on the first column coordinates, the vertical box diagram presents 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, the points of the same color indicate an origin of the same group, and the distance reflects the similarity of the samples.

C. Scatter plot of two sets of NMDS analysis results. Each point in the figure represents a sample. The distances from the response sample are similar. The greater the distance between two points, the greater the difference in the community structure between the two groups.

In the plot, blue represents the esophageal mucosa group, and red corresponds to the esophageal cancer postoperative tissue group.

Fig. 2 Species composition analysis

A. Cladogram view of the representative microbial structure from the two groups.

B. Differential bacteria between the two groups through LEfSe analysis

Notes: Dominant microbial classes for specific groups are represented by different colors. The diameter of each dot is proportional to the OTU abundance. Red and green represent the HC and ESCC groups. Significantly different logarithmic LDA scores were set to 2. Only groups with significant differences are shown in the figure.

Fig.3 Comparison of the flora between groups and selection of species markers

A. Species importance map

Notes: The abscissa reflects the importance level, and the ordinate indicates the species name sorted according to importance. The figure reflects the genus of the bacteria in the classifier that plays a major role in the classification effect, which is arranged from the largest to the smallest. Error rate refers to the error rate of the random forest method for predicting classification with the features below. High values indicate that the classification accuracy based on the genus characteristics is low, and differences in genus characteristics may not be obvious between the groups. For example, when all levels are considered, the top 60 species are plotted

Fig.3B ROC curve analyzing the clinical accuracy of using differential bacteria obtained from the ESCC and HC groups for the diagnosis of ESCC

Notes: The point closest to the upper left of the ROC graph is the critical value with the greatest sensitivity and specificity.

Fig.S1 Venn diagram showing the distribution of operational taxonomic units (OTUs) detected in the two groups

Notes: Group A (blue) is the esophageal cancer postoperative tissue group, and Group B (yellow) is the esophageal mucosa group.

Fig. S2 Comparison of the two groups of phenotypic classification based on BugBase

Note: A: Gram-positive bacteria and B: Gram-negative bacteria. Group A: postoperative tissue group. Group B: esophageal mucosa group

Figures

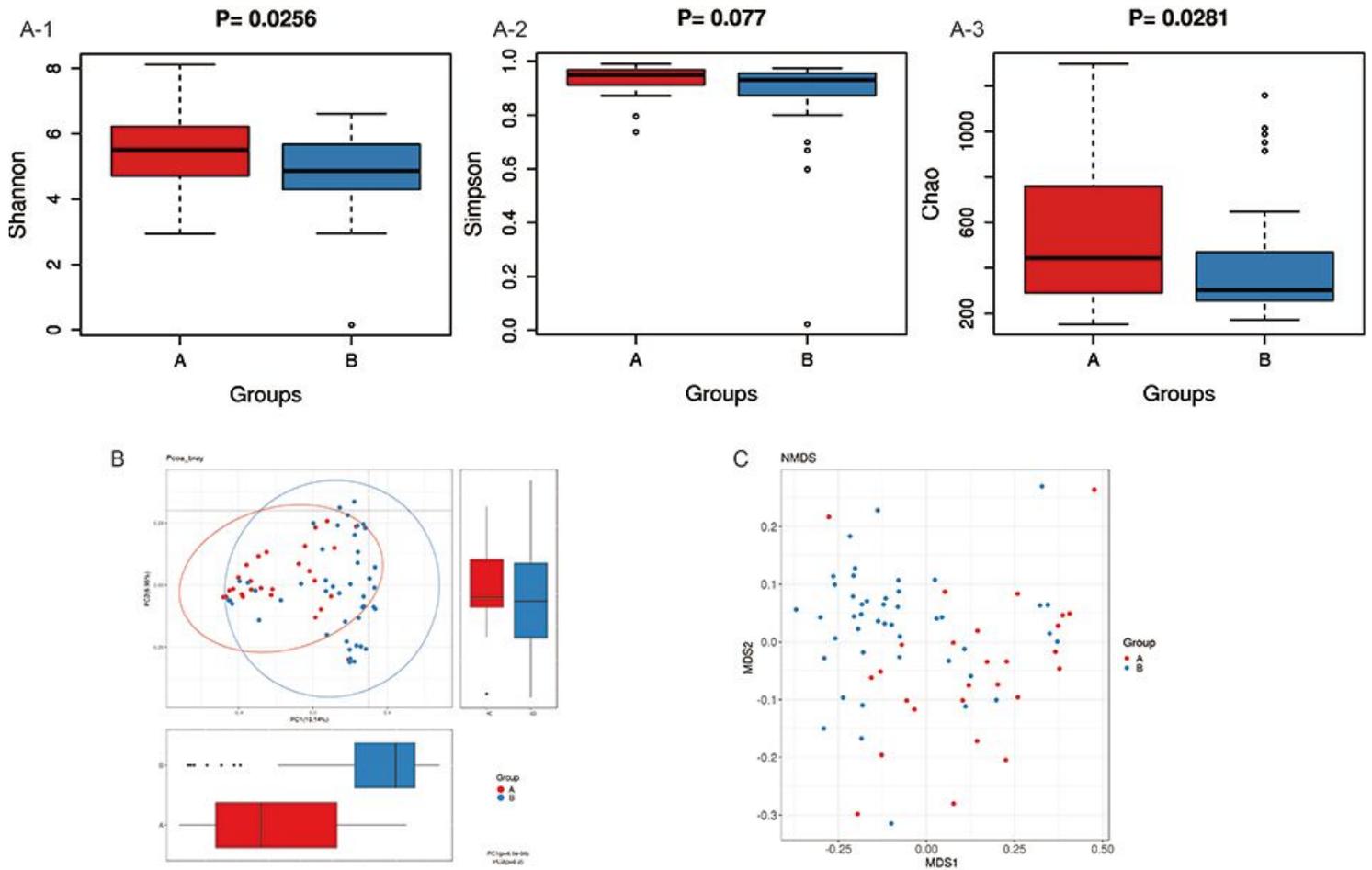


Figure 1

Comparison of alpha and beta diversity of the esophageal flora after esophageal surgery (Group A) and in the esophageal mucosa group (Group B) Notes: A: A-1 to A-3 represent the Shannon, Simpson, and Chao1 indices , respectively, P-value is marked at the top of each figure. The abscissa contains the name of the group, and the ordinate shows the alpha diversity index of the different groups. The box chart can display five statistics (i.e., minimum value, first quartile, median, third quartile, and maximum value) as five lines from bottom to top. Outliers are marked "o". P-value corresponds to the value of the rank sum test corresponding to the alpha diversity index. B. PCoA plots of the unweighted Unifrac distances of the variation in microbiota composition detected in the ESCC and HC groups (sample name was removed). The horizontal box diagram shows the distribution of the values grouped on the first column coordinates, the vertical box diagram presents 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, the points of the same color indicate an origin of the same group, and the distance reflects the similarity of the samples. C. Scatter plot of two sets of NMDS analysis results. Each point in the figure represents a sample. The distances from the response sample are similar. The greater the distance between two points, the greater the difference in the community structure between the two

groups. In the plot, blue represents the esophageal mucosa group, and red corresponds to the esophageal cancer postoperative tissue group.

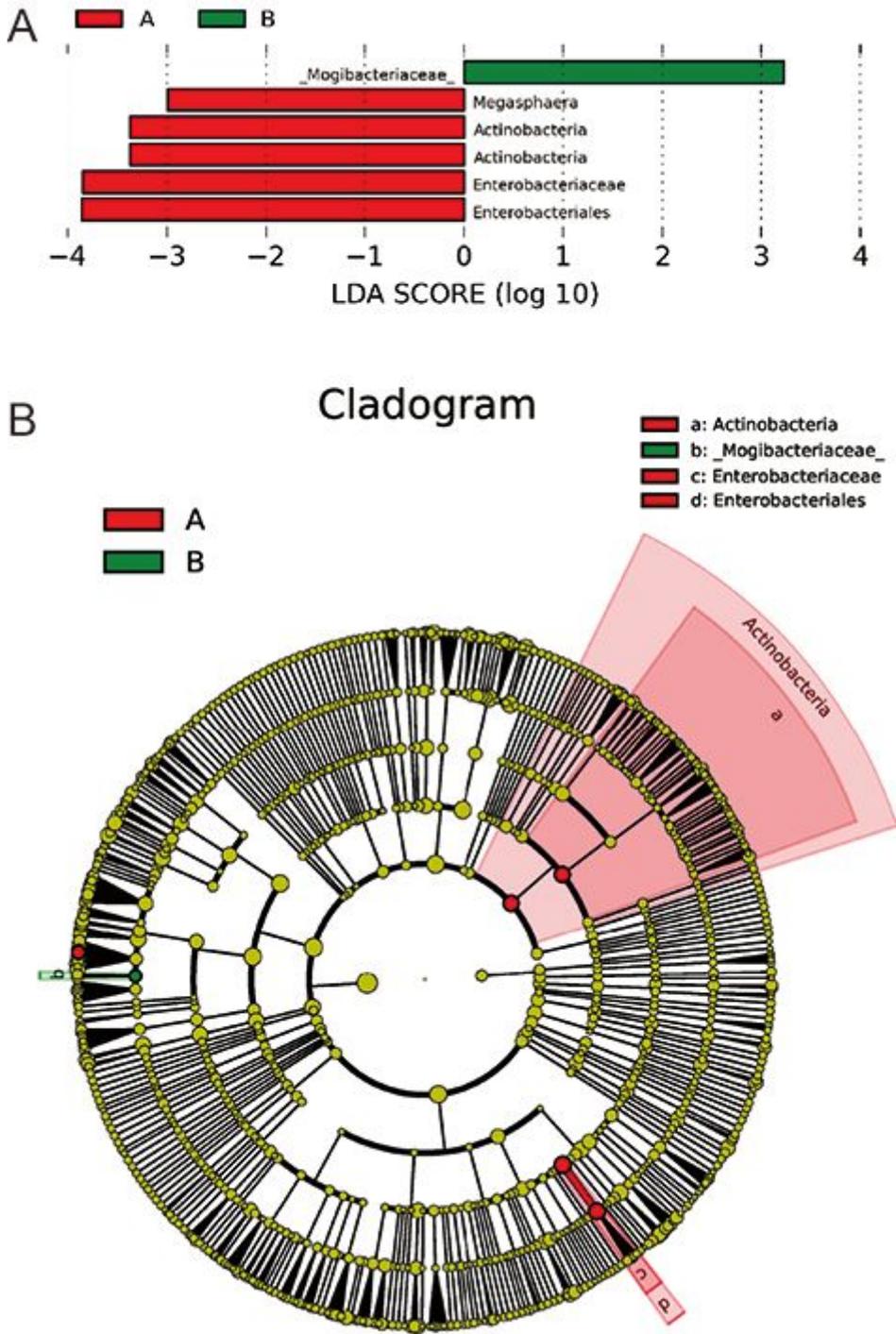


Figure 2

Species composition analysis A. Cladogram view of the representative microbial structure from the two groups. B. Differential bacteria between the two groups through LefSe analysis Notes: Dominant microbial classes for specific groups are represented by different colors. The diameter of each dot is proportional to the OTU abundance. Red and green represent the HC and ESCC groups. Significantly

different logarithmic LDA scores were set to 2. Only groups with significant differences are shown in the figure.

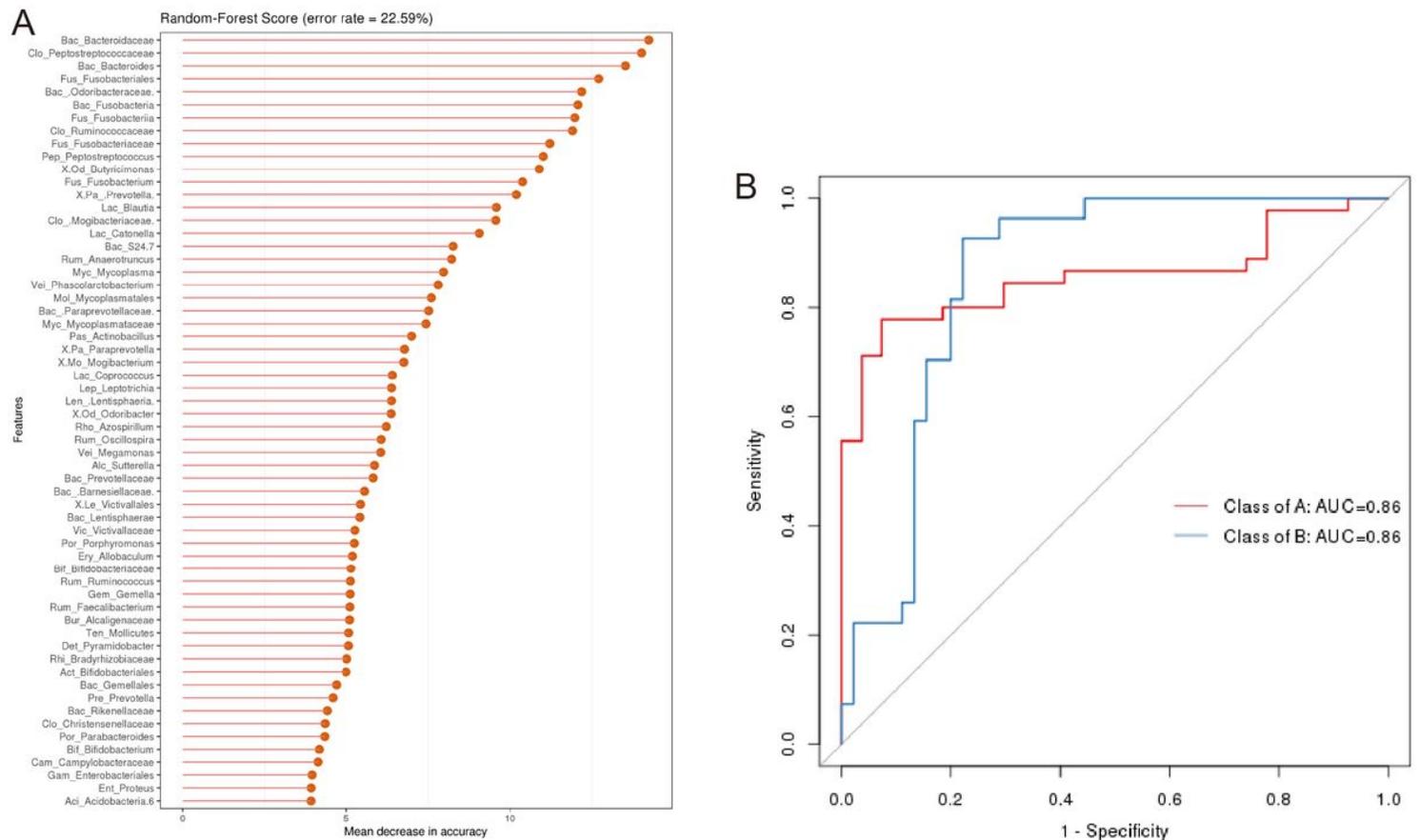


Figure 3

Comparison of the flora between groups and selection of species markers A. Species importance map Notes: The abscissa reflects the importance level, and the ordinate indicates the species name sorted according to importance. The figure reflects the genus of the bacteria in the classifier that plays a major role in the classification effect, which is arranged from the largest to the smallest. Error rate refers to the error rate of the random forest method for predicting classification with the features below. High values indicate that the classification accuracy based on the genus characteristics is low, and differences in genus characteristics may not be obvious between the groups. For example, when all levels are considered, the top 60 species are plotted Fig.3B ROC curve analyzing the clinical accuracy of using differential bacteria obtained from the ESCC and HC groups for the diagnosis of ESCC Notes: The point closest to the upper left of the ROC graph is the critical value with the greatest sensitivity and specificity.

Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS2.tif](#)

- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)