



Article

# The Core and Distinction of the Gut Microbiota in Chinese Populations across Geography and Ethnicity

Deng Lin <sup>1,†</sup>, Ran Wang <sup>1,2,†</sup> , Junjie Luo <sup>1</sup> , Fazheng Ren <sup>1,2</sup>, Zhenglong Gu <sup>3</sup>, Yiqiang Zhao <sup>1,\*</sup> and Liang Zhao <sup>1,2,4,\*</sup>

<sup>1</sup> Beijing Advanced Innovation Center for Food Nutrition and Human Health, Department of Nutrition and Health, China Agricultural University, Beijing 100083, China; apiaolin@163.com (D.L.); wangran8609@126.com (R.W.); luojj@pku.edu.cn (J.L.); renfazheng@cau.edu.cn (F.R.)

<sup>2</sup> Key Laboratory of Precision Nutrition and Food Quality, Ministry of Education, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

<sup>3</sup> Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA; zg27@cornell.edu

<sup>4</sup> Key Laboratory of Functional Dairy, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

\* Correspondence: yiqiangz@cau.edu.cn (Y.Z.); lzhaol@cau.edu.cn (L.Z.)

† These authors contributed equally to the work.

Received: 9 August 2020; Accepted: 12 October 2020; Published: 14 October 2020



**Abstract:** The diversity of the human gut microbiota constitutes a fundamental health indicator of different populations. The relative importance of geographical location and ethnicity on the gut microbiota, however, has not been previously addressed. Due to unique ethnic distributions across China, we recruited distinct minority ethnic groups, including Han populations, in each of the seven cities that were explored in this study. We investigated the gut microbiota of 394 healthy subjects (14 groups) from these seven different cities using 16S rRNA sequencing. Our results indicated that both geographical location and ethnicity were major factors. However, geographical location exhibited greater influence than ethnicity on both the composition and diversity of the gut microbiota. In addition, a total of 15 shared biomarkers at the genus level were identified in three distinct locations, including seven in Inner Mongolia, seven in Xinjiang and one in Gansu. Furthermore, 65 unique biomarkers were found in 14 population groups, which indicated specific communities in different populations. Based on the gut microbiota species, two main enterotypes—namely Prevotella (ETP) and Bacteroides (ETB), which consist of *Prevotella* and *Bacteroides* as the core bacterial genus, were observed in Chinese populations. Our unique experimental design using the same ethnic group—Han, as a control in different locations, enables delineating the importance of geographical location and ethnicity on the gut microbiota, and provides the fundamental characteristics of gut microbiota diversity in Chinese populations.

**Keywords:** gut microbiota; Chinese populations; geography; ethnicity; Han nationality

## 1. Introduction

The gut microbiota is considered an internalized “microbial organ” that exerts far-reaching impacts on human health [1]. It helps to extract energy from a wide range of host-indigestible carbohydrates, to produce vitamins, short-chain fatty acids, to promote immune homeostasis and to prevent colonization of the gut by pathogens [2–4]. These functions are health-relevant and are considered as ecosystem services that are provided by the gut microbiota for the benefit of the host [5].

Healthy individuals generally maintain a relatively stable structure of the gut microbiota, while microbiome dysbiosis is associated with adverse health conditions that includes inflammatory

bowel disease [6], type 2 diabetes mellitus [7], obesity [8], asthma, autism and rheumatoid arthritis [9]. The specific compositions of the microorganisms that constitute the gut microbiota, related to diseases incidence and progression, have been previously considered as bacterial biomarkers [10,11], which might serve as important targets for disease intervention.

Identification of specific microorganisms as key ecosystem service providers and a deeper understanding of the interactions among microorganisms with their hosts, are fundamental issues deserving of further study in human microbiome research [12]. Numerous studies of the gut microbiota composition showed that multiple factors drive substantial differences across individuals in terms of the composition and diversity of the gut microbiota [13,14]. These factors include host genome composition, geographical location, diet and lifestyle [13,15].

There are studies that have focused on whether the existence of stable bacterial biomarkers from the gut microbiota can separate populations from different geographical locations or distinct ethnicities [14,16]. Studies have shown that ethnic background was strongly associated with a variety of microbial taxa, gene families and metabolic pathways; however, this could be explained as being associated with differential dietary intake and other habits common to those sub-populations of individuals from distinct geographical locations [15,17]. It remains unclear which specific factor plays a dominant role in shaping the gut microbiome [18].

China occupies a significant land mass with extensive differences in both climate and environment, and has developed a diverse dietary culture over time [12]. The Han Chinese people and another 55 different ethnic minority groups constitute approximately 20 percent of the world's population. Traditionally, each ethnic group dwelled in a distinct geographical location and preserved their own dietary habits and specific lifestyles [12,19], especially in the context of Mongolians, Uyghurs and Tibetans. Ethnicity might thus contribute to differences in the diversity of the gut microbiota. However, there are relatively few studies that have analyzed gut microbiota in different geographical locations and ethnicities in China, especially studies that have explored differences between Hans and non-Hans in the same geographic location.

In this study, we investigated the gut microbiota of healthy subjects from seven different geographical locations in China. In each location, we sampled both Han Chinese and one representative ethnic group. With this design, we assessed the impact of geographical location, ethnicity and other factors on the constitution of the gut microbiota, and further quantified predominant factors. We also report on core communities of the Chinese population and bacterial biomarkers of various ethnic groups in different geographic locations. This work might help advance our understanding of gut microbiota diversity in the Chinese population.

## 2. Materials and Methods

### 2.1. Study Groups

We recruited 394 healthy individuals, aged from 17 to 49 years that were unrelated volunteers, consisting of 14 groups from 7 cities in different provinces across China. The subjects were identified as Han and another 7 minority ethnic groups (i.e., the Bai, Hui, Korean, Miao, Mongol, Tibetan and Uyghur, Table S1A). Each non-Han minority ethnic group was recruited from distinct geographical locations with different natural environmental conditions and economic status. The identity and ethnicity of each volunteer was confirmed by evaluating consistency within the census register, spoken language, dressing style and lifestyle with that of the corresponding ethnic group. For comparison, we also recruited Han volunteers from matched locations where each non-Han minority ethnic group was recruited.

None of the volunteers had gastrointestinal tract disorders or had taken any antibiotics for at least 2 months prior to the sampling. The study was carried out in accordance with the recommendations of the local Ethics Review Board on Biomedical Research Involving Human Subjects of the National Health Commission of the People's Republic of China, with written informed consent from all subjects

in accord with the Declaration of Helsinki. The protocol was approved by The Ethics Committee of the China Agricultural University (Research Project Identification No. 2011/23).

The data and sample collection were performed from May to September in 2012. Subjects were required to record their defecation frequency, stool consistency [20], and diet over the past three days. Subjects were instructed to collect fresh fecal samples by themselves with a pre-provided sampling kit (including a foam box, a screw cap tube with RNAlater solution, a tube rack, a pre-cooling ice pack), and immerse the specimen in the screw cap tube with RNAlater solution. Samples were transferred to the local collaboration laboratory within 4 h and stored at 4 °C. After all samples were collected, they were transported on ice to a Beijing laboratory, and stored at 4 °C until analysis.

Geographical locations and environmental data (Table S1B) were obtained from global climate data (WorldClim) and used in bioclimatic modeling (CliMond). WorldClim is recognized by the international community as a regional and accurate climatological database [21]. CliMond provides different formats of environmental data, modeling tools, and meteorological expertise for ecological studies [22].

### 2.2. Analysis of Fecal Microbiota by 16S rRNA Sequencing

We amplified 16S rRNA genes using a broad-range, bacteria-specific primer pair 967F: CAACGCGAAGAACCCTTACC; 1046R: CGACAGCCATGCANACCT for the V6 hypervariable 16S RNA region [23]. Sequencing was carried out by using the Illumina High-seq 2000 platform according to standard procedures.

Operational taxonomic unit (OTU) classification, chimera removal, tree construction and taxonomic assignment were all performed using the QIIME2 package [24], after removing low-quality sequences from the raw datasets according to the provided manual by Illumina. The OTUs mapped against the SILVA 138 SSU/132 LSU database [25], and KEGG Orthologies (KOs) were annotated according to an analysis pipeline of the Tax4Fun R package [26].

### 2.3. Feature Selection

There were 122 factors that might influence the composition of gut microbiota from the questionnaire and WorldClim information (Table S1B). We divided these factors into eight categories that included the following: population indicators, climatic factors, personal physical conditions, cereal foods, foods with high protein levels, vegetables, fruits and drinks. The population indicators included geographical location and ethnicity and the climatic factors were acquired from WorldClim according to the latitude and longitude coordinates. The personal factors reflected the physical conditions of healthy respondents. The other five categories are self-explanatory. In order to explore the interactions between geographical location and ethnicity on dietary habits, we employed quantitative variables that were decomposed by variable decomposition. In order to explore the most important factors that influenced the composition of the gut microbiota, all factors were screened by a redundancy analysis (RDA) and a forward feature selection strategy. Finally, independent factors with the highest R-square were retained.

To further study the effects of factors at taxonomic levels (core OTUs, core genera, OTUs, genera, families, orders, classes and phyla) of the gut microbiota and their Kos, we elected to use permutational multivariate analysis of variance (PERMANOVA) to retrieve the explanatory variance.

### 2.4. The $\alpha$ -Diversity and $\beta$ -Distance

Shannon Wiener indices were calculated with the aim of evaluating diversity in all individuals based on OTUs. Differences in the Shannon Wiener index among different groups were tested by Turkey-HSD.

In order to verify the clustering of 14 groups at the genus level, we calculated the Bray–Curtis distance for all group-pairs. Similarly, the weighted Unifrac distance and Euclidean distance were used to measure the difference among groups at the OTU and KO levels, respectively. The 196 combinations

(14 by 14) were further grouped into four categories: (1) pairs of different ethnic groups and different locations; (2) pairs of same ethnic group but different locations; (3) pairs of different ethnic groups but same location; and (4) pairs of same ethnic group and same location. Differences in distances among the four categories were tested by Turkey-HSD. All analyses were performed with the “vegan” R package [27] and the “GuniFrac” R package [28].

### 2.5. Enterotypes and Bacterial Biomarkers Analysis

Based on the taxonomic composition at the genus level, gut microbiota structures were divided into three enterotypes, which included *Prevotella* (ETP), *Bacteroides* (ETB), and *Firmicutes* (ETF), as described previously [29,30]. The chi-square test was used to explore whether the distribution of enterotypes was influenced by geographical location and ethnicity, i.e., to explore the influence of geographic locations by testing Hans from different geographical locations, and to explore the influence of ethnicity by testing Hans and non-Hans from the same geographical location.

In order to pick up informative bacterial biomarkers across different groups, the relative abundance of the bacterium was compared among groups by Linear discriminant analysis Effect Size (LEfSe), which generates a composite output of a bacterial biomarker list based on three statistical methods including: the Kruskal–Wallis test, Wilcoxon rank sum test and the linear discriminant analysis [31]. LEfSe was used for comparing the taxonomic characteristics among Han as compared non-Han populations; Hans as compared Hans in different locations; non-Hans as compared non-Hans in different locations; and Hans to non-Hans in the same location. Please refer to the supplemental material for detailed methodological steps and data (Figure S4).

All genera were assessed for their impact on community composition by RDA. We selected the top 50 RDA with the largest impact on community composition. Meanwhile, all genera were analyzed by Spearman’s correlation analysis to derive correlated genera with an absolute value of the Spearman’s rho coefficient  $> 0.5$  and  $p$ -value  $< 0.05$ . Venn diagrams were created to show overlaps for bacterial biomarkers, core genera, the top 50 RDA and the correlated genera.

## 3. Results

### 3.1. Overall Gut Microbiota Distribution in Chinese Populations

At the level of the Operational Taxonomic Unit (OTU), a total of 41,572 OTU types were identified. Han individuals from Guizhou province contain the highest average of OTU types (i.e., 7615 types identified in 360,229 OTU reads, Table S1D), while Han individuals from Xinjiang had the lowest average of OTU types (i.e., 4045 types identified in 101,026 reads, Table S1D). Moreover, 159 core OTUs were detected in all individuals (average of proportion equal to 73.04%, and ranging from 42.18–86.07%, Figure S1A, Table S1E).

At the genus level, 85 core genera were detected in all individuals (mean proportion equal to 85.43%, and ranging from 74.30–93.23%, Figure S1B, Table S1F). *Prevotella*, including 19% of *Prevotella\_9*, 2.7% of *Prevotella\_2* and 0.22% of *Prevotella\_1* (Table S1F), as well as *Bacteroides* (average proportion equal to 14.37%, Table S1F) were the two most abundant genera found in the gut microbiota.

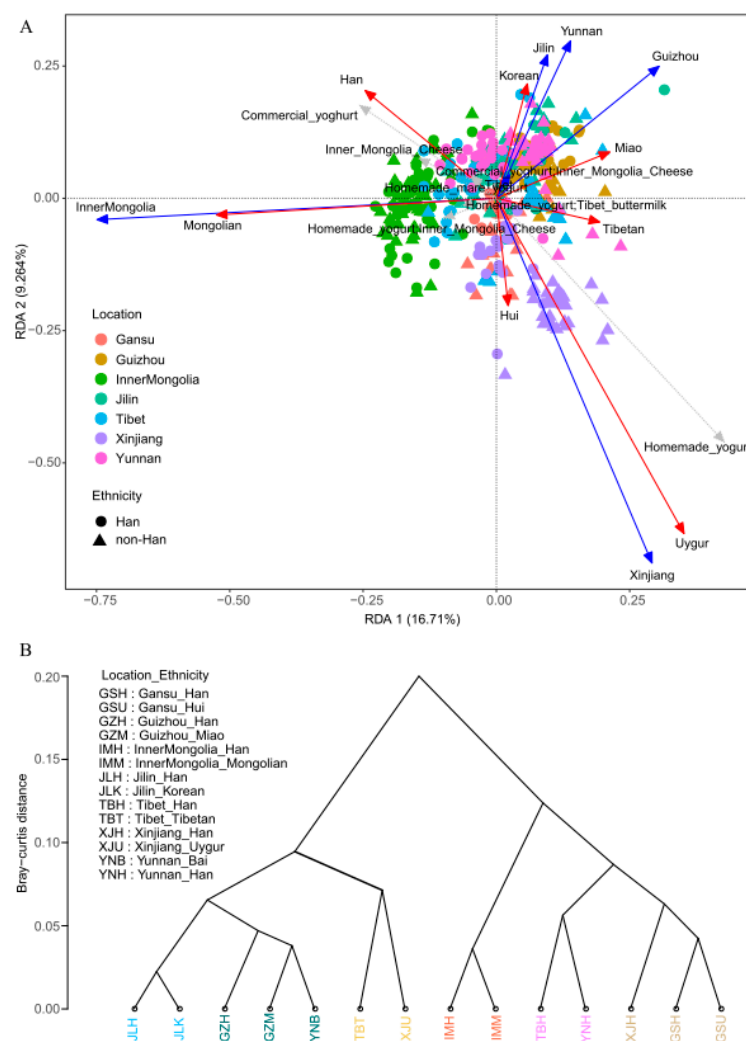
At the phyla level, *Firmicutes*, *Bacteroidetes*, and *Proteobacteria* constituted the three most dominant bacterial phyla, with 49.08%, 42.06% and 5.39% of the total sequences identified, respectively. The ratio of *Firmicutes*:*Bacteroidetes* showed considerable variation among ethnic groups ( $1.23 \pm 0.34$ , Table S1G), which was consistent with previous studies conducted in Asian, African and Western populations [4,12,17].

At the genetic-function level, a total of 6410 KEGG Orthologies (KOs) were detected when using Tax4Fun (Table S1H). K02014, K06147 and K02004 were three of the most abundant KOs (with an average abundance of 0.0116, 0.00883 and 0.00601, respectively) in all 14 groups (Table S1H). These three KOs were annotated to regulate the iron complex outer membrane receptor proteins, ATP-binding

cassette (ABC) subfamily B, and the bacterial putative ABC transport system permease proteins, respectively (Table S1H).

### 3.2. Geographical Location and Ethnicity Were Major Factors Influencing Composition of the Gut Microbiota

We further investigated the factors influencing the composition of gut microbiota. The redundancy analysis (RDA) showed that geographical location and ethnicity influenced the composition of the gut microbiota to a greater extent than all other factors (the adjusted R square was 0.187 and 0.163, respectively, Table S2A). According to variance decomposition, these two factors might indirectly influence the microbiota by affecting dietary habits, especially in cereal foods (Tables S3B,C and S4A). In addition, types of fermented dairy products significantly influenced the microbiota with a relatively small adjusted R square of 0.061. This is interesting because the Hans prefer consumption of commercial yogurt, while homemade yogurt is more favored by minorities such as Uyгур (Figure 1A, Table S3J).



**Figure 1.** (A) Redundancy analysis (RDA) biplot (RDA on a covariance matrix, scaling 2) showing the two first axes of a canonical ordination of 394 samples. Circles and triangles represent different ethnicity [Han and other minorities (non\_Han)] with 7 colors used to identify geographic location; only three factors (geographic location by using blue arrows, ethnicity by using red arrows and types\_of\_fermented\_dairy\_products by using grey arrows) were retained from 122 potential influencing variables after the forward selection. (B) Complete linkage agglomerative clustering dendrogram of Bray-Curtis distance of genera among different groups. The X-axis labels in different colors represent different groups.

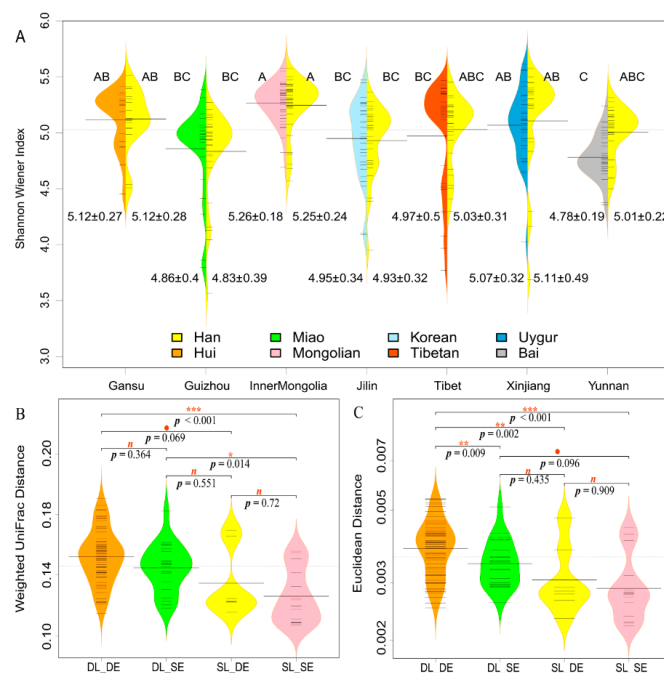
At each taxonomic level and KOs, it was found that a combination of geographical location and ethnicity explained the incidence of more microbiota variations than could be explained when geography and ethnicity were considered separately (Figure S2, Table S4A). Moreover, geography exerted a greater impact on the top 20 percent of abundant bacteria than did ethnicity, which indicated its important effects on the microbiota (Figure S2, Table S4A).

Although we showed that geographical location and ethnicity both had significant influences on the microbiota at different taxonomic levels and KOs (Figure S3, Table S4B–E)—when using one-way analysis of covariance (ANCOVA) or one-way permutational multivariate analysis of variance (PERMANOVA)—it was also found that two-way ANCOVA or two-way PERMANOVA analyses only showed a significant influence for the geographic location, but not for ethnicity (Figure S3, Table S4F–H).

Taken together, our results indicate that geography and ethnicity might jointly explain the diversity of the gut bacterial composition, while geographical location has a greater influence than ethnicity on the dominant gut bacterial community.

### 3.3. Geographical Location Has a Greater Influence than Ethnicity on $\alpha$ -Diversity and $\beta$ -Distance

We found that the  $\alpha$ -diversity of the gut microbiota, as measured by the Shannon Wiener indices, significantly varied among geographical locations ( $F = 12.70$ ,  $p < 0.001$ ), but no significance was observed among different ethnic groups ( $F = 1.13$ ,  $p = 0.34$ , Figure 2A, and Table S5A), indicating that geographic location more greatly influenced  $\alpha$ -diversity of gut microbiota than ethnicity. Hans ( $5.26 \pm 0.18$ ; Figure 2A) and Mongolians ( $5.26 \pm 0.24$ , Figure 2A) from Inner Mongolia had the highest  $\alpha$ -diversity value, while Bai populations from Yunnan showed the lowest value of  $\alpha$ -diversity ( $4.78 \pm 0.19$ , Figure 2A).



**Figure 2.** (A) Beanplot of the Shannon Wiener index. Different letters mean significant difference at the 5% level, and the same letter means no significant difference. From “A” to “C”, the Shannon Wiener index decreased significantly. Tail of each plot notes mean value  $\pm$  standard deviation. (B) and (C) show, respectively, the difference of weighted UniFrac distance of operational taxonomic units (OTUs) and the Euclidean distance of KOs among different combinations with geographic location and ethnicity (at the X-axis, DL = Different Location; SL = Same Location; DE = Different Ethnicity; SE = Same Ethnicity); “p” represents p-value of Tukey tests between two groups connected by horizontal lines. \*\*\*:  $p < 0.001$ ; \*\*:  $0.001 \leq p < 0.01$ ; \*:  $0.01 \leq p < 0.05$ ; •:  $0.05 \leq p < 0.1$ ; n:  $p \geq 0.1$ , not significant.

As shown by the dendrogram constructed on the basis of the Bray–Curtis distance at the genus level, the 14 ethnic groups were divided into six clusters: (1) Han/Mongolian from Inner Mongolia; (2) Han/Korean from Jilin; (3) Han/Hui from Gansu, and Han from Xinjiang; (4) Han/Miao from Guizhou, and Bai from Yunnan; (5) Han from Tibet/Yunnan; and (6) Tibetan from Tibet and Uygur from Xinjiang (Figure 1B). The first five clusters were most likely affected by geographical factors with adjacent locations.

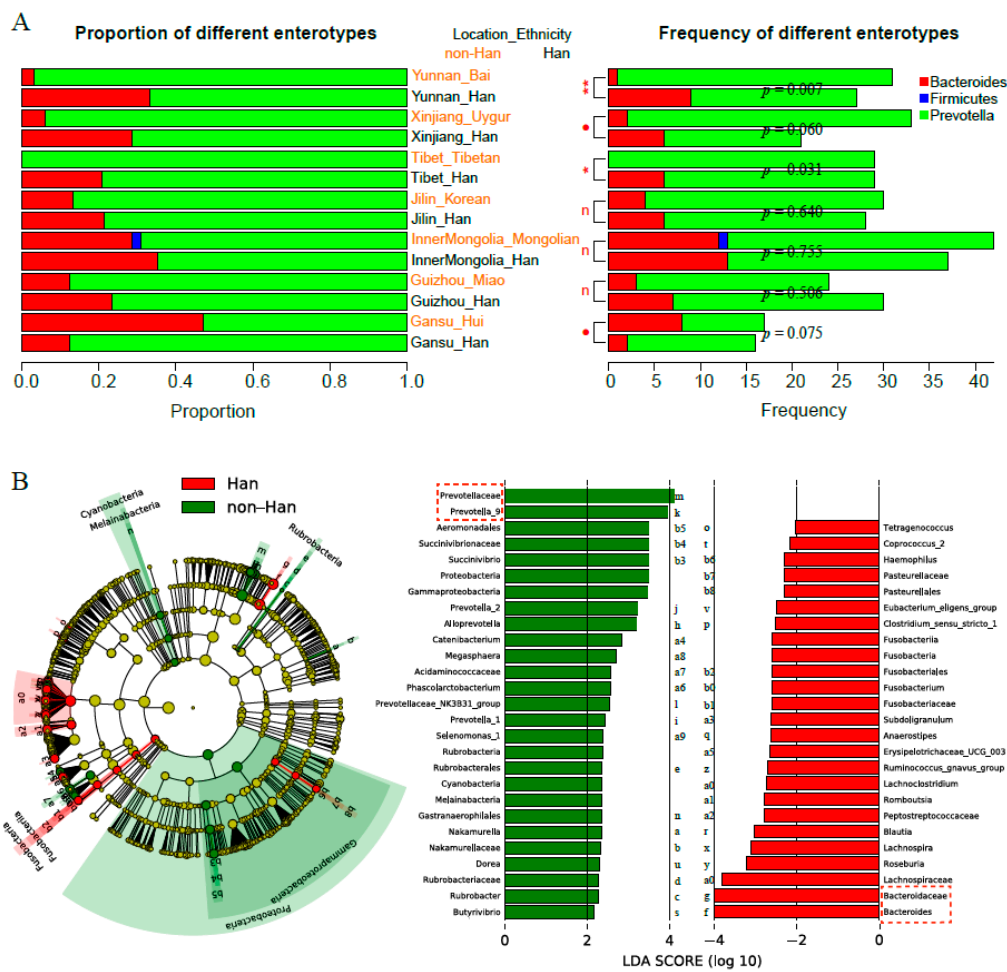
The influence of geographic location and ethnicity on the composition of gut microbiota and KOs were further assessed by weighted UniFrac distances and the Euclidean distance, respectively. At the OTU level, different ethnic groups from different locations (DL\_DE) had the highest weighted UniFrac distances; DL\_DE and DL\_SE were significantly higher than the same ethnic group from the same geographic location (SL\_SE) (Figure 2B, and Table S5B). Similar results were obtained in the comparison of KOs by Euclidean distance. In addition, DL\_DE was significantly higher than the other three categories (Figure 2B, and Table S5C). There was a tendency that indicated that geographic location exerted a greater influence than ethnicity did.

#### 3.4. *Prevotella* and *Bacteroides* Contribute to Differences of Enterotype, Community Composition and Diversity of Gut Microbiota

Previous studies showed that Enterotypes of *Prevotella* (ETP), *Bacteroides* (ETB), and Firmicutes (ETF) were the three primary enterotypes found in different human populations [29,30]. Our results indicated that the enterotype ETF was rare in Chinese populations, and only one Mongolian among all subjects studied belonged to ETF (Figure 3A). Moreover, both ETP and ETB were two predominant enterotypes, and ETP had a higher frequency than ETB in all groups studied (Figure 3A).

There was no significant difference among Han populations from different geographical locations ( $\chi^2 = 4.779$ ,  $p$ -value = 0.572) with respect to enterotypes. Interestingly, the ratio of enterotypes was found to vary among different ethnicities (Figure 3A). Furthermore, within the same geographical location, the percentage of ETB enterotypes in Hans was consistently higher than that found in respective minorities, with the exception of Hui in Gansu province, with some locations reaching statistical significance (e.g., Bai and Tibetan, Yunnan:  $\chi^2 = 7.179$ ,  $p = 0.007$ , Tibet:  $\chi^2 = 4.647$ ,  $p = 0.031$ , Figure 3A).

As shown in Figure 3B, distributions of *Prevotella\_9* and *Bacteroides* varied between the Han and non-Han minority ethnic groups when assessing the data for overall comparisons. Specifically, the scores for *Prevotella\_9* and Prevotellaceae in the Han groups were significantly lower than those of non-Han groups (Figure 3B, Linear discriminant Analysis: LDA score (log10) > 4), whereas patterns for *Bacteroides* and Bacteroidetes were in the opposite direction (Figure 3B, LDA (log10 score) < -4). In addition, we found that the ratio of *Prevotella\_9*:*Bacteroides* was negatively correlated with the Shannon Wiener index ( $\rho = -0.724$ ,  $p < 0.001$ ). Additionally, it was found that Hans and Mongolians from Inner Mongolia had the highest Shannon Wiener indices with a relatively low *Prevotella\_9*:*Bacteroides* ratio (Figure 2A). It suggested that the closer the relative abundance of the two dominant genera, then the greater the  $\alpha$ -diversity of the gut microbiota will be.

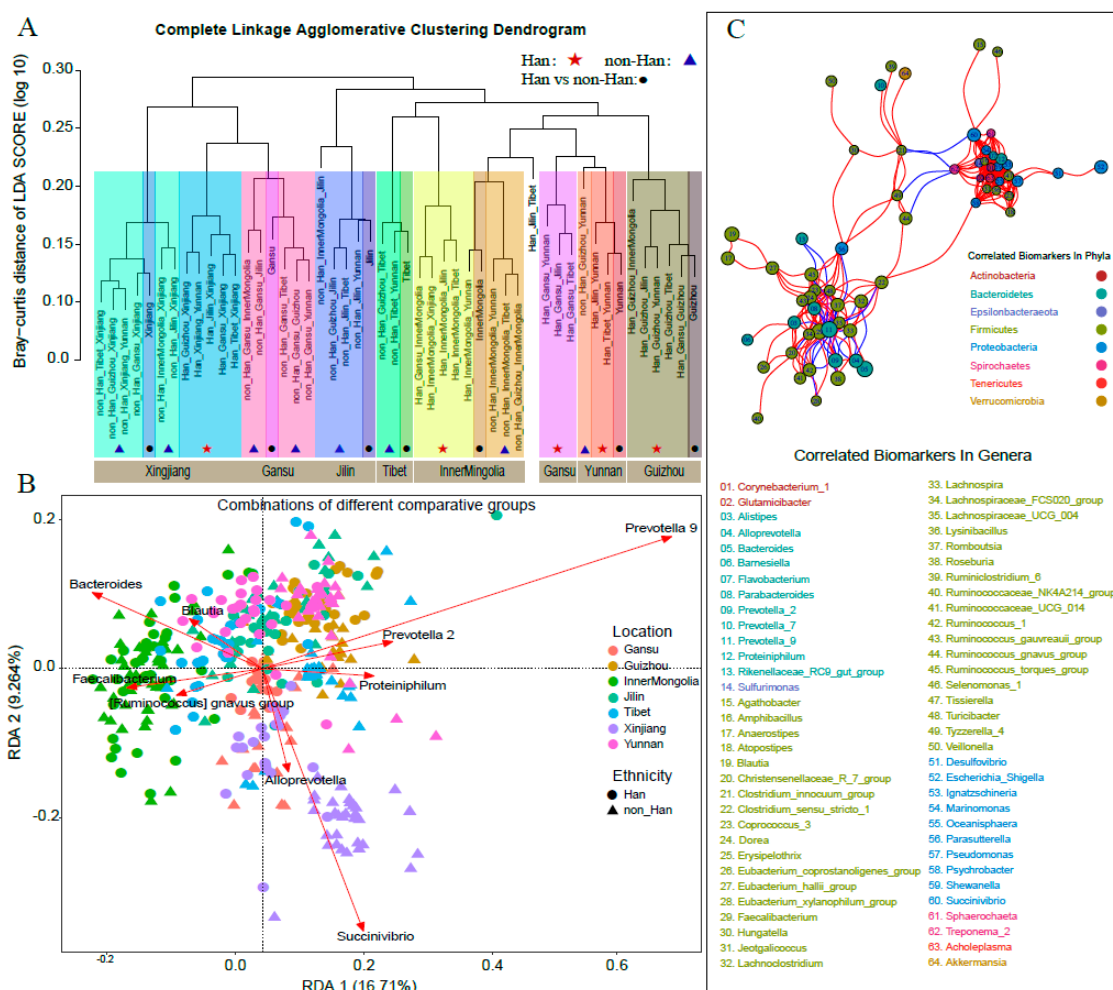


**Figure 3.** (A) Gut microbiota structures were divided into three enterotypes, Prevotella (ETP); Bacteroides (ETB); Firmicutes (ETF), based on their taxonomic composition at genus level; “p” represents the p-value of chi-square tests between two groups connected by longitudinal lines. \*\*: 0.001 ≤ p < 0.01; \*: 0.01 ≤ p < 0.05; •: 0.05 ≤ p < 0.1; n: p ≥ 0.1, no significant. (B) Plot\_Cladogram of LefSe analysis. Taxonomic representation of statistically and biologically consistent differences between Han groups and non-Han groups. Differences are represented in the color of the most abundant class (red indicating Han groups, green indicating non-Han groups and yellow non-significant). Each circle’s diameter is proportional to the taxon’s abundance. The genera, which is significantly different between Hans and non-Hans, and its taxonomic rank are both highlighted—for example, multiple differentially abundant sibling taxa consistent with the variation of the parent clade; the right list of LefSe provides features that were differential among conditions of interest with statistical and biological significance, ranking them according to the effect size.

### 3.5. Bacterial Biomarkers of Various Ethnicity in Different Geographical Location

Subjects from the same geographical location or same ethnicity tended to cluster together in the dendrogram based on LDA scores by the LDA effect size analysis (LEfSe) model. This suggested that subjects within the same geographical location or from the same ethnicity might share specific biomarkers (Figure 4A, Table S6A).





**Figure 4.** (A) Complete linkage agglomerative clustering dendrogram of the Bray–Curtis distance of the LDA score (log 10) between combinations of different comparative groups. The X-axis in brown marks the major geographic location effects. Different geographic or ethnic groups are represented in the other shading colors. ★, ▲ and ● represented corresponding to the comparison results of different categories. (B) RDA biplot (RDA on a covariance matrix, scaling 2) showing the two first axes of a canonical ordination of 394 samples. Circles and triangles represent different ethnicities [Han and other minorities (non\_Han)] with 7 colors to identify geographic location; 655 genera (just showed Top 9 significant genera by using red arrows). (C) Co-occurrence network deduced from all genera (Table S6) enriched in 64 bacterial biomarkers which were correlated each other. Sizes of the nodes represent the logarithm of the total number of each genus sequence. Red edges, Spearman’s rank correlation coefficient > 0.5, adjusted  $p < 0.05$ ; blue edges, Spearman’s rank correlation coefficient < -0.5, adjusted  $p < 0.05$ . Different colors of nodes represent the corresponding phyla. Each node with a number annotated to the genera name were given the corresponding annotation.

All bacterial biomarkers were divided into shared and unique biomarkers according to whether those were detected for the Hans in all, or in particular geographical locations. According to the relative abundance, they were divided into bacterial biomarkers with greater quantity and bacterial biomarkers with decreased quantity. In this way, we evaluated the influence of geographical and ethnic factors on these biomarkers.

In total, 158 bacterial biomarkers including 78 bacterial biomarkers at the genus level were detected by LefSe (Table S6A). In particular, these included 41 shared biomarkers and 120 unique biomarkers (Table S6A). In terms of the relative abundance, the biomarkers comprised 129 with higher quantity,

14 biomarkers with decreased quantity, and 15 biomarkers with opposite quantitative directions in different locations (Table S6A). In addition, most bacterial biomarkers were found in Xinjiang and Inner Mongolia by LEfSe (Figure 4A and Figure S4, Table S6A), indicating the characteristics of the microbiota in these two provinces.

A total of 15 shared biomarkers at the genus level were detected, including seven (i.e., Coprococcus3, Ruminococcus1, Romboutsia, Parasutterella, Faecalibacterium, Alistipes, Prevotella\_9) in Inner Mongolia, seven3 (Pseudomonas, Tissierella, Shewanella, Ignatzschineria, Acholeplasma, Proteiniphilum, Blautia) in Xinjiang, and only one (Prevotella\_7) in Gansu (Table S6A). It was found that there was no significant difference in the relative abundance of the majority of shared biomarkers (Table S6A). It suggested that non-Hans individuals showed stronger regional characteristics of shared biomarkers than did Hans in the geographical location. Meanwhile, a total of 65 unique biomarkers for 14 groups at the genus level were detected (Table S6A). We found that the unique biomarkers were mainly derived from Bacteroidetes, Firmicutes and Proteobacteria.

### 3.6. Bacterial Biomarkers Were Correlated with Each Other in Each Communities

Using RDA, it was shown that bacterial biomarkers at the genus level with overlapping top nine RDAs were capable of highly explanatory evidence aligned to the community composition (Figure 4B and Figure S1C). For example, *Prevotella\_9*, as the most dominant genus, was distributed in all subjects; *Bacteroides* showed higher RDA explanatory power in the microbial community for Inner Mongolia and *Succinivibrio* showed higher RDA explanatory power in Xinjiang (Figure 4B, Table S6C). These results indicated that the bacterial biomarkers not only showed differences among themselves in various groups, but also influence on the whole microbial community structure.

We found that 64 bacterial biomarkers were correlated with each other, including 267 pairs of a positive correlation (Spearman's  $\rho > 0.5$ ), and 36 pairs that showed a negative correlation relationship (Spearman's  $\rho < -0.5$ , Figure 4C, and Table S6D–F). The negative correlations were found mostly for comparisons between Bacteroidetes and Firmicutes. This suggested that higher abundance bacteria may restrain low abundance bacteria. However, some bacteria with lower abundances like *Actinobacteria*, *Epsilonbacteraeota* and *Spirochaetes*, showed a positive correlation with each other (Figure 4C, and Table S6D–F), which might suggest an additional mechanism of cooperation in these bacteria.

## 4. Discussion

We collected stool samples of Hans and non-Hans individuals from seven cities of different provinces across China. Based on high throughput sequencing data, we analyzed the composition of gut microbiota and the potential influence of 122 factors. We found that the geographic location and ethnicity most significantly influenced the gut microbiota. Moreover, we further detected core bacteria and bacterial biomarkers among different geographical locations and ethnicities in China.

We found that differences in the gut microbiota were mainly influenced by geographical location. Many prior studies of different populations had shown that both composition and diversity of the gut microbiota could be influenced by many factors, including underlying diseases, dietary factors, host gene interactions, life styles, geographical location, and climate [4,15,17,32]. Screening out the most critical factor from many factors requires appropriate dimension reduction methods. Geographical location and ethnicity were screened out as the most critical factors by two rounds of RDA and forward selection analysis (Figure 1A, and Table S2). We further showed that geography explained most of the observed variations in the gut microbiota, as well as other diversity measures. These results are consistent with some prior reports [33]. These studies also verified that geographic factors more strongly influenced the gut microbiota than did genetic, ethnic or other factors among Israeli, Moroccan, Turk, Ghanaian, and Surinamese populations, using the kinship test, PERMANOVA, multidimensional scaling etc. [14,16]. Geographic factors included climatological factors, topography, and regional dietary habits, which need to be further studied.

We consider two potential factors may drive the ethnic differences in gut microbiota composition. First, genotype may contribute to the microbiota changes. Host genotype had important effects on gut microbiota, as shown in previous studies. This study found that Tibetans and Uygurs both had significant difference in gut microbial composition, compared with Hans who were from Tibet or Xinjiang respectively. As shown in the genetic relationship, Tibetan and Uygur had great genetic distance to Han [15,34]. This could contribute to different bacterial colonization in the gut [15,34]. Second, dietary habits may be involved in the ethnic differences. We found that home-made yoghurt and Tibet buttermilk correlated with microbial changes in Uygur and Tibetan populations, respectively. These are traditional fermented foods containing lactic acid bacteria, which could affect the microbial profile. However, ethnicity might be a blend factor, and further studies should identify the potential effective driven elements behind the ethnicity.

The main enterotypes in the Chinese population were ETP and ETB, which was consistent with studies conducted on African populations, but it was different from studies conducted on European or American populations in which ETB and ETF were major enterotypes [15,34]. The ratio of ETP was higher than that of ETB for all ethnic groups or geographical locations. Dietary habits could significantly contribute to enterotypes, since ETP was adept at digesting carbohydrates and fibers [35], while ETF was adept at digesting animal protein and saturated fats in a “Western” diet [36]. Compared with the Western population, high carbohydrate/fiber intake in the Chinese population might formally account for ETP [37].

We identified two key bacterial biomarkers that distinguished Hans and non-Hans populations—including *Prevotella\_9* and *Bacteroides*—which corresponded to the two major enterotypes, ETP and ETB. In addition, *Prevotella* and *Bacteroides* might directly lead to different community compositions and diversity. *Prevotella* tends to change the host environment in terms of digesting carbohydrates and fiber [35], while *Bacteroides* tend to form a community for digesting fats and proteins [38]. According to an extension of the niche theory, if the two dominant genera reach a certain balanced state, it will release increased abundance of the niche allowing the associated bacterium to survive in the host gut micro-environment [5,30,39]. Thus, the overall diversity of the gut microbiota will be increased.

The microbiota composition of the local Hans population was different across different geographical locations but became closer to the non-Hans population in the same location. However, there were no shared biomarkers between the Tibetans and the Hans in Tibet, an observation which was consistent with previous studies in which there was a large difference in the gut microbiota between these two populations [40,41]. It may be due to the highland environment that prevents dietary communication between Tibetans and Hans. Tibetans prefer highland barley as cereal food, while the local Hans do not necessarily (significant intaking frequency between Tibetans and Hans,  $t = 12.932$ ,  $p < 0.001$  by two sample  $t$ -test methods, Table S1B). Meanwhile, genetic differences between Tibetans and Hans could also contribute to the microbial differences [40,41]. In general, however, with increased residence time, biomarkers with regional or ethnic characteristics are becoming similar, due to the mutual acceptance of a dietary culture.

In conclusion, through the comparative analysis between Hans and non-Hans populations, this study found that geography had the most important influence on gut microbiota. Moreover, there was a shared influence among different factors. Therefore, further studies are warranted to explore the internal and external factors of variations in the gut microbiota, and to establish network interactions of various potential factors. We found *Prevotella\_9* and *Bacteroides* to be the two dominant genera of gut bacteria in Chinese people. These two genera, and other associated bacterium as bacterial biomarkers, built a correlated microbiotic system, and then formed the two main enterotypes, ETP and ETB, which collectively provide a theoretical reference for regional intestinal diagnoses and treatment. In addition, by using this approach, more bacterial biomarkers could be found, with their applications gradually evolving from classification to targeted medical uses.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-2607/8/10/1579/s1>, Figure S1: Venn diagram, Figure S2: Results of Variation decomposition, Figure S3: Results of ANOCOV, Figure S4: LEfSe, Table S1: Base information, Table S2: Redundancy analysis and forward selection component analysis, Table S3: Results of chi-square test among geographical location and ethnicity, Table S4: Results of Variation decomposition, ANOCOV, and PERMANOVA, Table S5: Tukey test for multiple comparisons, Table S6: Bacterial Biomarker, core genera, RDA TOP50, correlated genera.

**Author Contributions:** D.L. and R.W. carried out most experiments. D.L., L.Z. and Y.Z. acquired data and analyzed data. L.Z., Z.G., Y.Z. and F.R. came up the study concept. L.Z., Z.G. and F.R. designed the study. J.L., Z.G., Y.Z. and F.R. contributed to the scientific discussion. D.L. wrote the manuscript. L.Z., Y.Z., Z.G. and F.R. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the ‘111’ Project from the Education Ministry of China (Grant No. B18053), and the Key-Area Research and Development Program of Guangdong Province (2018B020203001).

**Acknowledgments:** We thank the support of the high-performance computing platform of the State Key Laboratory of Agrobiotechnology.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Clemente, J.C.; Ursell, L.K.; Parfrey, L.W.; Knight, R. The impact of the gut microbiota on human health: An integrative view. *Cell* **2012**, *148*, 1258–1270. [[CrossRef](#)] [[PubMed](#)]
2. Flint, H.J.; Scott, K.P.; Louis, P.; Duncan, S.H. The role of the gut microbiota in nutrition and health. *Nat. Rev. Gastroenterol. Hepatol.* **2012**, *9*, 577–589. [[CrossRef](#)] [[PubMed](#)]
3. Gill, S.R.; Pop, M.; Deboy, R.T.; Eckburg, P.B.; Turnbaugh, P.J.; Samuel, B.S.; Gordon, J.I.; Relman, D.A.; Fraser-Liggett, C.M.; Nelson, K.E. Metagenomic analysis of the human distal gut microbiome. *Science* **2006**, *312*, 1355–1359. [[CrossRef](#)] [[PubMed](#)]
4. Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J.H.; Chinwalla, A.T.; Creasy, H.H.; Earl, A.M.; FitzGerald, M.G.; Fulton, R.S.; et al. Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486*, 207–214. [[CrossRef](#)]
5. Costello, E.K.; Stagaman, K.; Dethlefsen, L.; Bohannan, B.J.; Relman, D.A. The application of ecological theory toward an understanding of the human microbiome. *Science* **2012**, *336*, 1255–1262. [[CrossRef](#)]
6. Li, E.; Hamm, C.M.; Gulati, A.S.; Sartor, R.B.; Chen, H.; Wu, X.; Zhang, T.; Rohlf, F.J.; Zhu, W.; Gu, C.; et al. Inflammatory bowel diseases phenotype, *C. difficile* and NOD2 genotype are associated with shifts in human ileum associated microbial composition. *PLoS ONE* **2012**, *7*, e26284. [[CrossRef](#)]
7. Qin, J.; Li, Y.; Cai, Z.; Li, S.; Zhu, J.; Zhang, F.; Liang, S.; Zhang, W.; Guan, Y.; Shen, D.; et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **2012**, *490*, 55–60. [[CrossRef](#)]
8. Fei, N.; Zhao, L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J.* **2013**, *7*, 880–884. [[CrossRef](#)]
9. Hall, A.B.; Tolonen, A.C.; Xavier, R.J. Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* **2017**, *18*, 690–699. [[CrossRef](#)]
10. Gilbert, J.A.; Quinn, R.A.; Debelius, J.; Xu, Z.Z.; Morton, J.; Garg, N.; Jansson, J.K.; Dorrestein, P.C.; Knight, R. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **2016**, *535*, 94–103. [[CrossRef](#)]
11. Gilbert, J.A.; Blaser, M.J.; Caporaso, J.G.; Jansson, J.K.; Lynch, S.V.; Knight, R. Current understanding of the human microbiome. *Nat. Med.* **2018**, *24*, 392–400. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, J.; Guo, Z.; Xue, Z.; Sun, Z.; Zhang, M.; Wang, L.; Wang, G.; Wang, F.; Xu, J.; Cao, H.; et al. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J.* **2015**, *9*, 1979–1990. [[CrossRef](#)] [[PubMed](#)]
13. Turnbaugh, P.J.; Hamady, M.; Yatsunencko, T.; Cantarel, B.L.; Duncan, A.; Ley, R.E.; Sogin, M.L.; Jones, W.J.; Roe, B.A.; Affourtit, J.P.; et al. A core gut microbiome in obese and lean twins. *Nature* **2009**, *457*, 480–484. [[CrossRef](#)] [[PubMed](#)]
14. Rothschild, D.; Weissbrod, O.; Barkan, E.; Kurilshikov, A.; Korem, T.; Zeevi, D.; Costea, P.I.; Godneva, A.; Kalka, I.N.; Bar, N.; et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **2018**, *555*, 210–215. [[CrossRef](#)] [[PubMed](#)]

15. Yatsunenکو, T.; Rey, F.E.; Manary, M.J.; Trehan, I.; Dominguez-Bello, M.G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R.N.; Anokhin, A.P.; et al. Human gut microbiome viewed across age and geography. *Nature* **2012**, *486*, 222–227. [CrossRef] [PubMed]
16. Deschasaux, M.; Bouter, K.E.; Prodan, A.; Levin, E.; Groen, A.K.; Herrema, H.; Tremaroli, V.; Bakker, G.J.; Attaye, I.; Pinto-Sietsma, S.J.; et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat. Med.* **2018**, *24*, 1526–1531. [CrossRef]
17. De Filippo, C.; Cavalieri, D.; Di Paola, M.; Ramazzotti, M.; Poulet, J.B.; Massart, S.; Collini, S.; Pieraccini, G.; Lionetti, P. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14691–14696. [CrossRef] [PubMed]
18. Gupta, V.K.; Paul, S.; Dutta, C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. *Front. Microbiol.* **2017**, *8*, 1162. [CrossRef]
19. Dehingia, M.; Devi, K.T.; Talukdar, N.C.; Talukdar, R.; Reddy, N.; Mande, S.S.; Deka, M.; Khan, M.R. Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Sci. Rep.* **2015**, *5*, 18563. [CrossRef]
20. O'Donnell, L.J.; Virjee, J.; Heaton, K.W. Detection of pseudodiarrhoea by simple clinical assessment of intestinal transit rate. *BMJ* **1990**, *300*, 439–440. [CrossRef]
21. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Clim.* **2005**, *25*, 1965–1978. [CrossRef]
22. Kriticos, D.J.; Webber, B.L.; Leriche, A.; Ota, N.; Macadam, I.; Bathols, J.; Scott, J.K. CliMond: Global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol. Evol.* **2012**, *3*, 53–64. [CrossRef]
23. Armougom, F. Exploring Microbial Diversity Using 16S rRNA High-Throughput Methods. *J. Comput. Sci. Syst. Biol.* **2009**, *2*. [CrossRef]
24. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **2019**, *37*, 852–857. [CrossRef] [PubMed]
25. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef]
26. Asshauer, K.P.; Wemheuer, B.; Daniel, R.; Meinicke, P. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* **2015**, *31*, 2882–2884. [CrossRef]
27. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Szoecs, E.; Stevens, M.H.; et al. Vegan: Community Ecology Package. In *R Package Version 2.5-6*; 2019; Available online: <https://CRAN.R-project.org/package=vegan> (accessed on 1 September 2019).
28. Chen, J.; Bittinger, K.; Charlson, E.S.; Hoffmann, C.; Lewis, J.; Wu, G.D.; Collman, R.G.; Bushman, F.D.; Li, H. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **2012**, *28*, 2106–2113. [CrossRef]
29. Costea, P.I.; Hildebrand, F.; Arumugam, M.; Backhed, F.; Blaser, M.J.; Bushman, F.D.; de Vos, W.M.; Ehrlich, S.D.; Fraser, C.M.; Hattori, M.; et al. Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* **2018**, *3*, 8–16. [CrossRef]
30. Arumugam, M.; Raes, J.; Pelletier, E.; Le Paslier, D.; Yamada, T.; Mende, D.R.; Fernandes, G.R.; Tap, J.; Bruls, T.; Batto, J.M.; et al. Enterotypes of the human gut microbiome. *Nature* **2011**, *473*, 174–180. [CrossRef]
31. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**, *12*, R60. [CrossRef]
32. Tyakht, A.V.; Kostyukova, E.S.; Popenko, A.S.; Belenikin, M.S.; Pavlenko, A.V.; Larin, A.K.; Karpova, I.Y.; Selezneva, O.V.; Semashko, T.A.; Ospanova, E.A.; et al. Human gut microbiota community structures in urban and rural populations in Russia. *Nat. Commun.* **2013**, *4*, 2469. [CrossRef] [PubMed]
33. He, Y.; Wu, W.; Zheng, H.M.; Li, P.; McDonald, D.; Sheng, H.F.; Chen, M.X.; Chen, Z.H.; Ji, G.Y.; Zheng, Z.D.; et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **2018**, *24*, 1532–1535. [CrossRef] [PubMed]
34. Yin, Y.; Fan, B.; Liu, W.; Ren, R.; Chen, H.; Bai, S.; Zhu, L.; Sun, G.; Yang, Y.; Wang, X. Investigation into the stability and culturability of Chinese enterotypes. *Sci. Rep.* **2017**, *7*, 7947. [CrossRef] [PubMed]

35. Sandberg, J.; Kovatcheva-Datchary, P.; Bjorck, I.; Backhed, F.; Nilsson, A. Abundance of gut *Prevotella* at baseline and metabolic response to barley prebiotics. *Eur. J. Nutr.* **2019**, *58*, 2365–2376. [[CrossRef](#)] [[PubMed](#)]
36. Mejia-Leon, M.E.; Barca, A.M. Diet, Microbiota and Immune System in Type 1 Diabetes Development and Evolution. *Nutrients* **2015**, *7*, 9171–9184. [[CrossRef](#)] [[PubMed](#)]
37. Wang, L.C.; Fang, F.S.; Gong, Y.P.; Yang, G.; Li, C.L. Characteristics of repaglinide and its mechanism of action on insulin secretion in patients with newly diagnosed type-2 diabetes mellitus. *Medicine* **2018**, *97*, e12476. [[CrossRef](#)] [[PubMed](#)]
38. Roth, C.; Petricevic, M.; John, A.; Goddard-Borger, E.D.; Davies, G.J.; Williams, S.J. Structural and mechanistic insights into a *Bacteroides vulgatus* retaining N-acetyl-beta-galactosaminidase that uses neighbouring group participation. *Chem. Commun.* **2016**, *52*, 11096–11099. [[CrossRef](#)]
39. Shmagel, A.; Demmer, R.; Knights, D.; Butler, M.; Langsetmo, L.; Lane, N.E.; Ensrud, K. The Effects of Glucosamine and Chondroitin Sulfate on Gut Microbial Composition: A Systematic Review of Evidence from Animal and Human Studies. *Nutrients* **2019**, *11*, 294. [[CrossRef](#)]
40. Li, L.; Zhao, X. Comparative analyses of fecal microbiota in Tibetan and Chinese Han living at low or high altitude by barcoded 454 pyrosequencing. *Sci. Rep.* **2015**, *5*, 14682. [[CrossRef](#)]
41. Li, K.; Dan, Z.; Gesang, L.; Wang, H.; Zhou, Y.; Du, Y.; Ren, Y.; Shi, Y.; Nie, Y. Comparative Analysis of Gut Microbiota of Native Tibetan and Han Populations Living at Different Altitudes. *PLoS ONE* **2016**, *11*, e0155863. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).