Supplementary information

Exploring the links between gut microbiome changes and irritable bowel syndrome in Han populations in the Tibetan Plateau

Xingguang ZHANG[1,2], Wei XU[2], Weilong ZHONG[1], Wencheng ZHANG[2], Cheng YANG[2], Lisa DUAN[2], Haiyan NIU[2], Yanmei DONG[2], Taotao LIU[2], Shihai XIA[2], Bangmao WANG[1]

[1]Department of Gastroenterology and Hepatology, Tianjin Medical University General Hospital, Tianjin Institute of Digestive Diseases, Tianjin Key Laboratory of Digestive Diseases, Tianjin 300052, China
[2]Department of Gastroenterology, Characteristic Medical Center of Chinese People's Armed Police Force, Tianjin Key Laboratory of Hepatopancreatic Fiberosis and Molecular Diagnosis & Treatment, Tianjin 300162, China

16S rRNA gene sequencing

1 Genome DNA extraction

The genome DNA of the samples was extracted by CTAB/SDS method, the purity and concentration of the extracted DNA were detected using 1% agarose gel electrophoresis, followed by taking an appropriate amount of DNA in a centrifuge tube and diluting it to 1 ng/μL with sterile water.

1.1 Amplicon generation

16S rRNA/18SrRNA/ITS genes in distinct regions (16S V4/16S V3/16S V3-V4/16S V4-V5, 18S V4/18S V9, ITS1/ITS2, Arc V4) were amplified with specific primer (e.g. 16S V4: 515F-806R, 18S V4: 528F-706R, 18S V9: 1380F-1510R, et al.) and barcodes. All PCR mixtures contained 15 μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs), 0.2 μmol/L of each primer and 10 ng target DNA, and cycling conditions consisted of a first denaturation step at 98 ℃ for 1 min, followed by 30 cycles at 98 ℃ (10s), 50 ℃ (30s) and 72 ℃ (30s) and a final 5 min extension at 72 ℃.

1.2 PCR products quantification and qualification

PCR products were detected by agarose gel electrophoresis at 2% concentration, and equal amounts were fully mixed and detected again using agarose gel electrophoresis at 2% concentration, and the target bands were recovered using the gel recovery kit provided by Qiagen Gel Extraction Kit (Qiagen, Germany).

1.3 Library construction and up-sequencing

The libraries were constructed using the NEBNext® Ultra™ IIDNA Library Prep Kit (Cat No. E7645), and the constructed libraries were evaluated on the Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Finally, the library was sequenced on an Illumina NovaSeq platform and 250 bp paired-end reads were generated.

2 Data analysis

2.1 Paired-end reads merged and quality control

Paired-end reads basing on their unique barcodes were assigned to samples and were truncated by cutting off the barcodes and primer sequences, and the reads were merged using FLASH (Version 1.2.11, http://ccb.jhu.edu/software/FLASH/) (Magoč and Salzberg, 2011) to obtain Raw Tags. The obtained Raw Tags were then quality controlled using the fastp (Version 0.20.0) software to obtain high quality Clean Tags. Finally, Clean Tags were compared with the database (Silva database https://www.arb-silva.de/ for 16S/18S, unite database https://unite.ut.ee/ for ITS) using Vsearch (Version 2.15.0) to detect chimeras and remove them to obtain the Effective Tags (Haas et al., 2011).

2.2 ASVs denoise and species annotation

For the obtained Effective Tags previously, the DADA2 module in QIIME2 software (Version QIIME2-202006) was used for noise reduction, and sequences with abundance less than 5 were filtered out to obtain the final ASVs (Amplicon Sequence Variants) and feature tables (Li et al., 2020b). Subsequently, the obtained ASVs were compared with the database using the classify-sklearn module in QIIME2 software to obtain the species information of each ASV. According to the annotated results of ASVs and the characterization table of each sample, the table of species abundance at the level of kingdom, phylum, class, order, family, genus and species were obtained.

2.3 Alpha diversity

In order to analyze the diversity, richness and uniformity of the communities in the sample, QIIME2 software was used to calculate Observed_otus, Chao1, Shannon, Simpson, Dominance, Good's coverage and Pielou_e indices, and to draw dilute Rarefaction Curve and species accumulation box plots. The intergroup differences in alpha diversity were analyzed.

2.4 Beta diversity

QIIME2 software was used to calculate the weighted and unweighted UniFrac distance and R software (Version 3.5.3) was used to plot the PCA, PCoA and NMDS downscaling maps. Among them, PCA and PCoA would call the ade4 and ggplot2 packages in R software. Subsequently, the adonis and anosim functions in QIIME2 software were called to analyze the significance of community structure differences between groups. LEfSe analysis (LDA score threshold: 4) of significantly different species biomarkers between groups was completed using LEfSe software (Version 1.0). MetaStat analysis was performed using R software to obtain p-values for the two comparison groups at six taxonomic levels (phylum, class, order, family, genus and species), and the species with *p*-values less than 0.05 were selected as significantly different species between groups. T-tests analysis was also performed using R software to achieve analysis of significant species differences at each taxonomic level. Further, PICRUSt2 software (Version 2.1.2-b) was applied for functional annotation analysis to study the different functions of the communities in different samples.

References

Haas BJ, Gevers D, Earl AM, et al., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res*, 21(3):494-504. https://doi.org/10.1101/ gr.112730.110

Li MJ, Shao DT, Zhou JC, et al., 2020. Signatures within esophageal microbiota with progression of esophageal squamous cell carcinoma. *Chin J Cancer Res*, 32(6):755-767. https://doi.org/10.21147/j.issn.1000-9604.2020.06.09

Magoč T, Salzberg SL, 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957-2963. https://doi.org/10.1093/bioinformatics/btr507

Table S1 Statistical results obtained at each step during data processing by the illumina NovaSeq Sequencing

Group	Raw PE	Nochime	Base (nt)	AvgLen (nt)	GC (%)
GH.3 M	114542±9718	93881±8280	23761651±2095538	253	53.05±0.71
GH.6 M	102856±17624	94212±18175	23830964±4588420	253	52.31±0.88
GH.12 M	110242±6195	87832±8129	22217317±2058079	253	52.93±1.06
Z.3 M	101265±17029	86059±12355	21779327±3133816	253	53.02±0.51
Z.6 M	102699±24868	95797±23685	24238746±5986063	253	53.01±0.37
Z.12 M	111350±6331	86184±6807	21798639±1721363	253	53.10±0.19
J.0 M	108926±6157	79777±8878	20191534±2267877	253	53.09±0.94

Notes: Raw PE: Raw data for sequencing. Nochime: Splitting, splicing and chimera filtering of raw data to get effective data for analysis (Effective Tags). Base: Base is the number of bases of Effective Tags. AvgLen(nt): Average length of Effective Tags. GC%: GC base content in Effective Tags.

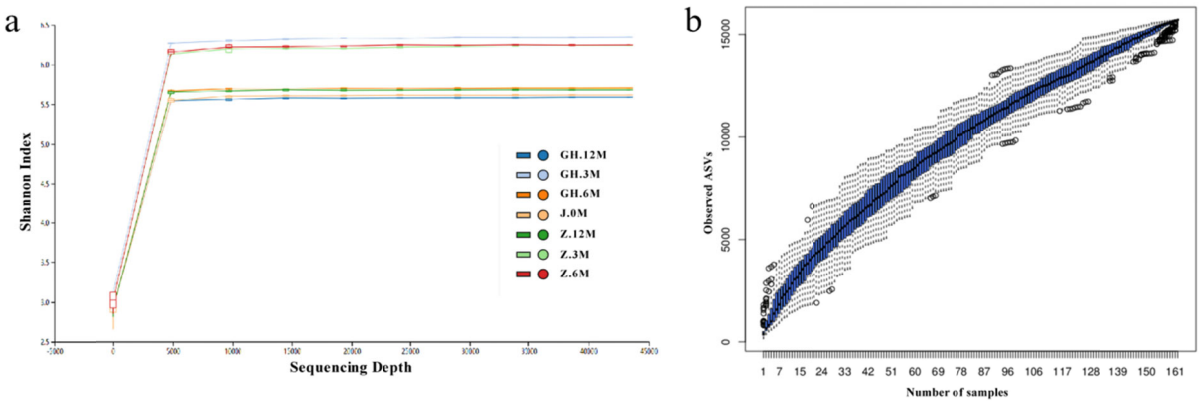

Fig. S1 Species diversity curves. (a) Species Rarefaction Curve, the horizontal axis represents the amount of sequencing data, and the vertical axis represents the corresponding shannon index for each group. (b) Species Accumulation Boxplot, the horizontal coordinate was sample size and the vertical coordinate was the number of ASVs after sampling, and the overall results reflected the rate of new ASVs emergence under continuous sampling.

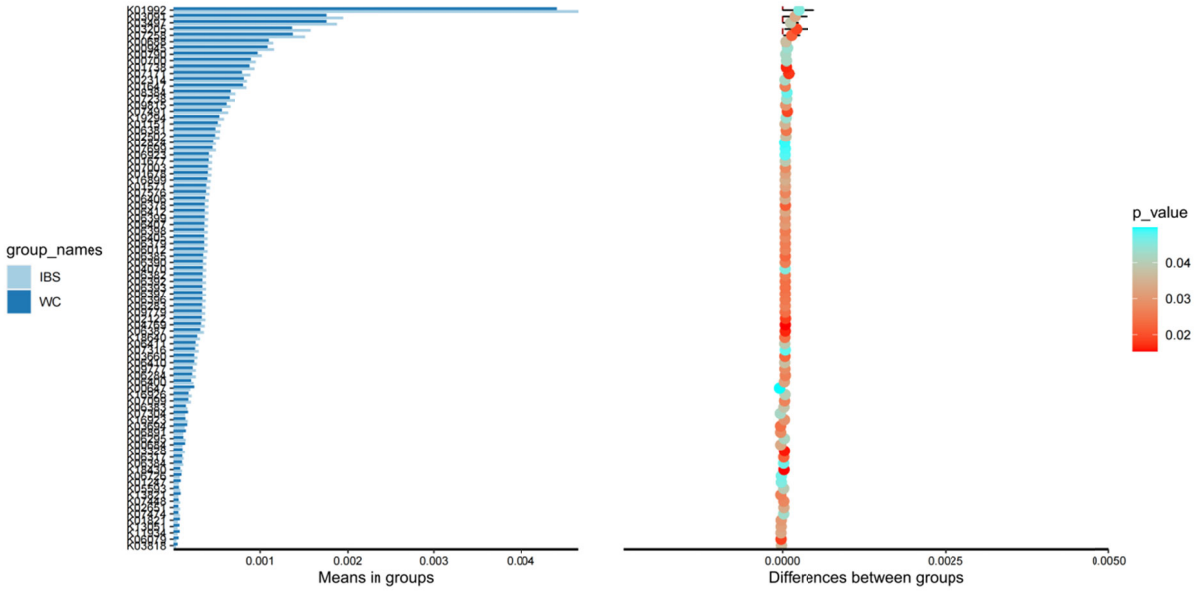

Fig. S2 Functional prediction of differential genes between the IBS and WC groups based on KO database.