

SCIENTIFIC REPORTS



OPEN

Unique Features of Ethnic Mongolian Gut Microbiome revealed by metagenomic analysis

Wenjun Liu^{1,*}, Jiachao Zhang^{1,*}, Chunyan Wu^{2,*}, Shunfeng Cai², Weiqiang Huang¹, Jing Chen², Xiaoxia Xi¹, Zebin Liang², Qiangchuan Hou¹, Bing Zhou², Nan Qin³ & Heping Zhang¹

Received: 20 April 2016
Accepted: 19 September 2016
Published: 06 October 2016

The human gut microbiota varies considerably among world populations due to a variety of factors including genetic background, diet, cultural habits and socioeconomic status. Here we characterized 110 healthy Mongolian adults gut microbiota by shotgun metagenomic sequencing and compared the intestinal microbiome among Mongolians, the Hans and European cohorts. The results showed that the taxonomic profile of intestinal microbiome among cohorts revealed the Actinobacteria and *Bifidobacterium* were the key microbes contributing to the differences among Mongolians, the Hans and Europeans at the phylum level and genus level, respectively. Metagenomic species analysis indicated that *Faecalibacterium prausnitzii* and *Coprococcus comes* were enriched in Mongolian people which might contribute to gut health through anti-inflammatory properties and butyrate production, respectively. On the other hand, the enriched genus *Collinsella*, biomarker in symptomatic atherosclerosis patients, might be associated with the high morbidity of cardiovascular and cerebrovascular diseases in Mongolian adults. At the functional level, a unique microbial metabolic pathway profile was present in Mongolian's gut which mainly distributed in amino acid metabolism, carbohydrate metabolism, energy metabolism, lipid metabolism, glycan biosynthesis and metabolism. We can attribute the specific signatures of Mongolian gut microbiome to their unique genotype, dietary habits and living environment.

The gut microbiota (GM) is recognized as a human co-evolutionary partner that facilitates nutritional acquisition and immune modulation, and helps maintain host homeostasis in response to profound lifestyle changes^{1–3}. Researchers have delineated the structural and functional configurations of gut metagenomes in different world populations by next generation sequencing^{4–7}. The functional GM layout may represent host genotypic characteristics and reflect an adaptive ecosystem response to the host diet and cultural habits.

The microbiota of the Mongolian population is of particular interest to researchers, because Mongolia encompasses a uniquely wide range of environmental conditions and ethnogeographical cohorts. Moreover, the Mongol Empire was the world's largest contiguous empire, exerted a major influence that greatly enhanced the cultural exchange between Mongolia and Europe that took place during the Middle Ages, when Mongolians intermingled with various populations in the Eurasian continent. Importantly, it has been reported that genetically, more than 20% people in the world are related to Mongolians⁸. Mongolians living in pastureland, such as Khentii Province, maintains a traditional nomadic lifestyle (Fig. 1A), whereas those in urban areas, such as Ulan Bator (the capital of Mongolia) and TUW Province (the suburbs of the capital), have adopted an urban lifestyle due to rapid modernization and economic development. Chinese Mongolians living in Hohhot and Xilingol have inter-married with ethnic Han Chinese, which may generate some differences from their brethren in the north at the genomic level. Nevertheless, all ethnic Mongolians consume similar diet, including traditional cheese (Fig. 1B) and red meat (Fig. 1C) as their main food.

In our previous study, by employing 454 pyrosequencing technology, we described the profiles of gut microbiota of Mongolians from Mongolia and China. At the genus level, *Prevotella* of the *Firmicutes* phylum was the

¹Key Laboratory of Dairy Biotechnology and Engineering, Education Ministry of P. R. China, Department of Food Science and Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China. ²RealBio Genomic Institute, Shanghai 200050, China. ³State Key Laboratory for Diagnosis and Treatment of Infectious Disease, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital, Zhejiang University, Hangzhou 310003, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to N.Q. (email: nqin@zju.edu.cn) or H. Z. (email: hepingdd@vip.sina.com)



Figure 1. Country-wide sampling highlights distinct features of Mongolia gut microbiota. (A) Traditional Mongolian yurt in pasturing area. (B,C) Mongolian's staple foods of such as cheese and red meat are drying in the air.

most abundant genus, and the amounts of *Bacteroides*, *Faecalibacterium*, *Oscillibacter*, *Roseburia*, *Clostridium*, *Coprococcus*, *Ruminococcus*, *Alistipes*, *Parabacteroides*, *Catenibacterium*, *Subdoligranulum* and *Eubacterium* all exceeded 1%, and we found seasonal effects on intestinal microbiota were more distinct in rural Mongolians^{9,10}. From previous research, we found the intestinal microbiota of Mongolians exhibited some unique features. However, little is known about the functional gene and metabolic pathway of the intestinal microbiota of Mongolians^{11,12}. It has been reported that prevalence of cardiovascular and cerebrovascular diseases in Mongolian adults was significantly higher than individuals of other ethnic groups and that these diseases became the leading cause of mortality in Mongolians. Some previous research attributes the aforementioned chronic metabolic diseases to the Mongolian dietary tradition of high consumption of red meat and liquor¹. With the development of human microbiome project and next generation sequencing, great research effort has been dedicated to elucidating the relationship between intestinal microbiota and various human metabolic diseases^{13,14}. The approach may also generate great insight by exploring the potential connection between cardiovascular and cerebrovascular diseases and the intestinal microbiota.

Here we characterized 110 healthy Mongolian adults gut microbiota by shotgun metagenomic sequencing and compared the intestinal microbiome among Mongolians, the Hans and European cohorts. (i) The taxonomic profile of intestinal microbiome among cohorts revealed the *Actinobacteria* and *Bifidobacterium* were the key microbes contributing to the differences among Mongolians, the Hans and Europeans at the phylum level and genus level, respectively. (ii) We constructed a non-redundant microbial catalogue for Mongolian contained 1,491,813 genes, and merged present catalogue with other published healthy gut microbial catalogues of the Hans and Europeans cohorts. The final merged non-redundant catalogue contained 5,127,164 genes, and 283,401 genes were common for all individuals. (iii) At the functional level, a unique microbial metabolic pathway profile was present in Mongolian's gut which mainly distributed in amino acid metabolism, carbohydrate metabolism, energy metabolism, lipid metabolism, glycan biosynthesis and metabolism. We can attribute the specific signatures of Mongolian gut microbiome to their unique genotype, dietary habits and living environment.

Results

Taxonomic characterization. In present research, we characterized the fecal microbiota of 110 Mongolian individuals and compared the structure of intestinal microbiota among the cohorts from Hohhot, Xinlingol, TUW, Ulan Bator and Khentii. The sampling sites included a substantial part of relative densely populated places of ethnic Mongolian settlements, including areas in China and Mongolia. As a result, five representative Mongolian settlements were chosen to collect samples, including Ulan Bator city (35 healthy adults), TUW province (16 healthy adults) and the Khentii pasturing area (12 healthy adults) of Mongolia, Hohhot city (22 healthy adults) and Xilingol pasturing area (25 healthy adults) of China. Principal coordinates analysis (PCoA) based on the Bray–Curtis distance of species abundance (Fig. 2A) revealed the structure of intestinal microbiota between the Han Chinese's and the European's were similar, but significantly different from the Mongolian's. The bacteria alpha diversity based on Shannon index (Fig. 2C) indicated the abundance of intestinal microbiota in Mongolian's and European's gut was significantly higher than Han Chinese's. Importantly, we observed that there was no significant difference of these global parameters among the Mongolians located in Hohhot, Xinlingol, TUW, Ulan Bator and Khentii (Fig. 2D).

To further explore the gut microbial community features of Mongolian population, we compared the results with that of previously reported samples^{2,10,11,15} and carried out taxonomic assignment for the metagenomic data using Metaphlan2 programme. Results were presented in the phylogenetic tree in Fig. 3. For all of these Mongolian subgroups, *Actinomyces*, *Rothia*, *Catenibacterium*, *Phascolarctobacterium*, *Sutterella* and *Lactobacillus* genera were remarkably enriched. On the contrary, *Holdemania*, *Peptostreptococcaceae*, *Potyvirus* or *Anaerostipes* were rarely detected in all Mongolian samples (Fig. 3). Taxonomic profile of the gut microbiomes among the populations (Fig. 4) also revealed that *Actinobacteria* and *Bifidobacterium* were the key components contributing to the differences among Mongolians, the Hans and Europeans at the phylum level and genus level, respectively.

Metagenomic species (MGS) associated with Mongolians. To explore features of the gut microbiota in Mongolians, we compared the genes abundance in 110 Mongolians and 268 Hans, whereby we identified 115,786 genes with significant abundance differences (Wilcoxon rank-sum, Benjamin–Hochberg q -value < 0.001): 78,345 were more abundant in the Hans, and 37,438 in the Mongolians. Likewise, the similarity

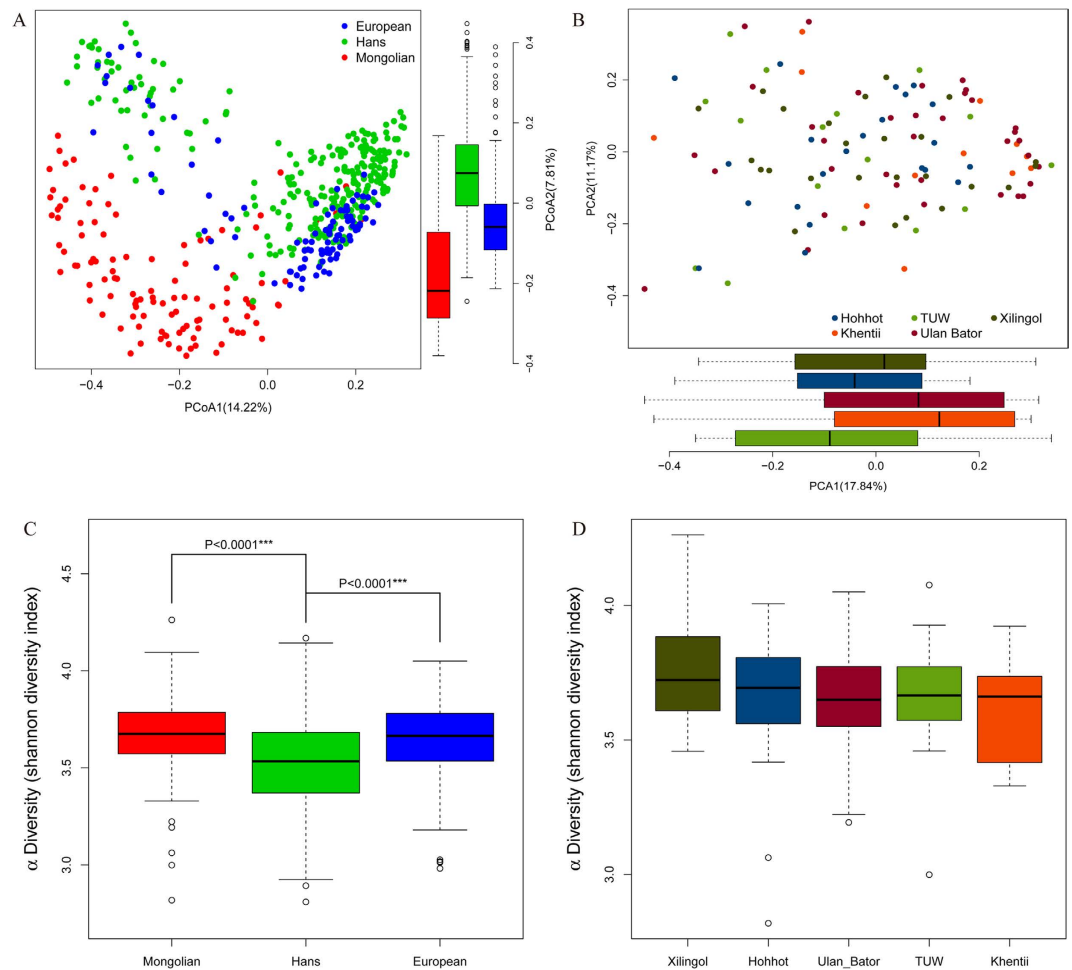


Figure 2. Comparison of the gut microbiomes of different area individuals. (A) PCoA plot with Bray–Curtis distances generated from species abundance of Mongolians, Europeans, and the Hans. (B) PCoA plot with Bray–Curtis distances generated from species abundance of the five arears Mongolian. (C) Shannon diversity index of Mongolians, Europeans, and the Hans at the species level. (D) Shannon diversity index of Mongolians, Europeans, and the Hans at the species level.

comparison between Mongolians and Europeans was also carried out, which identified 127,540 genes: 66,030 were enriched in the Europeans and 61,510 in the Mongolians.

The differentially abundant genes with high correlation were clustered into MGS^{16,17} based on their abundance in all samples. We found 50 MGS to be differentially abundant between the two groups, 26 and 24 enriched in Hans individuals and Mongolian subjects, respectively (Fig. 5 and Supplementary Table 3). Based on the taxonomic characterization of the MGS significantly enriched in Mongolians, we found three *Faecalibacterium prausnitzii*, two identified at strain level and one at species level, which has anti-inflammatory properties and found to be associated with ‘healthy’ microbiome in previous research^{16,18}. One other MGS enriched in Mongolian was identified at species level as *Coprococcus comes* ATCC 27758, which might contribute to gut health through butyrate production. Interestingly, *Collinsella aerofaciens*, which is enriched in Mongolians, was reportedly enriched in patients with symptomatic atherosclerosis¹². For the comparison of Mongolians and Europeans, we found 29 and 16 MGS were enriched in European individuals and Mongolian subjects, respectively (Fig. S2). Remarkably, like the comparison between the Hans and Mongolians, *Faecalibacterium prausnitzii*, *Bifidobacterium adolescentis* and *Collinsella aerofaciens* were enriched in Mongolians.

Another intriguing question we focused on was the correlation between the Mongolians diet and their intestinal microbiota. By transforming food frequency questionnaire to the nutrition information, we calculated the Spearman’s rank correlation coefficient between the Mongolians-enriched MGS and the dietary nutrition factors (Fig. 6 and Fig. S3). From the figure, a significant positive ($P < 0.05$) correlation can be observed between the protein, K, Zn, Fe and VB1 and the species *Collinsella aerofaciens*. Additionally, the elements Se and Mg exhibited significant positive ($P < 0.05$) correlation with the species *Bifidobacterium adolescentis*.

Gene catalogue of gut microbes. We sequenced the fecal metagenome of 110 Mongolian individuals using illumina platform and obtained 725 gigabases (Gb) raw data, with an average of 6.54 Gb for each sample. After quality control, we got an average of 6.33 Gb data for each, totalling 696 Gb of high-quality data which was free of adaptor and human DNA contaminants. Based on these data, we constructed a Mongolian gene

Monglia Microbime Project

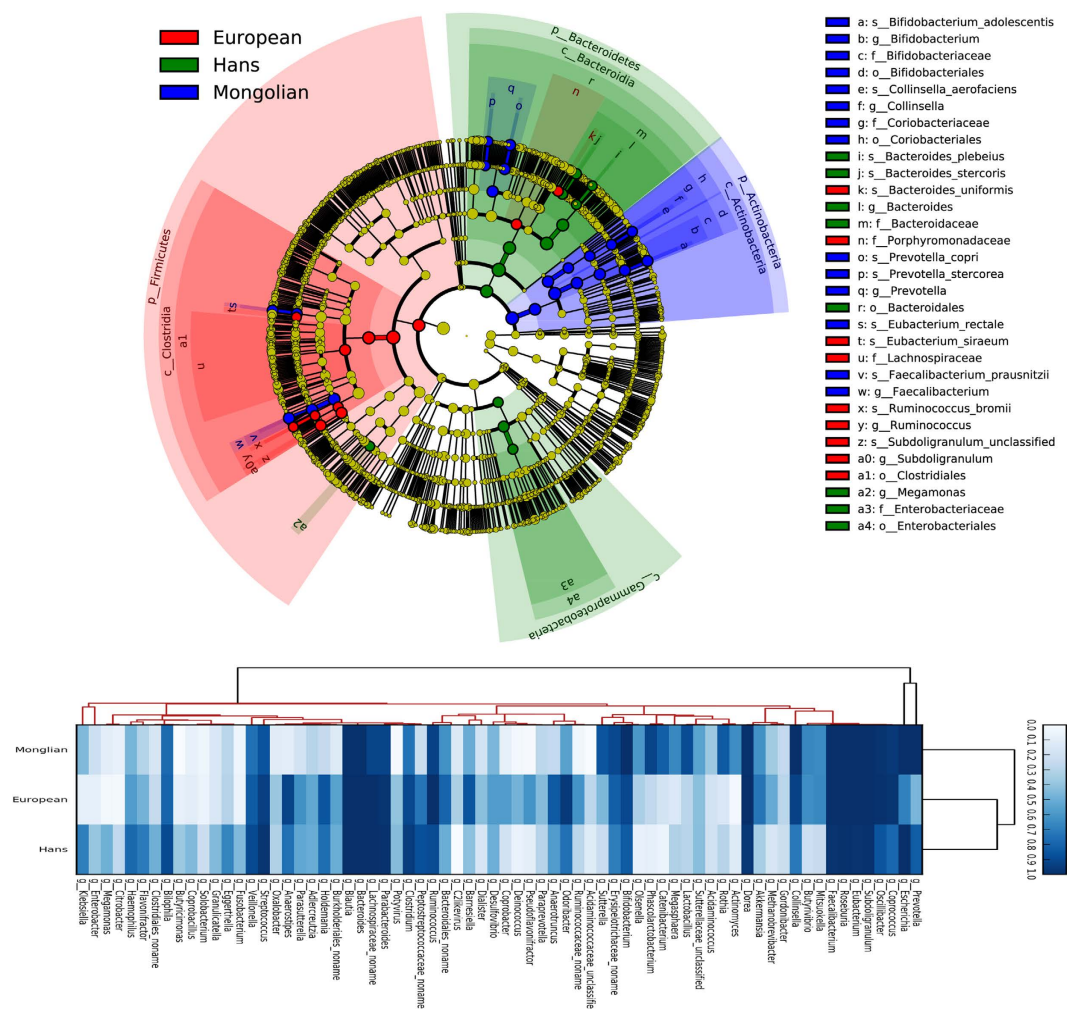


Figure 3. The Comparison of the Mongolian fecal microbiome project to others. All samples data about the three groups (European, Mongolia and Hans) have been detection taxonomic information by use Metaphlan2 programme, the cladogram derived from LEfSe analysis (<http://huttenhower.org/galaxy/>), the heatmap is the detection frequency for the microbial taxa in all samples (100% is deepblue, 0% is blank).

catalogue distributed in five areas (Khentii pasturing area, TUV province, Ulan Bator, Hohhot City and Xilingol pasturing area) using the methodology developed by MetaHIT. Samples from Ulan Bator had the most abundant gene catalogue, containing 768,497 genes. Khentii set contained 400,178 genes, representing the fewest sample. Xilingol, Hohhot and TUV catalogue contained 512,067,702,521 and 433,515 genes, respectively. The merged non-redundant catalogue contained 1,491,813 genes. The Hohhot, Khentii, TUV, Ulan Bator and Xilingol catalogue sets contained 137,496, 99,760, 110,428, 277,758 and 242,637 unique genes, respectively.

Furthermore, we compared the Mongolian catalogue with four other published healthy gut microbial catalogues: ShenZhen (T2D), HangZhou (LC), American (HMP), European (MetaHIT)^{2,10,11,15}. To facilitate this comparison, we predicted genes of the four healthy groups above using the same criteria. The Mongolian gene catalogue contained 1,418,578 genes, the American catalogue 2,243,320 genes, the European catalogue 1,711,862 genes, the HangZhou catalogue 1,183,368 genes and the ShenZhen catalogue 2,067,885 genes. In total, 283,401 genes were common to all catalogues. The Mongolian, American, European, HangZhou, and ShenZhen sets contained 599,370, 1015,117, 591,857, 233,036 and 777,465 unique genes, respectively. The final merged non-redundant catalogue contained 5,127,164 genes (Fig. S1).

Functional characterization. To explore the functional feature of the gut microbiota in Mongolian, we annotated the gene catalogue by KEGG and eggNOG. At the functional level, it can be observed the functional structures of intestinal microbiota among Mongolian, the Hans and Europeans were significantly different (Fig. 7A), as a unique KEGG orthologue group (KO) profile was present in Mongolian's gut. Similarly to taxonomic diversity, Mongolian's gut and European's gut show higher diversity than Han Chinese's in functional level (Fig. 7B). At pathway level, Mongolians enriched KOs was mainly distributed in amino acid metabolism,

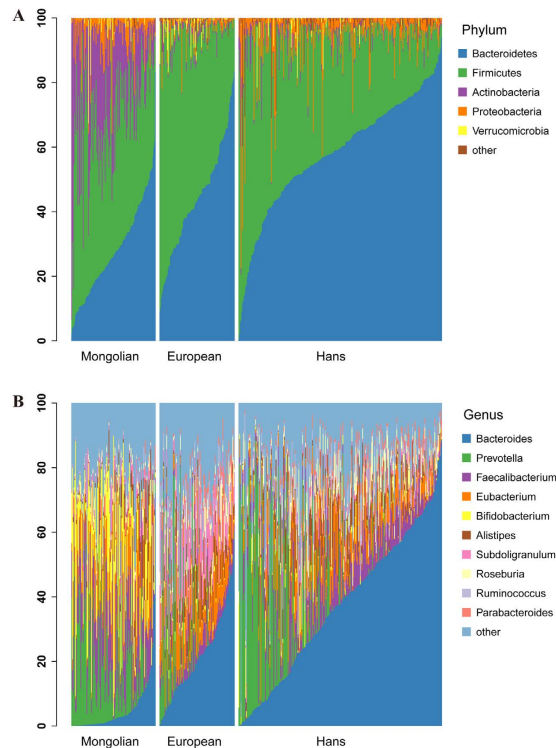


Figure 4. Relative taxa abundance comparison among three different area populations. (A) Phylum level. (B) Genus level. Individuals are showed along the horizontal axis, and relative taxa proportion is represented by the vertical axis.

carbohydrate metabolism, energy metabolism, lipid metabolism, glycan biosynthesis and metabolism (Fig. 8A,C). In the amino acid metabolism, mainly enriched pathway were alanine, aspartate and glutamate metabolism, cysteine and methionine metabolism, lysine biosynthesis and lysine degradation (Fig. 8B,D). Combined the abundance analysis of amino acid metabolic pathways for each species, *Faecalibacterium prausnitzii*, *Prevotella copri* which were enriched in Mongolians, seemed to be the primary contributors to the amino acid metabolism, especially the L-lysine biosynthesis and L-isoleucine biosynthesis (Fig. 9). It potentially indicated some functional enrichment might be related to the enrichment of some specific species. The enriched amino acid metabolism and energy metabolism functions were associated with Mongolians' high meat and fermentation uptake (Fig. S6). Moreover, we identified significantly different (reporter score > 1.6) abundant metabolic models between Mongolian and the Hans or Europeans by calculating the reporter scores. The majority of the models that were differentially abundant in the Mongolian subjects coincided with the observed changes in the metabolic pathway which mainly belong to the metabolic of carbohydrate, amino and nucleotide sugar, phosphotransferase and aminoacyl-tRNA and the biosynthesis of the methane and sulfide (Fig. S4).

Since we observed the carbohydrate metabolic between Mongolians and the Hans was significantly different. We further explored the potential for complex carbohydrate degradation in Mongolian and the Hans gut metagenomes, we screened for carbohydrate-active enzymes (CAZymes) in assembled contigs. As a result, Mongolian GM showed a higher diversity of CAZymes than that of the Hans GM ($P < 0.0001$, Wilcoxon rank-sum test; Fig. S5). This demonstrates an increased capacity for complex carbohydrate metabolism in the Mongolian GM.

Discussion

In present research, we observed several distinct features present in the gut microbiota of ethnic Mongolian from two countries, Mongolia and China. Taxonomically, the genus *Collinsella* was significantly enriched in Mongolians, and a significant positive correlation was observed between protein uptake in diet with this genus (Fig. 6). It was reported that the *Collinsella* was considered as the biomarker in symptomatic atherosclerosis patients¹⁹. In previous study, the intestinal microbiota of patients of symptomatic atherosclerotic and healthy controls were compared. In genus level, it is indicated the abundance of the *Collinsella* was increased in patients with symptomatic atherosclerosis, whereas *Roseburia* and *Eubacterium* were enriched in healthy controls¹⁹. Additionally, the *Collinsella*, *Enterobacteriaceae* and *Streptococcus* were considered as the potential pro-inflammatory GM components enriched in the T2D patients, which resulted in the rapid decline of the SCFA in gut²⁰. Accordingly, we can infer the high protein uptake in Mongolians diet correlates with robust growth of the intestinal *Collinsella*, which may increase the risk of the morbidity of cardiovascular and cerebrovascular diseases in Mongolian adults.

On the other hand, the *Lactobacillus* and *Bifidobacterium* commonly considered as probiotics were largely detected in Mongolians gut^{21,22}. As we known, the *Lactobacillus* and *Bifidobacterium* are major benefit microbes

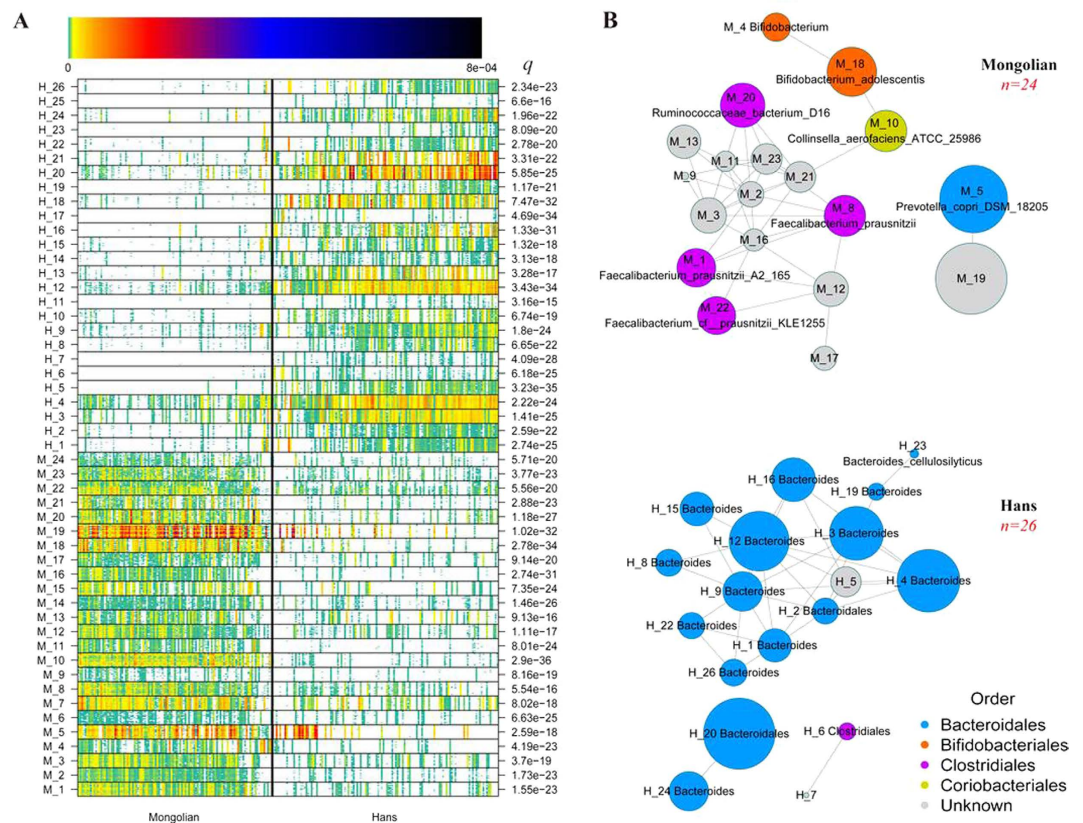


Figure 5. Gut MGS in the Mongolian and the Hans individuals. (A) The Heatmap of 25 ‘tracer’ genes abundance for each MGS in the Mongolians (110 individuals) and the Hans (268 individuals). Individuals are represented along the horizontal axis, sorted by increasing abundance of the Hans enriched MGS. Abundance of genes in rows is indicated by color gradient (white, not detected), and the enrichment significance is shown with Benjamin–Hochberg q value. (B) Co-occurrence network of enriched MGS in Mongolian ($n = 24$) and the Hans individuals ($n = 26$), respectively. Each node represents one MGS, and two nodes are linked if Spearman’s rank correlation > 0.6 , using the edge width to represent the correlation strength. The node size is proportional to the mean relative abundance of MGS in the respective population. Nodes were colored based the phylogenetic order level.

in human gut. They exhibit a range of health-promoting effects, including the regulation of intestinal microbial homeostasis¹⁹, the inhibition of pathogens in the intestinal mucosa²³, the modulation of local and systemic immune responses, the production of vitamins²⁴, and the bioconversion of a number of dietary compounds into bioactive molecules. Various species of *Lactobacillus* played a key role in yoghurt fermentation. And in previous study, researchers had demonstrated that the *Lactobacillus* was able to promote the colonization of *Bifidobacterium*²³. Thus, the widely existence of the genera *Lactobacillus* and *Bifidobacterium* may be related to Mongolians dietary habits of fermented milk consumption. Additionally, we found that the genus *Bifidobacterium* exhibited significant positive correlation with the dietary selenium (Se). The interaction between the Se and the genus *Bifidobacterium* had been reported in previous study, and the supplementary Se in MRS ager was able to promote the growth of the *Bifidobacterium*²⁵.

The increased pathway abundance of carbohydrate, amino acid and nucleotide sugar, and phosphotransferase and aminoacyl-tRNA of intestinal microbiota in Mongolians suggested the more vigorous microbial growth and metabolic stage in Mongolian’s gut compared with that of Han Chinese. The aminoacyl-tRNA was able to combine with its corresponding animal acid, and transferred the animal acid to the ribosome²⁶. In present research, we found the biosynthesis of aminoacyl-tRNA was enhanced which indicated an overall enhanced level of microbial gene expression in the gut of Mongolians. Moreover, the up-regulation of the metabolic pathway of phosphotransferase provide the source of energy for microbe’s proliferation²⁷. On one hand, the robust metabolism of gut microbes in Mongolians was able to promote the digestion and absorption of food in intestine. On the other hand, attributed to the stronger metabolism, intestinal microbes would produce more metabolic products including beneficial or harmful factors which had a greater impact on human health.

Accordingly, we found that the biosynthesis of the methane and sulfide in intestinal microbiota of Mongolians were significantly higher than that of other populations. The two compounds were recognized as the typical inflammatory factor in human gut. Previous researches had demonstrated the perniciousness of the methane and sulfide in intestine of T2D²⁸ and IBD patients^{29,30}. As such, it is tempting to speculate that the high morbidity of

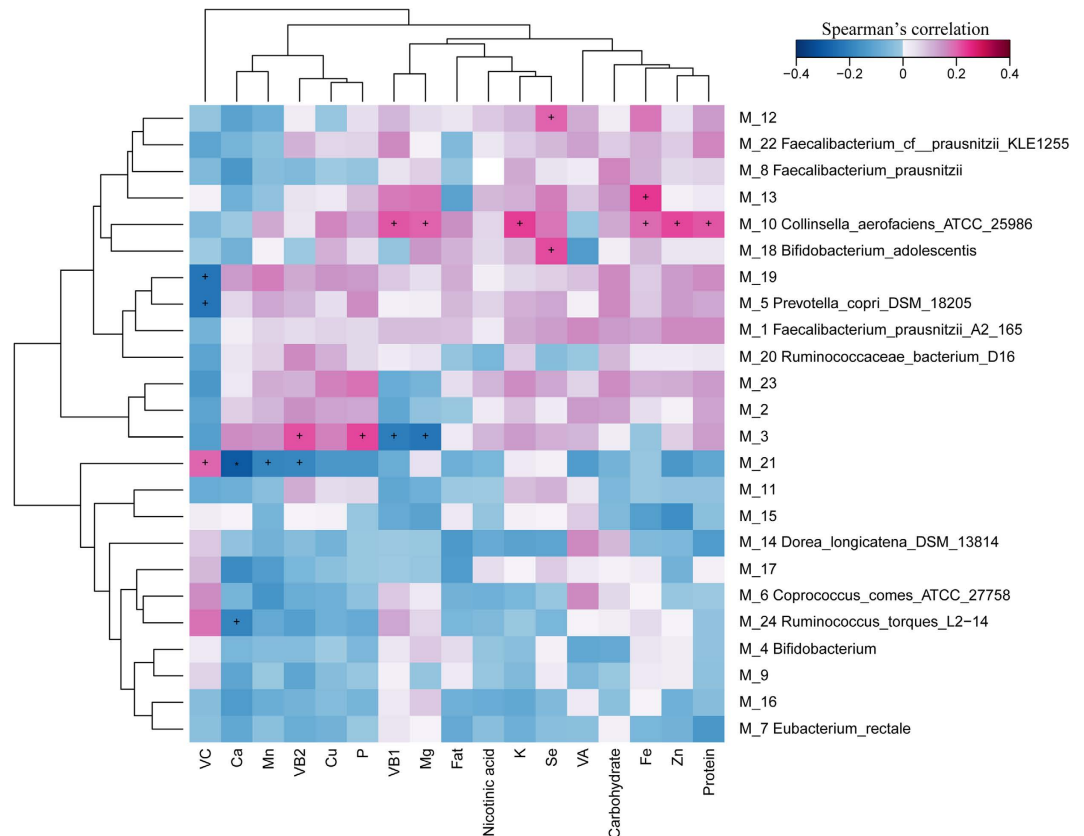


Figure 6. Numerical correlation between dietary indices and Mongolian enriched MGS in the comparison of Mongolians and the Hans. Spearman's rank correlation coefficients is indicated by color gradient, red represents positive correlation; blue represents negative correlation. + Denotes $P < 0.05$; * denotes $P < 0.01$.

cardiovascular and cerebrovascular diseases in Mongolian adults were pertinent to their relatively strong biosynthesis of the methane and sulfide in intestinal microbiota.

In Mongolian diet, meat (including mutton and beef) is the major source of protein, and the Mongolians consume a significantly greater amount of animal protein than Han Chinese. Intestinal microbiota played the key role in animal protein metabolism. Within this microflora, the genera *Clostridium* and *Bacteroides*, which were extensively detected in the Mongolians gut, have been mainly associated with the metabolism of animal proteins, a variety of amino acids and saturated fat acids^{16–18}. Accordingly, we noticed the pathways for degradation of protein and the biosynthesis of the amino acid were enriched in the intestinal microbiota of Mongolians.

Materials and Methods

Sample information. In present research, a total of 110 Mongolians were recruited for sampling. Among these participants, 63 Mongolians were from Mongolia and the other 47 were from Inner Mongolia of China. For participants from Mongolia, 35 volunteers lived a typical modern lifestyle in Ulan Bator, the capital of Mongolia, 16 volunteers lived in the TUV province which covers the suburbs of Ulan Bator, and 12 volunteers lived in the Khentii pasturing area, a typical Mongolian steppe. For the 47 volunteers from Inner Mongolia of China, 22 volunteers lived in Hohhot, the capital of Inner Mongolia, which lived an urban lifestyle. The others lived in Xilingol pasturing area, a typical Mongolian grassland like Khentii. After obtaining the written informed consent, we collected habitual long-term dietary information from all participants using a food frequency questionnaire and transformed them to the nutrition information (Supplementary Table 1). The study protocol was approved by the Ethical Committee of the Inner Mongolia Agriculture University (Hohhot, China). Sampling and all subsequent steps described in the Materials and Methods have been conducted in accordance with the approved guidelines.

DNA sequencing. DNA was extracted from faecal samples using the QIAampDNA Stool MiniKit as previously described¹⁰. The metagenomic DNA libraries were constructed with 2 μ g genome DNA according to the Illumina TruSeq DNA Sample Prep v2 Guide, with an average of 500 bp insert size. The quality of all libraries was evaluated using an Agilent bioanalyser with a DNA LabChip 1000 kit. Sequencing was performed by Illumina HiSeq2500.

Metagenomic data for comparison. The public gut microbial metagenomic data used in this study include: (i) 83 healthy Hans fecal samples (Hangzhou), which were downloaded from EBI under accession number ERP005860¹⁴; (ii) 185 healthy Hans fecal samples (Shenzhen), which were downloaded from NCBI under

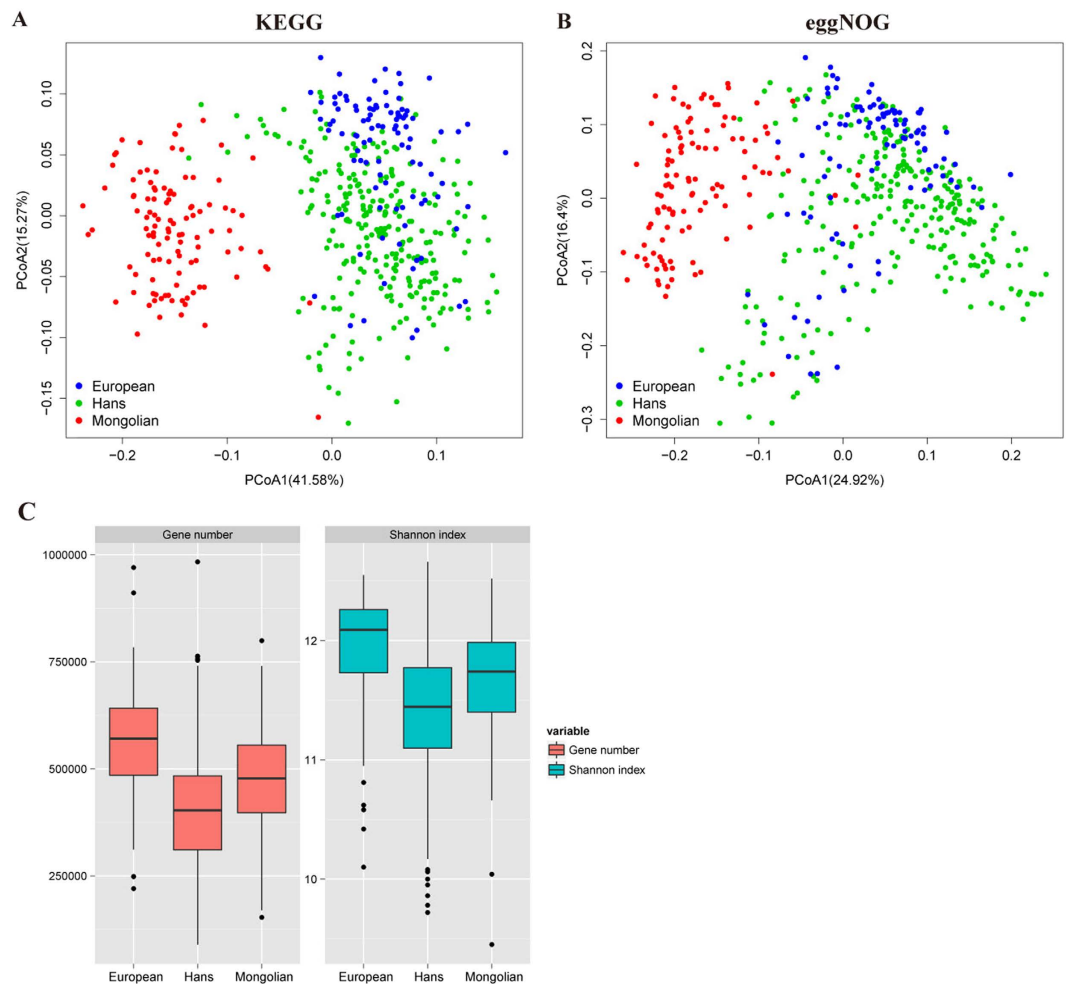


Figure 7. Function of Mongolian gut microbiomes distinctive to the European and the Hans population. (A) PCoA plot with Bray–Curtis distances generated from KEGG Orthologue (Enzyme) profiling. (B) PCoA plot with Bray–Curtis distances generated from eggNOG Orthologue profiling. (C) Gene diversity of Mongolians, Europeans, and the Hans.

accession number SRA045646¹⁴; (ii) 99 healthy European fecal samples from the MetaHIT project, which were downloaded from EBI under accession number ERA000116².

Quality control of reads. Raw paired-end reads were processed quality control using the following criteria: (1) reads with adaptor were removed; (2) reads were trimmed from the 3' end using a quality threshold of 30; (3) reads containing more than 50% bases with low quality (Q30) were removed; (4) Reads shorter than 70 bp were removed; (5) reads that mapped to human genome were removed. For alignment, SOAPaligner 2.21 27 was used by parameter of '-m 100 -x 1000', which are "reporting all repeat hits, Minimal insert size of 100 bp, Maximal insert size of 1000". For detailed information on the meaning of each option above, we refer to <http://soap.genomics.org.cn/soapaligner.html>. The resulting high-quality reads were used for the further analysis.

De novo assembly and gene catalogue construction. High-quality reads were used to assembly with SOAPdenovo 28 (version 2.04) using parameters of '-M 3 -u -L 100 -d 1 -F -k 53', which are "mergeLevel of 3, un-masking contigs with high/low coverage before scaffolding, minContigLen of 100 and KmerFreqCutoff of 1, filling gaps in scaffold, k-mer length of 53". For detailed information on the meaning of each option above, we refer to <http://soap.genomics.org.cn/soapdenovo.html>. The resulting scaffolds were cut into contigs at ambiguous Ns and the contigs longer than 500 bp were saved. All these contigs were applied for gene prediction.

We use MetaGeneMark³¹ (version 3.26) to identify ORFs from the contigs of each sample using a length threshold of 100 bp (Supplementary Table 2). Then the non-redundant gene catalogue was constructed by pairwise comparison of all the predicted ORFs with CD-HIT³² (version 4.5.7) and the redundant genes were removed using a sequence identity cut-off of 0.95 and aligned length covered over 90% of the shorter sequence. The final non-redundant Mongolian gene catalogue contains 1,491,813 microbial genes, with an average length of 808 bp.

In light of some low-abundance microbes that were not detected in the limited sequencing data, we combined with the previously constructed gene catalogue, including MetaHIT gene catalogue², HMP gene catalogue³, T2D gene catalogue¹⁴ and LC gene catalogue¹³ to build a non-redundant gene catalogue for further analyses. The ORFs

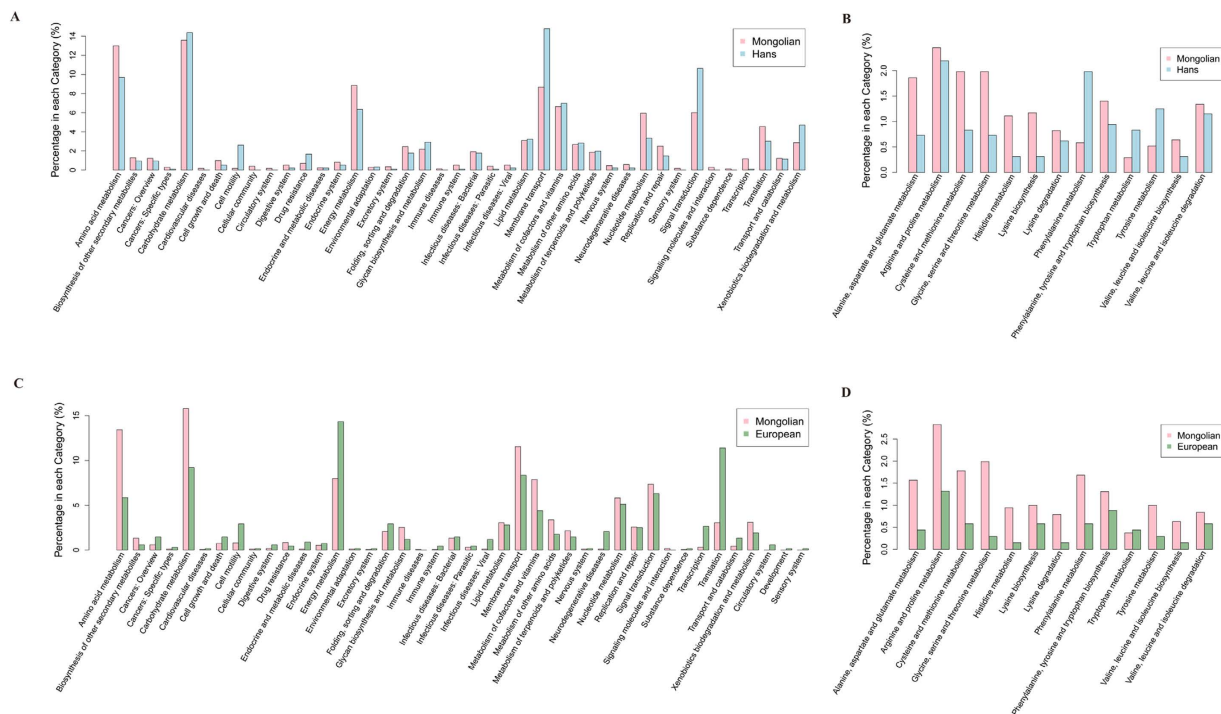


Figure 8. The distribution of KEGG functional categories of KO markers. (A) Comparison between the Mongolians-enriched and the Hans-enriched KO markers on level 2 of KEGG functional category. (B) Comparison between the Mongolians-enriched and the Hans-enriched KO markers on pathways in “Amino acid metabolism”. (C) Comparison between the Mongolians-enriched and Europeans-enriched KO markers on level 2 of KEGG functional category. (D) Comparison between the Mongolians-enriched and Europeans-enriched KO markers on pathways in “Amino acid metabolism”.

in the contigs of healthy individuals from above studies were predicted using MetaGeneMark and merged with Mongolian gene catalogue into a final one with the same criteria of 90% coverage of shorter gene and 95% identity. Finally, we got a non-redundant human gut gene catalogue which contains 4,998,380 genes, with an average length of 763 bp.

Taxonomic and gene profiling. We use MetaPhlan2³³ to produce organism abundance profiling with default parameters, which relied on about 1 million unique clade-specific marker genes identified from about 17,000 reference genomes. 1,036 microorganisms were mapped to the marker genes, include 9 Archaea, 963 Bacteria, 18 Eukaryota and 46 Viruses.

Relative abundances of the genes were determined with the procedure introduced in Qin N *et al.* Nature 2014¹³. When calculating the abundance of genes, the high quality reads from each sample were aligned against the gene catalogue by using SOAPalign 2.21¹⁵ with parameters of ‘-r 2 -m 100 -x 1000’ and only the both paired-end reads which could be mapped to a same gene were accepted. For alignment, SOAPaligner 2.21 was used by parameter of ‘-m 100 -x 1000’, which are “reporting all repeat hits, Minimal insert size of 100 bp, Maximal insert size of 1000”. For detailed information on the meaning of each option above, we refer to <http://soap.genomics.org.cn/soapaligner.html>.

MGS analysis. For the comparison of the faecal microbiome between Mongolians and the Hans, gene markers with differentially abundance were identified (Benjamin–Hochberg q -value < 0.001). The comparison between Mongolians and Europeans was carried out in the same way. To cluster genes into Metagenomic Species (MGS), we followed the method described by Le Chatelier³³ and Nielsen¹⁵. We clustered the genes with Spearman correlation coefficient (ρ) > 0.8 using single-linkage clustering and then fused the clusters with more than 25 genes which had a Spearman correlation coefficient (ρ) > 0.8.

The taxonomically annotation of MGS were performed as previously describe¹³. MGS was assigned to a taxonomical level from strain to super kingdom level when >90% of its genes had a best hit to the same phylogenetic group using blast with >95% identity and >90% overlap of query.

To constructed the co-occurrence network of MGS, we computed the Spearman correlation coefficient between MGS using their abundances and clustered the MGS according to the Spearman’s correlation. The co-occurrence network of MGS was then visualized by Cytoscape3.0.2.

KEGG, EggNOG and CAZy analysis. Putative amino acid sequences of the predicted genes were aligned against the proteins in eggNOG 3.0 database³⁴ and KEGG database³⁵ (release 2014-12-09) using BLAST. Each gene was assigned to one or more OG(s) or EC(s) by the highest annotated scoring hit(s) containing at least one

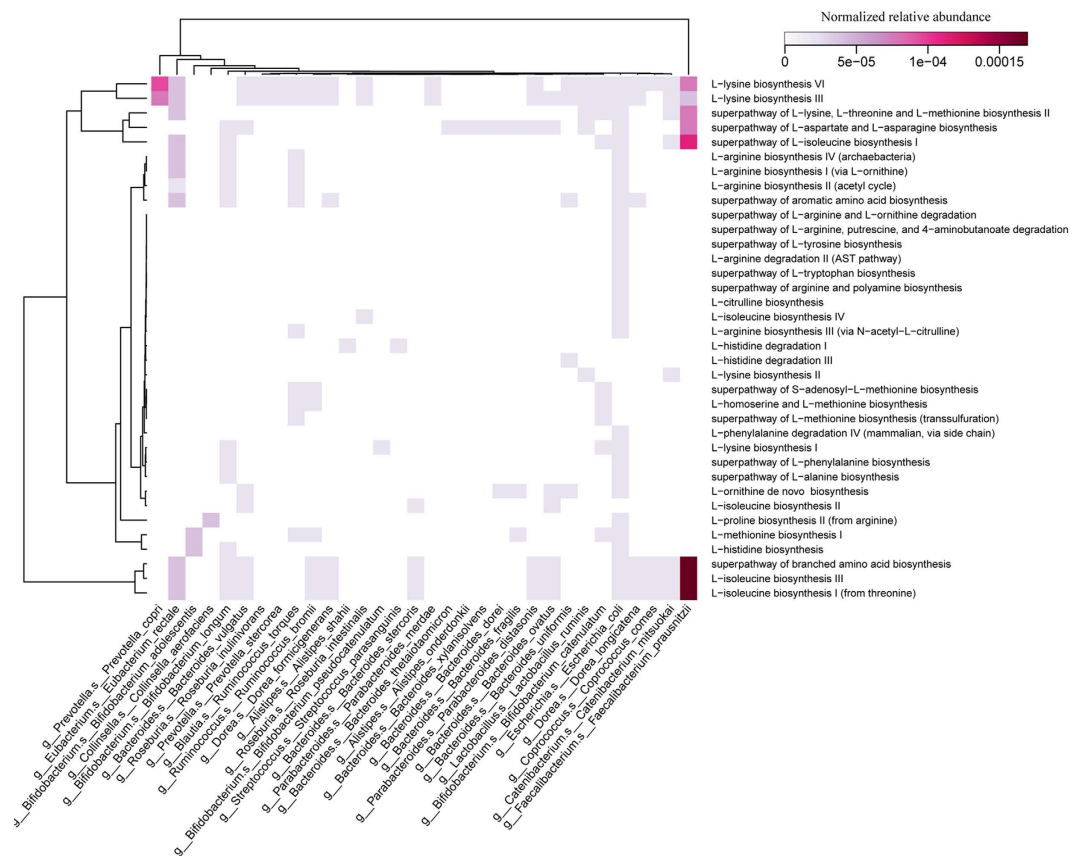


Figure 9. Heatmap of amino acid metabolic pathways of microbes in species level. The heatmap visualizes the abundance of each of the pathways in each of the species (median of Mongolians). The normalized relative abundance of each pathway of each specie was calculated by HUMAnN2 (<http://huttenhower.sph.harvard.edu/humann2>).

HSP scoring over 60 bits. To construct the profiling of OGs in eggNOG and ECs in KEGG database, we accumulated the relative abundance of genes from the same OG or EC using the methods introduced in Qin J. *et al.* Nature 2010². To identify the differentially enriched KO modules, we computed their reporter scores from the Z-scores of individual KOs³⁶. A module with a reporter score of $Z > 1.6$ was defined as differentially enriched module.

CAZymes (Carbohydrate-Active Enzymes) were predicted from amino acid sequences by against to family-specific HMM of CAZymes in dbCAN database³⁷ using Hmmscan program in HMMER 3.0 package³⁸.

The raw sequencing data for all samples have been deposited to Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>) under accession SRP080787.

References

- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023, doi: 10.1038/4441022a (2006).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65, doi: 10.1038/nature08821 (2010).
- David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563, doi: 10.1038/nature12820 (2014).
- Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227, doi: 10.1038/nature11053 (2012).
- Martinez, I. *et al.* The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep* **11**, 527–538, doi: 10.1016/j.celrep.2015.03.049 (2015).
- Rampelli, S. *et al.* Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Curr Biol* **25**, 1682–1693, doi: 10.1016/j.cub.2015.04.055 (2015).
- Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Sci Adv* **1**, doi: 10.1126/sciadv.1500183 (2015).
- Pederson, N., Hessel, A. E., Baatarbileg, N., Anchukaitis, K. J. & Di Cosmo, N. Pluvials, droughts, the Mongol Empire, and modern Mongolia. *Proc Natl Acad Sci USA* **111**, 4375–4379, doi: 10.1073/pnas.1318677111 (2014).
- Zhang, J. *et al.* The diversity of intestinal microbiota of Mongolians living in Inner Mongolia, China. *Benef Microbes* **4**, 319–328, doi: 10.3920/BM2013.0028 (2013).
- Zhang, J. *et al.* Mongolians core gut microbiota and its correlation with seasonal dietary changes. *Sci Rep* **4**, 5001, doi: 10.1038/srep05001 (2014).
- Uurtuya, S., Kotani, K., Yoshioka, H., Yamada, T. & Taniguchi, N. Determinants of carotid atherosclerosis in the general Mongolian population. *Ethnicity & disease* **20**, 257–260 (2010).
- Xu, J. *et al.* The predictive value of waist-to-height ratio for ischemic stroke in a population-based prospective cohort study among Mongolian men in China. *PLoS One* **9**, e110245, doi: 10.1371/journal.pone.0110245 (2014).
- Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64, doi: 10.1038/nature13568 (2014).

14. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60, doi: 10.1038/nature11450 (2012).
15. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714, doi: 10.1093/bioinformatics/btn025 (2008).
16. Voreades, N., Kozil, A. & Weir, T. L. Diet and the development of the human intestinal microbiome. *Front Microbiol* **5**, 494, doi: 10.3389/fmicb.2014.00494 (2014).
17. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108, doi: 10.1126/science.1208344 (2011).
18. Knights, D. *et al.* Rethinking “enterotypes”. *Cell Host Microbe* **16**, 433–437, doi: 10.1016/j.chom.2014.09.013 (2014).
19. Alexander, M. D. *et al.* Symptom differences and pretreatment asymptomatic interval affect outcomes of stenting for intracranial atherosclerotic disease. *AJNR. American journal of neuroradiology* **35**, 1157–1162, doi: 10.3174/ajnr.A3836 (2014).
20. Candela, M. *et al.* Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet. *The British journal of nutrition* **116**, 80–93, doi: 10.1017/S0007114516001045 (2016).
21. Oda, Y. *et al.* Improvement in Human Immune Function with Changes in Intestinal Microbiota by *Salacia reticulata* Extract Ingestion: A Randomized Placebo-Controlled Trial. *PloS One* **10**, e0142909, doi: 10.1371/journal.pone.0142909 (2015).
22. Pace, F., Pace, M. & Quartarone, G. Probiotics in digestive diseases: focus on *Lactobacillus GG*. *Minerva Gastroenterol Dietol* **61**, 273–292 (2015).
23. Inturri, R., Stivala, A. & Blandino, G. Microbiological characteristics of the probiotic strains *B. longum* BB536 and *L. rhamnosus* HN001 used in combination. *Minerva Gastroenterol Dietol* **61**, 191–197 (2015).
24. Hevia, A., Delgado, S., Sanchez, B. & Margolles, A. Molecular Players Involved in the Interaction Between Beneficial Bacteria and the Immune System. *Front Microbiol* **6**, 1285, doi: 10.3389/fmicb.2015.01285 (2015).
25. Pusztahelyi, T., Kovacs, S., Pócsi, I. & Prokisch, J. Selenite-stress selected mutant strains of probiotic bacteria for Se source production. *J Trace Elem Med Biol* **30**, 96–101, doi: 10.1016/j.jtemb.2014.11.003 (2015).
26. Hong, H. J. *et al.* Aminoacyl-tRNA synthetase-interacting multifunctional protein 1 suppresses tumor growth in breast cancer-bearing mice by negatively regulating myeloid-derived suppressor cell functions. *Cancer Immunol Immunother* **65**, 61–72, doi: 10.1007/s00262-015-1777-2 (2016).
27. Saier, M. H. Jr. & Zhang, Z. Control of Transposon-Mediated Directed Mutation by the *Escherichia coli* Phosphoenolpyruvate: Sugar Phosphotransferase System. *J Mol Microbiol Biotechnol* **25**, 226–233, doi: 10.1159/000375375 (2015).
28. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103, doi: 10.1038/nature12198 (2013).
29. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* **6**, 6528, doi: 10.1038/ncomms7528 (2015).
30. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208, doi: 10.1038/nature13828 (2015).
31. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**, 5623–5630, doi: 10.1093/nar/gkl723 (2006).
32. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659, doi: 10.1093/bioinformatics/btl158 (2006).
33. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**, 811–814, doi: 10.1038/nmeth.2066 (2012).
34. Powell, S. *et al.* eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**, D284–D289, doi: 10.1093/nar/gkr1060 (2012).
35. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**, D277–D280, doi: 10.1093/nar/gkh063 (2004).
36. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* **102**, 2685–2689, doi: 10.1073/pnas.0406811102 (2005).
37. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445–W451, doi: 10.1093/nar/gks479 (2012).
38. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**, 205–211 (2009).

Acknowledgements

We sincerely thank all the Mongolian volunteers for their participation. This work was supported by National Natural Science Foundation of China (NO. 31430066), National Natural Science Foundation of China (NO. 81301475) and Natural Science Foundation of Zhejiang Province (NO. LR15H030002).

Author Contributions

H.Z. led the project and H.Z., N.Q. and W.L. conceived and designed the project, J.Z., X.X., W.H. and Q.H. collected the samples. B.Z. performed DNA extraction experiments. B.Z. and C.J. performed library construction and sequencing. N.Q. designed the analysis. C.W., C.J., J.Z. and Z.L. analyzed the data. N.Q. and S.C. did the functional annotation analyses. J.Z. and C.W. wrote the initial manuscript with significant contributions from S.C., N.Q. and C.J. and critical input from all other authors.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, W. *et al.* Unique Features of Ethnic Mongolian Gut Microbiome revealed by metagenomic analysis. *Sci. Rep.* **6**, 34826; doi: 10.1038/srep34826 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>