RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2023.028

ScienceAsia 49 (2023): **391–402**

Effects of different ethnicities and *FOXC2* gene *C-512T* polymorphisms on the intestinal flora: A study on healthy Dai and Han populations in Yunnan

Lei Liang[a], Hongjiang Zhou[a], Elizabeth Rao[b], Xiaoyun Su[a], Hua Liu[c], Yandan Su[c], Xuxiang Zhang[a], Xin Nian[a,*]

[a] The First Affiliated Hospital of Kunming Medical University, Department of Endocrinology, Kunming 650032 China
[b] Incinta Fertility Center, 21545 Hawthorne Boulevard Pavilion B, Torrance, CA 90503 USA
[c] The First Affiliated Hospital of Kunming Medical University, Clinical laboratory, Kunming 650032 China

*Corresponding author, e-mail: nianxinkm@hotmail.com

Received 31 Jul 2022, Accepted 22 Dec 2022
Available online 15 May 2023

ABSTRACT: Present studies on ethnicity and intestinal flora show that different ethnic factors, such as diet, geography, and genetics, can make remarkable differences in the structure of intestinal microbiota. To the best of our knowledge, in this study, the differences in intestinal flora of healthy Dai and Han Chinese in the Yunnan region of China were determined for the first time by high-throughput sequencing, and the correlation between polymorphisms of *FOXC2 C-512T* and the structure of intestinal flora was analyzed. The study found that there were significant differences in α and β diversity in intestinal flora between healthy Dai and Han people, and the two ethnic groups had dominant microbiota, through analysis at gate and genus levels. Moreover, our study shows that the intestinal flora structure of *FOXC2 C-512T* locus *CT* is different from that of the other two genotypes in terms of α and β diversity and that *FOXC2 C-512T* polymorphism can affect the composition of human intestinal flora. This study mainly fills the gap in the field of ethnic-gut microbiome research of Dai people and reveals the effect of *FOXC2 C-512T* polymorphism on the structure of human intestinal flora.

KEYWORDS: race, intestinal flora, Dai Chinese, Han Chinese, *FOXC2* genes

INTRODUCTION

With the continuous progression of intestinal microbiota research, the association between race and intestinal flora has garnered increasing attention in recent years [1]. A cross-sectional study in Malaysia [2] showed that 16S amplification sequence analysis of stool samples from 214 community members in Malaysia (46 Malays, 65 Chinese, 49 Indians, and 54 Yakun) showed significant differences in the gut microflora of different ethnic groups ($p = 0.001$). In another wide-scaled study in the Netherlands [3], stool samples were collected from 2,084 people of six ethnic groups, namely Dutch, Ghanaian, Moroccan, African Surinamese, South Asian Surinamese, and Turks. The thorough analysis of α diversity showed that the gut flora is significantly diverse among different ethnic groups ($p < 2.2e-16$). Therefore, they concluded that the ethnicity of individuals should be considered in microbiome studies and their potential future applications should be studied in ethnically diverse societies.

China is a multi-ethnic country with 56 ethnic groups. The geographical, dietary preferences and culture among ethnic groups are distinct and varied, which is advantageous for studying the relationship between different ethnicities and intestinal flora. Liao et al [4] studied the healthy populations of Yao, Zhuang, and Han Chinese groups and reported significant structural differences in the fecal microflora of the three ethnic groups. This indicated that ethnicity significantly affects the gut microflora. Furthermore, Lin et al [5] found that there were significant differences in the intestinal flora of different ethnic groups of healthy populations in 7 different cities in China (Bai, Hui, Korean, Miao, Mongolian, Tibetan, and Uyghur). The study also showed that compared with ethnicity, geographical factors have a greater effect on the composition and diversity of intestinal microflora. The Dai people are a unique ethnic group in Southeast Asia and South Asia, and they are widely distributed in India, Vietnam, Indonesia, Thailand, Yunnan in China, and other countries. The Dai people have a global population of 68 million. Their population is about 1.26 million in China. The Dai ethnic group has a large and widely distributed population. Even though there is currently research worldwide on the effects and correlation between ethnicity and intestinal flora, there is no literature report on the intestinal flora of the Dai ethnic group in China.

The human fork-headed transcription factor (FOXC2) is a key regulator of fat cell metabolism, skeletal tissue development, lymphangiogenesis, and lung maturation. It also plays a major role in many gene regulatory pathways [6]. FOXC2 plays a multifaceted role in cancer. It promotes cell proliferation, epithelial-mesenchymal transformation, angiogenesis, and lymphangiogenesis [7]. The gut microflora is also closely related to the development

of cancer [8]. González-Loyola et al [9] found that the loss-of-function of *FOXC2* disrupts the intestinal epithelial barrier by changing the structure of intestinal flora and interfering with lymphatic function, thereby exerting an effect on various diseases such as inflammatory bowel disease, cancer, and lymphedema. Our previous research has confirmed that *FOXC2* is strongly associated with glycolipid metabolism [10, 11]. The polymorphism of *C-512T* in the 5′ untranslated region of *FOXC2* is significantly associated with insulin resistance (HOMA-IR) [12]. The intestinal flora is also a key factor affecting glycolipid metabolism and HOMA-IR [13]. Therefore, the polymorphism of *FOXC2 C-512T* can be closely associated with the composition and structure of gut microbiota.

In this study, the differences in intestinal microflora of healthy Dai and Han Chinese were studied for the first time by collecting fecal specimens from healthy Dai and Han populations in Yunnan. We did not consider geographical factors. We then simultaneously investigated the correlation between polymorphisms of *FOXC2 C-512T* and the intestinal microflora by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and 16s rRNA in these two ethnic groups.

MATERIALS AND METHODS

Participants

We recruited healthy subjects from the Dai and Han ethnic groups who underwent assessment at the physical examination center of the First Affiliated Hospital of Kunming Medical University. The subjects were asked to sign an informed consent form if they met the study inclusion indicators. The study was conducted in accordance with the Declaration of Helsinki, which was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University (Approval No. 2022L79).

The following were the study inclusion indicators: Han and Dai volunteers whose 3 generations of direct blood relatives belonged to the same ethnic group as that of the volunteers; Han and Dai volunteers who followed the customs and habits that were deemed unique to their specific ethnic groups.

The following were the study exclusion indicators: suffering from different acute and chronic diseases, inflammation and digestive system diseases; consumption of probiotics, antibiotics, and medications that affected their gut microbiota in the 3 months before sample collection, long-term smoking and alcohol abuse, breastfeeding or pregnant women; when more than one candidates belonged to the same family, one subject was randomly selected and the remaining were excluded from the study.

A total of 75 healthy subjects, which consisted of 27 Dai and 48 Han subjects, were included in the study. For all participants of the study, 1 ml of the whole blood was drawn in an anticoagulant tube containing EDTA and stored in a refrigerator at −80 °C. The feces of fasting subjects were collected in the early morning time to facilitate sterile fecal collection and stored in a −80 °C refrigerator within 1 h of collection.

Polymorphism analysis of *C-512T*

Whole blood DNA extraction was conducted by using the Proteinase K method [14]. The main reagents included a centrifugal column genomic NDA extraction kit, 2×Taq PCR Mastermix, DNA Marker (Beijing Tiangen Biochemical Co., Ltd., China); PCR amplified upstream primers: 5-′GCCGACGGATTCCTGCGCTC-3′, downstream primers: 5′-CGCTCCTCGCTGGCTCCA-3′ (Shanghai Jierui Biological Engineering Co., Ltd., China); agarose (Biowest Spain); sterile deionized water (Beijing Lanjieke Biotechnology Co., Ltd., China).

The PCR-RFLP method was conducted in accordance with the method proposed by Nian et al [12], albeit with some minor changes. The PCR amplification system (50 µl) was composed of a 2 µl DNA template, 1 µl upstream primer (20 mM), 1 µl downstream primer (20 mM), 32 µl sterile deionized water, and 14 µl of 2×Taq PCR Mastermix. After centrifugation for 2 s, the solution was placed in the PCR amplification instrument at the following setting: 95 °C predenaturation for 5 min, 95 °C denaturation for 30 s, 63 °C annealing for 30 s, 72 °C extensions for 30 s for 35 cycles, and 72 °C extensions for 10 min.

PCR product analysis: A 2.5% agarose gel was prepared and mixed with 5 µl of the PCR product and 1 µl of 6× loading buffer (Beijing Tiangen Biochemical Co., Ltd.). This mixture was added to the gel-dosing wells, with 6 µl of 500 bp Marker (Shanghai Bioengineering Co., Ltd., China) as the molecular weight standard. After electrophoresis, the presence of the 378 bp fragment was observed under the ultraviolet image analysis system to determine whether the experiment was successful. The successful samples were then sent to Shanghai Bioengineering Co., Ltd., for polymorphism sequencing of the *FOXC2* gene *C-512T*.

16S rRNA sequencing of stool samples

Extraction of fecal DNA

DNA was extracted from the feces by using the SDS method [15]; agarose gel electrophoresis was performed to detect DNA purity and concentration, and an appropriate amount of the sample was taken in a centrifuge tube and diluted to 1 ng/µl with sterile water.

PCR amplification

DNA (1 ng/µl) was diluted and used as the template. Depending on the selection of the sequencing area, specific primers with barcodes were used (Shanghai

Bioengineering Co., Ltd.). PCR was performed with highly efficient and high-fidelity enzymes to ensure amplification efficiency and accuracy. The PCR reaction system consisted of 15 μl of High-Fidelity PCR Master Mix (New England Biolabs, USA), 0.2 μM of forward and reverse primers (Shanghai Bioengineering Co., Ltd.), and 10 ng template DNA for a total of 30 μl. After centrifugation for 2 s, the solution was placed in a PCR amplification instrument at 98 °C for predenaturation for 1 min, 98 °C denaturation for 10 s, 50 °C annealing for 30 s, 72 °C extensions for 30 s for 30 cycles, and 72 °C extensions for 5 min.

PCR product purification

PCR products were detected by electrophoresis using agarose gel at a 2% concentration; an equal concentration was mixed according to the PCR product concentration, and 2% agarose gel electrophoresis was performed to detect the PCR products after mixing thoroughly.

Library building and hands-on sequencing

After building the library, library quantification and quality control were performed with Qubit before undertaking next-generation sequencing with the Illumina platform. The sequencing model used was Novaseq6000, the sequencing depth was 4w raw reads, and the sequencing strategy was pe250. The resulting raw image data file was converted into the Sequenced Reads by Base Calling analysis, and the data was stored in the FASTQ (fq for short) file format.

Statistical analysis

The measurement data of the population characteristics of this study conformed to the normal distribution and were expressed as the mean ± standard deviation. The inter-group comparison was based on the independent sample t-test. Counting data were expressed as the number of examples/%, while the inter-group comparison was tested by χ^2. SPSS 19.0 software was used for double-sided data processing. $p < 0.05$ was considered to indicate statistical significance.

Bioinformatic analysis

We used the DADA2 method [16] to perform mass filtering, denoising, stitching, and de-chiming of the original sequencing sequence. As per the DADA2 method, we performed deduplication operations on the valid data to obtain the deduplication ASV, applied the diversity core-metrics-phylogenetic command in the QIIME2 software for standardization, and set the standardized data depth to 95% of the minimum sample sequence quantity. Finally, we obtained the amplicon needed for the sequence variants (ASV).

We also used the R ggplot2 package to perform α-diversity analysis of ASVs and created a box diagram [17]. The indications for assessing diversity included the Chao1 index, Ace index, Shannon index, and Simpson index. The Chao1 index and Ace index measured the species abundance, with larger values indicating greater numbers of species. The Shannon index and Simpson index were used to measure species diversity, where larger values indicated more varieties of species.

β-diversity analysis was based on unifrac distance using vegan packages in the RStudio software [18]. In the β-diversity analysis, principal coordinates analysis (PCoA) extracted the most important elements and structures from multidimensional data through a series of eigenvalues and eigenvector sorting. The positions of the extracted elements and structures were then labeled on the three-dimensional coordinate space. The idea of dimensionality reduction was used to project these positions onto a two-dimensional diagram for better understanding. We located the most dominant coordinates in the distance matrix to observe differences between individuals or groups. The closer the sample distances in the figure, the greater the similarity in the species composition structure; hence, samples with high similarity in the community structure tended to clump together while those with large community differences were located far apart.

Anosim analysis, also known as similarity analysis, is a nonparametric test used to evaluate whether the difference between groups is significantly greater than the difference within a group so as to determine whether the grouping is meaningful. This analysis also tests for any significant difference in the β diversity between samples in different groups. When the R-value is near 1, the difference between the groups is greater than the difference within a group, and when the R-value is near 0, there is no evident difference within a group as opposed to between two groups.

LEfSe is the latest linear discriminant analysis (LDA) effect size-based method that essentially combines nonparametric statistical tests (Kruskal-Wallis multigroup tests with Wilcoxon rank-sum tests) with LDA to screen for the key biomarkers (such as the key functional taxa/species). This approach allows researchers to identify features of different abundances and associated classes, and it is not limited to the analysis of differences between different sample groups. LEfSe can also deeply analyze different subgroups to pick the key functional groups or species that are consistent throughout different subgroups. Accordingly, LEfSe has been widely applied in microbial amplicon analysis and metagenome analysis, making it particularly suitable for determining biomarkers in medical research [19].

Table 1 Comparison of general information between Han and Dai volunteers.

	Han ($n = 48$)	Dai ($n = 27$)	p
Gender (male/female)	9/39	3/24	0.386
Age	51.88 ± 9.78	53.26 ± 9.67	0.545
BMI	25.07 ± 2.76	24.12 ± 3.44	0.195

Gender differences were tested by the chi-square test, and age differences were assessed using the t-test.

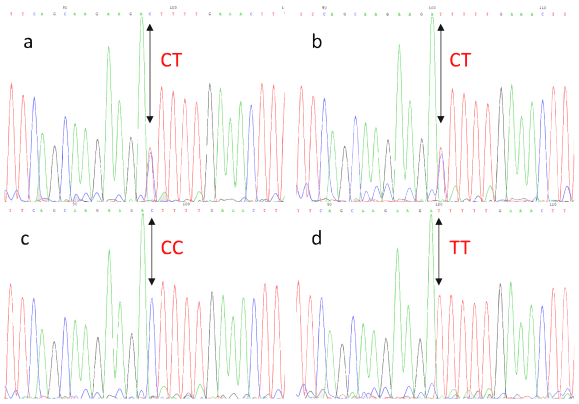

Fig. 1 (a) and (b), sequencing site maps of heterozygous mutant types (CT); (c) sequencing site maps of wild type (CC); (d), sequencing site maps of homozygous mutant types (TT).

RESULTS

Clinical features of the participants

The study included 75 Chinese subjects: 27 Dai and 48 Han. There were no significant differences in sex, age, or BMI between the Dai and Han subjects (Table 1).

Differences in the polymorphisms of C-512T between Dai and Han Chinese

Three genotypes (Fig. 1) consisting of a wild type (CC), a heterozygous mutant type (CT), and a homozygous mutant (CT) were detected at the C-512T locus of *FOXC2*. There were no significant differences in the CC, CT, and TT genotype distribution and C and T allele frequencies in the sample population consisting of 27 Dai and 48 Han subjects (Table 2). The genotype frequency distribution fulfilled the Hardy-Weinberg balance test (Han: $p = 0.683$; Dai: $p = 0.106$).

Differences in the intestinal flora between Dai and Han Chinese

Sequencing quality and analysis clustering results

We measured the average raw data of 115,938 reads for each sample, 114,924 reads after filtering, 111,216 reads after denoising, and 86,522 reads obtained by decimalization after splicing, with an efficiency of 74.63%. Deduplication of the resulting valid data yielded a deduplication ASV, and the normalized data depth was set at 95% of the minimum sample sequence amount. The normalized sample was sequenced at a depth of 36206 and many ASVs to 11081.

By randomly sampling data from a sample of a set number of sequencing data, counting the number of species they represent (the number of ASVs), the number of sequencing data extracted and the corresponding amount of species was used to construct a sparse curve (Fig. 2a). We found that the curve of the vast majority of samples tended to be flat, which indicated that the number of sequencing data was gradually reasonable, and more data would only produce a small number of new species. Moreover, we also plotted the aroma index curve for revalidation (Fig. 2b) and found that the aroma index curve also tended to be flat, which indicated that the amount of sequencing data was large enough that the sequencing results would not be greatly affected by the re-increase of the sample. However, this indicated that the sample already covered the vast majority of microbial species information.

Analysis of α diversity of intestinal flora

The analysis of the α diversity of Dai and Han Chinese showed that the Chao1 index (Fig. 2c) and Ace index (Fig. 2d) of the Dai subjects were significantly higher than the Chao1 index ($p < 0.001$) and Ace index of the Han subjects ($p < 0.001$), which indicated that the abundance of intestinal flora in the Dai population is significantly higher than that in the Han population. No significant difference was found between the Dai and Han populations ($p = 1$; $p = 0.7405$) in the Shannon index (Fig. 2e) and the Simpson index (Fig. 2f). This indicated that there was no significant difference in the diversity of intestinal flora between the Dai and Han populations.

Analysis of β diversity of intestinal flora

The Principal Co-ordinate Analysis Method (PCoA) computed using Weighted UniFrac and Unweighted UniFrac (Fig. 3a,b) showed that the β diversity of the Dai and Han populations were significantly different. The two axes of PCoA computed by Weighted UniFrac and Unweighted UniFrac were by performing the Kruskal-Wallis test (Fig. 3c-f). We found that in the PCoA calculated by Unweighted UniFrac, there were significant differences between Han and Dai populations on two axes ($p < 0.001$). In the PCoA calculated by Weighted UniFrac, the two groups on the PCoA1 axis were significant ($p < 0.01$), whereas no significant difference was found between the two groups on the PCoA2 axis ($p > 0.05$). We then performed an Anosim analysis to determine whether there was a significant difference in the β diversity of different grouped samples. The results showed that both Weighted

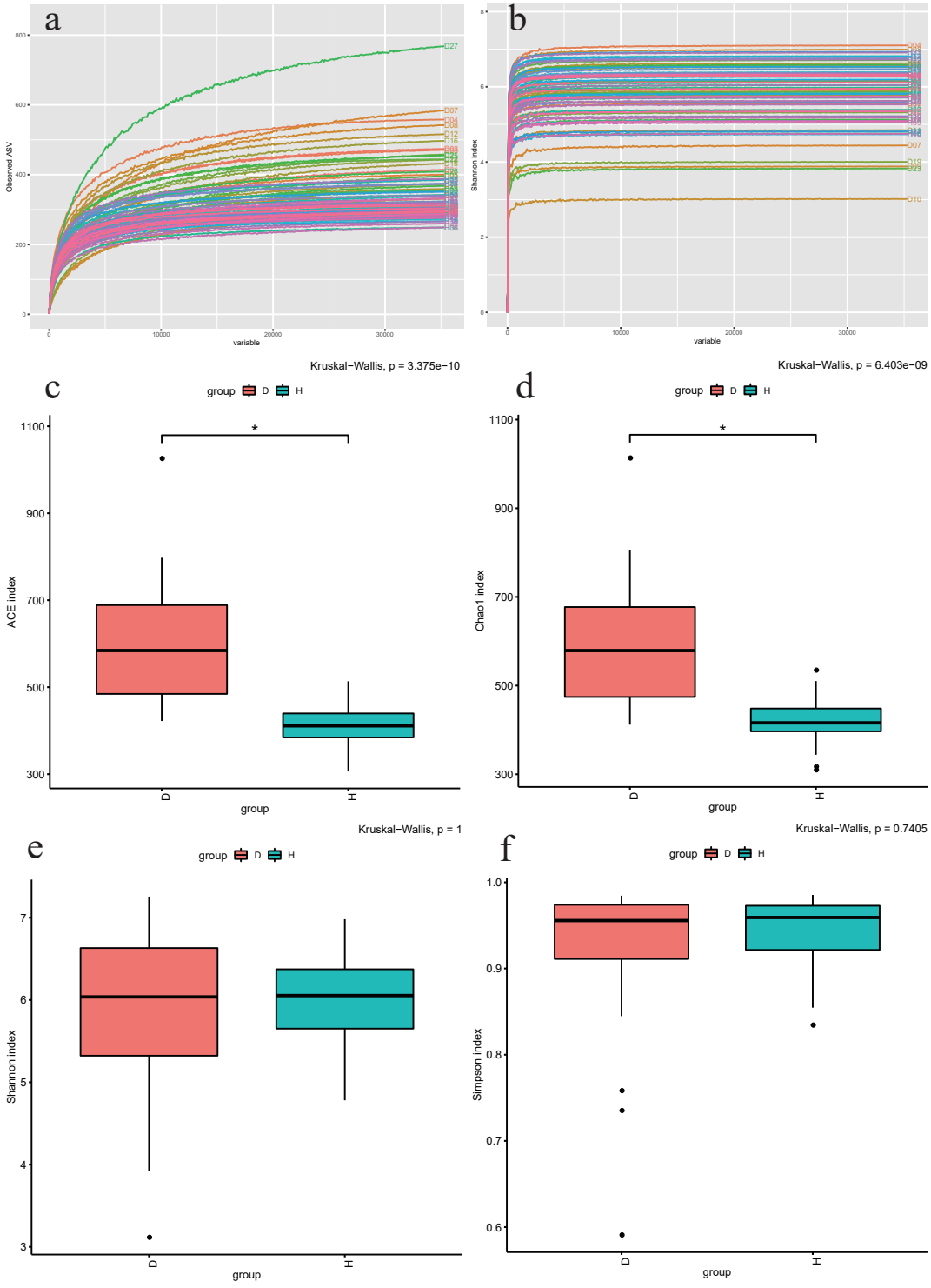

Fig. 2 (a), a sparse curve, which directly reflects the rationality of the amount of sequencing data, and indirectly reflects the richness of species in the sample, when the curve tends to be flat, indicating that the amount of sequencing data is gradually reasonable, and more data will only produce a small number of new species; (b), the Shannon index curve, similar to the sparse curve, also reflects sample diversity; (c) and (d), box plots of the Chao1 index ($p = 6.403e-9$) and the ACE index ($p = 3.375e-10$), respectively, which measure the abundance (the number) of species; (e) and (f), box plots of the Shannon index ($p = 1$) and the Simpson index ($p = 0.7405$), respectively, which measure species diversity.

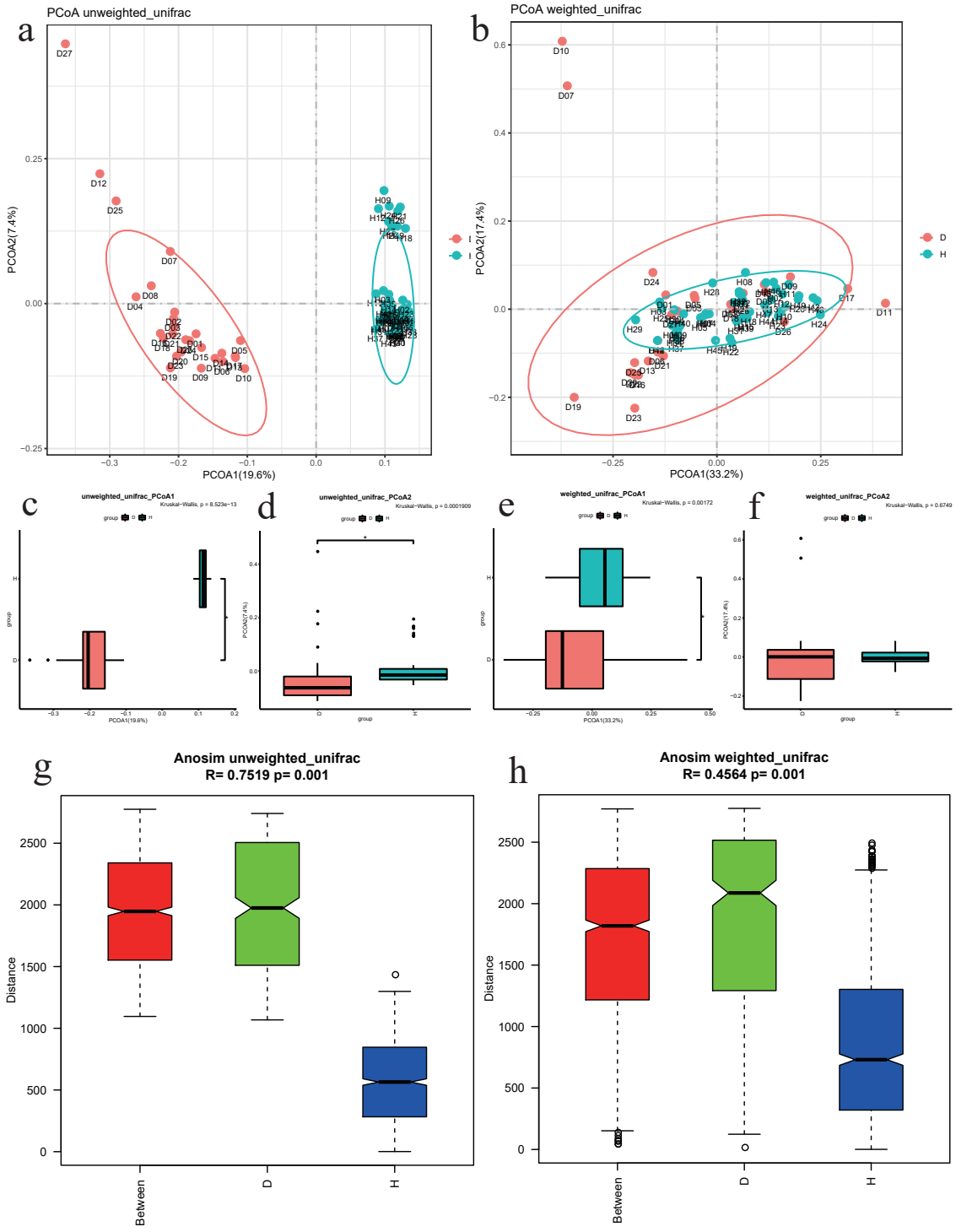

Fig. 3 (a), a PCoA plot calculated by Unweighted Unifrac; (b), a PCoA plot calculated by Weighted Unifrac, the closer the samples are, the more similar the species composition, so samples with high community structure similarity tend to congregate together, while samples with large community differences are far apart; (c), a box diagram ($p = 8.523e-13$) made by calculating the difference between two sets of PCoA1 axes using the Kruskal-Wallis test; (d), a box diagram of the PCoA2 axis ($p = 0.0001909$); (e), a box diagram ($p = 0.00172$) made by calculating the difference between two sets of PCoA1 axes in the (b) using the Kruskal-Wallis test; (f), the box-type diagram of the PCoA2 axis ($p = 0.6749$); (g), the Anosim analysis box plot calculated by Unweighted Unifrac ($p = 0.001$); (h), the Anosim analysis box plot calculated by Weighted Unifrac ($p = 0.001$).

Table 2 Comparison of genotypes and alleles of *C-512T* polymorphism.

Ethnic	Number	Genotype			*p*	Allele		*p*
		CC (%)	CT (%)	TT (%)		C (%)	T (%)	
Han	48	11(22.9)	21(43.8)	16(33.3)	0.079	43(44.8)	53(55.2)	0.539
Dai	27	4(14.8)	19(70.4)	4(14.8)		27(50.0)	27(50.0)	

Differences in Genotype and Allele were tested by the chi-square test.

UniFrac and Unweighted UniFrac calculation methods confirmed a significant difference in the intestinal flora between Dai and Han ($p = 0.001$) (Fig. 3g,h). This indicated that the healthy Han and Dai populations have little similarity in the composition and structure of the intestinal microorganisms.

Differences in the intestinal flora of the Dai and Han Chinese populations

We compared the ASV sequence with the microbial reference database Silva (www.arb-silva.de) to obtain the corresponding species classification information of each ASV, then determined the composition of each sample community at the gate and genus level to generate a species abundance table at different taxonomic levels. We then used the R tool to draw a community structure map (Fig. 4a,b). By comparing these results to the database Silva, species annotations were made and statistics on different taxonomic levels were calculated. Of the 11081 ASVs, the number of ASVs that can be annotated to the database was 99.68%, the proportion of gate levels was 98.51%, and the proportion of genera levels was 75.38%. At the gate level, the dominant species were mainly Firmicutes, Bacteroidota, and Proteobacteria (Fig. 4c), and the dominant species at the genus level were *Bacteroides*, *Faecalibacterium*, and *Pseudomonas* (Fig. 4d).

Furthermore, we performed a LefSe analysis based on species abundance. At the gate level, the abundance of Proteobacteria and Actinobacteriota in the Han population was significantly greater than that of the Dai population, whereas Fusobacteriota was more dominant in the Dai population. At the genus level, the dominant bacteria in the Han population were *Pseudognonas*, *Bacteroides*, *Blautia*, and *Bifidobacterium*, and the dominant bacteria in the Dai population were *Closedtridium* and *Prevotella* (Fig. 4e,f).

Correlation between *C-512T* polymorphisms and intestinal flora

We categorized polymorphisms per the *C-512T* genotype sequencing results into CC, TT, and CT groups and performed a high-throughput sequencing analysis. The α diversity analysis indicated that indices, including Chao1, Shannon, and Simpson, overall showed no significant differences between the three groups (Fig. 5a,c,d); however, the intestinal flora abundance in the CT group was significantly higher than that in the TT group per the ACE index (Fig. 5b). During the β diversity analysis, the PCoA calculated using Unweighted UniFrac showed differences between the intestinal flora of the three groups (Fig. 5e). We calculated two coordinate surfaces of the PCoA separately by the Kruskal-Wallis test and found significant differences between the CT and TT groups on the PCoA1 axis ($p < 0.01$) (Fig. 5g) and between the CT, CC, and TT groups on the PCoA2 axis ($p < 0.05$) (Fig. 5h). The PCoA calculated by Weighted UniFrac showed no significant differences between the three groups (Fig. 5f,i,j). Moreover, the Anosim analysis showed that the *R*-value was close to 0, indicating no statistical differences between and within the three groups of *FOXC2* and suggesting that racial factors had no significant effect on the *C-512T* polymorphism grouping of *FOXC2* (Fig. 5k,l). Furthermore, the LefSe analysis did not show any different significant microflora between the three groups (Fig. 5m).

DISCUSSION

As research on intestinal flora progresses, the intestinal microbiome has become a "bridge of connection" between various biological systems. It is a key factor in neural [20], gastrointestinal [21], cardiovascular [22], respiration system [23], endocrine metabolism [13], and has been confirmed to be closely related to the nervous system through the gut-brain microbiota axis [24]. In the two review articles, they reported that the intestinal microbiota plays a vital role in human health and diseases, and it will become the core of the medical field in the future [25, 26]. Factors affecting the intestinal microbiota have already become a popular focus of the current research.

We analyzed the intestinal flora of the healthy populations of the Dai ethnic group and the Han nationality in Yunnan, China, to verify the effect of ethnic genes on the composition and structure of intestinal flora by excluding factors such as drugs, geography, and diseases as much as possible. The present study showed that due to dietary factors, significant differences were observed between the intestinal microflora of the Dai and Han populations. At the gate level, Actinobacteriota and Proteobacteria were significantly higher in the Han subjects than in the Dai subjects, and the Fusobacterium phylum was more dominant in the Dai subjects than in the Han subjects. Although

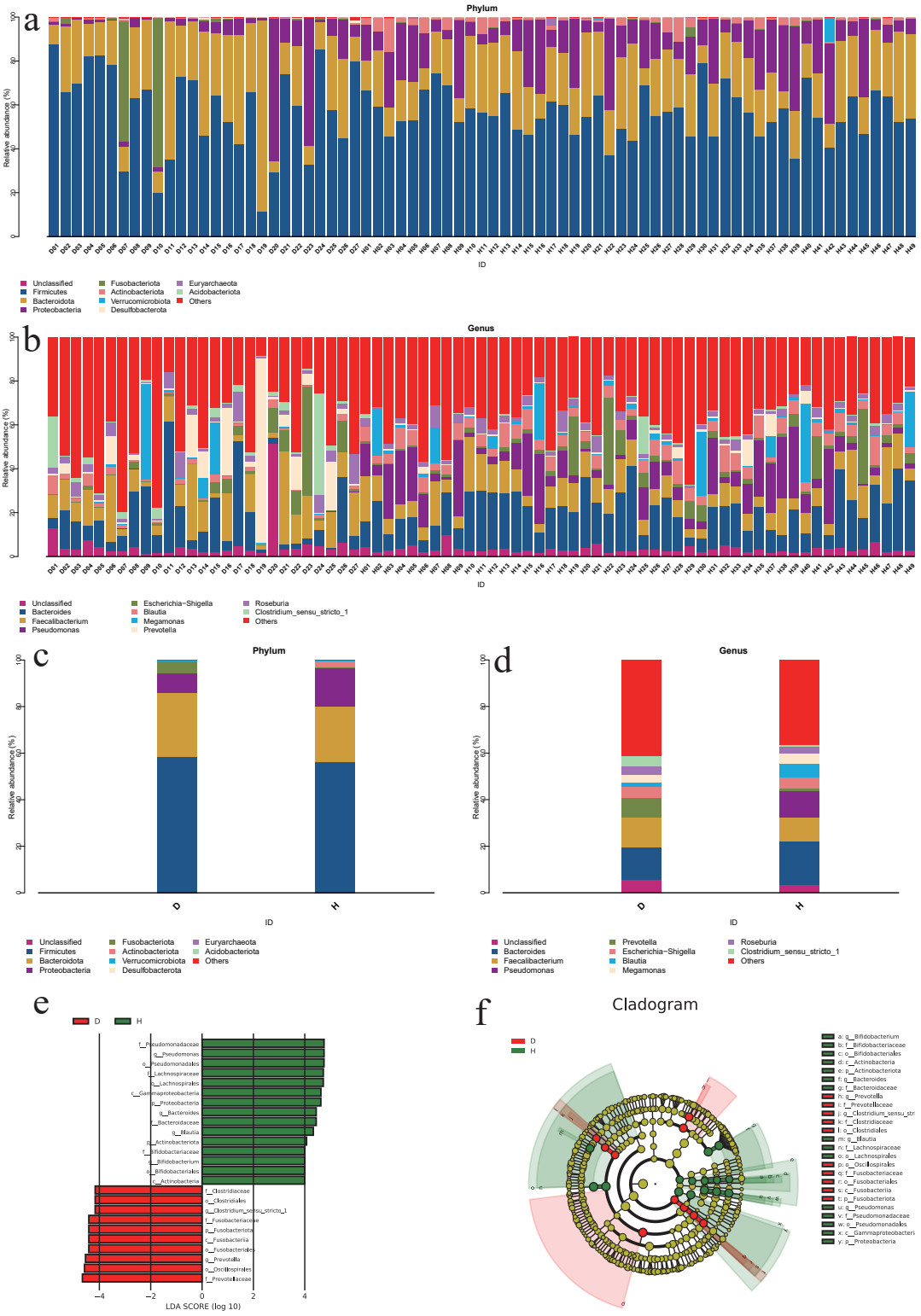

Fig. 4 (a), the structure of the intestinal flora of individual samples at the gate level; (b), the structure of the intestinal flora of each sample at the genus level; (c) the structure of the intestinal flora between the healthy population of the Dai ethnic group and the healthy group of Han people at the gate level; (d), the structure of the intestinal flora between healthy Dai populations and Han healthy groups at the genera level; (e), a histogram of LDA value distribution for LefSe analysis; (f), a branching plot of LEfSe analysis evolution.

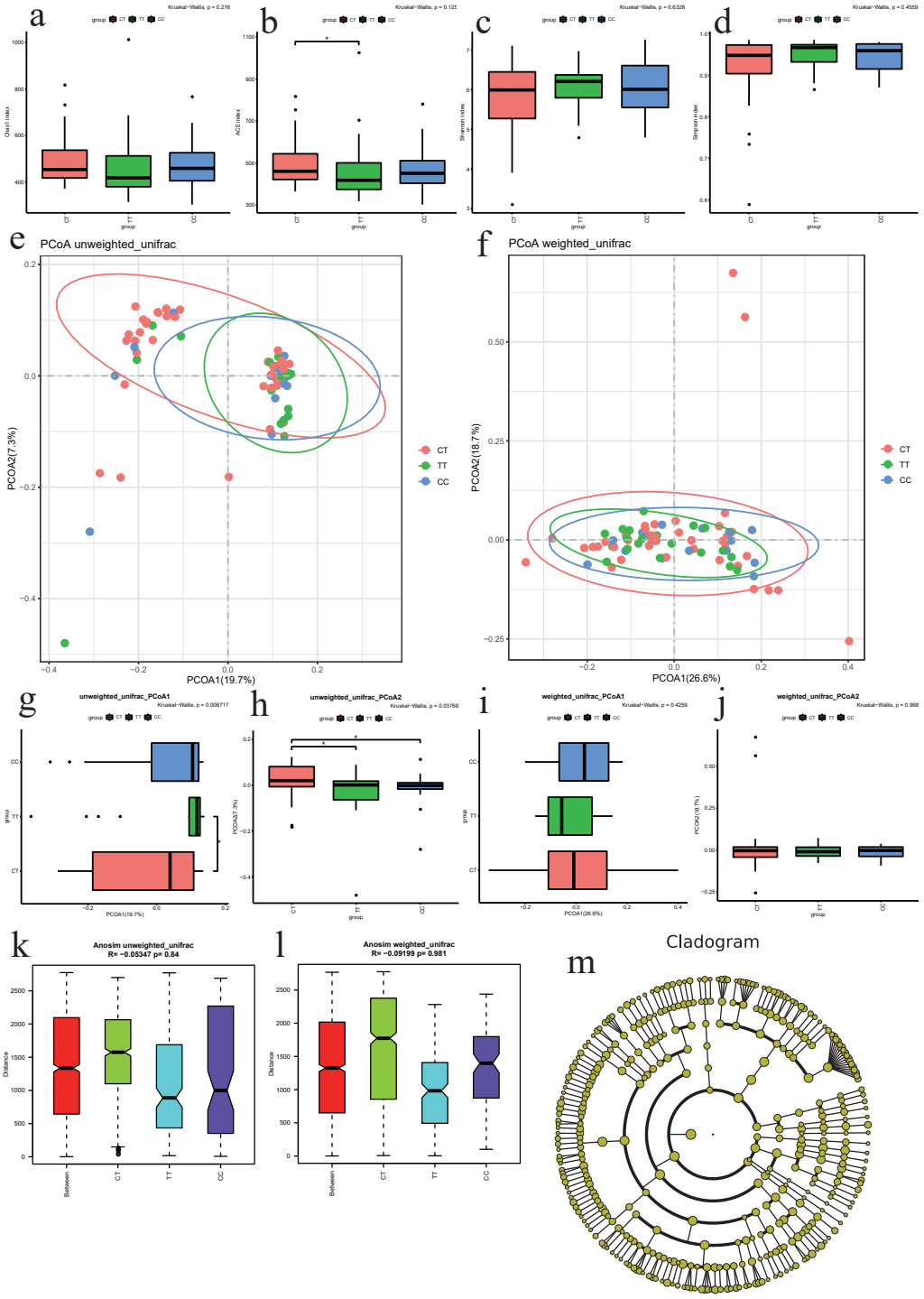

Fig. 5 α diversity analysis between the three genotypes of the *FOXC2* gene *C-512T* in (a), (b), (c) and (d) were box plots of the Chao1 index ($p = 0.216$), the ACE index ($p = 0.1257$), the Shannon index ($p = 0.6328$) and the Simpson index ($p = 0.4559$); (e), the PCoA graph calculated by Unweighted Unifrac; (f), the PCoA graph calculated by Weighted Unifrac; (g), a box-type diagram ($p = 0.008717$) made by calculating the difference between the three groups of genotypes in the PCoA1 axis of the (e) by the Kruskal-Wallis test; (h), the box-type diagram of the PCoA2 axis ($p = 0.03766$); (i), a box-type diagram ($p = 0.4259$) made by calculating the differences between three sets of PCoA1 axes using the Kruskal-Wallis test; (j), a box-type diagram ($p = 0.968$) of the PCoA2 axis; (k), the Anosim analysis box-type diagram calculated by Unweighted Unifrac ($p = 0.84$); (l), the Anosim analysis box-type diagram calculated by weighted Unifrac ($p = 0.981$); (m), a branching diagram of the evolution of LEfSe analysis.

no significant differences were observed between Firmicutes and Bacteroidota abundances in the Dai and Han healthy populations, at the genus level *Blautia* in Firmicutes and *Bacteroides* in Bacteroidota were more dominant in the Han subjects, whereas *Clostridium* in Firmicutes and *Prevotella* in Bacteroidota were more dominant in the Dai subjects. In addition, *Bifidobacterium* in Actinobacteriota and *Pseudomonas* in Proteobacteria were statistically higher in the Han subjects. Interestingly, *Fusobacteriota*, which was dominant in the Dai subjects, was identified as dominant bacteria at phylum, class, order, and family levels; however, it was not dominant at the genus level in the Dai subjects. One possibility is that *Fusobacteriota* is a relatively unknown bacterium; hence, no annotation was available for it in the database used in this study.

Regarding diets, the biggest difference between Han and Dai groups is that Han people prefer a carbohydrate and meat diet, whereas members of the Dai ethnic group prefer a vegetarian diet. *Blautia* and *Bifidobacterium*, short-chain fatty acid-producing bacteria, are significantly correlated with carbohydrate diets [27, 28], which may be one of the reasons for the higher abundance of *Blautia* and *Bifidobacterium* in the Han subjects. In addition, Lee et al [29] reported that the vegetarian diet culture of Southeast Asian countries, including Indonesia and Thailand, led to a low abundance of *Bacteroides* in their gut microbiota compared with those in North American countries. An animal protein diet increases the abundance of *Bacteroides* [30], and the relative abundance of *Prevotella* in Bacteroidota, a de-westernized genus, is inversely correlated with the relative abundance of *Bacteroides* [31, 32]. It is also closely related to carbohydrate and vegetarian diets [33]. Based on the data analyzed in the present study, we confirmed significant differences in intestinal flora between different ethnic groups. Hence, distinctive food cultures are key factors affecting intestinal microflora.

Interestingly, studies have confirmed that *Pseudomonas* abundance is strongly associated with a plant-based diet [34], and it reduces *Clostridium* abundance [35]. We found that *Pseudomonas* was significantly more dominant in the healthy Han populations than in the healthy Dai populations, whereas *Clostridium* was significantly more dominant in the Dai subjects than in the Han subjects, which was not consistent with previous studies. Our further analysis suggested that interactions between bacterial strains in the intestinal flora may alter the effects of a diet on the microbiota. Luo et al [36] found that *Bifidobacterium* could affect the abundance of 38 bacterial genera in the intestine, of which 23 were positively correlated with *Bifidobacterium*, whereas 15 were negatively correlated with it. The positively correlated bacteria included *Blautia*, whereas the negatively correlated bacteria included *Clostridium* and *Prevotella*. In the present study, *Bifidobacterium* abundance in the healthy Han population was significantly higher than that in the healthy Dai population, which was likely to promote *Blautia* growth and inhibit *Clostridium* and *Prevotella*. Hence, we infer that communication and interactions between different species within intestinal flora possibly affect the structure of intestinal flora.

The human microbiome project data showed that host genetics may affect the α diversity of the gut microbiota and confirmed the association between *LCT* and *Bifidobacterium* abundance [37]. Another study [38] recruited 416 pairs of twins (171 pairs of fraternal twins and 245 pairs of identical twins) and showed that the diversity and abundance of the intestinal flora of the identical twins were more similar than those of the fraternal twins. This indicates that genes significantly affect the gut microbiota. Wang et al [39] performed the genome-wide association analysis of 82 subjects from Germany to determine changes in the intestinal microbiota affected by several factors, including vitamin D receptors; the results showed that the effect of host genes on the intestinal microbiota accounted for about 10%, providing a theoretical basis for the study of genes related to the intestinal microbiota.

Nakayama et al [40] performed the high-throughput sequencing of feces samples collected from Asian countries and regions, including China, Thailand, Japan, Indonesia, and Taiwan, and found that the intestinal flora of people in China, Japan, and Taiwan was dominated by *Bacteroides* and *Bifidobacterium*, whereas that of Thailand and Indonesia was dominated by *Prevotella*. Lee et al [29] also found relatively low abundance of *Bacteroides* in Indonesian and Thai populations. The results of these two studies [29, 40] of the Thai population are consistent with the results of the healthy Dai population in the present study, confirming the effect of genes on the composition of the gut microbiota in the absence of slight differences between various factors, including diets and cultures.

In this study, we confirmed the correlation between the *FOXC2 C-512T* polymorphism and the intestinal microbiota. The study involved the intestinal flora of both ethnic groups; however, as the Dai CC and TT groups consisted of only 4 people, the sample size was small. We analyzed the Dai and Han populations together considering that the genotype frequency distribution of the Han and Dai populations was consistent with the Hardy-Weinberg balance test and that no significant differences existed in the genotype distribution and allele frequencies of the Dai and Han populations.

The α diversity analysis showed that the abundance of intestinal flora in the CT group was significantly higher than that in the TT group, whereas no significant difference was observed between the

α diversities of the three groups. The β diversity analysis showed that the PCoA calculated by Unweighted UniFrac showed differences between the gut microflora of the three groups; however, the LefSe analysis did not show any different microflora between the three groups.

FOXC2 is the main regulatory gene of human lymphatic production [41], which can directly affect the balance between lymphatic functions and intestinal flora, change the barrier effect of intestinal epithelial cells and the colonization environment of intestinal microorganisms, and affect several diseases, including inflammatory bowel disease and cancer [9]. In addition, it can regulate the inflammatory response of the intestine by affecting lymphatic functions [42], which may be one of the reasons for the effect of *FOXC2* on human metabolic disorders. Although no direct reports on the correlation between *FOXC2* and the intestinal microbiota are currently available, we analyzed whether *FOXC2 C-512T* polymorphisms affect the structure of intestinal flora, especially differences between the CT type and the structure of the intestinal flora of the other two types. However, due to the small sample size, this difference was not noticeable. In addition, the differences in the flora of the Han and Dai populations may affect the microbiota analysis between the groups, resulting in a certain deviation in the LefSe analysis results. Despite this, the outcomes of the present study are sufficient to guide our subsequent research and provide a theoretical basis for future studies.

CONCLUSION

We confirmed that significant differences were observed between the structures of the intestinal flora of the healthy Dai and Han people in Yunnan, China. The findings help provide a basis for understanding the effect of ethnic factors on intestinal flora and filled the existing gap in research on the Dai ethnic minority in China. In addition, we also found that the CT genotype of the *FOXC2 C-512T* locus differed from the other two genotypes regarding α and β diversities, suggesting that the composition and structure of the human intestinal flora may be related to the *FOXC2 C-512T* polymorphism. We will explore the underlying mechanisms in future studies.

Data Availability Statement

Raw data of intestinal microbiology in healthy Dai and Han Chinese populations has been uploaded to the SRA database on the NCBI website (Bioproject number PRJNA861680; SRA number SRP387939). Other datasets presented in this article are available with requests directed to L.L., The First Affiliated Hospital of Kunming Medical University, Department of endocrinology.

Acknowledgements: Thanks to the support of the Physical Examination Center and the Laboratory of the First Affiliated Hospital of Kunming Medical University. This work was supported by grants from Yunnan Provincial Clinical Medical Research Center for Metabolic Diseases (No. 202102AA100056), Joint Research Found of Yunnan Provincial Science and Technology Department-Kunming Medical University (No. 202201AY070001-066), and 2022 Postgraduate Innovation Fund of Kunming Medical University (No. 2022S192).

REFERENCES

1. Brooks AW, Priya S, Blekhman R, Bordenstein SR (2018) Gut microbiota diversity across ethnicities in the United States. *PLoS Biol* **16**, e2006842.
2. Dwiyanto J, Hussain MH, Reidpath D, Ong KS, Qasim A, Lee SWH, Lee SM, Foo SC, et al (2021) Ethnicity influence the gut microbiota of individuals sharing a geographical location: a cross-sectional study from a middle-income country. *Sci Rep* **11**, 2618.
3. Verhaar BJH, Collard D, Prodan A, Levels JHM, Zwinderman AH, Backhed F, Vogt L, Peters MJ, et al (2020) Associations between gut microbiota, faecal short-chain fatty acids, and blood pressure across ethnic groups: the HELIUS study. *Eur Heart J* **41**, 4259–4267.
4. Liao M, Xie Y, Mao Y, Lu Z, Tan A, Wu C, Zhang Z, Chen Y, et al (2018) Comparative analyses of fecal microbiota in Chinese isolated Yao population, minority Zhuang and rural Han by 16S RNA sequencing. *Sci Rep* **8**, 1142.
5. Lin D, Wang R, Luo J, Ren F, Gu Z, Zhao Y, Zhao L (2020) The core and distinction of the gut microbiota in Chinese populations across geography and ethnicity. *Microorganisms* **8**, 1579.
6. Chen X, Wei H, Li J, Liang X, Dai S, Jiang L, Guo M, Qu L, et al (2019) Structural basis for DNA recognition by FOXC2. *Nucleic Acids Res* **47**, 3752–3764.
7. Wang T, Zheng L, Wang Q, Hu YW (2018) Emerging roles and mechanisms of FOXC2 in cancer. *Clin Chim Acta* **479**, 84–93.
8. Baffy G (2020) Gut microbiota and cancer of the host: Colliding interests. *Adv Exp Med Biol* **1219**, 93–107.
9. González-Loyola A, Bovay E, Kim J, Lozano TW, Sabine A, Renevey F, Arroz-Madeira S, Rapin A, et al (2021) FOXC2 controls adult lymphatic endothelial specialization, function, and gut lymphatic barrier preventing multiorgan failure. *Sci Adv* **7**, eabf4335.
10. Zhang X, Wang Y, Zuo F, Su Y, Liu H, Wu B, Nian X (2020) Effect of insulin-regulated FOXC2 expression in adipocyte differentiation and insulin resistance. *Diabetes Metab Syndr Obes* **13**, 2801–2809.
11. Zhang XX, Wang YM, Su YD, Zuo F, Wu B, Nian X (2021) MiR-26a regulated adipogenic differentiation of ADSCs induced by insulin through the CDK5/FOXC2 pathway. *Mol Cell Biochem* **476**, 1705–1716.
12. Nian X, Zhang X, Wang Y, Li H, Li J, Liu H, Qin L (2016) Correlations of FOXC2 gene expression and polymorphism with type 2 Diabetes Mellitus. *Clin Lab* **62**, 78–91.
13. Cao Y, Ren G, Zhang Y, Qin H, An X, Long Y, Chen J, Yang L (2021) A new way for punicalagin to alleviate insulin resistance: regulating gut microbiota and autophagy. *Food Nutr Res* **65**, 5689.
14. Ahmad NN, Cu-Unjieng AB, Donoso LA (1995) Modification of standard proteinase K/phenol method for DNA isolation to improve yield and purity from frozen blood. *J Med Genet* **32**, 129–130.

15. Nahar A, Baker AL, Nichois DS, Bowman JP, Britz ML (2021) Benchmarking DNA extraction methods for phylogenomic analysis of Sub-Antarctic *Rhodococcus* and *Williamsia* species. *Microorganisms* **9**, 1253.
16. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583.
17. Maag JLV (2018) gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2. *F1000Res* **7**, 1576.
18. Kambura AK, Mwirichia RK, Kasili RW, Karanja EN, Makonde HM, Boga HI (2016) Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiol* **16**, 136.
19. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60.
20. Ding W, You Z, Chen Q, Yang L, Doheny J, Zhou X, Li N, Wang S, et al (2021) Gut microbiota influences neuropathic pain through modulating proinflammatory and anti-inflammatory T cells. *Anesth Analg* **132**, 1146–1155.
21. Karakan T, Ozkul C, Akkol EK, Bilici S, Sobarzo-Sanchez E, Capasso R (2021) Gut-brain-microbiota axis: Antibiotics and functional gastrointestinal disorders. *Nutrients* **13**, 389.
22. Witkowski M, Weeks TL, Hazen SL (2020) Gut microbiota and cardiovascular disease. *Circ Res* **127**, 553–570.
23. Chunxi L, Haiyue L, Yanxia L, Jianbing P, Jin S (2020) The gut microbiota and respiratory diseases: New evidence. *J Immunol Res* **2020**, 2340670.
24. Mörkl S, Butler MI, Holl A, Cryan JF, Dinan TG (2020) Probiotics and the microbiota-gut-brain axis: Focus on psychiatry. *Curr Nutr Rep* **9**, 171–182.
25. Gomaa EZ (2020) Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113**, 2019–2040.
26. Sharma S, Chaubey KK, Singh SV, Gupta S (2022) Symbiotic microbiota: A class of potent immunomodulators. *ScienceAsia* **48**, 855–865.
27. Satti M, Modesto M, Endo A, Kawashima T, Mattarelli P, Arita M (2021) Host-diet effect on the metabolism of Bifidobacterium. *Genes (Basel)* **12**, 609.
28. Zhou MS, Zhang B, Gao ZL, Zheng RP, Marcellin DFHM, Saro A, Pan J, Chu L, et al (2021) Altered diversity and composition of gut microbiota in patients with allergic rhinitis. *Microb Pathog* **161**, 105272.
29. Lee YK, Conway P, Pettersson S, Nair GB, Surono I, Egayanti Y, Amarra MS (2017) The gut, its microbes and health: relevance for Asia. *Asia Pac J Clin Nutr* **26**, 957–971.
30. Rinninella E, Cintoni M, Raoul P, Lopetuso LR, Scaldaferri F, Pulcini G, Miggiano GAD, Gasbarrini A (2019) Food components and dietary habits: Keys for a healthy gut microbiota composition. *Nutrients* **11**, 2393.
31. Tett A, Pasolli E, Masetti G, Ercolini D, Segata N (2021) Prevotella diversity, niches and interactions with the human host. *Nat Rev Microbiol* **19**, 585–599.
32. Arumugam M, Raes J, Pelletier E, Paslier DL, Yamada T, Mende DR, Fernandes GR, Tap J, et al (2011) Enterotypes of the human gut microbiome. *Nature* **473**, 174–180.
33. Ghosh TS, Rampelli S, Jeffery IB, Santoro A, Neto M, Capri M, Giampieri E, Jennings A, et al (2020) Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries. *Gut* **69**, 1218–1228.
34. Soto-Giron MJ, Kim JN, Schott E, Tahmin C, Ishoey T, Mincer TJ, Dewalt J, Toledo G (2021) The edible plant microbiome represents a diverse genetic reservoir with functional potential in the human host. *Sci Rep* **11**, 24017.
35. Pagliai G, Russo E, Niccolai E, Dinu M, Pilato VD, Magrini A, Bartolucci G, Baldi S, et al (2020) Influence of a 3-month low-calorie Mediterranean diet compared to the vegetarian diet on human gut microbiota and SCFA: the CARDIVEG Study. *Eur J Nutr* **59**, 2011–2024.
36. Luo J, Li Y, Xie J, Gao L, Liu L, Ou S, Chen L, Peng X (2018) The primary biological network of *Bifidobacterium* in the gut. *FEMS Microbiol Lett* **365**, fny057.
37. Kolde R, Franzosa EA, Rahnavard G, Hall AB, Vlamakis H, Stevens C, Daly MJ, Xavier RJ, et al (2018) Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Med* **10**, 6.
38. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Treuren WV, et al (2014) Human genetics shape the gut microbiome. *Cell* **159**, 789–799.
39. Wang J, Thingholm LB, Skieceviciene J, Rausch P, Kummen M, Hov JR, Degenhard TF, Heinsen FA, et al (2016) Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* **48**, 1396–1406.
40. Nakayama J, Watanabe K, Jiang J, Matsuda K, Chao SH, Haryono P, La-Ongkham O, Sarwoko MA, et al (2015) Diversity in gut bacterial community of school-age children in Asia. *Sci Rep* **5**, 8397.
41. Wu X, Liu NF (2011) FOXC2 transcription factor: a novel regulator of lymphangiogenesis. *Lymphology* **44**, 35–41.
42. Becker F, Potepalov S, Shehzahdi R, Bernas M, Witte M, Abreo F, Traylor J, Orr WA, et al (2015) Downregulation of FoxC2 increased susceptibility to experimental Colitis: Influence of lymphatic drainage function? *Inflamm Bowel Dis* **21**, 1282–1296.