Characteristic and Interplay of Esophageal Microbiota in Esophageal Squamous Cell Carcinoma

Zheng Lin  
Fujian Medical University

Wenqing Rao  
Fujian Medical University

Zhisheng Xiang  
Fujian Medical University

Qiaoyan Zeng  
Fujian Medical University

Shuang Liu  
Fujian Medical University

Kaili Yu  
Fujian Medical University

Jinsong Zhou  
Fujian Medical University

Jianwen Wang  
Anxi County Hospital

Weilin Chen  
Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Yuanmei Chen  
Fujian Cancer Hospital: Fujian Provincial Cancer Hospital

Xiane Peng  
Fujian Medical University

Zhijian Hu (✉ huzhijian@fjmu.edu.cn)  
Fujian Medical University  https://orcid.org/0000-0003-1334-5334

Research

Keywords: esophageal squamous cell carcinoma(ESCC); microbiota, co-occurrence network, PICRUSt2

DOI: https://doi.org/10.21203/rs.3.rs-615147/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background**: Esophageal squamous cell carcinoma (ESCC) is a severe cancer in the world. The role of esophageal microbiota for ESCC is still uncertain. In the current study, 120 paired tissues from ESCC patients were collected, and 16s rRNA sequencing was performed to explore the esophageal microbiota.

**Results**: The present investigation shows that the diversity and composition of the microbiota in ESCC cancerous tissues and para-cancerous tissues is significantly different, this variation between subjects beta diversity mainly explained by regions and sampling seasons. Species *R.Mucilaginosa*, *P.Endodontalis*, unclassified species in genus *Leptotrichia*, genus *Phyllobacterium*, and genus *Sphingomonas* were enriched in cancerous tissue. On the other hand, class Bacilli, *N.Subflava*, *H.Pylori*, *A.Parahaemolyticus*, *A.Rhizosphaerae*, unclassified species in genus *Campylobacter* and genus *Haemophilus* were increased in para-cancerous tissue. Compared with the co-occurrence network in cancerous tissue, a denser and more complex association network was formed in para-cancerous tissue. Moreover, the above differential taxa also participated in both co-occurrence network but played quite different roles. Finally, the functional association analyses revealed the altered signaling pathways in ESCCs were correlated to esophageal microbiota.

**Conclusion**: Compared with para-cancerous tissues, microbiota in cancerous tissues showed significant differences in diversity and composition. The alterations in microbial co-occurrence network and functional pathways in ESCC tissues may be involved in carcinogenesis and the maintenance of local microenvironment for ESCC. These discoveries of the esophageal microbiota for ESCC patients may contribute to the etiology for ESCC prevention, diagnosis, early intervention, and treatment.

Background

Esophageal cancer is one of malignant cancer and contributes to diseases burden especially in developing countries. In 2020, there were 604,100 new cases of esophageal cancer globally[1]. China had the highest number of incident cases, accounting for up to 53% of new cases[2]. There are two main histological subtypes of esophageal cancer: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC), among which ESCC accounts for more than 85% of global esophageal cancer cases[3]. ESCC develops from the squamous epithelial cells that make up the inner lining of the esophagus. The etiology of ESCC is multifactorial and still not been well elucidated, which restricts the effective prevention of this disease. Thus, there is an urgent need to characterize to clarify the pathogenesis of ESCC and to explore new diagnostic and therapeutic possibilities.

Microbiome is an emerging area in the etiological exploration of various diseases. A balance of microbiome is strongly associated with a variety of human disease: the gastrointestinal microbiota plays an indispensable role in maintaining the health of humans[4]. The normal microenvironment is conducive to maintaining the steady state of the esophagus, and the dysbiosis microbiota can lead to the occurrence and development of ESCC[5, 6]. However, there are insufficient studies on the changes of esophageal microbiota in ESCC, and the findings among these studies are also inconsistent[7, 8]. The earliest findings are that lower microbial richness in upper digestive tract is independently associated with esophageal diseases[9]. However, these upper digestive microbiota are not colonized in the esophagus possibly. Then, Yang investigates the differences of esophageal microbiota in ESCC(n=18) and patients with physiological normal esophagus(n=11) by 16S rRNA profiling[8]. In our previous studies, we detected the esophageal microbiota associated with alcohol consumption[a well-known risk factor][10] and prognosis[11] in ESCC. Shao characterizes the microbial communities of paired tumor and non-tumor samples from only 67 patients with ESCC in the north of China[7]. Though the above studies explore the esophageal microbiota, the sample sizes were too small to find differences in diversity with adequate power [12] and did not take other factors that influenced microbiota into consideration.

Recently, the factors including regions[13], sampling season[14], and the interactions between microbes[15] have been shown to influence microbiota composition. He et al determined that host location showed the strongest associations with microbiota variations[13]. Moreover, there are significant shifts in human microbiome composition across different seasons[14]. Besides, microbes indulged in a variety of behaviors involving complex cooperation and communication may involve in the development of human diseases, no studies have reported the interaction network of esophageal microbiota so far. Hence, it is essential to consider these factors in microbiome study.

Therefore, to identify the impact of different regions, sampling seasons on esophageal mucosal microbiota of ESCC, we conducted a study and performed high-throughput profiling of the esophageal mucosal microbiota in paired cancerous and para-cancerous tissues in the same ESCC cases. Additionally, we detected the microbial interaction by co-occurrence networks and their functional effects on the host.

Results

**Basic characteristics of all participants**

The demographic and clinical data of the study population are reported in Table 1. Total 120 ESCC patients were recruited from two independent clinical centers. There were 89 males and 31 females. The median age was 61 years. The majority of the tumors originated from the middle(47.50%) and lower(43.33%) esophagus. Most of the patients were with advanced tumor(60.00%). About 70.83% had a high risk index.

**Esophageal microbial diversity analysis**

For alpha diversity, four characteristic metrics were evaluated. The alpha diversity in para-cancerous was significantly higher than cancerous tissue except for Pielou evenness index(Faith's phylogenetic diversity $P<0.001$, Observed OTUs $P=0.001$, Shannon index $P=0.027$) (Figure 1A). The difference of alpha diversity between paired cancer and para-cancerous tissue only observed in regions(Faith's phylogenetic diversity $P<0.001$, Observed ASVs $P<0.001$) and sampling seasons(Faith's phylogenetic diversity $P<0.001$, Observed ASVs $P=0.003$) (Figure 1B).
In multivariate Adonis test, the beta diversity of cancerous tissue was significantly different compared with para-cancerous tissue whichever measured by Jaccard distance (P=0.004), Unweighted unifrac distance (P=0.004) or Weighted unifrac distance (P=0.028) (Figure 2A). About 1%-2% of the variance in beta diversity was explained by tissue type. What is more, the differences of beta diversity between within paired cancerous and para-cancerous tissue were associated with regions (P=0.032) and sampling seasons (P=0.042) based on Unweighted unifrac distance (Figure 2B).

**Microbial composition analysis**

Pair-wise PCoA results were displayed in Figure 4A. Based on Bray Curtis, Jaccard, and Unweighted unifrac distance, microbiota from the two tissue types were separated into clusters. Total 9,453 features were found after the sequence denoised in all samples (Figure 6A). More taxa were observed in para-cancerous compared with cancerous tissue (8,133 vs 6,533) and about 5,213 taxa were detected in both tissue (Figure 3B). The composition of esophageal microbiota between cancerous and para-cancerous tissue in level phylum and genus was shown in Figure 3C and Figure S1, respectively. To elucidate the phylogenetic relationship of the cancerous and para-cancerous microbiota, the heat trees of microbiota (the relative abundance >0.1%) were plotted in Figure 3D. The relative abundance of correspondant branches of bacteria in phylum Proteobacteria, Bacteroidetes, Fusobacteria and Firmicutes was similar between cancerous and para-cancerous tissue. However, some bacteria were enriched in different tissues (Figure 3E). In class Alphaproteobacteria, the branches of bacteria of order Rhizobiales and Sphingomonadales were enriched in cancerous tissue while order Rhodospirillales was higher in para-cancerous tissue. The most bacteria of class Alphaproteobacteria and its branch were enriched in par-cancerous tissue except family Klebsiella. Moreover, the microbiota from phylum TM7 and its branches of bacteria, class Bacilli, family Helicobacteracea and its genus Helicobacter and its species pylori were higher in para-cancerous tissue.

**Differential abundance analysis**

A total of 56 differential taxa were picked out by ANCOM2 algorithm (Figure 3G). As the host region exerted the strongest effect on microbiota, we divided all participants into Zhangzhou group and other regions group (Table S1). There were 32, 36, 16 differential abundance taxa between cancerous and para-cancerous tissue in all regions group, Zhangzhou city group and other regions group, respectively (Figure 3F). There were only three shared differential bacteria in cancerous and para-cancerous tissue from different regions named family Enterobacteriacea, unclassified species from genus Sphingomonas and genus Phyllobacterium (Figure 3G). It was clearly ascertainment differential bacteria in phylum Firmicutes and Proteobacteria were enriched in para-cancerous tissue from Zhangzhou city. Moreover, we observed an interesting alteration that the unclassified species in genus Mycoplanae was enriched in cancerous tissue from other regions whereas enriched in para-cancerous tissue from Zhangzhou city. Sampling seasons were another powerful factor that influenced host microbiota (Figure 3G). The microbiota in phylum Cyanobacteria in cancerous tissue sampled in summer had a higher relative abundance. The relative abundance of bacteria in phylum Proteobacteria were significantly enriched in par-cancerous tissue when sampled in spring and summer from Zhangzhou city.

Hence, the dominant candidate differential taxa (Table 2) between cancerous and para-cancerous tissue were selected according to the grand means of relative abundance were exceeded 0.1% from above 56 differential taxa. They were species *R. mucilaginosa, P. Endodontalis*, unclassified species in genus *Leptotrichia*, unclassified species in genus *Phyllobacterium*, and unclassified species in genus *Sphingomonas*, which enriched in cancerous tissue. On the other hand, class Bacilli, *N. Subflava, H. Pylori, A. Parahaemolyticus, A. Rhizosphaerae*, unclassified species in genus *Campylobacter* and unclassified species in genus *Haemophilus* were enriched in para-cancerous tissue. Next, to explore the confounding effect of regions, sampling seasons, tumor location and risk index on host microbiota, we added these four covariates into ANCOM2 analysis. The results showed that the relative abundance of unclassified species in genus *Leptotrichia*, unclassified species in genus *Sphingomonas* and *A. Rhizosphaerae* could be influenced by regions; the relative abundance of unclassified species in genus *Campylobacter* could be influenced by sampling seasons and tumor location. However, none of the differential taxa were significantly various in either low or high risk index with ESCC.

**Microbial co-occurrence networks**

To understand the interaction among esophageal microbiota in cancerous and para-cancerous tissue, we illustrated the microbial co-occurrence networks of two groups. There were 3,089 positive and 348 negative correlations in cancerous tissue, while the para-cancerous samples had 3,761 positive and 355 negative correlations. Obviously, the microbial co-occurrence networks were distinct between the cancerous and para-cancerous tissue (Figure 3G). The centrality of co-occurrence networks was described with three dimensions which were degree, betweenness, and closeness centrality, respectively. Interestingly, the degree and betweenness centrality of co-occurrence networks were from phylum Proteobacteria and certain differential taxa did not emerged in the co-occurrence networks. The role of differential taxa in different microbiota networks was described with three dimensions which were degree, betweenness, and closeness centrality, respectively. Interestingly, the degree and betweenness centrality of co-occurrence networks were from phylum Proteobacteria and certain differential taxa did not emerged in the co-occurrence networks.

To quantify such difference, we counted the number of node and its centrality in the microbial networks under different tissue type. As expected, the numbers of interacted microbiota in para-cancerous tissue (273 nodes) were higher than that in cancerous tissue (201 nodes) (Figure 4A and B). The centrality of co-occurrence networks was described with three dimensions which were degree, betweenness, and closeness centrality, respectively. Interestingly, the degree and closeness of shared nodes between cancerous and para-cancerous tissue were quite different. Next, the edges of the networks were evaluated (Figure 4C). Despite having a few overlapped edges, the distribution of rank of overlapped edges varies in cancerous and para-cancerous tissue.

To reveal the importance of above candidate differential taxa in the network, the heatmap was shown in Figure 4D. The most of differential taxa in the network were from phylum Proteobacteria and certain differential taxa did not emerge in the co-occurrence networks. The role of differential taxa in different networks (cancer vs para-cancer) was also various. The discrepancies of microbial co-occurrence networks in two groups may be attributed to the specific metabolism in different tissue types.

**The association between esophageal microbiota and predict function**

Most of the differential taxa in para-cancer were negatively associated with EC 2.7.10.1 which regulated the EGFR, ERBB2, ERBB4, and FGFR1 signaling pathways. Different from this, the differential taxa in cancer were positively associated with EC2.1.1.107 and EC3.1.3.16 in the MET and PTEN signaling pathways.
Among the eight MetaCyc metabolic pathways which were significantly differed between cancer and para-cancer tissues (Figure 5B, Table S2), the relative abundance of PWY-3661 and PWY-7431 was increased in cancerous tissue, and other six pathways (PWY-1882, PWY-5265, PWY-6565, PWY-6731, PWY-6906, and PWY-7391) enriched in para-cancerous tissue. Notably, the differential taxa played various roles in different tissues. For example, unclassified species from genus *Phylllobacterium* and *Sphingomonas* were positively associated with PWY-6906 and PWY-1882 in cancer, whereas negatively with PWY-6731 and PWY-6565 in para-cancer. Similarly, the correlation was demonstrated between taxa which were *A. Parahaemolyticus* and *A. Rhizosphaerae* and pathways including PWY-7431, PWY-7391, PWY-6731, and PWY-1882 in cancer, but PWY-6565 in para-cancer. In addition, the unclassified species in genus *Haemophilus* only associated with the pathways in cancer.

**Discussion**

To characterize the esophageal microbiota in the ESCC, we compared the microbial diversity and composition of paired cancerous and para-cancerous tissues for ESCC in this study. Moreover, 56 candidate differential taxa were determined, of which 13 taxa have more abundant. Besides, we suggested that the microbial co-occurrence network was another critical aspect of the microbial community. Furthermore, the relationship between differential taxa and predicted functional ESCC-related gene pathway was investigated.

We observed the differences of microbial diversity between the cancerous and para-cancerous tissues for ESCC, which was similar to findings of Yang[8] and Yu[9]. The microbial structure is modified due to the altered esophageal microenvironment in the progression of carcinogenesis. Interestingly, in general linear regression model, the within-pair differences in microbial diversity were only impacted by regions and sampling seasons. It is in agreement with many studies that environmental factors including sampling regions or seasons are the relevant factors affecting microbiota. Daphna et al. report that more than one-fifth of the difference in microbiome beta diversity can be inferred from diet and lifestyle for different seasons[16]. An animal experiment also indicated that seasonal variation plays a major role in the temporal changes of the microbial community structure[17]. Moreover, Samuel et al. found that the human fecal microbiota altered due to season cycling[18]. He et al. illustrated that geographical factor could strongly alter the composition of gut microbiota[13]. However, there are no other studies considering the effect of environmental factors on the esophageal flora yet, except our previous research[10]. This study demonstrated again that the impact of sampling regions or seasons on the diversity of esophageal microbiota cannot be ignored. Therefore, further studies involving microbiomics are needed to be taken into consideration the effect of the environmental factors on microbiota to ensure the generalizability.

Several differential bacteria were detected in para-cancerous tissue for ESCC. Of note, strains of *R. Mucilaginosa* were differentially abundant in oral wash with head and neck squamous cell carcinoma compared to healthy controls[19]. It could be inferred that higher *R. Mucilaginosa* was always presented in cancer patients. What is more, it is reported *R. Mucilaginosa* upregulated TNF-α and upregulation of CD36 in all cell lines in cocultures[20] and increased abundance of *R. Mucilaginosa* exhibited the ability to produce acetaldehyde[21, 22]. The above functional could explain the findings of the present study that higher *R. Mucilaginosa* in esophageal cancerous tissue than that in para-cancerous. Another differential bacterium, *P. Endodontalis*, was dominantly associated with oral diseases[23]. Interestingly, *P. Endodontalis* was the most representative taxa which were abundant in gastric adenocarcinoma microbiome[24]. In accordance with our results, Richard et al. identified colorectal mucosal genus *Sphingomonas* particularly enriched in colitis-associated cancer[25]. Moreover, *Sphingomonas* were higher in the bladder mucosa cancerous tissues[26]. This bacteria were called passenger bacteria according to Wang's hypothesis[27]. If microbiota is involved in carcinogenesis, we can expect that it will be enriched in the para-cancerous tissue. Many studies indicated *N. subflava* is significantly more present in healthy control, as found in our study[28, 29]. In addition, similar observations were previously reported *H. Pylori* was increased in adjacent compared to tumor tissues for gastric cancer, which is in the line with our results[30–32]. It is been proven that *H. Pylori* is pathogenic bacteria again. *Campylobacter* usually linked to gingivitis and periodontitis, has recently been associated with gastrointestinal diseases[33]. Leung and Wang demonstrated *Campylobacter* as the key genera of the tumor mucosa in colorectal cancer[27, 34]. Sandra investigated *Campylobacter* species appear to be more prevalent and abundant in the esophagus of patients in the Barrett's esophagus[35], but its abundance decreased with the progression to EAC[36]. Our current study suggested this finding in ESCC.

A disease-associated microenvironment could be affected by a specific microbial network. Several studies have shown ecological interaction networks of microbiota in cancer and adjacent mucosal tissues in gastric[32] and colorectal cancer[27]. The esophagus is a part of the gastrointestinal tract, we highlight the esophageal microbial co-occurrence network for ESCC in this study. Our microbial co-occurrence network analysis suggests bacterial compositions in different sample types show specific correlation patterns. Compared with the cancerous group, more microbial taxa were involved in the para-cancerous group. And, they formed a denser and more complex association network. Interestingly, strong co-occurrence interactions formed by family Lachnospiraceae, *C. aerofaciens*, and undefined species from genus *Blaufia*, showed the centralities of these taxa in both two networks. This suggests that hub taxa may have a predominant impact on the structure of the microbiota in ESCC patients, which deserves further investigations. Notably, the interactions of *B. adolescentis* in the cancerous samples may contribute to the formation of a disease-specific interaction network. Further, the markedly increased interactions of *Roseburia* and its species *R. faecis* in the para-cancerous network might also contribute to the maintenance of the esophageal microenvironment. Moreover, the above differential taxa also participated in the co-occurrence network and had various importance. It is implicated that the microbial interaction is complex, the full consideration of microbial interact network should be taken in further microbiome studies.

We have preliminarily revealed differences in the predicted microbiota functions in cancerous and para-cancerous tissues. The enriched differential taxa in cancerous tissue were positively associated with the MET(EC2.1.1.107) and PTEN(EC3.1.3.16) pathway. The increased differential taxa in para-cancerous tissue were negatively associated with the EGFR, ERBB2, ERBB4, FGFR1(EC2.7.10.1), and MTOR(EC2.1.11.1) pathway. Mutations in PTEN, ERBB2, ERBB4, MET, and MTOR are found in ESCCs. Interestingly, regardless of the presence or absence of mutations in these genes in ESCCs, their products (ERBB2, MET, and MTOR) are often overexpressed in tumor tissues, indicating their vital carcinogenic role in ESCC[37]. It is possible that the abnormal profiled abundance of differential bacteria causes changes in expression of PTEN, ERBB2, ERBB4, MET, and MTOR and stimulates tumor growth. Additionally, the enriched...
differential MetaCyc pathways in cancerous tissue are linked to amino acid degradation. It has been implicated that microbiota in dysplasia tissues perturbed amino acid metabolism probably involving in the process of tumor development[38]. On the other hand, a large proportion of increased pathways in para-cancerous tissue are associated with bacteria-related biomolecules synthesis. This could explain why the more abundant microbiota were detected in para-cancerous. We found that \textit{H. pylori} was negatively associated with the polyamine biosynthesis which was in line with the Takashima[39] investigation that \textit{H. pylori} significantly inhibited proliferation of enzymes of polyamine biosynthesis of cells. What is more, Rousseau[40] explored that \textit{Haemophilus} in animals created changes in peptidoglycan metabolism, it is also consistent with our results. To develop a deeper understanding of esophageal carcinogenesis, further studies are needed to examine the significance of microbial functional variations in the ESCC microenvironment.

**Advantages and Limitation**

In this study, esophageal mucosa samples were obtained undergoing surgery from ESCC patients, thus avoiding possible oral microbial contamination that may occur during upper digestive endoscopy sampling. Our work provided insights into the composition, function and interaction network of the mucosa-associated bacterial community in the tumor microenvironment in ESCC.

However, our study had certain limitations. Firstly, our experiment protocol of extract the DNA of bacteria did not include a bead-beating step. This could lead to an overestimation of the proportion of Gram negative bacteria in the results. Additionally, this study did not include esophageal tissues from individuals without ESCC for comparison, the application of these results was limited. Furthermore, PICRUSt2 is a predictive tool to assessment the microbial function, this method may not accurately reflect the biological functions of microbiota.

**Conclusion**

Compared with para-cancerous tissues, microbiota in cancerous tissues showed significant differences in diversity and composition. The alterations in microbial co-occurrence network and functional pathways in ESCC tissues may be involved in carcinogenesis and the maintenance of local microenvironment for ESCC. Future research would concentrate on verification using a larger number of samples and multiple centric populations, as well as expanding our work into cell culture systems and animal models to examine the pathogenic roles of microbiota in ESCC. These discoveries of the esophageal microbiota for ESCC patients may contribute to the etiology for ESCC prevention, diagnosis, early intervention, and treatment.

**Methods**

**Studying population**

We performed a hospital-based retrospective study of 120 patients pathologically diagnosed with primary ESCC between February 2013 and October 2017 at Fujian Provincial Cancer Hospital and Zhangzhou Municipal Hospital(Figure 6A). Subjects were chosen according to the following criteria. Inclusion criteria: (a) underwent esophagectomy surgery; (b) pathologically diagnosed with primary ESCC; (c) tumor stage clarified with number of harvested lymph nodes (HLN) ≥ 20; (d) undergoing neither preoperative radiotherapy nor chemotherapy; (e) no antibiotic use through preoperative two months; (g) no record of other infectious diseases; and (h) resident of Fujian province for more than 10 years. Exclusion criteria: (a) incomplete clinicopathological data and non-availability of tissue samples; (b) metastatic malignancy or recurrent esophageal cancer; (c) received pharmacotherapy (such as oral, intramuscular, and intravenous antibacterial drugs, various probiotics or other drugs affecting the microbiota,) within a month. Written informed consent was obtained from all the patients. The study was approved by the Ethics Committee of Fujian Medical University (approval no. 201495).

**Demographic and clinical information**

The basic information of all the participants was collected through a detailed questionnaire comprising of sociodemographic status, dietary habits, daily physical activity, smoking status, alcohol consumption, family history of cancer and gastrointestinal symptoms. Clinicopathological features (viz. differentiation status, location, and tumor, node, and metastasis (TNM) stage) for each patient were also collected from their respective medical records.

**Sample collection and preservation**

Paired cancerous and para-cancerous tissue samples were obtained from each patient immediately after surgical resection in the operating room. The para-cancerous tissue samples were from an area at a distance of 3 cm from the cancerous tissue. The tissues samples were cut into small pieces and placed in autoclaved cryovials, stored in liquid nitrogen for 1 h, and, then transferred to a -80°C refrigerator for storage. All samples were evaluated by pathological haematoxylin-eosin (HE) staining.

**Bacterial DNA extraction and 16S rRNA sequencing**

The sodium dodecyl sulphate (SDS) method was used to extract bacterial DNA from the samples. The extracted DNA was quantitatively detected by Qubit fluorometer(Invitrogen, America), and the results were acceptable. Each extraction was performed with a blank buffer control to detect contaminants from either reagents, or other unintentional sources. However, the negative controls detected too few DNA to prepare library and hence were not sequenced.

Amplification of the 16S rRNA gene used primers targeting regions V3–V4, which included forward primer (341F: 5’-CCTAYGGRBGCASCAG-3’) and reverse primer (806R: 5’-GGACTACNNGGTATCTAA-3’). The sequencing platform was the HiSeq2500 PE250(Illumina, America).

**Sequence data processing**
Raw sequencing data from patients with ESCC were imported into Quantitative Insights Into Microbial Ecology (QIIME2-2020.02) and processed using the DEBLUR algorithm to denoise and then inferred exact amplicon sequence variants (ASVs). The detailed analysis workflow was presented in Figure 6B. The curated ASVs were aligned and annotated by the Naïve Bayes classifier using the Greengenes (version 13.5) database, and were used for the subsequent construction of the phylogenetic tree. ASVs were submitted to a pre-trained Naïve Bayes Classifier (NBC) trained on full-length 99% Greengenes reference for the taxonomic classification. Before diversity analysis, the threshold for rarefaction depth was decided by minimizing sequences loss while maximizing the number of samples. At the depth of 10000 sequences per sample, the richness of observed communities had tended to be saturated, with 113 pairs of cancer and para-cancer samples were kept for the Alpha and Beta diversity metrics construction. For the purpose of providing better resolution and limiting false discovery rate (FDR) penalty on statistical tests, the low abundance features (with total ASVs counts less than 100 or detected in less than 20 samples) had been filtered before the differential abundance analysis (955 features left) (Table S3).

Statistical analysis

Questionnaires and clinicopathological data were double-entered into EpiData (version 3.1, Denmark). The demographic and baseline clinical features were displayed using n(%). The individuals’ risk index of ESCC was calculated by variables included age, smoking, drinking, eating speed, hot food, pickled food, and fruit from questionnaire (Sup_file 1). All statistical analyses were evaluated using R software (R version 4.0.2), and two-tailed P<0.050 was considered statistically significant.

Since the samples were paired, the Wilcoxon sign rank test was applied for comparisons of Alpha diversity (observed ASVs, Shannon index, and Faith’s Phylogenetic Diversity) between cancer and matched para-cancer tissues. For Beta diversity (Bray-Curtis, Jaccard, unweighted UniFrac and weighted UniFrac distances), the Adonis action of PERMANOVA tests were implemented to evaluate whether the variation of distances could be explained by other controlled variables (ESCC risk scores, sampling seasons, residential regions, tumor locations, and TNM stages).

The general linear models were used to test whether the aforementioned controlled variables could impact the diversity metrics. For Alpha diversity, the paired differences of diversity metrics within each pair sample were calculated by subtracting the diversity values from cancer tissues to the corresponding para-cancer tissues and served as dependent variables. While for the Beta diversity, the dependent variables were defined as the pairwise distances between cancer and matched para-cancer tissues (within-subject distances). Before linear regression, all dependent variables were checked for normality, and the natural logarithmic transformation was applied for reducing skewness. All P values for controlled variables were corrected by the Benjamini-Hochberg FDR procedure.

PCoA plotting which based on the Bray-Curtis, Jaccard, unweighted UniFrac and weighted UniFrac distances were used to depict the microbiome composition. The ANCOM2 tests were performed to detect the differential abundance in different tissue groups. It had been documented that the regional variation limits applications of microbial-disease association models, therefore, the data were split into Zhangzhou group (50 pairs of samples) and other Cities from Fujian Province group (70 pairs of tissues) according to patients’ residential regions.

For the functional prediction, the PICRUSt2 pipeline was used to generate predictions for EC numbers and MetaCyc pathways. The strength of edges of microbial co-occurrence network were assessed by SPARCC algorithm, and the interaction network diagram visualized with Cytoscape. The top hub taxa were assessed by plugin cytoHubba in Cytoscape. Then, we selected four well established pathogenesis enzyme genes in ESCC according to Lin[37] from KEGG database. Then, Spearman correlation was performed to explore the association between the differential taxa and four enzyme genes. DESeq algorithm was applied to calculate the differential MetaCyc pathways and visualized by volcano plot, then the Spearman correlation between differential taxa and differential MetaCyc pathways was calculated.

List Of Abbreviations

ESCC: oesophageal squamous cell carcinoma; EAC: esophageal adenocarcinoma; HLN: harvested lymph nodes; TNM: tumor, node, and metastasis; HE: haematoxylin-eosin; SDS: sodium dodecyl sulphate; NBC: Naïve Bayes Classifier; QIIME: Quantitative Insights Into Microbial Ecology; ASVs: amplicon sequence variants; FDR: false discovery rate.

Declarations

Ethical approval

All procedures performed in studies involving human participants were carried out by the ethical standards of the institutional and/or national research committee, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from participants, and the study was approved by the Institutional Review Board of Fujian Medical University (number: 201495).

Consent for publication

NO

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of interest

The author(s) declare that there are no conflicts of interest.
Funding

Funding was obtained from the National Key R&D Program of China (No.2017YFC0907100), Medical Innovation project of Fujian Province (No.2018-CX-38), Central government-led local science and technology development special project (No.2020L3009) and Startup Fund for scientific research, Fujian Medical University (No.2020H2008).

Authors and contributors

Zheng Lin: Writing – Review and Editing, Conceptualization, Methodology and Software. Wenqing Rao: Writing – Original Draft Preparation, Writing – Review and Editing, Conceptualization, Investigation and Methodology. Zhisheng Xiang: Formal analysis, Conceptualization, Investigation and Methodology. Qiaoyan Zeng: Data Curation, Methodology and Investigation. Shuang Liu: Software, Validation and Visualization. Kaili Yu: Investigation and Visualization. Jinsong Zhou: Data Curation and Investigation. Jianwen Wang: Software. Weilin Chen: Software. Yuanmei Chen: Resources. Xiane Peng: Methodology, Validation. Zhijian Hu: Methodology, Validation, Writing - Review and Editing, Supervision, Project administration and Funding. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by a grant from the National Key R&D Program of China (No.2017YFC0907100), Medical Innovation project of Fujian Province (No.2018-CX-38), Central government-led local science and technology development special project (No.2020L3009) and Startup Fund for scientific research, Fujian Medical University (No.2020H2008). There was no potential financial or personal conflict of interest. We are grateful to the Fujian Provincial Cancer Hospital Affiliation to Fujian Medical University for data collection.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jamal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021. doi:10.3322/caac.21660.
2. Cao W, Chen H-D, Yu Y-W, Li N, Chen W-Q. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J (Engl). 2021;134:783–91. doi:10.1097/CM9.0000000000001474.
3. Arnold M, Laversanne M, Brown LM, Devesa SS, Bray F. Predicting the Future Burden of Esophageal Cancer by Histological Subtype: International Trends in Incidence up to 2030. Am J Gastroenterol. 2017;112:1247–55. doi:10.1038/aig.2017.155.
4. Fischbach MA. Microbiome: Focus on Causation and Mechanism. Cell. 2018;174:785–90. doi:10.1016/j.cell.2018.07.038.
5. 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. doi:10.1053/j.gastro.2009.04.046.
6. Liu N, Ando T, Ishiguro K, Maeda Q, Watanabe O, Funasaka K, et al. Characterization of bacterial biota in the distal esophagus of Japanese patients with reflux esophagitis and Barrett's esophagus. BMC Infect Dis. 2013;13:130. doi:10.1186/1471-2334-13-130.
7. 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. doi:10.1002/cncr.32403.
8. Yang W, Chen C-H, Jia M, Xing X, Gao L, Tsai H-T, et al. Tumor-Associated Microbiota in Esophageal Squamous Cell Carcinoma. Front Cell Dev Biol. 2021;9:641270. doi:10.3389/fcell.2021.641270.
9. Yu G, Gail MH, Shi J, Klepac-Ceraj V, Paster BJ, Dye BA, et al. Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. Cancer Epidemiol Biomarkers Prev. 2014;23:735–41. doi:10.1158/1055-9965.EPI-13-0855.
10. Rao W, Lin Z, Liu S, Zhang Z, Xie Q, Chen H, et al. Association between alcohol consumption and oesophageal microbiota in oesophageal squamous cell carcinoma. BMC Microbiol. 2021;21:73. doi:10.1186/s12866-021-02137-x.
11. Liu Y, Lin Z, Lim Y, Chen Y, Peng X-E, He F, et al. Streptococcus and Prevotella are associated with the prognosis of oesophageal squamous cell carcinoma. Journal of medical microbiology. 2018;67:1058–68. doi:10.1099/jmm.0.000754.
12. Casals-Pascual C, González A, Vázquez-Baeza Y, Song SJ, Jiang L, Knight R. Microbial Diversity in Clinical Microbiome Studies: Sample Size and Statistical Power Considerations. Gastroenterology. 2020;158:1524–8. doi:10.1053/j.gastro.2019.11.305.
13. He Y, Wu W, Zheng H-M, Li P, McDonald D, Sheng H-F, et al. Characterization of bacterial biota in the distal esophagus of Japanese patients with reflux esophagitis and Barrett's esophagus. BMC Infect Dis. 2013;13:130. doi:10.1186/1471-2334-13-130.
14. Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y. Seasonal variation in human gut microbiome composition. PLoS One. 2014;9:e90731. doi:10.1371/journal.pone.0090731.
15. West SA, Duggie SP, Buckling A, Gardner A, Griffin AS. The Social Lives of Microbes. Annu. Rev. Ecol. Evol. Syst. 2007;38:53–77. doi:10.1146/annurev.ecolsys.091206.095740.
16. Rothschild D, Weisbrod O, Barkan E, Kulishikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555:210–5. doi:10.1038/nature25973.
17. Maurice CF, Knowles SCL, Ladau J, Pollard KS, Fenton A, Pedersen AB, Turnbaugh PJ. Marked seasonal variation in the wild mouse gut microbiota. ISME J. 2015;9:2423–34. doi:10.1038/ismej.2015.53.
18. Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science. 2017;357:802–6. doi:10.1126/science.aan4834.
47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

Tables

Table 1. Basic information of 120 ESCCs

| Variables     | n    | (%)   |
|---------------|------|-------|
| Gender        |      |       |
| Female        | 31   | (25.83) |
| Male          | 89   | (74.17) |
| Age           | 61 (56, 65) |
| ≤60           | 56   | (46.67) |
| >60           | 64   | (53.33) |
| Region        |      |       |
| Zhangzhou     | 50   | (41.67) |
| Fuzhou        | 11   | (9.17) |
| Putian        | 11   | (9.17) |
| Quanzhou      | 33   | (27.50) |
| Others        | 15   | (12.50) |
| Season        |      |       |
| Spring        | 39   | (32.50) |
| Summer        | 26   | (21.67) |
| Autumn        | 24   | (20.00) |
| Winter        | 31   | (25.83) |
| Tumor location|      |       |
| Upper thoracic| 11   | (9.17) |
| Middle thoracic| 57 | (47.50) |
| Lower thoracic| 52   | (43.33) |
| TNM           |      |       |
| Stage I       | 12   | (10.00) |
| Stage II      | 36   | (30.00) |
| Stage III     | 72   | (60.00) |
| Risk indexb   | 1.99 (1.44, 2.49) |
| Low           | 70   | (29.17) |
| High          | 170  | (70.83) |

\* The italic numbers were the medians, 25\textsuperscript{th} and 75\textsuperscript{th} percentiles, respectively.

\* The median value of risk index in controls (see supplementary file 1) were used as the cut-off for low and high categories definition.

Table 2. The candidate taxa with differential abundance between cancer and para-cancer tissues in ESCC\*.
| Phylum        | Class          | Order          | Family              | Genus       | Species         | Relative abundance (%) | FC<sup>c</sup> |
|--------------|----------------|----------------|---------------------|-------------|-----------------|-------------------------|----------------|
| Actinobacteria | Actinobacteria | Actinomycetales | Micrococcaceae      | Rothia      | Mucilaginosa     | 0.291                  | 6.767         |
| Bacteroidetes | Bacteroidia    | Bacteroidales  | Porphyromonadaceae  | Porphyromonas | Endodontalis    | 0.286                  | 10.593        |
| Firmicutes    | Bacilli        | -              | -                   | -           | -               | 0.009                  | 0.190         |
| Fusobacteria  | Fusobacteriia  | Fusobacteriales | Leptotrichiaceae    | Leptotrichia | UC              | 0.482                  | 3.257         |
| Fusobacteria  | Fusobacteriia  | Fusobacteriales | Leptotrichiaceae    | Leptotrichia | UC              | 0.287                  | 7.757         |
| Proteobacteria | α-Proteobacteria | Rhizobiales   | Phyllobacteriaceae  | Phyllobacterium | UC         | 8.805                  | 9.239         |
| Proteobacteria | α-Proteobacteria | Sphingomonadales | Sphingomonadaceae   | Sphingomonas | UC              | 2.736                  | 6.641         |
| Proteobacteria | β-Proteobacteria | Neisseriales  | Neisseraeae          | Neisseria   | Subflava      | 0.458                  | 0.482         |
| Proteobacteria | ε-Proteobacteria | Campylobacterales | Campylobacteraceae | Campylobacter | UC              | 0.253                  | 0.456         |
| Proteobacteria | ε-Proteobacteria | Campylobacterales | Helicobacteraceae  | Helicobacter | Pylori   | 0.322                  | 0.356         |
| Proteobacteria | γ-Proteobacteria | Pasteurellaeae | Pasteurellaceae      | Actinobacillus | Parahaemolyticus | 0.685                  | 0.472         |
| Proteobacteria | γ-Proteobacteria | Pasteurellaeae | Pasteurellaceae      | Haemophilus  | UC           | 0.068                  | 0.398         |
| Proteobacteria | γ-Proteobacteria | Pseudomonadales | Moraxellaeae        | Acinetobacter | Rhizosphaerae | 0.057                  | 0.245         |

<sup>a</sup> The candidate differential taxa were selected according to the following two conditions: 1) pass the ANCOM tests conducted in patients from Zhangzhou City (50 pairs) or the other regions (70 pairs), or the pooled population (120 pairs) (Fig 4G); 2) the grand means of relative abundance were exceeded 0.1%.

<sup>b</sup> Mean relative abundance were presented.

<sup>c</sup> Fold change (FC) = mean relative abundance in cancer/ mean relative abundance in para-cancer.

<sup>d</sup> The differential abundant taxa between cancer and para-cancer were selected by the ANCOM2 algorithm under detected cut-off at 0.7, and were adjusted for sex, age, risk index, TNM, season, tumor location and regions. The normalized ANCOM's W statistics were calculated by divided the W over the number of total taxa which were identified as none-structural zero.

<sup>e</sup> The ANCOM2 algorithm was applied for association detection. The detected cut-off at 0.7 was adopt for all analyses. The variables included in ANCOM2 comparisons were in line with those included in differential abundant taxa detection. TRUE or FALSE indicated that the relative abundance of candidate taxa could be or not be influenced by specific factors, respectively.

UC, unclassified; CA, cancer; PC, para-cancer; FC, fold change; RI, risk index.

**Figures**
**Figure 1**

Microbial comparison for alpha diversity between cancer and para-cancer. (A) Alpha diversity based on Pielou evenness index ($P=0.887$), Faith’s phylogenetic diversity ($P<0.001$), Observed ASVs ($P<0.001$), and Shannon index ($P=0.027$). (B) The general linear regression analysis to detect the pairwise difference for alpha diversity adjusting gender, age, risk index, region, sampling season, tumor location, and TNM stage.
Figure 2
Microbial comparison for beta diversity using multivariate Adonis test between cancer and para-cancer. (A) Beta diversity based on Bray Curtis distance (P=0.052), Jaccard distance (P=0.004), Unweighted unifrac distance (P=0.004), and Weighted unifrac distance (P=0.028). (B) The general linear regression analysis to detect the within pairs difference for beta diversity adjusting gender, age, risk index, region, sampling season, tumor location, and TNM stage.
Figure 3

The profile of esophageal microbiota between cancer and para-cancer. (A) Principal coordinate analysis (PCoA) plots based four distances between cancer and para-cancer. (B) The overlap of microbiota features between cancer and para-cancer. (C) Microbial relative abundances at the level phylum in cancer and para-cancer. (D) The heat tree plot of relative abundance (higher than 0.1%) of microbiota in cancer and para-cancer. (E) The heat trees of microbiota for log2 ratio of median relative abundance between cancer and para-cancer by univariate Wilcoxon rank sum test. (F) The differential taxa between cancer and para-cancer in All regions (panel A), Zhangzhou city (panel B) and Other regions (panel C) after multivariate adjustment using ANCOM2. (G) Heatmap of relative abundance of differential microbiota (the red words of right side of heatmap is enriched differential taxa in cancer, the green is that in para-cancer).
Figure 4

The analyses of microbial co-occurrence networks (A) Co-occurrence network of the microbiota in cancer and para-cancer tissue (only Sparcc absolute r>=0.4 was shown). (B) Discrepancies of co-occurrence network nodes between cancer and para-cancer tissue and its measurement dimensions included degree, betweenness, and closeness centrality. (C) Discrepancies of co-occurrence edges between cancer and para-cancer tissue. (D) Discrepancies of the importance of differential taxa in the co-occurrence network.
Figure 5

The association between esophageal microbiota and predict function. (A) The association between the differential microbiota and ESCC-related functional enzymes. (B) Volcano plot showing the differential MetaCyc metabolic pathways between cancerous and para-cancerous tissue. X-axis indicates the log2(fold change), and Y-axis indicates log10(FDR). The significant pathways which enriched in cancer (FDR < 0.05 and log2 fold change>2) are colored red dots, that are increased in para-cancer(FDR < 0.05 and log2 fold change<-2) are colored green dots. (C-D) The heatmap indicates the association between the differential MetaCyc metabolic pathways and microbiota in cancerous tissue(C) and para-cancerous tissue(D).
Figure 6

Flowcharts of inclusion criteria of participants and process of sequencing data. (A) Inclusion procedures for participants with ESCC. (B) Analysis workflow for sequencing data.

Supplementary Files

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

- FigureS1.pdf
- TableS1.docx
- TableS2.docx
- TableS3.xlsx
- supfile.docx