

RESEARCH ARTICLE

# Oral Microbiota and Risk for Esophageal Squamous Cell Carcinoma in a High-Risk Area of China

Xingdong Chen<sup>1,2,3</sup>, Björn Winckler<sup>1</sup>, Ming Lu<sup>3,4\*</sup>, Hongwei Cheng<sup>5</sup>, Ziyu Yuan<sup>2,3</sup>, Yajun Yang<sup>2,3</sup>, Li Jin<sup>2,3</sup>, Weimin Ye<sup>1,3</sup>

**1** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **2** Ministry of Education Key Laboratory of Contemporary Anthropology and State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China, **3** Fudan-Taizhou Institute of Health Sciences, Taizhou, China, **4** Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, China, **5** Taixing People's Hospital, Taixing, China

\* [lvming@sdu.edu.cn](mailto:lvming@sdu.edu.cn)



**OPEN ACCESS**

**Citation:** Chen X, Winckler B, Lu M, Cheng H, Yuan Z, Yang Y, et al. (2015) Oral Microbiota and Risk for Esophageal Squamous Cell Carcinoma in a High-Risk Area of China. PLoS ONE 10(12): e0143603. doi:10.1371/journal.pone.0143603

**Editor:** Qing-Yi Wei, Duke Cancer Institute, UNITED STATES

**Received:** September 4, 2015

**Accepted:** November 6, 2015

**Published:** December 7, 2015

**Copyright:** © 2015 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by a grant from Natural Science Foundation of China (81072358, 81273151), and a grant from Science and Technology Committee of Shandong Province (2012GSF11857). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Poor oral health has been linked with an increased risk of esophageal squamous cell carcinoma (ESCC). We investigated whether alteration of oral microbiota is associated with ESCC risk. Fasting saliva samples were collected from 87 incident and histopathologically diagnosed ESCC cases, 63 subjects with dysplasia and 85 healthy controls. All subjects were also interviewed with a questionnaire. V3–V4 region of 16S rRNA was amplified and sequenced by 454-pyrosequencing platform. Carriage of each genus was compared by means of multivariate-adjusted odds ratios derived from logistic regression model. Relative abundance was compared using Metastats method. Beta diversity was estimated using Unifrac and weighted Unifrac distances. Principal coordinate analysis (PCoA) was applied to ordinate dissimilarity matrices. Multinomial logistic regression was used to compare the coordinates between different groups. ESCC subjects had an overall decreased microbial diversity compared to control and dysplasia subjects ( $P < 0.001$ ). Decreased carriage of genera *Lautropia*, *Bulleidia*, *Catonella*, *Corynebacterium*, *Moryella*, *Peptococcus* and *Cardiobacterium* were found in ESCC subjects compared to non-ESCC subjects. Multinomial logistic regression analyses on PCoA coordinates also revealed that ESCC subjects had significantly different levels for several coordinates compared to non-ESCC subjects. In conclusion, we observed a correlation between altered salivary bacterial microbiota and ESCC risk. The results of our study on the saliva microbiome are of particular interest as it reflects the shift in microbial communities. Further studies are warranted to verify this finding, and if being verified, to explore the underlying mechanisms.

## Introduction

The positive association between alcohol use, tobacco smoking and the risk of esophageal squamous cell carcinoma (ESCC) has been well established, especially in Western countries. However, in areas with high incidence of ESCC, including the so-called “Asian esophageal cancer belt”, the major factors contributing to ESCC are yet to be established.[1] Recently, an association between indicators of poor oral hygiene and ESCC has been reported in studies from several high-risk areas of China,[2] India,[3] Iran[4], and from other areas including Latin America[5] and Japan.[6] Furthermore, poor oral health was reported as a risk factor for the precursor lesion of ESCC, *i.e.* esophageal squamous dysplasia,[7] and it may act synergistically in increasing the risk of ESCC with other risk factor (*e.g.* gastric atrophy).[8] There is reason to assume that poor oral health and hygiene are critical risk factors for ESCC in high-risk areas.

The underlying mechanism for the associations between oral health status and ESCC risk is not completely understood. It is well established that the oral microbiome plays a critical role in the maintenance of a normal oral physiological environment and in development of oral diseases, including periodontal diseases and tooth loss. Although little studied, the oral microbiome may be important in cancer and other chronic diseases, through direct metabolism of chemical carcinogens (*e.g.* nitrite, ethanol) [9, 10] and through systemic inflammatory effects [11]. We assumed that a stronger underlying association of ESCC risk with oral microbiome profiles would exist. Although some specific bacterial species in tissue and saliva have been linked to an elevated risk of ESCC by targeted approach,[12, 13] to date few studies have systematically investigated the relation between oral microbiota and ESCC risk. In current study, we aim to investigate the potential association between oral microbiota in saliva and ESCC risk using 16S rRNA amplicon sequencing approach, based on a large case-control study conducted in Taixing, an area with a high incidence of ESCC.

## Materials and Methods

### Study base

A case-control study on esophageal cancer was conducted during October of 2010 and March of 2012 in Taixing of Jiangsu Province, China. Briefly, cases were recruited mainly from endoscopy units at the four largest hospitals of Taixing (the People's Hospital of Taixing, the Second People's Hospital of Taixing, the Third People's Hospital of Taixing and the Hospital of Traditional Chinese Medicine of Taixing). More than 90% of the patients in this area are referred to these hospitals. Subjects who were suspected to have esophageal cancer under endoscopy were asked to participate in the study. Case recruitment was also supplemented by additional linkage to the local Cancer Registry, and sample collection of the supplementary cases was conducted at the end of the same year. Control subjects, frequency matched to the cases of ESCC on sex and age in 5-year groups and randomly selected from the Taixing population register, were enrolled into the study during the same period with cases. All subjects in the study were restricted to local inhabitants who have lived in Taixing for at least 5 years prior to diagnosis date for cases or interview date for controls.

The current study is a sub-project of the case-control study focusing on the relation between oral microbiota and ESCC risk. Cases were those who were recruited from endoscopy room during the period from October of 2010 to August of 2011 ( $N = 331$ ), and controls were those who were recruited during June of 2011 and August of 2011 ( $N = 400$ ). In order to avoid possible confounding which might affect diversity of oral microbiota, *i.e.* ambient temperatures and dietary habits in different seasons, cases collected during November of 2010 and March of

2011 were excluded (N = 124). Cases without histopathological confirmation, complete questionnaire or saliva sample were also excluded (N = 36). The study base of the current study thus included 171 ESCC cases and 400 controls. Since saliva collection was performed after endoscopy for ESCC cases, while for control subjects it was performed only after fasting overnight, we could not exclude the possibility of contamination by endoscopy which would affect the diversity of oral microbiota. Therefore, we included 80 subjects with a suspected diagnosis of esophageal dysplasia who also underwent endoscopy during the whole study period as another “control” group in the current study.

### Data collection by interview

All subjects underwent face-to-face interviews by trained interviewers using a standardized questionnaire. The questionnaire covered detailed information on age, sex, education, smoking, alcohol drinking, family history of ESCC and other potential confounders of interest. Dietary habits 10 years before interview were collected using a food frequency questionnaire specifically designed for this population.[14] The trained personnel counted each subject's number of teeth, recorded the number of missing and filled teeth (the sum of which was the MFT score) [4] and oral hygiene habits (times of tooth brushing per day).

### Saliva DNA extraction and subject selection

About 2~3 mL of saliva was collected from each participant after overnight fasting. Saliva collection was before antitumor treatment of cases, and for both cases and controls, no prescription of antibiotics one month before interview was reported. Saliva sample was mixed with 3mL lysis buffer (50mM Tris, pH 8.0, 50mM EDTA, 50mM sucrose, 100mM NaCl, and 1% sodium dodecyl sulfate). The mixture was delivered to the laboratory same day of collection and stored in the -20°C freezers. A modified high-salt DNA extraction method was used to extract DNA from saliva samples. Thirty microliters of proteinase K (20mg/mL, Sigma) and 150uL of 10% sodium dodecyl sulfate were added to 2mL of the mixture, which was then incubated overnight at 53°C in a shaking water bath. After addition of 400uL of 5M NaCl and incubation for 10 min on ice, the mixture was distributed equally into 2-mL centrifuge tubes and centrifuged for 10min at 13,000 rpm in an Eppendorf 5415D centrifuge. The supernatant from each tube was transferred to a new tube to which 800uL of isopropanol was added; the tubes were then incubated 10 min at room temperature and centrifuged for 15min at 13,000 rpm. The supernatants were discarded, and the sediments were washed twice with 500uL of 70% ethanol; then the sediments were dried and dissolved in 30uL of double-distilled water. DNA concentration of each sample was measured by the NanoDrop spectrophotometer.

We first selected study subjects according to the DNA quality standards (DNA concentration:  $\geq 20$ ng/uL; A260/280: 1.8~2.0; total amount:  $\geq 400$ ng) set by the BGI Company (Shenzhen, China) which conducted sequencing for current study. Eventually, 100 of 171 ESCC cases, 70 of 80 dysplasia controls and 312 of 400 controls met the standards. We thus enrolled 100 ESCC cases and 70 dysplasia control subjects, and for healthy controls we selected 100 controls frequency matched to the ESCC cases by sex and age in 5-year groups. Meanwhile, pathological sections were re-reviewed by an experienced pathologist, and one case from ESCC group was re-diagnosed as esophageal adenocarcinoma (excluded), three subjects from the dysplasia group were re-diagnosed as ESCC (regrouped into ESCC). Finally, 102 ESCC cases (ESCC group), 67 dysplasia control subjects (Dysplasia group) and 100 healthy controls (control group) were included in the current study.

## Sequencing, data processing and statistical analysis

16S ribosomal RNA (16S rRNA) amplicons covering hypervariable regions V3 to V4 were generated using primers (341\_F- CCTACGGGNGGCWGCAG and 805\_R- CTACCRGGGTATC TAATCC) incorporating Roche 454 FLX Titanium adapters (Branford, CT) and sample barcode sequences.[15] Amplicons were sequenced using single-read sequencing method following the manufacturer's specifications on the 454 Roche FLX Titanium pyrosequencing platform. Laboratory personnel were blinded to the case-control status. All the procedures except DNA extraction were conducted by the BGI Company.

Amplicon reads with mismatches in either primer or barcode were discarded and the remaining reads were stripped of barcode and primers. The *fastq\_filter* command of USEARCH 7.0.1001[16] was used to discard reads with more than one expected error as well as to truncate reads to a length of 300 nucleotides. Shorter reads were discarded. The quality-filtered reads were abundance sorted and clustered into operational taxonomic units (OTUs) using the USEARCH *cluster\_otus* command with 97% sequence identity. Singleton reads were ignored in the *cluster\_otus* command to avoid spurious OTUs. Chimera removal was performed as part of the OTU clustering step and by using the USEARCH *uchime\_ref* command against the "Gold" ChimeraSlayer reference database (r20110519).[17] Abundance tables were created by aligning the quality-filtered reads against the database of OTUs with the *usearch\_global* command.

QIIME 1.7.0[18] was used to assign taxonomy and to generate a phylogenetic tree after aligning the reads and filtering alignments. The scripts used in this step were: *assign\_taxonomy\_rdp.py* to assign taxonomy against the Greengenes database (v12\_10)[19] with the RDP classifier (v2.2), *make\_phylogeny.py* to build a phylogenetic tree using FastTree (v2.1.3),[20] *align\_seqs\_pynast.py* to align with PyNAST[21] against the Greengenes core reference alignment, and *filter\_alignment.py* to filter the PyNAST alignment.

Data analysis and visualization was performed using R (v3.0.1) and the package phyloseq (v1.4.5).[22] Samples with less than 1000 depth were discarded before analysis to ensure that sufficient biological diversity was captured. Alpha diversity and UniFrac[23] analyses were done after subsampling to even depth to reduce bias due to the dependence of these measures on sampling depth. For analyses on phylum and genus level the following steps were taken: 1) Greengenes suggested taxa assignments were heeded (*e.g.* [Prevotella] was treated as *Prevotella*), 2) spurious taxa with mean abundance under 0.01% were removed, 3) each sample was normalized to relative abundance by dividing by total abundance, and finally 4) unclassified taxa at the given rank were removed from further analysis.

Carriage (presence or absence; *i.e.* prevalence) of each genus was compared in three groups, and odds ratios (ORs) were calculated for genus, based on unconditional logistic regression modelling, adjusting for age, sex, education, smoking, alcohol drinking, family history of ESCC, MFT score (the number of missing and filled teeth), times of tooth brushing per day, daily consumption of pickled vegetables and daily consumption of fresh fruits. Relative abundance of each genus was compared using the Metastats package.[24] False discovery rate (FDR) adjustment was used to correct for multiple comparisons. Principal coordinate analysis (PCoA) was applied to ordinate dissimilarity matrices. A multinomial logistic regression model was used to compare the first 10 coordinates from PCoA among groups of study subjects.

## Ethical considerations

The study was approved by the Institutional Review Board of School of Life Sciences, Fudan University and the Institutional Review Board of Qilu Hospital, Shandong University ().

Written informed consent was obtained from all participants before interview and sample collection.

## Results

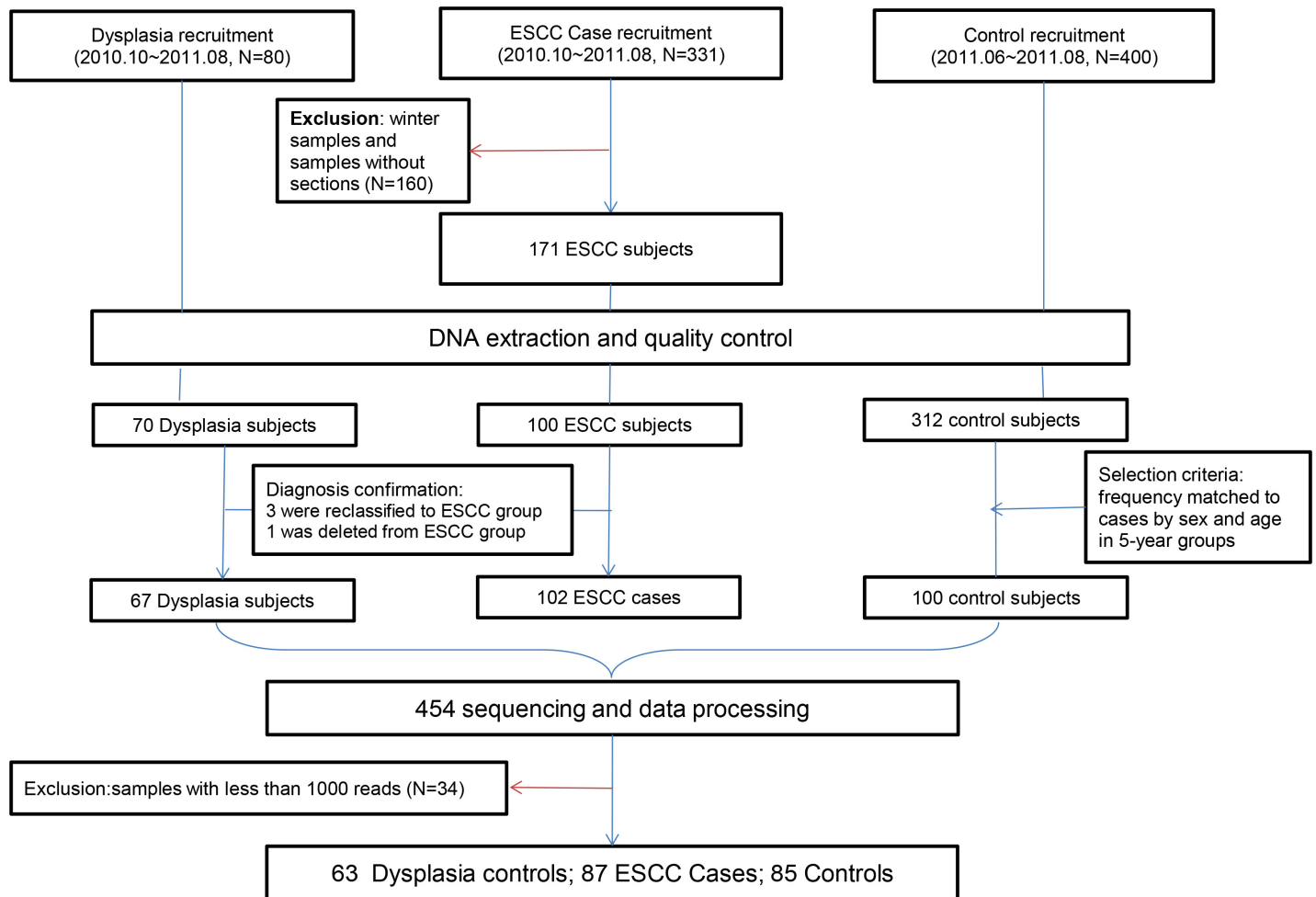
Multiplexed, barcoded sequencing data were deconvoluted and a total of 1.7M amplicon reads was obtained. Thirty-four samples had less than 1000 reads and were excluded (235 samples were left. [Fig 1](#)). Approximately 52% of all reads were discarded due to insufficient quality or read length less than 300 bp (384K short reads and 471K low quality reads), leaving 800K reads with 3402 average good quality reads per sample. The final data contained 32,192 unique reads and clustered into 489 OTUs.

Finally, 87 patients with ESCC (ESCC group, 59 males and 28 females), 63 patients with dysplasia (Dysplasia group, 41 males and 22 females), and 85 control subjects (Control group, 62 males and 23 females) remained for further analysis. The clinical parameters including age, sex, education, smoking status, drinking status, MFT score, times of tooth brushing per day, family history of ESCC and daily consumption of pickled vegetables and fruits are shown in [Table 1](#). Times of tooth brushing per day and daily consumption of pickled vegetables were significantly different among three groups ( $P < 0.05$ ); ESCC patients consumed more pickled vegetables and brushed teeth less often compared to Dysplasia and control subjects.

The sequencing reads were assigned to 437 OTUs in the ESCC group, 446 OTUs in the Dysplasia group, and 471 OTUs in the Control group. To evaluate the diversity and richness of bacterial types in the samples, Chao1 and Shannon indices were calculated. Observed mean values of Chao1 and Shannon indices were 120.8 and 3.4 for the ESCC group, 129.1 and 3.6 for the Dysplasia group, and 147.2 and 3.7 for the Control group, respectively. Tests for difference in OTU diversity and richness, measured by both mean Chao1 and Shannon indices showed significant differences for ESCC vs Control ( $P < 0.001$ ) and ESCC vs Dysplasia ( $P < 0.01$ ) ([Table 2](#)). At low depth, indices of Chao1, Shannon and mean numbers of OTUs increased sharply in all groups, however, the curves leveled off gradually with the increasing sequencing depth. The differences were always significant among three study groups, even at lower depth ([S1 Fig](#)).

For the overall bacterial community, OTUs were assigned to 12 Phyla and 44 Genera based on the Greengenes database (removing the rare taxa of mean abundance  $< 0.01\%$ ). Phyla composition was consistent across three groups. The five most abundant phyla were the following: *Bacteroidetes* (49.5%, 42.6% and 38.9% in the ESCC, Dysplasia and Control groups, respectively), *Firmicutes* (34.9%, 30.3% and 35.7%), *Proteobacteria* (5.9%, 15.1% and 12.7%), *Fusobacteria* (4.1%, 6.0% and 4.6%) and *Actinobacteria* (1.8%, 2.6% and 3.1%). [Fig 2](#) summarizes the most abundant phyla, representing 80~95% of the bacteria in each sample among the three groups.

In a genus-based analysis, we compared ESCC, dysplasia and control subjects for presence and relative abundance of each specific genus ([Table 3](#)). Compared to healthy controls, decreased carriage of several genera in ESCC subjects were found, including *Megasphaera*, *Aggregatibacter*, *Atopobium*, *Lautropia*, *Actinobacillus*, *Bulleidia*, *Catonella*, *Filifactor*, *Corynebacterium*, *TG5*, *Acholeplasma*, *Moryella*, *Butyrivibrio*, *Dialister*, *Peptococcus*, and *Cardiobacterium*. These differences remained for genera, *Lautropia*, *Bulleidia*, *Catonella*, *Corynebacterium*, *Moryella*, *Peptococcus* and *Cardiobacterium*, when comparing ESCC vs. dysplasia subjects. Based on relative abundance, the most predominant genera ( $> 5\%$  in the healthy control group) were *Prevotella* (42.4%, 38.4% and 36.1% in the ESCC, Dysplasia and Control groups, respectively), *Streptococcus* (21.9%, 14.5% and 16.1%), *Veillonella* (7.3%, 10.5% and 13.2%), and *Porphyromonas* (8.9%, 7.0% and 6.5%). Testing for differences in



**Fig 1. Flowchart of sample selection.**

doi:10.1371/journal.pone.0143603.g001

genera abundance revealed an increased level of colonization of members of *Prevotella* ( $P_{\text{FDR-adjusted}} < 0.001$ ), *Streptococcus* ( $P_{\text{FDR-adjusted}} < 0.001$ ) and *Porphyromonas* ( $P_{\text{FDR-adjusted}} < 0.001$ ), and a decreased level of most other genera in the ESCC group compared to non-ESCC groups, although for those genera with a low relative abundance the patterns were not clear.

We compared the overall bacterial community composition using weighted and unweighted UniFrac distance matrices, and applied PCoA to ordinate the matrix. The first 10 coordinates explained 53% and 85% of the variance for Unifrac and weighted Unifrac distances, respectively (S2 Fig). Correlations between coordinates and genera were calculated, and a multinomial logistic regression model was applied to compare the first 10 coordinates in three groups. For Unifrac distance, except for coordinates 7–9, all other coordinates showed significant differences between ESCC and healthy control group, while only coordinates 3, 4, 5 and 8 were significant when comparing ESCC with the Dysplasia group. Results were similar after adjustment for age, sex, education, smoking, alcohol drinking, family history of ESCC, MFT, times of tooth brushing per day, daily consumption of pickled vegetables and daily consumption of fresh fruits (Table 4). For weighted Unifrac distances, highly significant differences (coordinates 2 and 3) emerged when comparing ESCC with both healthy control and Dysplasia groups (Table 4). We further visualized the Unifrac or weighted Unifrac distances using the two most



**Table 1. Basic characteristics of subjects in esophageal squamous cell carcinoma (ESCC), dysplasia control and healthy control (Control) groups.**

	ESCC (%) (N = 87)	Dysplasia (%) (N = 63)	Control (%) (N = 85)	P *
Age (years, mean ± SD)	64.8±8.0	65.5±7.6	66.0±7.3	0.56
Sex				
Men	59(67.8)	41(65.1)	62(72.9)	
Women	28(32.2)	22(34.9)	23(27.1)	0.57
Education				
Illiteracy	26(29.9)	19(30.6)	19(22.4)	
1~5 years	39(44.8)	28(45.2)	42(49.4)	
≥5 years	22(25.3)	15(24.2)	24(28.2)	0.79
Smoking				
Never	36(41.4)	22(36.1)	26(30.6)	
Ever	51(58.6)	39(63.9)	59(69.4)	0.34
Alcohol drinking				
Never	37(42.5)	28(45.9)	51(60.0)	
Ever	50(57.5)	33(54.1)	34(40.0)	0.06
Family history of ESCC				
No	56(64.4)	45(71.4)	63(74.1)	
Yes	31(35.6)	18(28.6)	22(25.9)	0.36
MFT†				
None	11(12.9)	15(24.6)	17(20.0)	
1~4	27(31.8)	12(19.7)	25(29.4)	
≥4	47(55.3)	34(55.7)	43(50.6)	0.29
Times of tooth brushing per day				
<2	73(84.9)	50(82.0)	53(63.9)	
≥2	13(15.1)	11(18.0)	30(36.1)	<0.01
Daily consumption of pickled vegetables				
<10 g	18(22.2)	14(26.4)	37(43.5)	
≥10 g	63(77.8)	39(73.6)	48(56.5)	0.01
Daily consumption of fresh fruits				
<25 g	41(52.6)	32(57.1)	35(42.7)	
≥25 g	37(47.4)	24(42.9)	47(57.3)	0.21

\* P values were based on Wilcoxon rank-sum test for continuous variables, and chi-squared test for categorical variables (two-sided).

† MFT referred to sum of missing and filled teeth.

doi:10.1371/journal.pone.0143603.t001

significant coordinates (the ones which explained most variances and with small P values) found in multinomial logistic regression analyses (coordinates 1 and 3 for Unifrac distance; coordinates 2 and 3 for weighted Unifrac distance). For both Unifrac and weighted Unifrac distances, ESCC and healthy control subjects tended to cluster in opposite directions, while dysplasia subjects were located between the two groups (Fig 3).

## Discussion

Increasing evidence indicates a key role for the bacterial microbiota in carcinogenesis.[25] Our study was based on one of the largest sets of 16S rRNA gene sequences from the human oral cavity to evaluate the association between oral microbiota and ESCC risk. We found that ESCC subjects had decreased overall microbial diversity compared to dysplasia and healthy control

**Table 2. Measurements and comparisons of alpha diversity in esophageal squamous cell carcinoma (ESCC), dysplasia control and healthy control (Control) groups.**

Group	OTUs	Mean Shannon	Mean Chao1
ESCC	437	3.4	120.8
Dysplasia	446	3.6	129.1
Control	471	3.7	147.2
<i>P</i> value*			
ESCC vs. Control		<0.0001	<0.0001
ESCC vs. Dysplasia		0.0005	0.0869
Dysplasia vs. Control		0.0813	0.0007

\* *P* values of pairwise comparisons were based on Wilcoxon rank-sum test. *P* value adjustment method was Holm.

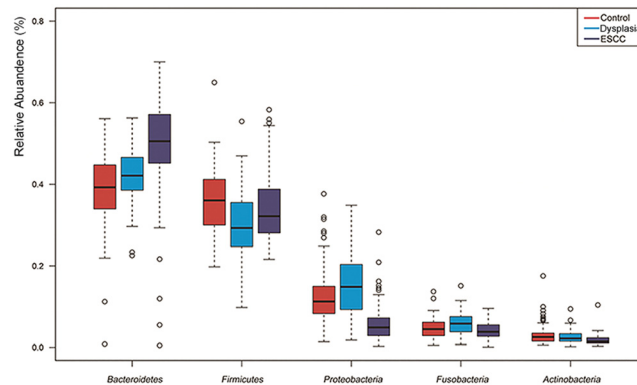
doi:10.1371/journal.pone.0143603.t002

subjects. In addition, decreased carriage of several genera, including *Lautropia*, *Bulleidia*, *Catonella*, *Corynebacterium*, *Moryella*, *Peptococcus* and *Cardiobacterium*, were found in the ESCC cases compared to non-ESCC subjects. Higher relative abundance of *Prevotella*, *Streptococcus* and *Porphyromonas*, but lower or similar relative abundance of most other genera were observed in ESCC group compared to non-ESCC groups. Multinomial logistic regression analyses on PCoA coordinates of Unifrac or weighted Unifrac distances also revealed that ESCC subjects had significantly different levels of several coordinates compared to non-ESCC subjects.

The current study is a sub-study of a case-control study on upper gastrointestinal cancers. For this study, every effort was made to enroll all of the incident ESCC cases in the study area. The frequency matched healthy controls were randomly selected from the general population. The participation rates for both cases and controls were more than 75%. Since we tried to enroll cases before histopathological diagnosis being made, we were also able to include another control group, *i.e.* dysplasia patients. The similar directions of associations when using different control groups strengthened the validity of our findings. In addition, histopathological diagnosis of ESCC and dysplasia by a single pathologist, saliva collection after overnight fasting for both cases and controls, and collection of extensive information on potential confounders (*e.g.* smoking, alcohol drinking, and other lifestyle factors) were among the strengths of the current study.

Our study also has several limitations. One of the main limitations was that we did not add beads during the process of saliva DNA extraction which might affect the composition and diversity of the oral microbiota. Nevertheless, since the same procedure of saliva DNA preparation was applied in all three groups, the observed differences between groups were still valid, although we could not draw conclusion on those hard-to-break bacteria. As a large fraction of samples did not pass the quality control and were excluded, while there might be underlying factors being masked in these subjects, we compared included subjects and excluded subjects for the clinical parameters including age, sex, education, smoking status, drinking status, MFT score, times of tooth brushing per day, family history of ESCC and daily consumption of pickled vegetables and fruits, and no significant differences were found. In addition, saliva samples for ESCC cases were collected after endoscopy, which was different from healthy controls. This might raise a concern that the differences of oral microbiota between ESCC cases and healthy controls might be due to contamination during endoscopy. In order to determine whether endoscopy has led (or not) to the changes in the microbiome found in the ESCC population, a small cohort including 30 individuals were enrolled in one of the study hospitals and their





**Fig 2. Abundances of microbiota in esophageal squamous cell carcinoma (ESCC), dysplasia control and healthy control (Control) groups at phylum level.**

doi:10.1371/journal.pone.0143603.g002

saliva samples were obtained both before and after endoscopy. The overall bacterial community composition was analyzed using the same pipeline. No significant difference in bacterial community composition was found between saliva samples collected before and after endoscopy (S3 Fig). Moreover, for subjects in another control group, *i.e.* dysplasia subjects, saliva samples were also collected after endoscopy. The similar directions of associations when comparing to different control groups somewhat allayed such a concern. Another limitation is that comparison of the composition and diversity of oral microbiota might be biased by the non-consistent sampling seasons among three groups, even if samples collected in winter in ESCC group were excluded due to the difference of temperature and dietary pattern. However, we compared the composition and diversity between different sampling seasons, but did not find any significant difference between different sampling seasons (data not shown). Finally, due to case-control study design, our results could not distinguish whether decreased microbial richness causes ESCC or is an effect of the cancer status, *e.g.* the oral microbiota may be modified by the confounding effect of restricted food intake due to symptoms from extensive lesions and/or dry mouth in the cancer group. Currently, we are conducting a prospective cohort study, Taizhou Longitudinal Study, [26] in which saliva samples were collected in baseline survey and can be used to prospectively assess the relationship between oral microbiota and the development of ESCC and other gastrointestinal cancers.

In the present study, we found that ESCC subjects had low salivary microbial diversity compared to healthy controls and dysplasia subjects. Similarly, lower bacterial diversity was observed in some other habitats such as the stomach with gastritis [27] and the intestines with colorectal cancer [28]. Most recently, a cross-sectional study in China showed that a decreased microbial richness in the upper digestive tract was associated with cancer-predisposing conditions of the stomach and esophagus (*i.e.* low serum pepsinogen I/II ratio and esophageal squamous dysplasia). [29]

Overall, the abundant bacterial groups found in our study are similar to those found in most other studies. The most common phyla in our samples were *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Fusobacteria* and *Actinobacteria*. However, our data suggest that the most abundant phylum and genus were *Bacteroidetes* and *Prevotella*, and this might be a little different from other studies which showed *Firmicutes* and *Streptococci* were dominant in oral microbiota. [30] The shift may be due to the different DNA extraction methods and different broad-range PCR primers applied. An alternative explanation for this inconsistency may be the impact of diet and oral hygiene. It has been reported that the composition of oral microbiota is

**Table 3. Prevalence and relative abundance of genera in esophageal squamous cell carcinoma (ESCC), dysplasia control (Dysplasia) and healthy control (Control) groups.**

Genus	% Positive (carriage) *					Abundances % †				
	ESCC	Dysplasia	Control	OR (95%CI) ‡		ESCC	Dysplasia	Control	P §	
				ESCC vs. Control	ESCC vs. Dysplasia				ESCC vs. Control	ESCC vs. Dysplasia
<i>Prevotella</i>	100	100	100	NA	NA	42.4	38.4	36.1	<0.01	0.04
<i>Streptococcus</i>	100	100	100	NA	NA	21.9	14.5	16.1	<0.01	<0.01
<i>Veillonella</i>	100	100	100	NA	NA	7.3	10.5	13.2	<0.01	<0.01
<i>Porphyromonas</i>	98.9	100	100	NA	NA	8.9	7.0	6.5	<0.01	0.02
<i>Neisseria</i>	95.4	98.4	100	NA	NA	2.0	5.1	3.7	<0.01	<0.01
<i>Fusobacterium</i>	98.9	96.8	100	NA	NA	2.7	3.8	3.1	0.13	<0.01
<i>Haemophilus</i>	96.6	98.4	100	NA	NA	1.6	3.5	2.7	<0.01	<0.01
<i>Rothia</i>	97.7	100	100	NA	NA	1.3	1.9	2.4	<0.01	0.01
<i>Leptotrichia</i>	100	100	100	NA	NA	1.5	2.6	1.9	0.05	<0.01
<i>Capnocytophaga</i>	98.9	100	98.8	NA	NA	2.3	2.5	1.5	<0.01	0.70
<i>Treponema</i>	92.0	87.3	97.7	0.30 (0.05~1.75)	1.75(0.42~7.32)	0.8	1.0	1.5	<0.01	0.45
<i>Megasphaera</i>	82.8	88.9	97.7	0.09 (0.02~0.52)¶	0.64(0.21~1.95)	0.4	0.5	1.4	<0.01	0.21
<i>Campylobacter</i>	97.7	100	98.8	0.35 (0.02~5.25)	NA	0.6	1.2	1.3	<0.01	<0.01
<i>Selenomonas</i>	95.4	93.7	100	NA	1.03(0.21~5.01)	0.8	0.7	1.1	0.01	0.70
<i>Aggregatibacter</i>	75.9	82.5	94.1	0.18 (0.06~0.57)¶	0.61(0.24~1.58)	0.5	1.1	1.0	<0.01	<0.01
<i>Peptostreptococcus</i>	92.0	98.4	98.8	0.17 (0.02~1.78)	0.43(0.04~4.24)	0.6	0.7	0.9	0.01	0.35
<i>Oribacterium</i>	96.6	98.4	98.8	0.45 (0.04~5.45)	0.48(0.04~6.35)	0.5	0.8	0.7	<0.01	<0.01
<i>Actinomyces</i>	92.0	98.4	96.5	0.38 (0.08~1.86)	0.23(0.02~2.24)	0.5	0.7	0.7	0.03	0.11
<i>Moraxella</i>	24.1	15.9	21.2	1.48 (0.62~3.56)	1.39(0.56~3.47)	0.3	0.1	0.6	0.49	0.44
<i>Atopobium</i>	79.3	93.7	90.6	0.27 (0.09~0.80)¶	0.38(0.11~1.28)	0.2	0.4	0.4	<0.01	<0.01
<i>Lautropia</i>	35.6	63.5	84.7	0.07 (0.03~0.17)¶	0.17 (0.07~0.41)¶	0.0	0.1	0.4	<0.01	<0.01
<i>Actinobacillus</i>	48.3	60.3	71.8	0.42 (0.19~0.88)¶	0.54(0.25~1.17)	0.2	0.3	0.4	0.28	0.35
<i>Bulleidia</i>	39.1	88.9	96.5	0.02 (0.01~0.11)¶	0.10 (0.03~0.27)¶	0.1	0.4	0.2	<0.01	<0.01
<i>Parascardovia</i>	73.6	82.5	85.9	0.42 (0.17~1.06)	0.62(0.24~1.60)	0.2	0.2	0.2	0.08	0.44
<i>Catonella</i>	63.2	92.1	94.1	0.14 (0.05~0.43)¶	0.16 (0.05~0.53)¶	0.1	0.2	0.2	<0.01	<0.01
<i>Filifactor</i>	50.6	61.9	80.0	0.22 (0.10~0.50)¶	0.59(0.27~1.31)	0.1	0.1	0.2	<0.01	0.01
<i>Tannerella</i>	81.6	81.0	82.4	0.71 (0.28~1.83)	0.96(0.35~2.61)	0.2	0.2	0.2	0.86	0.96
<i>Corynebacterium</i>	32.2	61.9	78.8	0.11 (0.05~0.25)¶	0.33 (0.15~0.74)¶	0.0	0.2	0.2	<0.01	<0.01
TG5	49.4	54.0	72.9	0.33 (0.15~0.72)¶	0.95(0.43~2.10)	0.1	0.1	0.2	<0.01	0.46

(Continued)

Table 3. (Continued)

Genus	% Positive (carriage) *					Abundances % †				
	ESCC	Dysplasia	Control	OR (95%CI) ‡		ESCC	Dysplasia	Control	P §	
				ESCC vs. Control	ESCC vs. Dysplasia				ESCC vs. Control	ESCC vs. Dysplasia
<i>Mycoplasma</i>	70.1	58.7	55.3	2.37 (1.10~5.13)	1.96(0.87~4.40)	0.2	0.2	0.1	0.68	0.96
<i>Elizabethkingia</i>	78.2	85.7	84.7	0.66 (0.27~1.60)	0.54(0.19~1.52)	0.2	0.3	0.1	<0.01	0.96
<i>Acholeplasma</i>	12.6	23.8	31.8	0.30 (0.12~0.77)¶	0.40(0.15~1.08)	0.0	0.0	0.1	0.04	0.77
<i>Moryella</i>	40.2	69.8	71.8	0.18 (0.08~0.41)¶	0.18 (0.08~0.43)¶	0.1	0.1	0.1	0.18	0.06
<i>Butyrivibrio</i>	51.7	47.6	72.9	0.27 (0.12~0.63)¶	1.18(0.55~2.52)	0.1	0.1	0.1	0.02	0.96
<i>Dialister</i>	25.3	34.9	70.6	0.10 (0.04~0.25)¶	0.54(0.23~1.29)	0.0	0.0	0.1	<0.01	0.10
<i>Streptobacillus</i>	42.5	30.2	32.9	1.03 (0.48~2.22)	1.71(0.75~3.87)	0.1	0.1	0.1	0.34	0.76
<i>Peptococcus</i>	28.7	50.8	68.2	0.20 (0.09~0.44)¶	0.32 (0.14~0.71)¶	0.0	0.1	0.1	<0.01	<0.01
<i>Abiotrophia</i>	39.1	33.3	54.1	0.49 (0.24~1.02)	1.24(0.56~2.75)	0.1	0.0	0.1	0.22	0.05
<i>Hylemonella</i>	20.7	38.1	37.7	0.41 (0.18~0.96)	0.43(0.19~1.01)	0.0	0.0	0.1	0.20	0.54
<i>Lactobacillus</i>	31.0	27.0	21.2	1.77 (0.73~4.30)	0.93(0.39~2.19)	0.1	0.0	0.0	0.28	<0.01
<i>Bacteroides</i>	44.8	30.2	51.0	0.54 (0.25~1.15)	1.75(0.76~4.01)	1.1	0.1	0.0	0.03	0.09
<i>Cardiobacterium</i>	12.6	41.3	37.7	0.21 (0.08~0.52)¶	0.23 (0.09~0.62)¶	0.0	0.0	0.0	<0.01	<0.01
<i>Odoribacter</i>	18.4	31.8	18.8	1.13 (0.44~2.90)	0.69(0.28~1.70)	0.0	0.0	0.0	0.73	0.13
<i>Herbaspirillum</i>	1.6	34.9	1.2	NA	NA	0.0	0.1	0.0	0.62	<0.01

\* Percentage of ESCC and control subjects who carried the specific taxon. NA = not applicable.

† Median relative abundance of the specific taxon in people who carry the taxon.

‡ Odds ratios (ORs) and 95% confidence intervals (CIs) were derived from unconditional logistic regression models with non-carriers as the referent, adjusting for sex, age, education, smoking, alcohol drinking, family history of ESCC, MFT, times of tooth brushing per day, daily consumption of pickled vegetables and daily consumption of fresh fruits.

§ P values were derived from the Metastats package. P value adjustment method was FDR (false discovery rate).

¶ FDR-adjusted P values were P less than or equal to 0.05.

doi:10.1371/journal.pone.0143603.t003

different between rural and urban districts.[31] De Filippo C. *et al.*[32] investigated human intestinal microbiota from children characterized by a modern western diet and a rural diet, and found that children in rural Africa showed a significant enrichment in *Bacteroidetes* and depletion in *Firmicutes*. Individuals with excellent oral hygiene typically harbor a relatively simple flora dominated by gram-positive cocci and rods, mostly comprised of *Streptococci*, however in individuals who do not maintain good oral hygiene, the flora shifts to become more diverse and complex and is dominated by anaerobic gram-negative bacteria, including *Prevotella*. [33] In the present study, even the healthy controls had relatively poor oral hygiene (as indicated by high indices of MFT and few times of tooth brushing per day). All the above

**Table 4. Multinomial logistic regression analysis of the first ten coordinates based on Unifrac and weighted Unifrac distances.**

Coordinate	Correlation with genus (Top 5 genera)					ESCC vs. Control		ESCC vs. Dysplasia	
						P *	P †	P *	P †
<b>Unifrac distance</b>									
Coordinate 1	<i>Treponema</i>	<i>Filifactor</i>	TG5	<i>Acholeplasma</i>	<i>Aggregatibacter</i>	0.0016	0.0043	0.3315	0.2289
Coordinate 2	<i>Lactobacillus</i>	<i>Campylobacter</i>	<i>Mycoplasma</i>	<i>Bulleidia</i>	<i>Neisseria</i>	0.0150	0.0200	0.1313	0.1916
Coordinate 3	<i>Bulleidia</i>	<i>Megasphaera</i>	<i>Veillonella</i>	<i>Atopobium</i>	<i>Porphyromonas</i>	<0.0001	<0.0001	<0.0001	<0.0001
Coordinate 4	<i>Prevotella</i>	<i>Lautropia</i>	<i>Elizabethkingia</i>	<i>Haemophilus</i>	<i>Actinomyces</i>	0.0305	0.0384	0.0012	0.0030
Coordinate 5	<i>Selenomonas</i>	<i>Hylemonella</i>	<i>Mycoplasma</i>	<i>Lactobacillus</i>	<i>Filifactor</i>	0.0133	0.0351	0.0029	0.0044
Coordinate 6	<i>Capnocytophaga</i>	<i>Acholeplasma</i>	<i>Treponema</i>	<i>Leptotrichia</i>	<i>Filifactor</i>	0.0076	0.0022	0.2457	0.0609
Coordinate 7	<i>Selenomonas</i>	<i>Corynebacterium</i>	<i>Leptotrichia</i>	<i>Lactobacillus</i>	<i>Mycoplasma</i>	0.0677	0.0135	0.3450	0.3626
Coordinate 8	<i>Elizabethkingia</i>	TG5	<i>Streptobacillus</i>	<i>Corynebacterium</i>	<i>Mycoplasma</i>	0.1144	0.1451	0.0124	0.0121
Coordinate 9	TG5	<i>Corynebacterium</i>	<i>Mycoplasma</i>	<i>Catonella</i>	<i>Peptostreptococcus</i>	0.3489	0.7809	0.6814	0.6973
Coordinate 10	<i>Lautropia</i>	<i>Bulleidia</i>	TG5	<i>Butyrivibrio</i>	<i>Abiotrophia</i>	0.0037	0.0022	0.6490	0.9423
<b>Weighted Unifrac distance</b>									
Coordinate 1	<i>Prevotella</i>	<i>Streptococcus</i>	<i>Rothia</i>	<i>Peptostreptococcus</i>	<i>Haemophilus</i>	0.9674	0.9863	0.2661	0.2644
Coordinate 2	<i>Haemophilus</i>	<i>Neisseria</i>	<i>Prevotella</i>	<i>Streptococcus</i>	<i>Catonella</i>	<0.0001	<0.0001	<0.0001	<0.0001
Coordinate 3	<i>Veillonella</i>	<i>Porphyromonas</i>	<i>Megasphaera</i>	<i>Atopobium</i>	<i>Actinomyces</i>	<0.0001	<0.0001	<0.0001	<0.0001
Coordinate 4	<i>Fusobacterium</i>	<i>Tannerella</i>	<i>Treponema</i>	<i>Peptococcus</i>	<i>Leptotrichia</i>	0.8960	0.8940	0.0705	0.2545
Coordinate 5	<i>Streptococcus</i>	<i>Leptotrichia</i>	<i>Fusobacterium</i>	<i>Prevotella</i>	<i>Capnocytophaga</i>	0.0001	0.0003	0.0491	0.0700
Coordinate 6	<i>Leptotrichia</i>	<i>Treponema</i>	<i>Streptococcus</i>	<i>Campylobacter</i>	<i>Dialister</i>	0.9774	0.8119	0.0361	0.0930
Coordinate 7	TG5	<i>Filifactor</i>	<i>Treponema</i>	<i>Mycoplasma</i>	<i>Bacteroides</i>	0.0002	0.0001	0.0749	0.0396
Coordinate 8	<i>Capnocytophaga</i>	<i>Streptococcus</i>	<i>Filifactor</i>	<i>Treponema</i>	<i>Dialister</i>	0.0061	0.0160	0.8769	0.7731
Coordinate 9	<i>Capnocytophaga</i>	<i>Leptotrichia</i>	<i>Streptococcus</i>	<i>Rothia</i>	<i>Neisseria</i>	0.6133	0.4827	0.4699	0.5216
Coordinate 10	<i>Peptostreptococcus</i>	<i>Atopobium</i>	<i>Actinomyces</i>	<i>Moryella</i>	TG5	0.2250	0.1265	0.2563	0.1965

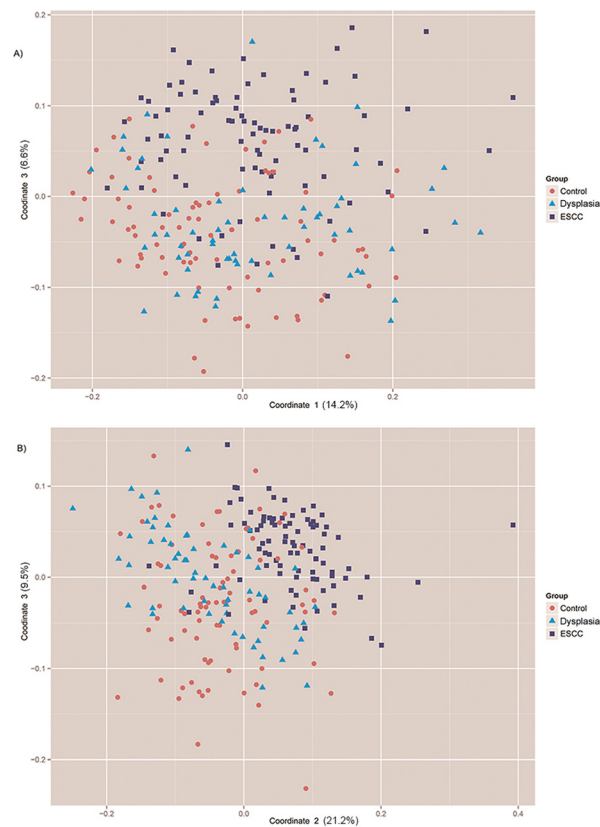
\* P values were derived from Wald chi-squared test, based on unadjusted regression models (10 coordinates were included simultaneously).

† P values were derived from Wald chi-squared test. The regression models included 10 coordinates, along with sex, age, education, smoking, alcohol drinking, family history of ESCC, MFT, times of tooth brushing per day, daily consumption of pickled vegetables and daily consumption of fresh fruits.

doi:10.1371/journal.pone.0143603.t004

reasons may contribute or partly explain the differences of the oral microbiota between different studies.

Our data show that decreased carriage of some genera, e.g. *Lautropia*, *Bulleidia*, *Catonella*, *Corynebacterium*, *Moryella*, *Peptococcus* and *Cardiobacterium*, are significantly associated with an increased risk of ESCC. However, due to low relative abundances of these genera (<0.5%), there is a danger that the observed differences were due to insufficient sampling depth. On the contrary, presence of genus *Mycoplasma*, which is unaffected by many common antibiotics and reported to be associated with several types of cancer, such as gastric cancer, [34] colon cancer [35] and prostate cancer, [36] was more common in ESCC cases than in healthy and Dysplasia control subjects, although the differences were not statistically significant after FDR adjustment. Higher relative abundance of *Prevotella* and *Streptococcus* were also observed in the ESCC group compared to non-ESCC groups. The proportion of these two genera accounted for nearly 65% of the overall community in ESCC subjects, which might to some extent explain the low diversity of oral microbiota in these patients. Although these genera seem to be non-pathogenic to the host, several studies have indicated that they might be associated with oral and upper digestive tract cancers. [37–39] Further studies are warranted to confirm these findings.



**Fig 3. Distribution of samples depicted by the two most significant coordinates for Unifrac (A) and weighted Unifrac distances (B).**

doi:10.1371/journal.pone.0143603.g003

Microbiota and host form a complex “super-organism” in which symbiotic relationships confer benefits to the host in many key aspects of life. A growing body of evidence implicates oral bacteria in the etiology of oral and gastrointestinal cancers.[29, 39] The oral microbiome may play an important role in cancer development, through direct metabolism of chemical carcinogens (e.g. activating alcohol and smoking-related carcinogens locally), and/or through systemic inflammatory effects.[30] Multi-disciplinary collaborations among various fields including epidemiology, microbiology, genetics, immunology, and bioinformatics will be needed to broaden our understanding of the relationship of oral microbiome and cancer risk, and to better understand of cancer etiology.

To summary, this is the first epidemiological study comparing the oral microbiota of ESCC and control subjects while controlling for potential confounders. We observed a correlation between altered salivary bacterial community structure and ESCC risk. The results of our study on the saliva microbiota are of particular interests given its modifiable nature. However, prospective and longitudinal cohort studies are required to verify this finding, along with functional studies, e.g. metagenomics and transcriptome studies. Establishment of the association between oral microbiome and ESCC risk may lead to significant advances in understanding of cancer etiology, potentially opening a new research paradigm for cancer prevention.

## Supporting Information

### S1 Fig. Rarefaction curves of indexes in three groups

(TIF)

### S2 Fig. Variations in distance matrix explained by the first 10 coordinates

(TIF)

### S3 Fig. Comparison of the overall bacterial community composition between groups of before and after endoscopy.

(TIF)

### S1 Checklist. STROBE Checklist.

(DOCX)

### S2 Checklist. Clinical Studies Checklist.

(DOCX)

## Acknowledgments

The authors would like to acknowledge the interviewers and technicians of Fudan-Taizhou Institute of Health Sciences for the data collection and sample preparation, the staff of Taixing Center for Disease Control and Prevention for the help in field work, and the staff of Taixing People's Hospital for the assistance of sample collection.

## Author Contributions

Conceived and designed the experiments: XC ML WY. Performed the experiments: XC ZY. Analyzed the data: XC BW WY. Contributed reagents/materials/analysis tools: ML YY LJ HC. Wrote the paper: XC BW WY ML LJ.

## References

1. Nyren O, Adami HO. Esophageal Cancer. In: Adami HO, David H, Dimitrios T, editors. *Textbook of cancer epidemiology*. New York: Oxford University Press; 2002. p. 137–61.
2. Abnet CC, Qiao YL, Mark SD, Dong ZW, Taylor PR, Dawsey SM. Prospective study of tooth loss and incident esophageal and gastric cancers in China. *Cancer Causes Control*. 2001; 12(9):847–54. PMID: [11714113](#)
3. Dar NA, Islami F, Bhat GA, Shah IA, Makhdoomi MA, Iqbal B, et al. Poor oral hygiene and risk of esophageal squamous cell carcinoma in Kashmir. *Br J Cancer*. 2013; 109(5):1367–72. doi: [10.1038/bjc.2013.437](#) PMID: [23900216](#)
4. Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan P, Aghcheli K, et al. Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2008; 17(11):3062–8. doi: [10.1158/1055-9965.EPI-08-0558](#) PMID: [18990747](#)
5. Guha N, Boffetta P, Wunsch FV, Eluf NJ, Shangina O, Zaridze D, et al. Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies. *Am J Epidemiol*. 2007; 166(10):1159–73. PMID: [17761691](#)
6. Sato F, Oze I, Kawakita D, Yamamoto N, Ito H, Hosono S, et al. Inverse association between toothbrushing and upper aerodigestive tract cancer risk in a Japanese population. *Head Neck*. 2011; 33(11):1628–37. doi: [10.1002/hed.21649](#) PMID: [21259377](#)
7. Sepehr A, Kamangar F, Fahimi S, Saidi F, Abnet CC, Dawsey SM. Poor oral health as a risk factor for esophageal squamous dysplasia in northeastern Iran. *Anticancer Res*. 2005; 25(1B):543–6. PMID: [15816626](#)
8. Nasrollahzadeh D, Malekzadeh R, Aghcheli K, Sotoudeh M, Merat S, Islami F, et al. Gastric atrophy and oesophageal squamous cell carcinoma: possible interaction with dental health and oral hygiene habit. *Br J Cancer*. 2012; 107(5):888–94. doi: [10.1038/bjc.2012.332](#) PMID: [22814581](#)



9. Nair J, Ohshima H, Nair UJ, Bartsch H. Endogenous formation of nitrosamines and oxidative DNA-damaging agents in tobacco users. *Crit Rev Toxicol*. 1996; 26(2):149–61. PMID: [8688158](#)
10. Salaspuro MP. Acetaldehyde, microbes, and cancer of the digestive tract. *Crit Rev Clin Lab Sci*. 2003; 40(2):183–208. PMID: [12755455](#)
11. Karin M, Lawrence T, Nizet V. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. *Cell*. 2006; 124(4):823–35. PMID: [16497591](#)
12. Morita E, Narikiyo M, Yano A, Nishimura E, Igaki H, Sasaki H, et al. Different frequencies of *Streptococcus anginosus* infection in oral cancer and esophageal cancer. *Cancer Sci*. 2003; 94(6):492–6. PMID: [12824872](#)
13. Narikiyo M, Tanabe C, Yamada Y, Igaki H, Tachimori Y, Kato H, et al. Frequent and preferential infection of *Treponema denticola*, *Streptococcus mitis*, and *Streptococcus anginosus* in esophageal cancers. *Cancer Sci*. 2004; 95(7):569–74. PMID: [15245592](#)
14. Zhuang M, Yuan Z, Lin L, Hu B, Wang X, Yang Y, et al. Reproducibility and relative validity of a food frequency questionnaire developed for adults in Taizhou, China. *PLoS One*. 2012; 7(11):e48341. doi: [10.1371/journal.pone.0048341](#) PMID: [23139777](#)
15. Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*. 2011; 5(10):1571–9. doi: [10.1038/ismej.2011.41](#) PMID: [21472016](#)
16. Edgar RC. UPPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013; 10(10):996–8. doi: [10.1038/nmeth.2604](#) PMID: [23955772](#)
17. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res*. 2011; 21(3):494–504. doi: [10.1101/gr.112730.110](#) PMID: [21212162](#)
18. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7(5):335–6. doi: [10.1038/nmeth.f.303](#) PMID: [20383131](#)
19. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006; 72(7):5069–72. PMID: [16820507](#)
20. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010; 5(3):e9490. doi: [10.1371/journal.pone.0009490](#) PMID: [20224823](#)
21. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 2010; 26(2):266–7. doi: [10.1093/bioinformatics/btp636](#) PMID: [19914921](#)
22. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013; 8(4):e61217. doi: [10.1371/journal.pone.0061217](#) PMID: [23630581](#)
23. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005; 71(12):8228–35. PMID: [16332807](#)
24. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS computational biology*. 2009; 5(4):e1000352. doi: [10.1371/journal.pcbi.1000352](#) PMID: [19360128](#); PubMed Central PMCID: PMC2661018.
25. Schwabe RF, Jobin C. The microbiome and cancer. *Nat Rev Cancer*. 2013; 13(11):800–12. doi: [10.1038/nrc3610](#) PMID: [24132111](#)
26. Wang X, Lu M, Qian J, Yang Y, Li S, Lu D, et al. Rationales, design and recruitment of the Taizhou Longitudinal Study. *BMC public health*. 2009; 9:223. doi: [10.1186/1471-2458-9-223](#) PMID: [19589173](#); PubMed Central PMCID: PMC2715397.
27. von Rosenvinge EC, Song Y, White JR, Maddox C, Blanchard T, Fricke WF. Immune status, antibiotic medication and pH are associated with changes in the stomach fluid microbiota. *ISME J*. 2013; 7(7):1354–66. doi: [10.1038/ismej.2013.33](#) WOS:000320852100011. PMID: [23466701](#)
28. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human Gut Microbiome and Risk for Colorectal Cancer. *J Natl Cancer Inst*. 2013; 105(24):1907–11. doi: [10.1093/nci/djt300](#) PMID: [24316595](#)
29. Yu GQ, Gail MH, Shi JX, 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(5):735–41. doi: [10.1158/1055-9965.Epi-13-0855](#) WOS:000335145300005. PMID: [24700175](#)
30. Ahn J, Chen CY, Hayes RB. Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes Control*. 2012; 23(3):399–404. doi: [10.1007/s10552-011-9892-7](#) PMID: [22271008](#)

31. Contreras M, Costello EK, Hidalgo G, Magris M, Knight R, Dominguez-Bello MG. The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology*. 2010; 156(Pt 11):3282–7. doi: [10.1099/mic.0.043174-0](https://doi.org/10.1099/mic.0.043174-0) PMID: [20847007](https://pubmed.ncbi.nlm.nih.gov/20847007/)
32. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci*. 2010; 107(33):14691–6. doi: [10.1073/pnas.1005963107](https://doi.org/10.1073/pnas.1005963107) PMID: [20679230](https://pubmed.ncbi.nlm.nih.gov/20679230/)
33. Socransky SS, Haffajee AD. Periodontal microbial ecology. *Periodontol 2000*. 2005; 38:135–87. PMID: [15853940](https://pubmed.ncbi.nlm.nih.gov/15853940/)
34. Yang H, Qu L, Ma H, Chen L, Liu W, Liu C, et al. Mycoplasma hyorhinis infection in gastric carcinoma and its effects on the malignant phenotypes of gastric cancer cells. *BMC gastroenterology*. 2010; 10:132. doi: [10.1186/1471-230X-10-132](https://doi.org/10.1186/1471-230X-10-132) PMID: [21062494](https://pubmed.ncbi.nlm.nih.gov/21062494/); PubMed Central PMCID: PMC2993648.
35. Mariotti E, Gemei M, Mirabelli P, D'Alessio F, Di Noto R, Fortunato G, et al. The percentage of CD133+ cells in human colorectal cancer cell lines is influenced by Mycoplasma hyorhinis infection. *BMC Cancer*. 2010; 10:120. doi: [10.1186/1471-2407-10-120](https://doi.org/10.1186/1471-2407-10-120) PMID: [20353562](https://pubmed.ncbi.nlm.nih.gov/20353562/)
36. Urbanek C, Goodison S, Chang M, Porvasnik S, Sakamoto N, Li C-z, et al. Detection of antibodies directed at M. hyorhinis p37 in the serum of men with newly diagnosed prostate cancer. *BMC Cancer*. 2011; 11:233. doi: [10.1186/1471-2407-11-233](https://doi.org/10.1186/1471-2407-11-233) PMID: [21663671](https://pubmed.ncbi.nlm.nih.gov/21663671/)
37. Sasaki H, Ishizuka T, Muto M, Nezu M, Nakanishi Y, Inagaki Y, et al. Presence of Streptococcus anginosus DNA in esophageal cancer, dysplasia of esophagus, and gastric cancer. *Cancer research*. 1998; 58(14):2991–5. PMID: [9679961](https://pubmed.ncbi.nlm.nih.gov/9679961/).
38. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *Journal of Translational Medicine*. 2005; 3:27. doi: [10.1186/1479-5876-3-27](https://doi.org/10.1186/1479-5876-3-27) WOS:000232161700001. PMID: [15987522](https://pubmed.ncbi.nlm.nih.gov/15987522/)
39. Pushalkar S, Mane SP, Ji XJ, Li YH, Evans C, Crasta OR, et al. Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS Immunol Med Microbiol*. 2011; 61(3):269–77. doi: [10.1111/j.1574-695X.2010.00773.x](https://doi.org/10.1111/j.1574-695X.2010.00773.x) WOS:000288211800004. PMID: [21205002](https://pubmed.ncbi.nlm.nih.gov/21205002/)