

# SCIENTIFIC REPORTS



OPEN

## Comparative analyses of fecal microbiota in Chinese isolated Yao population, minority Zhuang and rural Han by 16sRNA sequencing

Ming Liao<sup>1,4</sup>, Yuanliang Xie<sup>2,4</sup>, Yan Mao<sup>5</sup>, Zheng Lu<sup>2,4,6</sup>, Aihua Tan<sup>2,4</sup>, Chunlei Wu<sup>7</sup>, Zhifu Zhang<sup>8</sup>, Yang Chen<sup>2,3</sup>, Tianyu Li<sup>2,4,6</sup>, Yu Ye<sup>2,3,9</sup>, Ziting Yao<sup>2,4</sup>, Yonghua Jiang<sup>2,4</sup>, Hongzhe Li<sup>10</sup>, Xiaoming Li<sup>11</sup>, Xiaobo Yang<sup>2,4</sup>, Qiuyan Wang<sup>2,4</sup> & Zengnan Mo<sup>2,4</sup>

The gut microbiome in humans is associated with geography, diet, lifestyles and so on, but its relationship with some isolated populations is not clear. We used the 16sRNA technique to sequence the fecal microbiome in the Chinese isolated Yao population and compared it with the major minority Zhuang and the major ethnic Han populations living in the same rural area. Information about diet frequency and health status and routine serum measurements were collected. The unweighted UniFrac principal coordinates analysis showed significant structural differences in fecal microbiota among the three ethnic groups. Statistically significant differences were observed in the community richness estimator (chaos) and the diversity estimator (Shannon) among the three groups. At the genus level, the fecal samples of the isolated Yao population presented the lowest relative abundance of the *Megamonas* genus, which was potentially related to the high frequency of bean consumption in the diet. Two enterotypes were identified in the overall fecal microbiota in the three populations. In the isolated Yao population, a higher *Bacteroides* abundance was observed, but the *Prevotella* abundance decreased with increased alcohol consumption.

The human gut microbiota is closely associated with human health<sup>1</sup>. Understanding its abundance and diversity relative to different host characteristics is of great importance. The gut microbiota in humans can be affected by factors such as geography, dietary habits, living environment, age, traditions and changing lifestyles<sup>2,3</sup>. However, inconsistent results have been reported, probably due to the large variations in host properties and the habitat-dependent environmental factors across distinct human populations.

The effect of geographic or ethnic differences on microbiota diversity has not been confirmed. As shown in a survey across five northern European countries, the colonic microbiota structures did not present significant differences with respect to geographic origin, age or gender<sup>4</sup>. In another larger survey across six different countries covering Europe, Japan and the USA, the intestinal microbiota variations can be generally stratified to three enterotypes, but the enterotypes were independent of geography or ethnicity<sup>5</sup>. In another survey that included Malawian, Amerindian and USA residents, the phylogenetic composition of fecal microbiota was significantly

<sup>1</sup>Department of Reproductive Center, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530021, China. <sup>2</sup>Center for Genomic and Personalized Medicine, Guangxi Medical University, Nanning, Guangxi, 530021, China. <sup>3</sup>Guangxi Collaborative Innovation Center for Genomic and Personalized Medicine, Nanning, Guangxi, 530021, China. <sup>4</sup>Guangxi Key Laboratory of Genomic and Personalized Medicine, Nanning, Guangxi, 530021, China. <sup>5</sup>Oncology Department, Nanning Second People's Hospital, The third Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530031, China. <sup>6</sup>Institute of Urology and Nephrology, First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530021, China. <sup>7</sup>Department of Urology, First Affiliated Hospital of Xinxiang Medical College, Xinxiang, Henan, 453100, China. <sup>8</sup>Urology Department, Minzu Hospital of Guangxi Zhuang Autonomous Region, Affiliated Minzu Hospital of Guangxi Medical University, Nanning, Guangxi, 530001, China. <sup>9</sup>Department of Emergency Surgery, The first Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530021, China. <sup>10</sup>Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19014, USA. <sup>11</sup>Hezhou University, Hezhou, Guangxi, 542899, China. Correspondence and requests for materials should be addressed to Z.M. (email: [zengnanmo@hotmail.com](mailto:zengnanmo@hotmail.com))

	Han		Zhuang		Yao		P
N	47		28		59		
Male	14	(29.8%)	14	(50%)	25	(42.4%)	0.110
Age (yr)	50.0	(42.5–59.5)	55.5	(48.3–62.5)	44.0	(34.0–51.0)	0.007
BMI	23.6	(21.0–25.7)	22.9	(19.9–23.9)	21.3	(20.0–23.5)	0.030
weight (kg)	58.2	(51.4–66.5)	58.5	(49.5–63.3)	49.1	(45.8–54.8)	<0.001
height (cm)	157	(154–164)	158	(153–164)	152	(147–157)	<0.001
alcohol consumption							0.654
Never	30	(63.8%)	18	(64.3%)	32	(54.2%)	
Former	4	(8.5%)	1	(3.6%)	4	(6.8%)	
Current	12	(25.5%)	7	(25%)	22	(37.3%)	
cigarret smoking							0.461
Never	37	(78.7%)	17	(60.7%)	39	(66.1%)	
Former	2	(4.3%)	1	(3.6%)	0	(0%)	
Current	7	(14.9%)	4	(14.3%)	12	(20.3%)	
physical activity							0.125
> = 4 h/d	18	(38.3%)	13	(46.4%)	16	(27.1%)	
< 4 h/d	25	(53.2%)	12	(42.9%)	39	(66.1%)	
dietary habits							
have regular meals	42	(89.36%)	25	(89.29%)	49	(83.05%)	0.716
have regular breakfast	40	(85.11%)	27	(96.43%)	58	(98.31%)	0.020
drink unboiled water	19	(40.43%)	16	(57.14%)	31	(52.54%)	0.008
wash hands before meals	41	(87.23%)	17	(60.71%)	49	(83.05%)	0.013
wash teeth regularly	45	(95.74%)	28	(100%)	52	(88.14%)	0.110
prevalent taste							
salty	42	(89.36%)	24	(85.71%)	45	(76.27%)	0.233
spicy	42	(89.36%)	27	(96.43%)	54	(91.53%)	0.600
sugary	42	(89.36%)	24	(85.71%)	56	(94.92%)	0.353

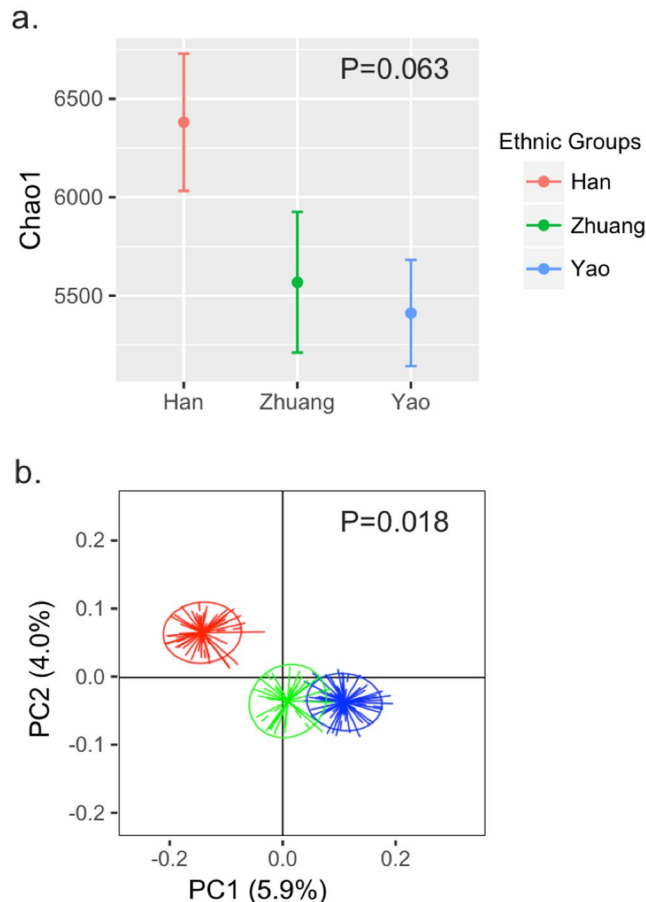
**Table 1.** Characteristics of the included samples\*. \*Continual variables were present as median values (25–75th percentile) and p was calculated by Kruskal-Wallis rank-sum test, whereas categorical variables were present as frequencies and proportions, and p was calculated by Fisher's exact test.

different between the individuals in different geographies, probably due to dietary differences. Similar results were observed in a Chinese survey that included 314 healthy young adults from seven ethnic origins<sup>6</sup>. A phylogenetically diverse set of gut microbiota may be common among distinct healthy populations, but phylogenetical diversity was mainly clustered by ethnicities/geography and less by lifestyles<sup>6</sup>. These inconsistent results from various studies need to be further clarified by substantially broader cross-cultural sampling<sup>7</sup>.

China has evolved into a multi-cultural society consisting of Han as the major ethnic group and 55 official ethnic minority groups. The Guangxi Zhuang Autonomous Region has the greatest proportion of the Zhuang minority in China and some other minorities as well. One minority, Yao, with a population of approximately 3 million living mainly in southern China<sup>8</sup>, has a unique genetic background and lifestyle<sup>9–11</sup>. One of its subpopulations, Tu Yao, with a population of 3 thousand, only lives in small villages high in the steep mountains or deep in the dense forests in the Hezhou province of Guangxi Zhuang Autonomous Region. Tu Yao is an isolated population because of a lack of traffic and little contact with the outside world. The original natives had no paddy fields, and slash-and-burn cultivation was always inadequate. This is the first study of this special population, and our study focuses on the fecal microbiota in this group. Because the Han population is the major ethnic group in China, and the Zhuang ethnic group is the major minority in Guangxi, we included fecal samples from Han and Zhuang populations in the mountains in the same area as Tu Yao to minimize the influences of genetic background and the environment. We carried out this comparative analysis in these three ethnic groups to characterize the fecal microbiota in Tu Yao and explore its potential relationship with diet, lifestyle and serum biomarkers and then examined the potential enterotypes<sup>12</sup> and their association with host properties.

## Results

**Baseline characteristics of the samples in the three ethnic groups.** A total of 47 Han, 28 Zhuang and 59 Tu Yao (simplified for Yao as follows) were included in our comparative analysis (Table 1). The proportion of males in the three ethnic groups was not statistically significant ( $P = 0.110$ ). The participants in Yao appeared significantly younger ( $P = 0.007$ ), lighter ( $P < 0.001$ ) and shorter ( $P < 0.001$ ) compared to the Han and Zhuang. Moreover, more people in the Yao group regularly had breakfast compared to the other two groups ( $P = 0.002$ ). The people in the Han group were more likely to wash their hands before meals ( $P = 0.013$ ), but were less likely to drink unboiled water ( $P = 0.008$ ) compared to the other two groups.

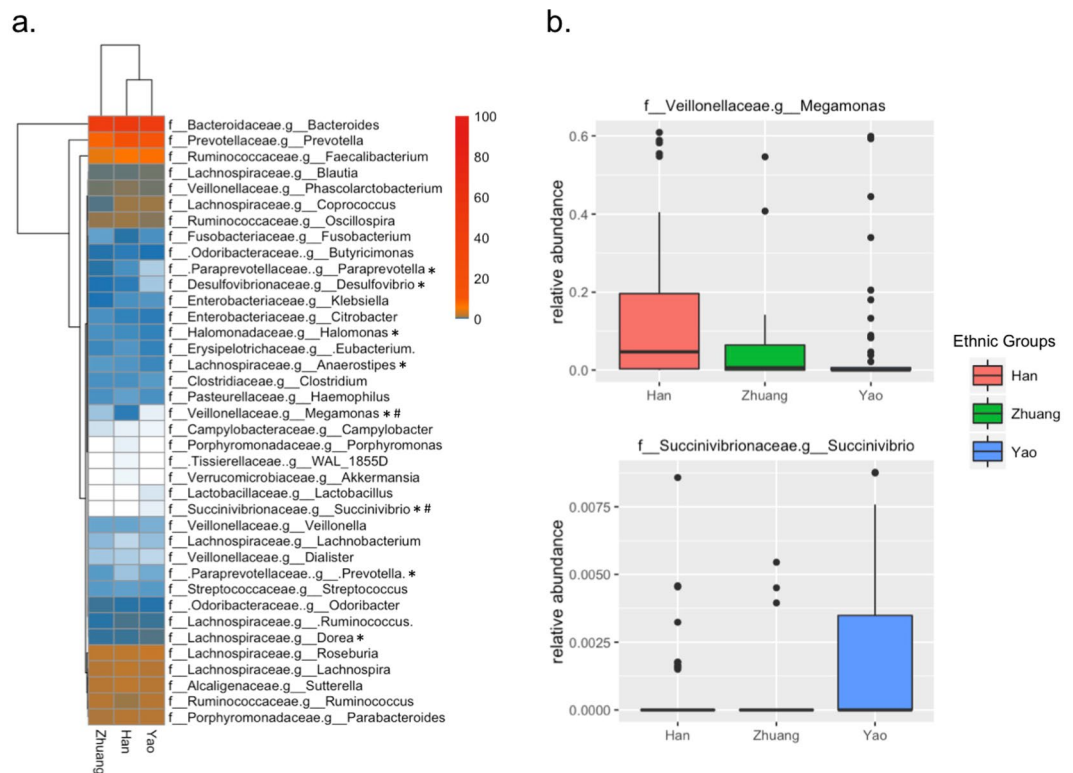


**Figure 1.** The analysis of alpha-diversity (Chao1 index) and beta-diversity (unweighted Unifrac) in the three ethnic groups. **(a)** The chao1 index at the 4200 sequences per sample in the rarefaction curve; P was calculated from the analysis of variance. **(b)** The unweighted UniFrac; P was calculated from the permutational multivariate analysis of variance.

**Alpha-diversity and beta-diversity analysis.** In total, 3,451,874 raw reads were obtained from all 134 fecal samples. After filtering, 2,270,148 high-quality sequences were produced, with an average of  $16,940 \pm 4064$  clean reads per sample. The total number of OTUs at the 97% similarity level was 687,280. The clean reads were highly correlated with the OTU counts in all samples (Fig. S1). The rarefaction curve of all samples almost reached a plateau at this sequencing depth, suggesting that the sequencing was appropriately deep. Alpha-diversity and beta-diversity were present in the three ethnic groups (Figs 1, S2). The chao1 index appeared to be higher in the Han group than in the other two ethnic groups, but this difference was not statistically significant (Fig. 1a). The MDS for unweighted UniFrac was performed to compare the overall structure of the gut microbiota of all samples based on the relative abundance of OTUs. The PERMANOVA revealed an obvious separation between the Han group and the other two groups (Fig. 1b). A statistically significant compositional difference was observed in the unweighted Unifrac analysis ( $P = 0.018$ ).

In terms of other estimators of alpha-diversity, the Shannon index appeared to be higher in the Zhuang group compared to the other two groups ( $P = 0.577$ ), and the index of PD\_whole\_tree was slightly different between the Yao group and the other two ethnicities ( $P = 0.432$ ). The observed species among the three groups were almost the same ( $P = 0.724$ ) (Fig. S2a). Other analysis methods (weighted Unifrac, Bray-Curtis and Jensen-Shannon) for beta-diversity revealed a modest separation between Yao and the other two ethnicities, but this difference was not statistically significant (Fig. S2b).

**Taxonomy-based comparisons at the genus level.** The overall microbiota compositions in the three ethnic groups at the family and genus levels are shown in Fig. 2. There were 22 families and 45 genera in the fecal samples. The dominant genera of the three groups were *Bacteroides* and *Prevotella* (Fig. 2a). Eight significant genera were variably distributed in the three groups, but only 2 genera remained statistically significant after Bonferroni adjustments (Fig. 2b). *Megamonas* was significantly more abundant in the fecal microbiota of the Han group compared to the Zhuang ( $p = 0.048$ ) and the Yao groups ( $p < 0.001$ ), and *Succinivibrio* was significantly more abundant in the fecal microbiota of the Yao group compared to the Han group ( $p = 0.054$ ) and the Zhuang group ( $p = 0.187$ ). The overall presence of microbiota in three ethnic groups at the family and genus levels is shown in Fig. S3. There were three clusters of microbiota, including highly prevalent, modestly prevalent and less prevalent groups. Accordingly, *Megamonas* was highly prevalent, but *Succinivibrio* was less prevalent.



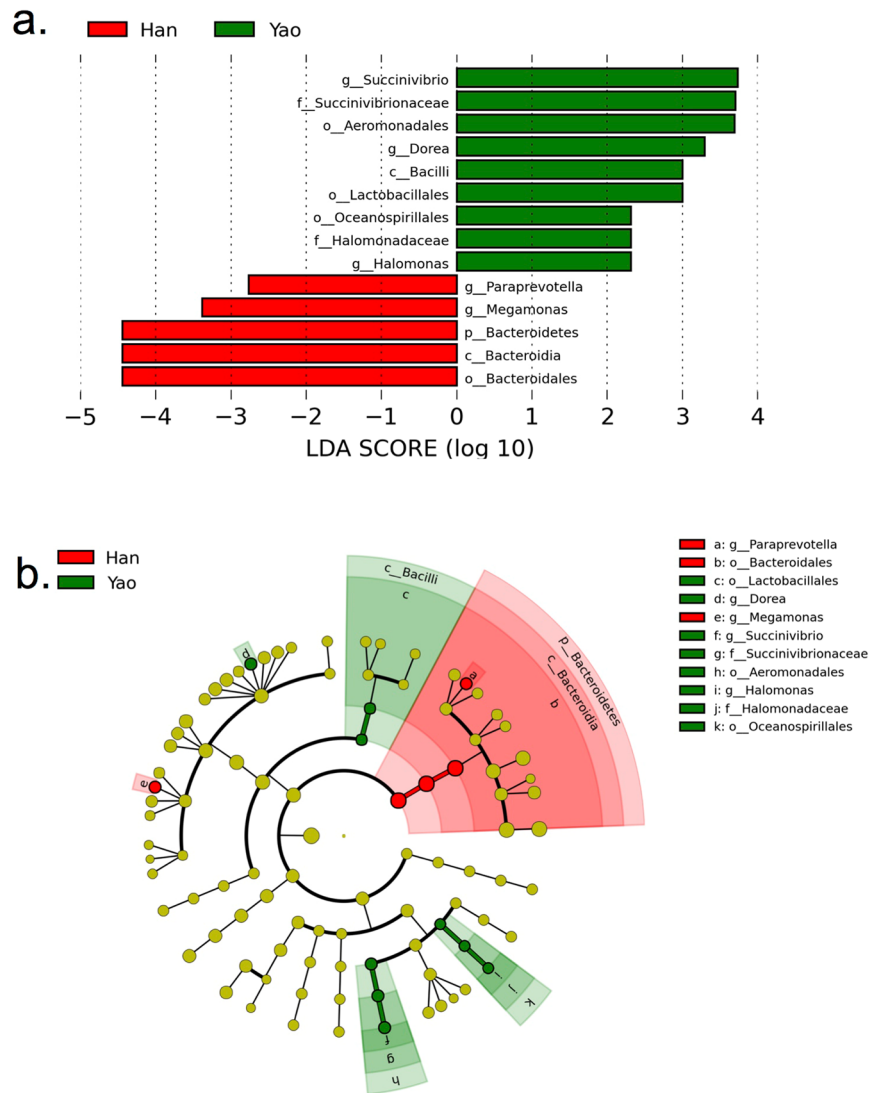
**Figure 2.** The relative abundance of detected genera in the three ethnic groups. (a) The pattern of median values of genus abundance in the three ethnic groups. (b) The relative abundance of two significant genera;  $p$  was calculated from the Kruskal-Wallis rank-sum test.

We also used the LEfSe algorithm to identify the specific taxa that are variably distributed in the three ethnic groups. By specifying ethnic statuses as distinct classes and using the default LDA cutoff of  $+/-2.0$ , we found that 5 taxa were over-represented (including the genus *Megamonas*) and 9 taxa were under-represented (including the genus *Succinivibrio*) in the Han group compared to the Yao group (Fig. 3). Additionally, 3 taxa were over-represented (including the genus *Megamonas*) and 1 taxa was under-represented in the Han group compared to the Zhuang group (Fig. S4).

**Diet frequency and serum biomarkers in the Yao group.** The diet frequencies of 21 specific foods in all samples are presented in Fig. 4. The diet pattern was slightly different in the Yao group compared to the other two groups (Fig. 4a). The diet frequencies of 10 specific foods were statistically different in all three groups ( $p < 0.05$ ). The differences in bean consumption frequency remained statistically significant after Bonferroni adjustments, which is potentially related to the relative abundance of either *Megamonas* or *Succinivibrio* (Fig. 4b). The *Megamonas* abundance increased ( $p = 0.034$ ) and the *Succinivibrio* abundance decreased with a higher frequency of bean consumption ( $p = 0.057$ ).

A total of 73 serum biomarkers were measured using routine blood tests for liver function, kidney function, sex hormones and immune factors. The predictive accuracy of these biomarkers to distinguish the Yao group from the other two groups was 83% (Fig. S5a). Only 15 biomarkers were significantly different among the three ethnic groups (Figs 5a and S5b). Yao people appeared to have relatively high levels of biomarkers in the routine blood tests, including MPV, PLCR, MCV, MCH and RDW (Fig. S5b). Meanwhile, they appeared to have relatively low levels of biomarkers of liver function, including ALB, AG, TBIL and DBIL (Fig. S5b). The three types of bilirubin were positively correlated to the log relative abundance of the *Megamonas* genus (Fig. 5b).

**Alcohol consumption and the common bacteria in various enterotypes.** The PAM analysis of the Jensen-Shannon distance favored partitioning into two clusters (Fig. 6a) but with quite low support (silhouette score 0.3), suggesting that clustering could be due to chance. The two clusters were distinguished by relatively high levels of the genera *Bacteroides* and *Prevotella* (Fig. 6b). The *Prevotella* cluster was distinguished by the additional presence of *Butyricimonas*, *Coprococcus* and *Paraprevotella* (Data not shown). The relative abundances of the *Bacteroides* and *Prevotella* genera were potentially associated with alcohol consumption in both Han and Yao populations, with greater alcohol intake corresponding to a higher abundance of *Bacteroides* and a lower abundance of *Prevotella* (Fig. 6c).



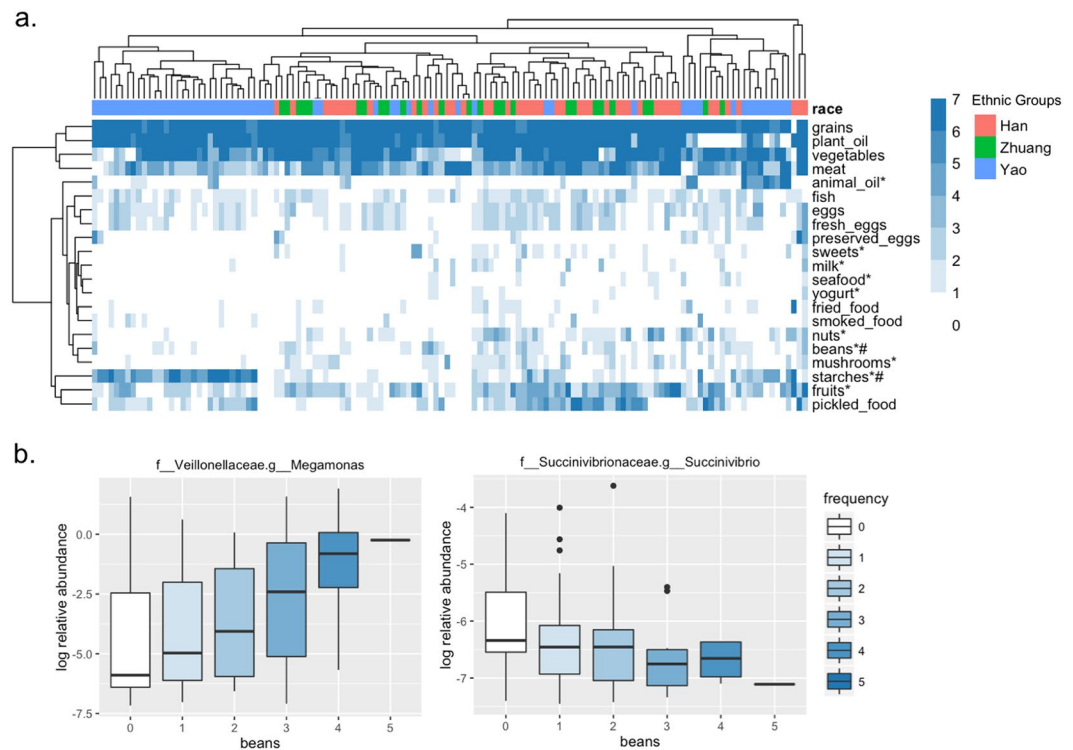
**Figure 3.** The LEfSe results of the Tu Yao and Han groups. **(a)** The linear discriminant analysis. **(b)** The cladograms report. Prefixes represent abbreviations for the taxonomic rank of each taxa: phylum (p\_\_\_), class (c\_\_\_), etc.

## Methods

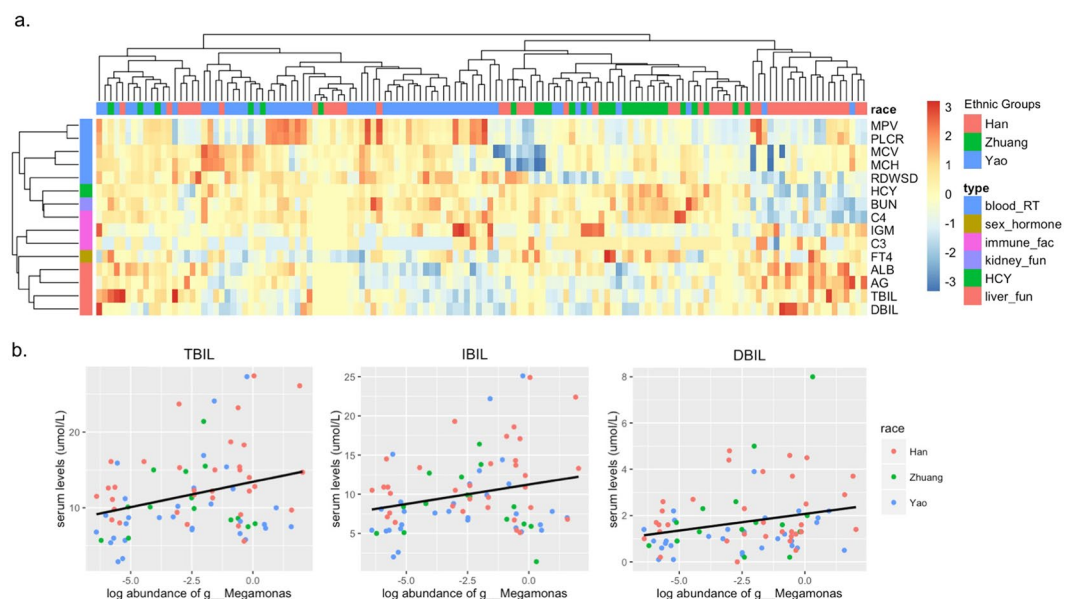
**Sampling and information collection.** We only enrolled the adults living in Babu district, Hezhou province, Guangxi Zhuang Autonomous Region. The Tu Yao population, one subpopulation of the Yao ethnicity, primarily lives in the villages of Etang town, high up in the mountains with little outside contact. Both the Han and Zhuang populations were living down the mountains in the same area. A comprehensive demographic and health survey was conducted in the form of routine physical examinations from June 2015 to July 2015. All participants included in our analysis were confirmed to have the same corresponding ethnicities within three generations. They had not taken any antibiotics or probiotics in the six months before the sampling dates. A fecal sample was collected from each subject. Fecal samples were frozen immediately after sampling and stored at  $-80^{\circ}\text{C}$ . Ultimately, we enrolled 47 Han, 28 Zhuang and 59 Yao adults in the current study. All the participants provided written informed consent. The survey received the approval from the ethics committee at Guangxi Medical University. All the experiments were performed in accordance with the approved guidelines and regulations.

Information on anthropometric measurements, lifestyles, dietary habits and health status were collected for each subject. The anthropometric measurements, including height and weight, were carried out by well-trained physicians from Guangxi Medical University. Weight was measured with the participants in light clothes but without shoes or heavy accessories using a digital weighing device. Height was measured without shoes, with the buttocks and the heels in close contact with the vertical board, and the movable headboard was placed on the top of the head with sufficient pressure to compress the hair. Body mass index (BMI) was calculated from weight in kilograms divided by height in meters squared. A standardized questionnaire was completed by the same physician after a comprehensive examination and a face-to-face interview with each participant to collect information about smoking status, alcohol consumption, dietary habits, and health status. Both cigarette smoking status and

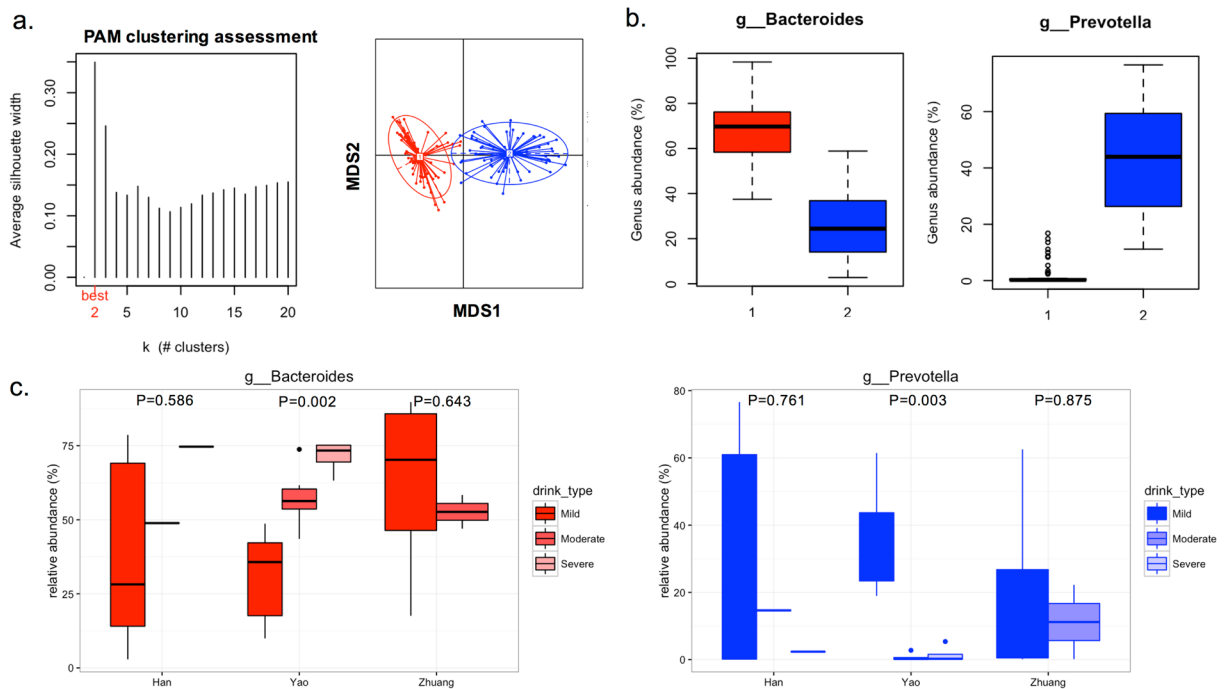




**Figure 4.** The dietary frequencies in the three ethnic groups. **(a)** The dietary patterns of 21 specific foods in the three ethnic groups; \*Indicates  $p < 0.05$  from Fisher's exact test, #Indicates  $p < 0.05$  after bonferroni adjustments. **(b)** The association of bean consumption frequency with two significant genera;  $p$  was calculated from the Kruskal-Wallis rank-sum test.



**Figure 5.** The significant serum biomarkers in the three ethnic groups. **(a)** The distribution pattern of serum biomarkers in the three ethnic groups. Both rows and columns were clustered by *Euclidean* distance, and the serum levels of biomarkers were normalized by z-scores. MPV: mean platelet volume, PLCR: platelet to large cell ratio, MCV: mean cell volume, MCH: mean cell hemoglobin, RDWSD: red cell distribution width - standard deviation, HCY: homocysteine, BUN: blood urea nitrogen, C3: complements 3, C4: complements 4, IGM: immunoglobulin M, FT4: free thyroxin, ALB: albumin, AG: albumin to globulin ration, TBIL: total bilirubin, DBIL: direct bilirubin. **(b)** The association between the serum levels of the three types of bilirubin and the log relative abundance of the *Megamonas* genus;  $p$  was calculated from the Spearman correlation analysis. TBIL: total bilirubin ( $p = ,0.0009$  adjusted R-squared = 0.083), IBIL: indirect bilirubin ( $p = 0.0082$ , adjusted R-squared = 0.050), DBIL: direct bilirubin ( $p = 0.0004$ , adjusted R-squared = 0.094).



**Figure 6.** Two common bacteria were associated with alcohol consumption. **(a)** Clustering using the Jensen-Shannon distance in the gut microbiota composition determined by 16S rRNA gene sequencing. The left panel shows that the data are naturally separated into two clusters by the PAM method. The x-axis shows the cluster number; the y-axis shows silhouette width, a measure of cluster separation. The right panel shows the clustering on the two results from the multidimensional scaling (MDS) analysis. **(b)** The relative abundance of common bacteria in each enterotype. Boxes represent the interquartile range (IQR), and the line inside represents the median. Whiskers denote the lowest and highest values within  $1.5 \times$  IQR. **(c)** The association between alcohol consumption and the relative abundance of the two common bacteria in the enterotypes in the three ethnic groups; p was calculated from the Kruskal-Wallis rank-sum test.

alcohol intake status were defined as never, former (cessation of smoking or drinking for at least six months), and current (daily smoking or drinking within six months). Physical activities were measured by total daily activities; those who regularly exercised for more than 4 h/d were considered physically active. The frequencies of having regular meals or breakfast, drinking unboiled water, washing hands before meals, and brushing teeth regularly were coded as never, sometimes, most times and every day. The taste preferences included salty, spicy or sugary. The diet frequencies of 21 kinds of foods were coded from 1 to 6, indexed as <1 time/month, 1–3 times/month, 1–3 times/week, 4–6 times/week, once/day, and 1–3 times/day respectively.

**Serum assay.** Briefly, approximately 10 mL venous blood specimens were obtained between 8 AM and 10 AM after an overnight fast and transported frozen to the Department of Clinical Laboratory at the First Affiliated Hospital of Guangxi Medical University in Nanning on the same day. Serums were centrifuged within 15–25 min and stored at  $-80^{\circ}\text{C}$  until analysis. The Beckman Coulter LH780 Hematology Analyzer was used for routine blood tests and the tests for serum lipids; liver function and kidney function were performed on the Hitachi 7600 auto-analyzer from Japan. Serum sex hormones, immunoglobulins, complements and other immune factors were measured on the COBA 6000 system E601 immunoassay analyzer (Roche diagnostics, IN, Germany).

**DNA extraction, Massively Parallel Sequencing and Sequence Processing.** Total DNA was extracted from the pellet using the QIAamp<sup>®</sup> Fast DNA Stool Mini Kit (QIAGEN) according to the manufacturer's protocol. The amount of DNA was determined by Nanodrop ND-2000 (Nanodrop, USA). The integrity and the size of the DNA were checked by 1% (w/v) agarose gel electrophoresis. All DNA samples were stored at  $-20^{\circ}\text{C}$  until further processing. Bacterial 16S rDNA amplification and library construction were performed according to the 16S Metagenomic Sequencing Library Preparation guide from Illumina (Forest City, CA) with minor modifications. Briefly, 2  $\mu\text{l}$  of total DNA was amplified using primers targeting the 16S V3 and V4 regions<sup>13,14</sup> (Illumina). Pooled V3-V4 amplicon libraries were sequenced using the Illumina MiSeq platform with a V3 reagent kit. As an added quality control measure, the software package MacQIIME (version 1.9.1) pipeline<sup>15</sup> was used to filter out and discard poor-quality reads using the default settings.

**Bioinformatic and Statistical Analyses.** Processed sequences were subjected to subsampled, open-reference, operational taxonomic unit (OTU) picking against Greengenes (version 13.8), with unmatched reads clustered by 97% identity into OTUs using UCLUST ([http://drive5.com/usearch/manual/uclust\\_algo.html](http://drive5.com/usearch/manual/uclust_algo.html)).

Taxonomies were assigned using the UCLUST consensus taxonomy assigner, and a phylogenetic tree was built using FastTree 2.1.3 (<http://meta.microbesonline.org/fasttree/>). MacQIIME (version 1.9.1) was used to calculate the parameters of alpha diversity, including observed species, chao1, shannon and phylogenetic diversity whole tree (PD\_whole\_tree), and the parameters of beta diversity, including weighted and unweighted UniFrac distances. The Bray-Curtis distance and the Jensen-Shannon distance were also calculated. The multidimensional scaling (MDS) and the permutational multivariate analysis of variance (PERMANOVA) analyses of the three ethnic groups were performed in R (version 3.2.2)<sup>16</sup>.

To determine which taxa were most likely to explain the differences between the three ethnic groups, taxa summaries generated in MacQIIME were reformatted for input into LEfSe via the Huttenhower Lab Galaxy Server (<https://huttenhower.sph.harvard.edu/galaxy/root>). This algorithm performed nonparametric statistical testing of whether individual taxa differed between the Han and Yao groups or the Han and Zhuang groups and then differentially ranked the abundant taxa by their linear discriminant analysis (LDA) log-scores. Differentially abundant taxa in the two corresponding groups that were statistically significant using an alpha of 0.05, and LDA log-scores exceeding  $\pm 2.0$  were visually represented as bar plots.

The median values of taxa abundance and the median percentages of taxa presence in the three ethnic groups were calculated, and the Manhattan distances were used for the clustering analysis. The Kruskal-Wallis rank-sum test was used to identify significant taxa abundance and Fisher's exact test was used to identify significant taxa presence in the three ethnic groups. Standardized serum biomarkers in the three ethnic groups were compared by the Kruskal-Wallis rank sum test. Random forest (RF) classification was used to evaluate the predictive accuracy of all the serum biomarkers to distinguish the three ethnic groups. Following the suggestion by Arumugam *et al.* that the human gut microbiome can be divided into enterotypes<sup>5</sup>, the partitioning around medoids (PAM) analysis was used to examine the potential clusters in our samples.

## Discussion

Growing evidence supports the diversity of the gut microbiome composition in different ethnic groups. We are the first to focus on one specific group, the Yao, who live in high mountains isolated from the outside world. However, this population is being gradually affected by modern society due to the paving of new roads in the region. Compared to the other two local ethnic groups, the Yao population had a significantly low abundance of the *Megamonas* genus, which may potentially be related to different dietary patterns or different patterns of serum biomarkers. Additionally, the common bacteria *Bacteroides* and *Prevotella* were potentially associated with alcohol consumption.

Studies on the gut microbiomes in different ethnic or demagogic groups have been conducted elsewhere in China, but no studies have focused on the isolated Yao population, which is on the edge of extinction. In a recent survey of 314 healthy young adults living in 9 provinces throughout China, the phylogenetic diversities of the gut microbiota in different ethnic groups were classified<sup>6</sup>. However, the study did not explore the specific genera or species in each ethnic group or examine the potential associations with diet or serum biomarkers. In our current study, both the alpha and the beta diversities in the Han population showed different patterns compared to the other two ethnic groups. The three ethnic groups can be significantly clustered by unweighted UniFrac distance. Moreover, we discovered two significant genera distributed variably in the three ethnic groups. The *Succinivibrio* genus was distinctive in the Yao group, which was in the g:f *Succinivibrionaceae*, h:o *Aeromonadales*. The *Megamonas* genus was distinctive in the Han group compared to the Zhuang and Yao groups, with a significantly low abundance in the Yao population. We further discovered that the relative abundance of the *Megamonas* genus may be related to the frequency of bean consumption and may be associated with serum levels of bilirubin in the three groups.

The relative abundance of the *Megamonas* genus in the gut has been reported in other studies. A higher abundance of the genera *Megamonas* was observed in Chinese populations compared to African populations<sup>17</sup>, with a lower abundance in Chinese centenarians than in younger elderlies<sup>18</sup>. *Megamonas* appeared to be related to some diseases, as its abundance was higher in obese Taiwanese individuals and lower in patients with Behcet's Disease or lower motor neuron bowel syndrome<sup>19–21</sup>. Moreover, *Megamonas* in vaginally delivered infants (both neonates and 2-month-olds) had more significant enrichment than in cesarean-delivered infants<sup>22</sup>. In our study, the relative abundance of the *Megamonas* genus was lowest in the Yao population compared to other two ethnic groups, which was probably related to specific dietary patterns. As shown in our study, a higher frequency of bean consumption was potentially associated with a higher relative abundance of *Megamonas* but a lower abundance of *Succinivibrio*. The low abundance of *Megamonas* in the Yao population may be related to the low frequency of bean consumption in this specific ethnic group. Since beans can improve serum lipids by increasing the colonic formation of short chain fatty acids<sup>23</sup>, which has been identified to benefit cardiovascular health<sup>24</sup>, the positive association between bean consumption and the *Megamonas* genus discovered in our study may implicate *Megamonas* as a beneficial microbe. Nevertheless, the effect of a single food on the abundance of a single microbe may be very weak, and the sample size in our study may be inadequate to discover the differences in diet consumption and differences in specific microbial groups. Therefore, we could not rule out the possibility that the association between bean frequency and *Megamonas* abundance was due to chance.

We also examined the association between serum biomarkers and the gut microbiota. The serum levels of bilirubin, whether direct, indirect or total bilirubin, appeared to be positively and significantly correlated with *Megamonas* abundance regardless of ethnicity. Serum bilirubin levels were reported to be positively associated with the greater consumption of fruits and vegetables. In our study, the frequency of fruits and vegetables consumption was potentially different among the three ethnic groups. The low level of serum bilirubin in the Yao population may be due to a lower frequency of fruits and vegetables consumption. Moreover, our discovery of a positive association between bilirubin levels and *Megamonas* abundance indicated that the effect of different dietary habits on bilirubin levels may be mediated by the *Megamonas* genus. Growing evidence supports serum



bilirubin as a novel biomarker in cardiovascular and metabolic diseases<sup>25</sup>, and our findings also support the modification of dietary habits or even the gut microbiome may influence bilirubin levels and potentially predict health statuses in humans.

The identification of gut microbial clusters in different populations was confirmed in our study. Similar to the enterotypes discovered by Arumugam *et al.*<sup>5</sup>, we found two clusters of fecal communities distinguished primarily by the levels of *Bacteroides* and *Prevotella*. One cluster was predominantly *Bacteroides* and the other cluster was predominantly *Prevotella*. The two clusters did not appear to be associated with the three ethnic groups in our study. However, the dominant genus in two clusters was potentially associated with alcohol consumption. In our subgroup analysis, those who consumed more alcohol had a higher relative abundance of *Bacteroides* but a lower abundance of *Prevotella*. *Prevotella*, which is common in non-Westerners who consume a plant-rich diet<sup>26</sup> or vegetarians in Western populations<sup>27</sup>, is considered a beneficial microbe<sup>28</sup>, but it is also linked with chronic inflammatory conditions<sup>29</sup>. Our finding correlated a lower *Prevotella* level with alcohol consumption, which can induce a chronic inflammatory status, supporting *Prevotella* as a beneficial microbe. *Bacteroides* and *Prevotella* may be anti-correlated in population-level microbiome surveys, in part because they are antagonistic<sup>28</sup>. Nevertheless, the role of diet in the human gut remains to be determined<sup>30</sup>. Collectively, our findings may support the beneficial role of *Prevotella* in alcohol consumption.

In our current study, we conducted a comparative analysis of fecal microbiota in three populations and discovered that *Megamonas* may be the distinctive genus in the isolated Chinese Yao population compared to the other two local ethnic groups, the minority Zhuang and the rural Han. The relative abundance of the *Megamonas* genus was positively associated with the frequency of bean consumption and the serum levels of bilirubin. Our data also suggested that the dominant genera in the potential enterotypes are associated with alcohol consumption. More studies are needed to clarify the association between enterotypes and alcohol consumption in different ethnic groups.

## References

- Karkman, A., Lehtimäki, J. & Ruokolainen, L. The ecology of human microbiota: dynamics and diversity in health and disease. *Annals of the New York Academy of Sciences*, <https://doi.org/10.1111/nyas.13326> (2017).
- Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome medicine* **8**, 51, <https://doi.org/10.1186/s13073-016-0307-y> (2016).
- Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227, <https://doi.org/10.1038/nature11053> (2012).
- Lay, C. *et al.* Colonic microbiota signatures across five northern European countries. *Applied and environmental microbiology* **71**, 4153–4155, <https://doi.org/10.1128/AEM.71.7.4153-4155.2005> (2005).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180, <https://doi.org/10.1038/nature09944> (2011).
- Zhang, J. *et al.* A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *The ISME journal* **9**, 1979–1990, <https://doi.org/10.1038/ismej.2015.11> (2015).
- Kwok, L. Y. *et al.* Characterization of fecal microbiota across seven Chinese ethnic groups by quantitative polymerase chain reaction. *PLoS One* **9**, e93631, <https://doi.org/10.1371/journal.pone.0093631> (2014).
- Chen, X. *et al.* MICB gene diversity and balancing selection on its promoter region in Yao population in southern China. *Human immunology* **77**, 1187–1193, <https://doi.org/10.1016/j.humimm.2016.08.015> (2016).
- Ge, K., Zhai, F. & Wang, Q. Effect of nationality on dietary pattern and meal behavior in China. *Am J Clin Nutr* **65**, 1290S–1294S (1997).
- Ruixing, Y. *et al.* Diet, lifestyle, and blood pressure of the middle-aged and elderly in the Guangxi Bai Ku Yao and Han populations. *Am J Hypertens* **21**, 382–387, <https://doi.org/10.1038/ajh.2008.1> (2008).
- Ding, X. L., Deng, Y. L., Zhang, J. & Miao, L. Y. Mutation-sensitive molecular switch method to detect CES1A2 mutation in the Chinese Han and Yao populations. *Genet Test Mol Biomarkers* **15**, 659–662, <https://doi.org/10.1089/gtmb.2011.0010> (2011).
- Knights, D. *et al.* Rethinking “enterotypes”. *Cell Host Microbe* **16**, 433–437, <https://doi.org/10.1016/j.chom.2014.09.013> (2014).
- Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**, e1, <https://doi.org/10.1093/nar/gks808> (2013).
- Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963, <https://doi.org/10.1093/bioinformatics/btr507> (2011).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336, <https://doi.org/10.1038/nmeth.f.303> (2010).
- McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290–297, [https://doi.org/10.1890/0012-9658\(2001\)082\[0290:FMMTCD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2) (2001).
- Alolga, R. N. *et al.* Significant pharmacokinetic differences of berberine are attributable to variations in gut microbiota between Africans and Chinese. *Sci Rep* **6**, 27671, <https://doi.org/10.1038/srep27671> (2016).
- Wang, F. *et al.* Gut Microbiota Community and Its Assembly Associated with Age and Diet in Chinese Centenarians. *J Microbiol Biotechnol* **25**, 1195–1204, <https://doi.org/10.4014/jmb.1410.10014> (2015).
- Shimizu, J. *et al.* Bifidobacteria Abundance-Featured Gut Microbiota Compositional Change in Patients with Behçet’s Disease. *PLoS One* **11**, e0153746, <https://doi.org/10.1371/journal.pone.0153746> (2016).
- Gungor, B., Adiguzel, E., Gursel, I., Yilmaz, B. & Gursel, M. Intestinal Microbiota in Patients with Spinal Cord Injury. *PLoS One* **11**, e0145878, <https://doi.org/10.1371/journal.pone.0145878> (2016).
- Chiu, C. M. *et al.* Systematic analysis of the association between gut flora and obesity through high-throughput sequencing and bioinformatics approaches. *Biomed Res Int* **2014**, 906168, <https://doi.org/10.1155/2014/906168> (2014).
- Kuang, Y. S. *et al.* Composition of gut microbiota in infants in China and global comparison. *Sci Rep* **6**, 36666, <https://doi.org/10.1038/srep36666> (2016).
- Finley, J. W., Burrell, J. B. & Reeves, P. G. Pinto bean consumption changes SCFA profiles in fecal fermentations, bacterial populations of the lower bowel, and lipid profiles in blood of humans. *J Nutr* **137**, 2391–2398 (2007).
- Tang, W. H., Kitai, T. & Hazen, S. L. Gut Microbiota in Cardiovascular Health and Disease. *Circ Res* **120**, 1183–1196, <https://doi.org/10.1161/CIRCRESAHA.117.309715> (2017).
- Vogel, M. E., Idelman, G., Konanah, E. S. & Zucker, S. D. Bilirubin Prevents Atherosclerotic Lesion Formation in Low-Density Lipoprotein Receptor-Deficient Mice by Inhibiting Endothelial VCAM-1 and ICAM-1 Signaling. *J Am Heart Assoc* **6**, <https://doi.org/10.1161/JAHA.116.004820> (2017).
- Martinez, I. *et al.* The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep* **11**, 527–538, <https://doi.org/10.1016/j.celrep.2015.03.049> (2015).

27. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108, <https://doi.org/10.1126/science.1208344> (2011).
28. Kovatcheva-Datchary, P. *et al.* Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of *Prevotella*. *Cell Metab* **22**, 971–982, <https://doi.org/10.1016/j.cmet.2015.10.001> (2015).
29. Hofer, U. Microbiome: pro-inflammatory *Prevotella*? *Nat Rev Microbiol* **12**, 5, <https://doi.org/10.1038/nrmicro3180> (2014).
30. Ley, R. E. Gut microbiota in 2015: *Prevotella* in the gut: choose carefully. *Nat Rev Gastroenterol Hepatol* **13**, 69–70, <https://doi.org/10.1038/nrgastro.2016.4> (2016).

## Acknowledgements

This research was supported in part by grants from the National Natural Science Foundation of China (#81272853, #81472414, and #81370857), the Guangxi Scientific Research and Technology Development Project (Guikegong 2013BC26299 and 1355005-3-17), and the Guangxi Natural Science Foundation (2015GXNSFBB139008, 2014GXNSFBA118201, and 2013GXNSFFA019002).

## Author Contributions

Ming Liao analyzed the data and wrote the manuscript. Yuanliang Xie, Aihua Tan, Zheng Lu, Chunlei Wu, Zhifu Zhang, Yang Chen, Tianyu Li, and Yu Ye collected the data. Yan Mao, Ziting Yao and Yonghua Jiang performed the experiments. Hongzhe Li conducted the data analysis. Xiaoming Li, Xiaobo Yang, Qiuyan Wang organized the data collection and the experiments. Zengnan Mo designed the study.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-17851-8>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018