



OPEN ACCESS

EDITED BY
Zheng Zhang,
Shandong University, China

REVIEWED BY
Jiyuan Zhou,
Guangzhou Medical University, China
Alexander N. Ignatov,
Peoples' Friendship University of
Russia, Russia

*CORRESPONDENCE
Huilin Sun
sun-hui-lin@126.com
Chao Liu
liuchaogzf@163.com
Ling Chen
lingczy@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 15 April 2022
ACCEPTED 01 July 2022
PUBLISHED 08 August 2022

CITATION
Huang L, Deng L, Liu C, Huang E,
Han X, Xiao C, Liang X, Sun H, Liu C
and Chen L (2022) Fecal microbial
signatures of healthy Han individuals
from three bio-geographical zones in
Guangdong.
Front. Microbiol. 13:920780.
doi: 10.3389/fmicb.2022.920780

COPYRIGHT
© 2022 Huang, Deng, Liu, Huang, Han,
Xiao, Liang, Sun, Liu and Chen. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Fecal microbial signatures of healthy Han individuals from three bio-geographical zones in Guangdong

Litao Huang^{1†}, Liting Deng^{2†}, Changhui Liu^{3†}, Enping Huang¹, Xiaolong Han³, Cheng Xiao¹, Xiaomin Liang¹, Huilin Sun^{2*}, Chao Liu^{3*} and Ling Chen^{1*}

¹Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, China, ²The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou, China, ³Guangzhou Forensic Science Institute, Guangzhou, China

Important forensic evidence traced from crime scenes, such as fecal materials, can help in the forensic investigation of criminal cases. Intestines are the largest microbial pool in the human body whose microbial community is considered to be the human "second fingerprint". The present study explored the potential for community characteristics of gut microbes in forensic medicine. Fecal microbiota profiles of healthy individuals from three representative Han populations (Guangzhou, Shantou and Meizhou) in Guangdong Province, China were evaluated using High-throughput sequencing of V3-V4 hypervariable regions of the 16SrRNA gene. Results of the present study showed that at the genus level, Shantou, Guangzhou, and Meizhou behaved as Enterotype1, Enterotype2, and Enterotype3, which were mainly composed of *Bacteroides*, *Prevotella*, and *Blautia*, respectively. Based on OTU abundance at the genus level, using the random forest prediction model, it was found that there might be potential for distinguishing individuals of Guangzhou, Meizhou, and Shantou according to their fecal microbial community. Moreover, the findings of the microbial community of fecal samples in the present study were significantly different from that of saliva samples reported in our previous study, and thus it is evident that the saliva and feces can be distinguished. In conclusion, this study reported the fecal microbial signature of three Han populations, which may provide basic data for the potential application in forensic practice, containing body fluid identification, and geographical inference.

KEYWORDS

forensic medicine, feces, gut microbiome, 16S rRNA gene sequencing, Guangdong Han individuals

Introduction

Human beings live in a world full of microbes, and co-evolution, co-adaptation as well as co-dependence are the relationships between them and indigenous microbiota (Turnbaugh et al., 2007; Blaser and Falkow, 2009). Microorganisms exist in many sites of the human body mainly in the intestines. They also have

a profound influence on human physiological metabolism and nutrition regulation (Hooper and Gordon, 2001). The human intestinal tract, a nutrient-rich microenvironment, carries 100 trillion (10^{14}) bacteria which are about 10 times more than the number of human cells (Hooper et al., 2002; Bäckhed et al., 2005). The colon is the main contributor to the total number of bacteria in the entire intestine with a density close to 10^{11} - 10^{12} cells/ml (Ley et al., 2006; Sender et al., 2016). Non-invasive fecal samples (the part in the middle of feces that is not in contact with the air and the ground) are usually considered representative of colonic microorganisms for they are easy to obtain and do not harm the subject (Davenport et al., 2017). Next-generation sequencing (NGS) using 16SrRNA gene sequence analysis overcomes the shortcomings that most traditional microorganisms cannot be cultivated and performed much deeper microbial community analysis at a low cost (Weinstock, 2012).

Feces have been specified as the important evidence for specific crimes, including burglary, robbery, and sexual cases. Particularly, in anal sexual assault cases, fecal traces of the victim may be left on a condom at the crime scene (Johnson et al., 2005). When analyzed, the microbial community in feces may help in individual identification and tracing the source of tissues and body fluids. Quak et al. distinguished individuals by researching the microbial profiles generated in fecal samples from 35 healthy volunteers of different ages. It was then proposed that individual identification can be carried out by applying the fecal microbial profile to the increase evidence value of the trace when there was no or only part of human STR in fecal samples (Quak et al., 2017). Microarray was also performed to analyze 175 samples from healthy individuals, successfully distinguishing and identifying the oral cavity, feces, and skin samples. The study noted that it might be beneficial for presenting important corroborating evidence for the scene left by the victim and/or suspect, aiding in the reconstruction of a case process (Quak et al., 2018).

Recent studies have shown that the human intestinal microbial community is not only affected by the host's own factors, but also by external factors (Wen and Duffy, 2017). In general, geography and environment have shown the main influence on intestinal microbes (He et al., 2018; Rothschild et al., 2018). Guangdong Province is located in the southernmost part of mainland China and is an important heritage site of Lingnan culture. Lingnan Han groups, consisting of the Guangfu, Hakka, and Chaoshan, account for a majority of Han people in Guangdong. They have a unique culture in terms of language, customs, and living habits. For instance, Guangfu people speak Cantonese, cook Cantonese cuisine, and live mainly in the Pearl River Delta area of Guangdong. Further, Hakkas people are concentrated in northern Guangdong, mainly Hakka dialect and Hakka cuisine, together with Chaoshan people living in eastern Guangdong have their own Chaoshan dialects and Chaoshan cuisine (Wang et al., 2010; Du et al.,

2019). Guangdong's three Han characteristic population was recognized as a branch of Han Chinese, and the gut microbiome characterization and forensic potential of these three groups are poorly defined or still need to be explored. The current study aimed to reveal the differences in fecal microbiota between the groups. Indigenous Han individuals from Guangzhou, Meizhou, and Shantou were selected as the representative of Guangfu, Hakka, and Chaoshan individuals, respectively. The fecal samples were collected and characterized through high-throughput sequencing of the samples in the V3-V4 region of the 16SrRNA gene. The prospect of forensic application of fecal microbiota was valued.

Materials and methods

Sample collection

This study was approved by the Biomedical Ethics Committee of Southern Medical University, Guangzhou, China. After obtaining informed consent, a total of 59 fecal samples were collected from healthy Han individuals (aged between 16 and 62) who had lived in Guangzhou, Meizhou, and Shantou for more than three generations in Guangdong Province, China. A total of 19, 20, and 20 samples from people in Guangzhou, Meizhou, and Shantou, respectively were collected. Participants were balanced by age and sex, divided into age1 (16–32 years old) and age2 (33–62 years old) groups, male and female groups. The participants received adequate training and guidance on the sample collection process before fecal collection and one sample was then collected per participant. The exclusion criteria were (1) participants who reported antibiotic use/other treatments within 3 months. (2) participants were diagnosed with any inflammation-related bowel disease or gastrointestinal disease within 3 months. (3) participants who lived <1 year or left the province within 1 month. According to the above criteria, a total of 59 healthy individuals from the three regions were included, and all fecal samples collected were named "F" (Guangzhou sample numbered from 1 to 19, Shantou sample numbers were from 20 to 39, Meizhou sample numbers were from 40 to 59). The participants used a sterile spoon to dig out a fallen scoop (about 3–5g) of fecal samples, collected them in a sterile plastic container, and immediately stored them in a refrigerator at -80°C in the laboratory awaiting extraction of the fecal bacterial genomic DNA.

DNA extraction, PCR amplification, and sequencing

Bacterial genomic DNA in the samples was extracted using QIAamp DNA Stool Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions.

The concentration and purity of DNA were quantified by using an ultraviolet spectrophotometer and DNA extraction quality is checked by 1% agarose gel electrophoresis. Qualified DNA samples were amplified using bacterial 16S rRNA corresponding DNA sequence V3-V4 region universal primers 338F (5'-ACTCCTACGGGAGGCAGCA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') which contained a unique sequence tag to barcode each sample. PCR enrichment was performed in a 25 μ l reaction containing 12.5 μ l of 2 \times Q5 Master Mix, 0.2 μ M of each primer, 120ng of the extracted DNA, and Nuclease-free water. PCR reaction amplification conditions were: initial denaturation at 98°C for 5 mins; followed by 15-21 cycles of denaturation at 98°C for 10 s, primer annealing at 57°C for 30 s, extension at 72°C for 30 s; and a final extension step at 72°C for 5 min. The PCR products were purified with AmpureXP beads and eluted in the Elution buffer. Libraries were built with NEB Next UltraTM DNA Library Prep Kit for Illumina (New England Biolabs Inc, Ipswich, USA). And then the validated libraries were used for sequencing on the Illumina MiSeq platform (Illumina Corporation, San Diego, USA). The sequencing data have been deposited in NCBI BioProject PRJNA824624 with the Biosample accessions SAMN27409411-SAMN27409469.

Bioinformatics analysis

The raw reads obtained by sequencing are filtered to obtain high-quality data (clean reads) for downstream analysis. Using the software FLASH (Magoč and Salzberg, 2011) (Fast Length Adjustment of Short reads, v1.2.11), the paired reads obtained by double terminal sequencing are assembled into a sequence, that is, a tag, by using the overlapping relationship. Use CUTADAPT (Martin, 2011) to remove tags containing primers, refer to the gold database (v20110519) chimera database, and use the UCHIME method in the VSEARCH (v2.3.4) (Rognes et al., 2016) software to remove the tags containing the chimera. Use VSEARCH (v2.3.4) software to cluster Tags with a similarity > 97% into an OTU, and get the OTU representative sequence. Use RDP classifier (v2.2) (Wang et al., 2007) software to compare OTU representative sequence with Silva (v128) database for species annotation. Alpha diversity is used to analyze the species diversity in the sample, using mothur (v1.39.5) (Schloss et al., 2009) software to calculate 5 indicators, including Chao, Ace, Shannon, and Simpson. Beta diversity is used to measure the diversity between samples, calculated using QIIME (v1.80) (Caporaso et al., 2010) software. The rest of the graphics are implemented using R package (v3.0.3). Use LefSe (LDA Effect Size) (v1.0) (Segata et al., 2011) to calculate the LDA score value. The significant flora must meet the threshold $p < 0.05$ and the LDA score value ≥ 2.0 (or ≤ -2.0). Through the use of QIIME (v1.80) (Schloss et al., 2009) software, the use of similarity

analysis (ANOSIM) for group comparison analysis, to find out the different components in the group.

Machine learning process

Random forest analysis was used to perform classification. This method constructed multiple decision trees by using the information contained in input features and predicted the classification of three regions by combining multiple weak classifiers (Breiman, 2001). According to the random forest method in the R package RandomForest (v4.6-14), the OTU data of intestinal microorganisms in the three regions was used to build a model for predicting the sample distribution in the areas. The RF classification method was divided into two steps: one was to build a decision tree based on randomly selected samples (the training set) which include 70% of the original data set (42 samples). The other one was to use the test set which was the remaining samples (17 samples) in the original data set to verify the decision tree (Svetnik et al., 2003). In addition, the receiver operating characteristic (ROC) curve was used to evaluate the constructed model, and the area under the ROC curve (AUC) was used to designate the ROC effect to evaluate the potential of intestinal microbial markers to predict different regions.

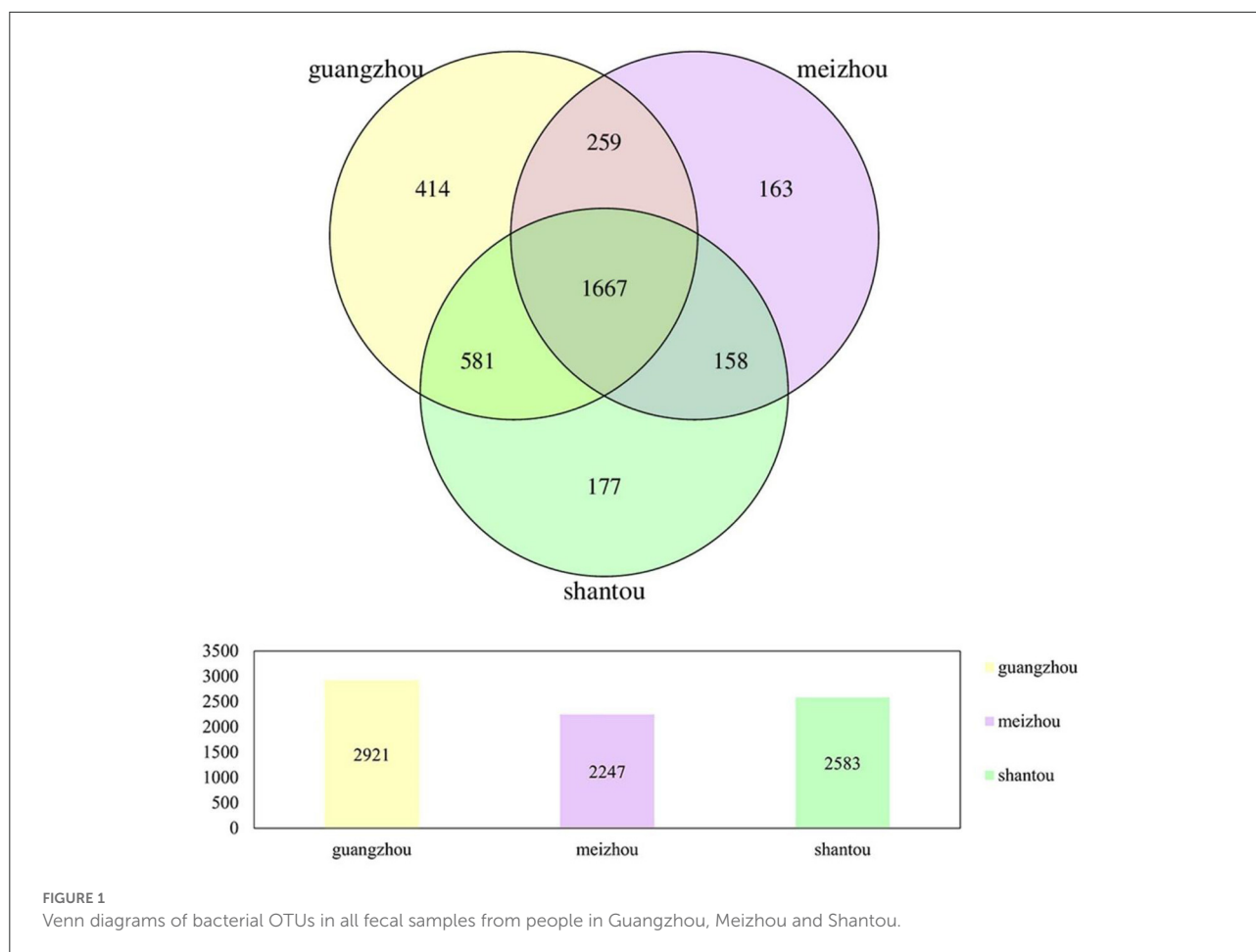
Results

Correlation with age and sex of the subjects

The present study explored the relationship between the composition of the gut microbial community and age as well as sex in the entire population. The results of ANOSIM analysis of the present study based on Bray-Curtis distance showed that there was no significant difference in the gut microbial community between age 1 and age 2 group ($p = 0.49$), and the male and female group ($p = 0.30$).

Whole sequencing data

Fecal samples of 59 healthy individuals from Guangzhou, Meizhou, and Shantou, Guangdong Province were subjected to high-throughput sequencing of 16SrRNA gene. After filtering, a data set consisting of 4256.44Mbp of effective and high-quality 16SrRNA gene sequences were generated, including 16,740,484 reads (median=221,912 reads, ranging from 79,892 to 599,496 reads; Supplementary Table 1). A cluster analysis of 97% similarity was performed to determine a total of 3,419 OTUs. All the valid sequences were annotated with species at different taxonomic levels, which yielded a total of 3,419 OTUs, belonging to 13 phyla, 15 classes, 21 orders, 35 families, 119



genera, and 22 species. The Venn diagram showed that the number of unique OTUs in Guangzhou, Meizhou, and Shantou was 414, 163, and 177, respectively, with 1667 OTUs shared by all the samples in the present study (Figure 1).

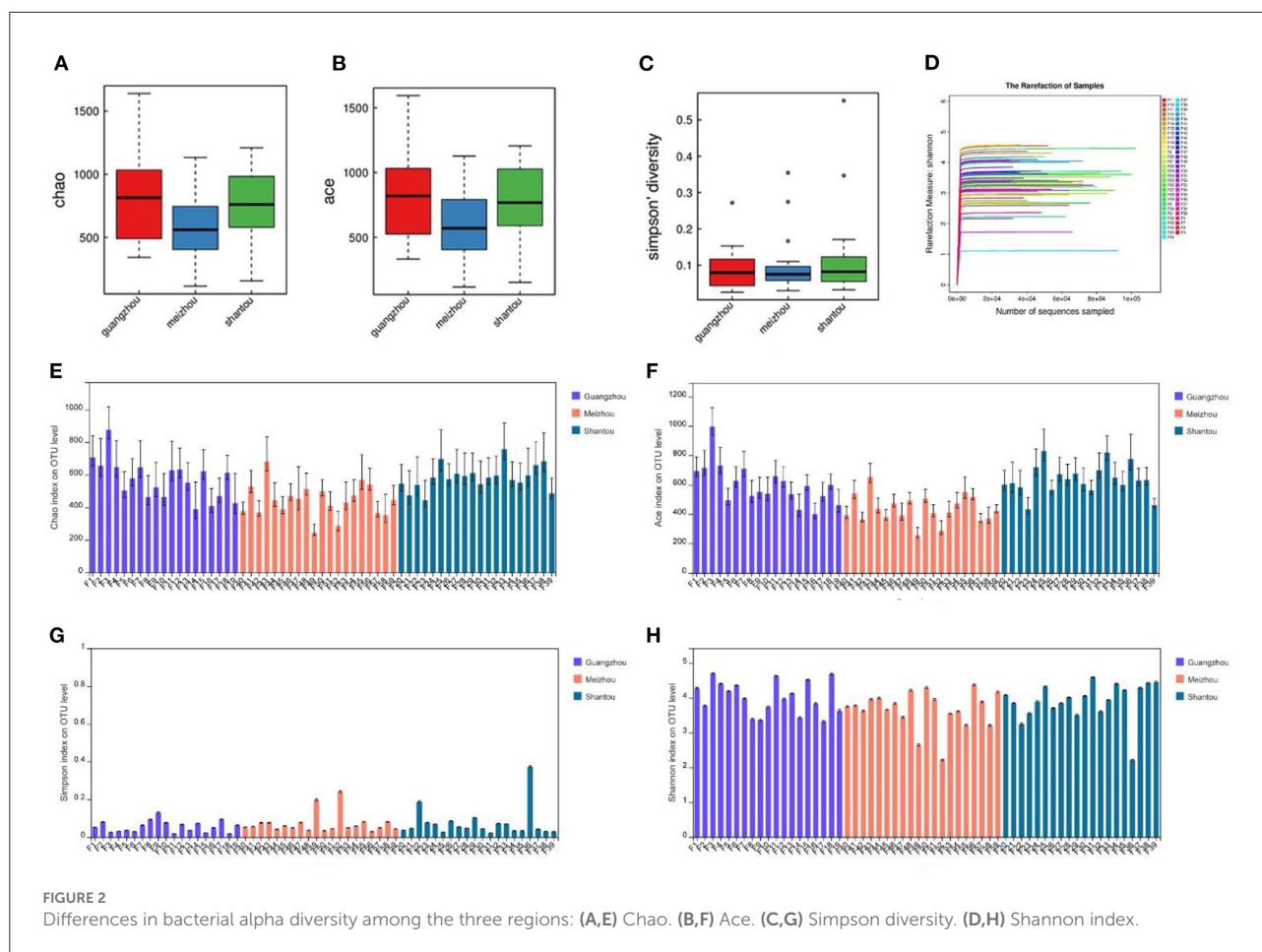
Richness and diversity of microbial communities

Microbial complexity in the feces was estimated based on alpha-diversity indices (Chao, Ace, Simpson, and Shannon), and the results showed that there was no significant difference in the diversity among all individuals in each group (Figures 2E–H). Pairwise diversity of the three groups in the present study, the indices of Chao and Ace represented the species richness. The results of the present study showed that individuals from Guangzhou and Shantou had significantly higher index values as compared with those from Meizhou (Figures 2A,B). Results of the Simpson diversity index in the current study revealed that the three regions had similar statistical index values, indicating no significant difference in species diversity (Figure 2C; $p >$

0.05). In addition, the sparse curve of the Shannon index showed a trend toward saturation as presented in Figure 2D which illustrated sufficient sequencing depth.

Overview of bacterial community composition

The average relative abundance of the three groups at the phylum and genus level was also evaluated to further intuitively uncover the microbial composition characteristics in the three regional groups as presented in Figure 3. It was found that phylum Firmicutes was the most predominant phyla in Guangzhou, Meizhou, and Shantou, with relative abundances of 46.7, 43.4, and 62.5%, respectively. This was followed by phylum Bacteroidetes, which contributed 43.1, 38.2, and 16.1% of the total sequences. Further, it was noted that Bacteroides had the highest abundance in the bacterial communities of fecal samples at the genus level, accounting for 28.7, 31.7, and 12.7% in Guangzhou, Meizhou, and Shantou, respectively. On the other hand, Faecalibacterium



accounted for 7.4, 7.9, and 8.9% in Guangzhou, Meizhou, and Shantou, respectively. The remaining top 10 bacterial genera were *Blautia*, *Eubacterium_rectale_group*, *Bifidobacterium*, *Roseburia*, *Prevotella_9*, *Megamonas*, *Escherichia-Shigella*, and *Fusobacterium*. Besides, it was found that the relative abundance of *Bifidobacterium* was 1.54%, 1.04%, and 5.09% in Guangzhou, Meizhou, and Shantou, respectively.

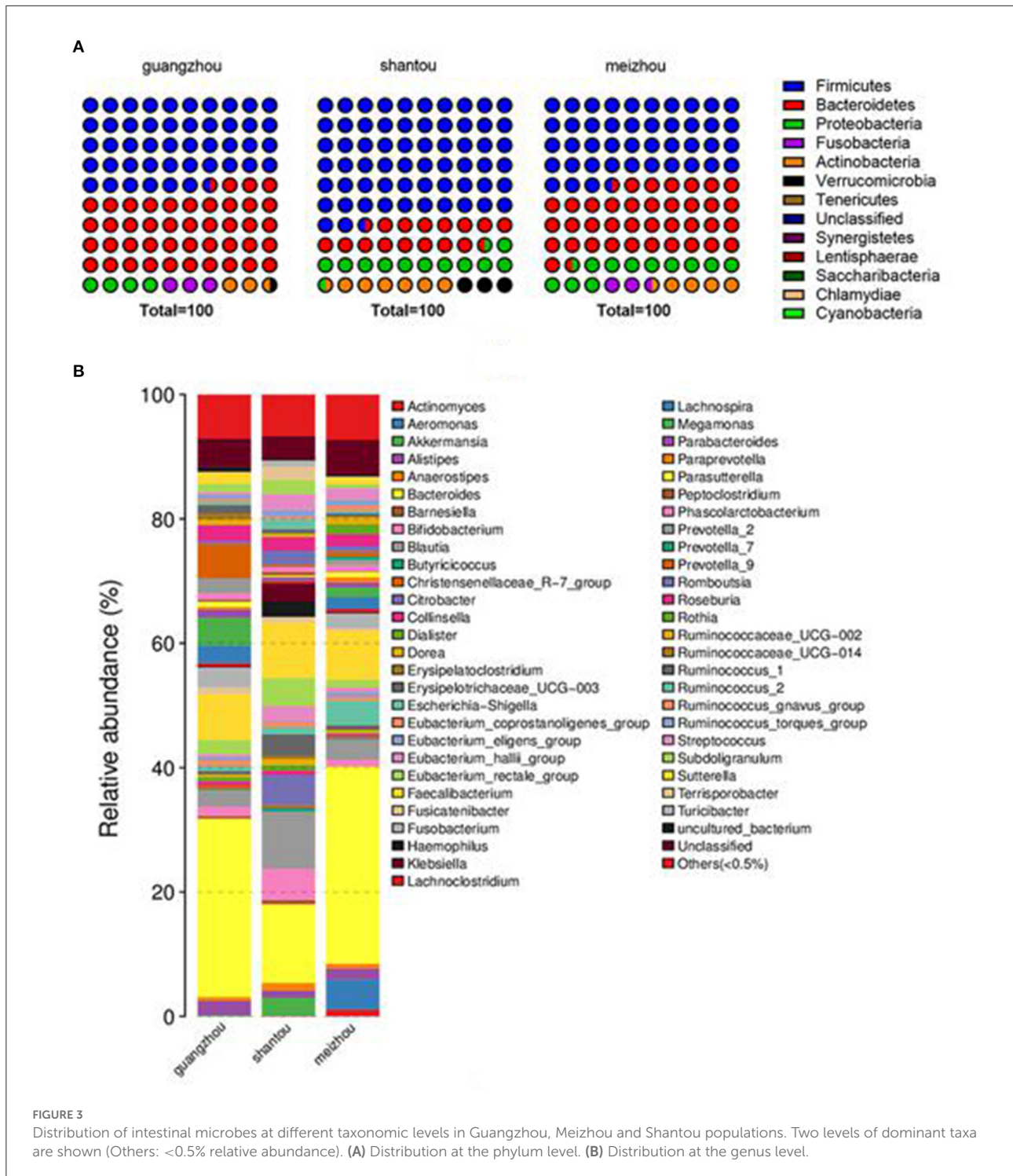
Genus-level core intestinal flora and comparison of feces and saliva

The intestinal core microbiome was determined at the genus level and defined as bacteria with >0.1% abundance in $\geq 80\%$ of the respective samples (Dehingia et al., 2015). It was found that there were six main genera in the fecal samples of all individuals, which constituted a genera-level phylogenetic core, including *Bacteroides*, *Blautia*, *Eubacterium_hallii_group*, *Faecalibacterium*, *Lachnoclostridium*, and *Roseburia* (Supplementary Table 2). Further, these fecal samples were used to compare with saliva samples we previously published (Yao et al., 2021) and the results of the comparisons were as

shown in Supplementary Figure 1. The data of the present study on principal coordinate analysis (PCoA) based on genus-level abundance revealed that there was a clear distinction between fecal samples and saliva samples. Further, the linear discriminant analysis (LDA) histogram reflected that at the genus level, the relative abundance of *Bacteroides*, *Faecalibacterium*, *Blautia*, and *Bifidobacterium* was higher in the fecal samples, whereas the relative abundance of *Streptococcus*, *Gemella*, *Porphyromonas*, and *Haemophilus* was higher in the saliva samples.

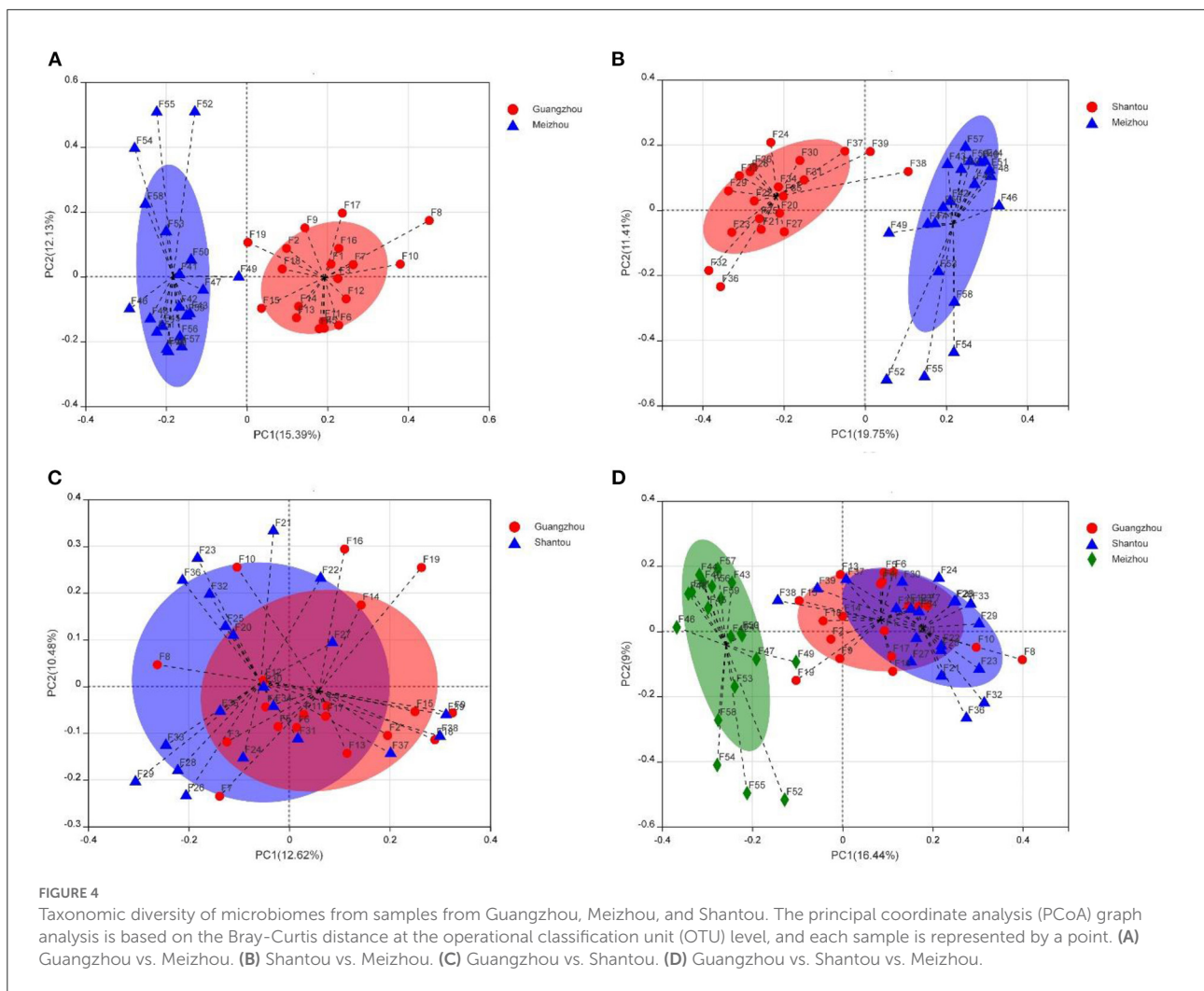
Beta diversity of bacterial communities

Beta diversity was assessed by PCoA and ANOSIM analysis using the Bray–Curtis distance method at the operational classification unit (OTU) level to further indicate the similarity between microbial communities. Although there were some slight overlaps in individual samples, the samples of Guangzhou and Meizhou groups, Shantou and Meizhou groups were roughly clustered. The similar structure of the intestinal microbiota community was found in the fecal samples between Guangzhou and Shantou, indicating



an overlap in community structure (Figures 4A–C). The samples of the Meizhou population formed an “out-group,” which was generally not confounding with the samples of the Guangzhou or Shantou populations (Figure 4D). The ANOSIM analysis was performed on the three geographical

groups (Supplementary Figure 2), and the results of this study demonstrated that the differences between the groups were greater than the differences within the groups, and the groupings were statistically significant ($R = 0.3254$, $p = 0.0010$).



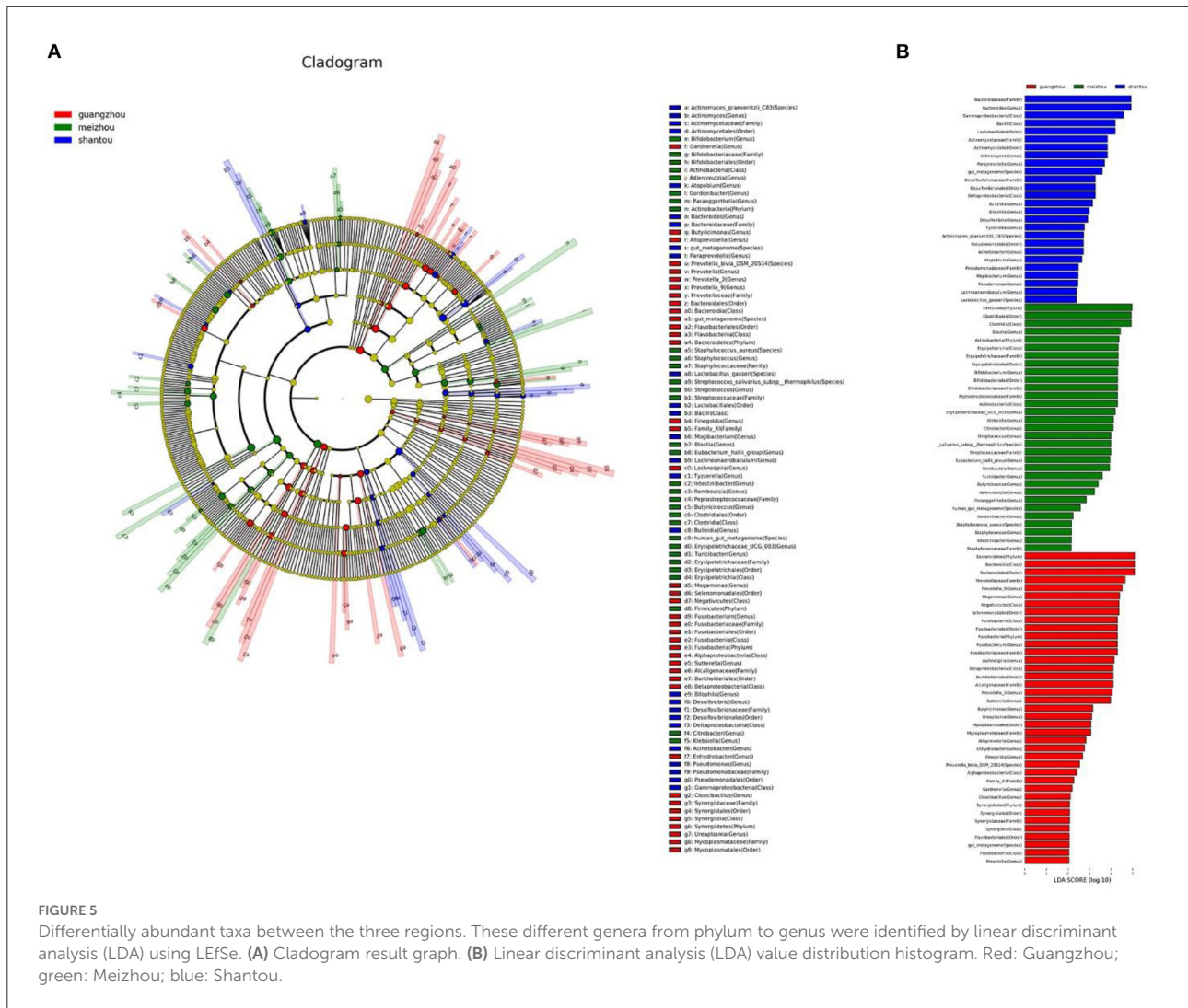
Comparison of differences among three regions

The linear discriminant analysis effect size (LEfSe) test for biomarkers was used to find the taxa with the strongest effect on region differentiation. The Cladogram chart showed that there were at least two significant differences in the phylum, class, order, family, genus, and species level in the fecal samples from Guangzhou and Meizhou (Figure 5A). The composition of the microbial community of the fecal samples from Shantou at the phylum level was not significantly different from that of Guangzhou and Meizhou. In addition, at least three significantly different microorganisms were found at the level of class, order, family, genus, and species levels. Further, a total of 96 differentially abundant taxa were found in the three regions shown in the histogram of LDA value distribution (Figure 5B). At the phylum level, the significant differences in the samples of the Guangzhou and Meizhou populations were mainly Bacteroidetes, and Firmicutes,

respectively. The top five microorganisms with significant differences at the genus level in the three regions included Prevotella-9, Megamonas, Fusobacterium, Lachnospira, and Prevotella_2 in Guangzhou, Bacteroides, Actinomyces, Paraprevotella, Bulleidia, Bilophila in Shantou, and Blautia, Bifidobacterium, Erysipelotrichaceae_UCG_003, Klebsiella, Citrobacter in Meizhou.

Random forest

During the construction of a random forest model based on the composition of gut microbes, top 230 OTUs markers were set as the best set. The markers performed well and were on the training set ($n = 42$, 14 samples in Guangzhou, Shantou and Meizhou). The validation set of the random forest model ($n = 17$, 5 Guangzhou samples, 6 Meizhou samples, and 6 Shantou samples) showed that 12 of the 17 validation samples were correctly classified, and 100% of the Meizhou samples were

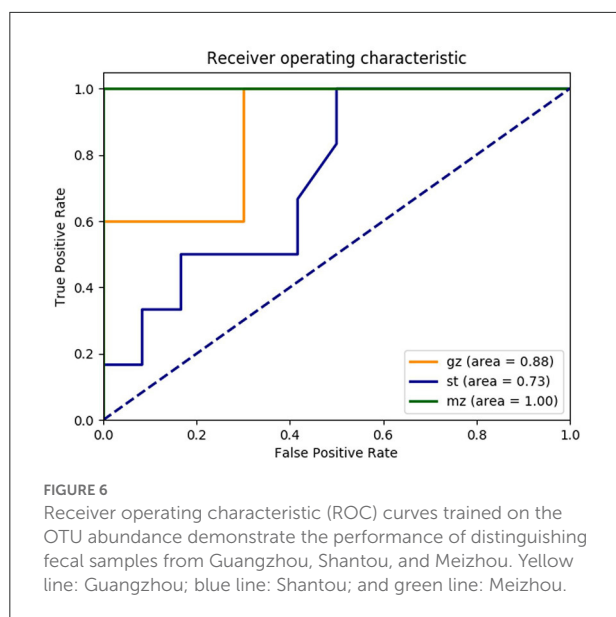


correctly predicted, whereas 2 Guangzhou samples (F7 and F8) were identified as Shantou samples and 3 Shantou samples (F23, F25, and F34) were identified as Guangzhou samples, with an overall accuracy of 70.59%. The performance of the model was evaluated using ROC analysis. The AUC of the area under the curve in Guangzhou, Shantou, and Meizhou were 0.88, 0.73, and 1.00, respectively (Figure 6).

Discussions

The present study explored the correlation between the gut microbiota of the entire population and age as well as sex. Further, the ANOSIM analysis showed that there were no statistical difference between the intestinal microbial community structures between 16 and 32 as well as between 33 and 62 years of age. Previous studies had shown that *Bifidobacterium* was dominant in infants and a larger

proportion of *Bacteroides* was dominant in elderly individuals (Claesson et al., 2011; Yatsunenکو et al., 2012). On the other hand, Firmicutes and Bacteroidetes as the dominant bacteria were mainly dominant in adults. The established microbiota composition remained unchanged when there was no change in long-term eating habits and pathophysiology (Adak and Khan, 2019). In the current study, the small difference between the two age groups could be associated with most young individuals in the current study (45 cases, 76.27% between 25–45 years old), with only one individual who was over 60 years old. In addition, it was evident from the results of this study that there were no statistical differences in fecal microbiota between males and females. The finding of the present study was consistent with the results of a study carried out by Arumugam et al. that found that sex had no effect on the structure of the gut microbes of individuals from six different countries (Arumugam et al., 2011). Moreover, several other studies have also shown that sex factors have less influence on the gut microbial community than



other factors (Kovacs et al., 2011; Human Microbiome Project Consortium, 2012).

The analyses performed at the phylum level in the present study showed that the intestinal microbiota of this research was made up of the four most important phyla, including Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria. It was evident that phylum Firmicutes and Bacteroidetes were the most abundant. This was similar to the results of previous studies (Jandhyala et al., 2015). Although the diversity of gut microbes at the phylum level was low, it was noted that they had significantly high diversity at the genus level. From the results of the current experiments, the predominant genera in all individuals was Bacteroides, followed by Faecalibacterium. A previous study reported that China had the highest abundance of Bacteroides at the genus level as compared with four other countries. This was consistent with the findings of the present study. Furthermore, the study reported that Japan had higher levels of Bifidobacterium whereas the abundance of Prevotella and Faecalibacterium was relatively higher in Korea (Nam et al., 2011). Previous studies had also indicated that Faecalibacterium was more dominant in the populations of Hadza, Italy, and the United States. Furthermore, Prevotella was a significant genus found among the Indian tribes, Mongolians, American Indians, and Malawi tribes (Dehingia et al., 2015). This difference in dominant genus originates from the variations in the intestinal microbiome, whereas the changes in the intestinal microbiome may be caused by geography and ethnicity among other factors (Dwiyanto et al., 2021).

One of the main interests of human gut microbial research was toward the core microbiota. The bacterial genera of Faecalibacterium, Eubacterium, Clostridium, Blautia, Ruminococcus, and Roseburia were found to be

the core gut microbiota in the representative populations of the world (Dehingia et al., 2015). In the current study, six genera-level core intestinal bacteria of the gut microbiota, ubiquitously in unrelated individuals from Guangdong, which were Bacteroides, Blautia, Eubacterium_hallii_group, Faecalibacterium, Lachnospirillum, and Roseburia. A microbial analysis report from nine provinces in China revealed a total of nine core bacteria (Blautia, Clostridium, Ruminococcus, Faecalibacterium, Subdoligranulum, Roseburia, Coprococcus, Bacteroides, and Phascolarctobacterium) (Zhang et al., 2015). In healthy western individuals, Bifidobacterium, Bacteroides, Faecalibacterium, Ruminococcus, Blautia, Dorea, Eubacterium, and Coprococcus were the core intestinal bacteria genus (Martínez et al., 2013). Further, the intestinal core flora shared by these people were Bacteroides, Blautia, and Faecalibacterium. In addition, more than 45% of the common bacterial genera could be detected in both feces and oral cavities (Segata et al., 2012). It is worth mentioning that the establishment of the intestinal saliva microbial communities was similar. According to a study by Schmidt et al., transmission to, and subsequent colonization of the large intestine by oral microbes commonly occurred in healthy individuals. Although it has been previously reported that *Streptococcus salivarius* and *S. mutans* were particularly found in saliva (Tagg and Ragland, 1991). A study conducted by Kai-NanZou et al. showed that the bacteria in the intestines overlapped with those in feces (Zou et al., 2016). These results indicated that the identification of sample types using a single microbial marker may be misjudged. The findings of fecal samples in the present study were compared with those of saliva samples in our previously published study (Yao et al., 2021). In addition, the results showed that fecal and saliva samples can be distinguished, which could avoid the defect of single microbial markers to identify both saliva and feces samples.

The PCoA displayed regional differences in intestinal microorganisms between Meizhou and the other two regions. Different geographic origins of humans may result in diverse compositions of the gut microbiome, due to distinctive genetic backgrounds or life environments (Li and Zhao, 2015). Guangfu and Chaoshan populations occupied the two rich areas of the Pearl River delta plain and Chaoshan plain, respectively. The barren and backward mountainous areas of northern and eastern Guangdong were the basic distribution areas of the Hakka people. Several studies have demonstrated that geographic location plays an important role in shaping the intestinal microbial community, and dietary habits could also affect the composition and distribution of intestinal microbes (De Filippo et al., 2010; Zhang et al., 2013; Singh et al., 2017). Through a return visit to the volunteers in the three regions, they simply recorded their eating habits. The Meizhou area was dominated by greasy food, whereas the Guangzhou and Shantou areas were dominated by intake of a light diet (Song et al., 2005; Zhong et al., 2017; Wang et al., 2019).

A high-fat diet had been shown to reduce the diversity and richness of human gut microbial communities, which was negatively correlated with the abundance of *Bifidobacterium*. Furthermore, Caesar et al. reported that *Bacteroides* increased in mice fed with lard (De Filippo et al., 2010; Caesar et al., 2015; Khine et al., 2019). The results of the present study suggested that the intestinal microbes in Meizhou had the lowest abundance of *Bifidobacterium* and microbial alpha diversity, whereas *Bacteroides* showed the highest abundance among the three regions. This might be related to the fact that Hakka ancestors lived in mountainous areas with inconvenient transportation, expended much physical strength on their daily labor, and needed to supplement foods with rich fat sources such as lard, developing a diet that preferred greasy foods. Therefore, diet may also be an important factor affecting the microbial differences in fecal samples from the three geographical regions. The dietary associations seen here paralleled a recent study comparing European and African, Europeans consuming high-fat foods formed a typical taxonomy dominated by *Bacteroidetes*, while Africans consuming low-fat diets had higher microbial diversity (De Filippo et al., 2010). At the same time, a study of American populations showed that the gut flora of individuals with a typical western diet high in animal fat and protein was dominated by *Bacteroides* (Wu et al., 2011). There are, of course, many differences between the three regions that might influence the gut microbiome, but dietary differences provide an attractive potential explanation.

According to the results of the ANOSIM analysis, there were significant differences in the intestinal bacterial community composition in samples from the three regions. A previous study identified three intestinal types: *Bacteroides* (Enterotype 1), *Prevotella* (Enterotype 2), and *Ruminococcus* (Enterotype 3) (Arumugam et al., 2011), which could afford a strong discriminatory classification ability among European individuals, although other studies had reported that Enterotype 3 was an uncertain bacterial composition (Liang et al., 2017). Hyun Seok et al. showed that structure of gut microbiota variations across the geographical location. The characterization of population distribution according to the three enterotype classifications showed that the distributions of Enterotype 2 and Enterotype 1 differed by region. Samples from the U.S. and Japan had large numbers of Enterotype 1, while samples from Amazon natives in Venezuela, as well as from Malawi and Tanzania in Africa had large numbers of Enterotype 2 (Oh et al., 2022). In the present study, linear discriminant analysis (LDA) using LEfSe showed that Shantou, Guangzhou, and Meizhou belonged to Enterotype 1, Enterotype 2, and Enterotype 3, which were mainly composed of *Bacteroides*, *Prevotella*, and *Blautia*, respectively.

The present study attempted to construct a prediction model on the basis of OTU abundance of a genus of intestinal microbes for biogeographic inference. According to the parameter importance ranking of random forest, the most

important characteristic differences in classification were mainly *Bacteroides*, *Lactobacillus*, and *Prevotella*-9. Similar to LEfSe analysis, it might be inferred that the main flora of intestinal microbes could be used as a factor in predicting geographic location. Likewise, a study conducted by De Filippo et al. found that Firmicutes and *Bacteroides* could distinguish children in rural Europe and Africa has significantly demonstrated that *Prevotella* was a powerful tool for discriminatory classification (De Filippo et al., 2010). The present study found that through verification, the accuracy of the predictions in the three regions was very high, especially in the Meizhou area, where the AUC was 1. All the samples from Meizhou in the verification set were correctly classified, whereas the performance of Guangzhou and Shantou was not satisfactory (the Guangzhou sample and the Shantou sample misjudged each other). Further, the finding of this study was similar to the results of PCoA. It might be possible that a combination of geography, dietary, and other factors play an important role (Yatsunenko et al., 2012). This needs to be understood by further research.

This study provides the first insight into the gut microbiome data of the three characteristic Han populations in Guangdong, which can enrich gut flora information of Chinese ethnic groups. And joint analysis of geography and diet might be helpful to provide enlightening information for forensic science. In addition, due to the complexity of the population composition and living environment of Guangdong Province, so the representativeness of researching samples from the selected three regions is limited. In our current study, individual differences need to be analyzed with large sample size, and the research is still limited to the relative abundance at the genus level. In the future, the sample size will be expanded, sample table information will be recorded in detail (recording used water sources, Food Frequency Questionnaire (FFQ), and other factors), and fecal microbiome analysis will be performed in depth based on microbial species level and sequence. In order to observe the flora differences in different regions of Guangdong Province, follow-up studies will further explore the gut flora of multi-ethnic and multiregional populations.

Conclusion

In conclusion, the current study used high-throughput sequencing methods to study the characteristics of the fecal microbial community of healthy Han individuals living in three regions of Guangdong Province. The results of the current study showed that the composition of intestinal microbes was mainly composed of *Bacteroides*, *Faecalibacterium*, and *Blautia* at the genus level. The feces could be significantly distinguished from saliva samples according to microbial differences at the genus level of both. Further, the populations in the three regions exhibited different enterotype classifications and the prediction model based on the random forest algorithm evidently showed

a significant effect in distinguishing individuals, which might be due to regional differences. In conclusion, microbial community information in feces may have the potential for forensic analysis of body fluid traceability and regionally specific.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SAMN27409411-SAMN27409469.

Ethics statement

The studies involving human participants were reviewed and approved by Biomedical Ethics Committee of Southern Medical University, Guangzhou, China. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LH: conceptualization, methodology, visualization, investigation, writing—original draft. LD and CL: validation, formal analysis. EH, XH, CX, and XL: resources, supervision, data curation. HS, CL, and LC: writing—review & editing. All authors discussed the results and contributed to the final manuscript.

Funding

This project was supported by the Open project of Natural Science Foundation of Guangdong Province (Grant no. 2020A1515010938), Science and Technology Program of

Guangzhou, China (Grant no. 2019030016 and Grant no. 202102080308), and Medical Science and Technology Research Foundation of Guangdong Province (A2019443). We are grateful to all volunteers who contributed samples for this study.

Acknowledgments

The authors would like to acknowledge the support of the specific colleagues and the collaboration effort of the wider project team, which included Southern Medical University, The First Affiliated Hospital of Guangdong Pharmaceutical University and Guangzhou Forensic Science Institute.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.920780/full#supplementary-material>

References

- Adak, A., and Khan, M. R. (2019). An insight into gut microbiota and its functionalities. *Cell Mol. Life Sci.* 76, 473–493. doi: 10.1007/s00018-018-2943-4
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180. doi: 10.1038/nature09944
- Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920. doi: 10.1126/science.1104816
- Blaser, M. J., and Falkow, S. (2009). What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* 7, 887–894. doi: 10.1038/nrmicro2245
- Breiman (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Caesar, R., Tremaroli, V., Kovatcheva-Datchary, P., Cani, P. D., and Bäckhed, F. (2015). Crosstalk between gut microbiota and dietary lipids aggravates WAT inflammation through TLR signaling. *Cell Metab.* 22, 658–668. doi: 10.1016/j.cmet.2015.07.026
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., et al. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl 1), 4586–4591. doi: 10.1073/pnas.1000097107
- Davenport, E. R., Sanders, J. G., Song, S. J., Amato, K. R., Clark, A. G., and Knight, R. (2017). The human microbiome in evolution. *BMC Biol.* 15, 127. doi: 10.1186/s12915-017-0454-7

- De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., et al. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14691–14696. doi: 10.1073/pnas.1005963107
- Dehingia, M., Devi, K. T., Talukdar, N. C., Talukdar, R., Reddy, N., Mande, S. S., et al. (2015). Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Sci. Rep.* 5, 18563. doi: 10.1038/srep18563
- Du, W., Wu, W., Wu, Z., Guo, L., Wang, B., and Chen, L. (2019). Genetic polymorphisms of 32 Y-STR loci in Meizhou Hakka population. *Int. J. Legal Med.* 133, 465–466. doi: 10.1007/s00414-018-1845-1
- Dwiyanto, J., Hussain, M. H., Reidpath, D., Ong, K. S., Qasim, A., Lee, S., et al. (2021). Ethnicity influences the gut microbiota of individuals sharing a geographical location: a cross-sectional study from a middle-income country. *Sci. Rep.* 11, 2618. doi: 10.1038/s41598-021-82311-3
- He, Y., Wu, W., Zheng, H. M., Li, P., McDonald, D., Sheng, H. F., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* 24, 1532–1535. doi: 10.1038/s41591-018-0164-x
- Hooper, L. V., and Gordon, J. I. (2001). Commensal host-bacterial relationships in the gut. *Science* 292, 1115–1118. doi: 10.1126/science.1058709
- Hooper, L. V., Midtvedt, T., and Gordon, J. I. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* 22, 283–307. doi: 10.1146/annurev.nutr.22.011602.092259
- Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature*. 486, 207–214. doi: 10.1038/nature11234
- Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., and Nageshwar, R. D. (2015). Role of the normal gut microbiota. *World J. Gastroenterol.* 21, 8787–8803. doi: 10.3748/wjg.v21.i29.8787
- Johnson, D. J., Martin, L. R., and Roberts, K. A. (2005). STR-typing of human DNA from human fecal matter using the QIAGEN QIAamp stool mini kit. *J. Forensic Sci.* 50, 802–808. doi: 10.1520/JFS2004428
- Khine, W., Zhang, Y., Goie, G., Wong, M. S., Liang, M., Lee, Y. Y., et al. (2019). Gut microbiome of pre-adolescent children of two ethnicities residing in three distant cities. *Sci. Rep.* 9, 7831. doi: 10.1038/s41598-019-44369-y
- Kovacs, A., Ben-Jacob, N., Tayem, H., Halperin, S., Iraqi, F. A., and Gophna, U. (2011). Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microb. Ecol.* 61, 423–428. doi: 10.1007/s00248-010-9787-2
- Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 124, 837–848. doi: 10.1016/j.cell.2006.02.017
- Li, L., and Zhao, X. (2015). Comparative analyses of fecal microbiota in Tibetan and Chinese Han living at low or high altitude by barcoded 454 pyrosequencing. *Sci. Rep.* 5, 14682. doi: 10.1038/srep14682
- Liang, C., Tseng, H. C., Chen, H. M., Wang, W. C., Chiu, C. M., Chang, J. Y., et al. (2017). Diversity and enterotype in gut bacterial community of adults in Taiwan. *BMC Genomics*. 18, 932. doi: 10.1186/s12864-016-3261-6
- Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Martínez, I., Muller, C. E., and Walter, J. (2013). Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS ONE*. 8, e69621. doi: 10.1371/journal.pone.0069621
- Nam, Y. D., Jung, M. J., Roh, S. W., Kim, M. S., and Bae, J. W. (2011). Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLoS ONE*. 6, e22109. doi: 10.1371/journal.pone.0022109
- Oh, H. S., Min, U., Jang, H., Kim, N., Lim, J., Chalita, M., et al. (2022). Proposal of a health gut microbiome index based on a meta-analysis of Korean and global population datasets. *J. Microbiol.* 60:533–49. doi: 10.1007/s12275-022-1526-0
- Quaak, F., van Duijn, T., Hoogenboom, J., Kloosterman, A. D., and Kuiper, I. (2018). Human-associated microbial populations as evidence in forensic casework. *Forensic Sci. Int. Genet.* 36, 176–185. doi: 10.1016/j.fsigen.2018.06.020
- Quaak, F. C., de Graaf, M. M., Weterings, R., and Kuiper, I. (2017). Microbial population analysis improves the evidential value of faecal traces in forensic investigations. *Int. J. Legal Med.* 131, 45–51. doi: 10.1007/s00414-016-1390-8
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 4, e2584. doi: 10.7717/peerj.2584
- Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 555, 210–215. doi: 10.1038/nature25973
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13, R42. doi: 10.1186/gb-2012-13-6-r42
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. doi: 10.1186/gb-2011-12-6-r60
- Sender, R., Fuchs, S., and Milo, R. (2016). Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell*. 164, 337–340. doi: 10.1016/j.cell.2016.01.013
- Singh, R. K., Chang, H. W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., et al. (2017). Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* 15, 73. doi: 10.1186/s12967-017-1175-y
- Song, F. Y., Toshiro, T., Li, K., Yu, P., Lin, X. K., Yang, H. L., et al. (2005). Development of a semi-quantitative food frequency questionnaire for middle-aged inhabitants in the Chaoshan area, China. *World J. Gastroenterol.* 11, 4078–4084. doi: 10.3748/wjg.v11.i26.4078
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. doi: 10.1021/ci034160g
- Tagg, J. R., and Ragland, N. L. (1991). Applications of BLIS typing to studies of the survival on surfaces of salivary streptococci and staphylococci. *J. Appl. Bacteriol.* 71, 339–342. doi: 10.1111/j.1365-2672.1991.tb03797.x
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature*. 449, 804–810. doi: 10.1038/nature06244
- Wang, M., Liang, B., Zhang, W., Chen, K., Zhang, Y., Zhou, H., et al. (2019). Dietary Lead Exposure and Associated Health Risks in Guangzhou, China. *Int. J. Environ. Res. Public Health*. 16. doi: 10.3390/ijerph16081417
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Wang, W. Z., Wang, C. Y., Cheng, Y. T., Xu, A. L., Zhu, C. L., Wu, S. F., et al. (2010). Tracing the origins of hakka and chaoshanese by mitochondrial DNA analysis. *Am. J. Phys. Anthropol.* 141, 124–130. doi: 10.1002/ajpa.21124
- Weinstock, G. M. (2012). Genomic approaches to studying the human microbiota. *Nature*. 489, 250–256. doi: 10.1038/nature11553
- Wen, L., and Duffy, A. (2017). Factors influencing the gut microbiota, inflammation, and Type 2 Diabetes. *J. Nutr.* 147, 1468S–1475S. doi: 10.3945/jn.116.240754
- Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 334, 105–108. doi: 10.1126/science.1208344
- Yao, T., Han, X., Guan, T., Zhai, C., Liu, C., Liu, C., et al. (2021). Exploration of the microbiome community for saliva, skin, and a mixture of both from a population living in Guangdong. *Int. J. Legal Med.* 135, 53–62. doi: 10.1007/s00414-020-02329-6
- Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature*. 486, 222–227. doi: 10.1038/nature11053
- Zhang, J., Guo, Z., Xue, Z., Sun, Z., Zhang, M., Wang, L., et al. (2015). A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J.* 9, 1979–1990. doi: 10.1038/ismej.2015.11
- Zhang, J., Zheng, Y., Guo, Z., Qiao, J., Gesudu, Q., Sun, Z., et al. (2013). The diversity of intestinal microbiota of Mongolians living in Inner Mongolia, China. *Benef. Microbes*. 4, 319–328. doi: 10.3920/BM2013.0028
- Zhong, Z., Liu, J., Li, B., Li, C., Liu, Z., Yang, M., et al. (2017). Serum lipid profiles in patients with acute myocardial infarction in Hakka population in southern China. *Lipids Health Dis.* 16, 246. doi: 10.1186/s12944-017-0636-x
- Zou, K. N., Ren, L. J., Ping, Y., Ma, K., Li, H., Cao, Y., et al. (2016). Identification of vaginal fluid, saliva, and feces using microbial signatures in a Han Chinese population. *J. Forensic Leg. Med.* 43, 126–131. doi: 10.1016/j.jflm.2016.08.003