Differences in gut microbiota and its metabolic function among different fasting plasma glucose groups in Mongolian population of China

Yanchao Liu
Inner Mongolia Medical University
Mingxiao Wang
Cspc Zhongqi Pharmaceutical Technology (Shijiazhuang) Co., Ltd
Wuyun-tana Li
Inner Mongolia Medical University
Yumin Gao
Inner Mongolia Medical University
Hailing Li
Inner Mongolia Medical University
Ning Cao
Inner Mongolia Medical University
Wenli Hao
Inner Mongolia Medical University
Lingyan Zhao (lingyan_zhao@163.com)
Inner Mongolia Medical University

Research Article

Keywords: type 2 diabetes, metabolic function, gut microbiota, Mongolian population, carotene intake

Posted Date: July 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1786308/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Microbiology on April 15th, 2023. See the published version at https://doi.org/10.1186/s12866-023-02852-7.

Abstract

Background Many studies reported the association between gut microbiota and type 2 diabetes(T2D) but it is still unclear which bacterial genus plays a key role and how the metabolic function of gut microbiota changes in the occurrence and development of T2D. Besides, there is insufficient evidence on the association between gut microbiota and T2D in Mongolian population. This study identified the main bacterial genus influencing the occurrence and development of T2D in Mongolian population, and analyzed the changes of metabolic function of gut microbiota. The association between dietary factors and the relative abundance of main bacterial genus or its metabolic function was also studied.

Methods Dietary surveys and gut microbiota analyses were performed on 24 Mongolian volunteers that were divided into T2D (6 cases), PRET2D (6 cases) and Control group (12 cases) according to the fasting plasma glucose (FPG) values. The relative abundance and metabolic function of gut microbiota from fecal samples of the volunteers were measured by metagenomic analysis. Correlation analysis was used to evaluate the association between dietary factors, and the relative abundance of the main bacterial genus or its metabolic function.

Results This study found that the relative abundance of gut microbiota and its metabolic function differed significantly across groups, and *clostridium* genus may be one of the key bacterial genus affecting the occurrence and development of T2D. First, the relative abundance of *clostridium* genus was significantly different among the three groups. Second, PRET2D or T2D group had a higher relative abundance of metabolic enzymes of most of the gut microbiota than Control group. Third, there was a strong correlation between the relative abundance of *clostridium* genus and that of many metabolic enzymes, which may be produced by the *clostridium* genus. Last, carotene intake was negatively correlated with the relative abundance of *clostridium* genus but positively correlated with tagaturonate reductase of *clostridium* genus, which catalyzed interconversions of pentose and glucuronate.

Conclusions The *clostridium* genus of gut microbiota plays an important role in the development of T2D, and it can be a potential biomarker for T2D in Mongolian population. Meanwhile, the metabolic function of gut microbiota has changed during the early stage of T2D, and the changes of the metabolic function of carbohydrate, amino acid, lipid or energy of *clostridium* genus may play a critical role. In addition, the carotene intake may affect the reproduction and metabolic function of *clostridium* genus.

Backgroud

Diabetes mellitus(DM)is one of the fastest-growing global health emergencies in the 21st century, and 90% of the DM is type 2 diabetes mellitus(T2DM/T2D). In 2021, estimatedly537 million people have diabetes, and this number is projected to reach 643 million by 2030[1]. China is a country with the largest number of diabetic patients, which brings a serious economic burden. The study of the cause of T2D will be beneficial to the prevention and treatment of T2D. Growing evidence indicates the gut microbiota plays an important role in the development of T2D [2], and the gut microbiota has changed before the sugar regulation was impaired [3]. Compared with the healthy people, the patients with diabetes have a moderate disturbance in the gut microbiota, and an increased number of pathogenic bacteria [2]. Similarly, PreT2D population has an abnormal gut microbiota [4]. This suggests that the gut microbiota changes dynamically during the occurrence and development of T2D. Therefore, we divided the population into T2D, preT2D and control group for comparative analysis. Dietary factors can affect the occurrence and development of T2D by influencing gut microbiota [5]. A cohort study revealed that higher fruit intake-associated gut microbiota was associated with a lower risk of T2D [6]. Numerous evidences show that the change in dietary pattern can influence the gut microbial composition and diversity [7]. Meanwhile, some metabolites of gut microbiota are linked with the risk of T2D [8]. In this study, we are concerned with how dietary factors and gut microbiota affect the occurrence and development

of T2D.This study focused not only on the changes of the gut microbiota itself, but also on the changes of the metabolic function of the gut microbiota through metagenomic analysis. Metagenomic analysis can truly reflect the composition and interaction of microbiota in a sample, and can be used to study the metabolic pathway and gene function at the molecular level [9–11]. The results of this study can be a reference for further research of the etiology of T2D, so as to promote the prevention and treatment of the T2D.

Methods

Participants

This study recruited Mongolian volunteers in Inner Mongolia of China, who met the following criteria. (1) Inclusion criteria: More than three generations of pure Mongolian; aged 18 to 79 years; meeting diagnostic criteria of T2D; T2D and PRET2D are the new diagnosed cases; The ratio of men to women is approximately 1:1; No diarrhea in the recent week; Have not received antibiotics for nearly a month; No exposure to radioactive substances and radiation in recent three months; No gastrointestinal diseases; voluntarily participate in the trial and sign the informed consent form. (2) Exclusion criteria: Other types of diabetes, such as type 1 diabetes, gestational diabetes and other special types of diabetes; Other types of endocrine disorders, such as primary aldosteronism, hyperthyroidism; Chronic infectious diseases such as chronic viral hepatitis and tuberculosis; Pregnant or lactating subjects; Subjects with mental diseases; Subjects with acute inflammation and trauma; Serious heart, brain, liver and kidney diseases, such as acute stroke, acute myocardial infarction and severe liver and kidney function damage; Taking hypoglycemic drugs and other drugs for a long time.

Dietary survey

This study collected information through face-to-face interviews. Food Frequency Questionnaire (FFQ) with 69-food items was used for the dietary survey, and duration of the diet retrospect was one year.

Fasting plasma glucose (FPG) test, physical examination and fecal samples collection

Fasting plasma glucose test was completed in a hospital. We Measured the height, weight, waistline, hipline, systolic blood pressure (SBP) and diastolic blood pressure (DBP) of the participants used unified standard measurement tools. Waist to hip ratio (WHR) and body mass index (BMI) were calculated. Meanwhile, fecal samples of the participants were collected for the metagenomic analysis.

Metagenomic analysis

1. Sequencing results pretreatment

The total DNA extracted from fecal samples was constructed into a library and sequenced by Illumina PE150.Preprocessing the Raw Data obtained from the Illumina HiSeq sequencing platform using Readfq (V8, https://github.com/cjfields/readfq) was conducted to acquire the Clean Data for subsequent analysis . The specific processing steps are as follows: 1) remove the reads which contains low quality bases (default quality threshold value < 38) above a certain portion (default length of 40 bp); 2) remove the reads in which the N base has reached a certain percentage (default length of 10 bp); 3) remove reads which shares the overlap above a certain portion with Adapter (default length of 15 bp) [12,13]. Considering the possibility of host contamination in the samples, Clean Data need to be performed sequence alignment with the host database which by default uses Bowtie2.2.4 software (Bowtie2.2.4, http://bowtie- bio.sourceforge.net/bowtie2/index.shtml) to filter the reads that are of host origin, the parameters [13] are as follows: –end-to-end, –sensitive, -I 200, -X 400.

2. Metagenome Assembly

(1) Single sample assembly

The Clean Data was assembled and analyzed [14] by SOAPdenovo software (V2.04, http://soap.genomics.org.cn/soapdenovo.html), the parameters [15] are set as follows: -d 1, -M 3, -R, -u, -F, -K 55. Then interrupt the assembled Scaftigs from N connection and leave the Scaftigs without N. All samples' Clean Data are compared to each Scaffolds respectively by Bowtie2.2.4 software to acquire the PE reads not used and the parameters are: --end-to-end, --sensitive, -I 200, -X 400[16,17].

(2) Mixed assembly

All the reads not used in the forward step of all samples are combined and then use the software of SOAPdenovo (V2.04) for mixed assembly with the parameters same as single assembly[13,14]. Break the mixed assembled Scaffolds from N connection and obtained the Scaftigs. Filter the fragment shorter than 500 bp in all of Scaftigs for statistical analysis both generated from single or mixed assembly.

3. Gene prediction and abundance analysis

(1) The Scaftigs (> 500 bp) assembled from both single and mixed are all predicted the ORF by MetaGeneMark (V2.10, http://topaz.gatech.edu/GeneMark/) software, and filtered the length information shorter than 100nt from the predicted result with default parameters [15].

(2) For ORF predicted, CD-HIT software (V4.5.8, http://www.bioinformatics.org/cd-hit)[18]is adopted to redundancy and obtain the unique initial gene catalogue (the genes here refers to the nucleotide sequences coded by unique and continuous genes), the parameters option are -c 0.95, -G 0, -aS 0.9, -g 1, -d 0[19].

(3) The Clean Data of each sample is mapped to initial gene catalogue using Bowtie2.2.4 and get the number of reads to which genes mapped in each sample with the parameter setting are -end-to-end, --sensitive, -I 200, -X 400[20]. Filter the gene which the number of reads < 2 in each sample and obtain the gene catalogue (Unigenes) eventually used for subsequently analysis.

(4) Based on the number of mapped reads and the length of gene, statistic the abundance information of each gene in each sample [21].

(5) The basic information statistic, core-pan gene analysis, correlation analysis of samples and venn figure analysis of number of genes are all based on the abundance of each gene in each sample in gene catalogue.

4. Taxonomy prediction

(1) DIAMOND [22] software (V0.9.9, https://github.com/bbuchfink/diamond/) is used to blast the Unigenes to the sequences of Bacteria, Fungi, Archaea and Viruses which are all extracted from the NR database (Version: 2018-01-02, https://www.ncbi.nlm.nih.gov/) of NCBI with the parameter setting are blastp, -e 1e-5. (2) For the finally aligned results of each sequence, as each sequence may have multiple aligned results, choose the result of which the e value < the smallest e value * 10 to take the LCA algorithm which is applied to system classification of MEGAN software to make sure the species annotation information of sequences [23].

(3) The table containing the number of genes and the abundance information of each sample in each taxonomy hierarchy (kingdom, phylum, class, order, family, genus, species) are obtained based on the LCA annotation result and

the gene abundance table. The abundance of a specie in one sample equal the sum of the gene abundance annotated for the specie. The gene number of a specie in a sample equal the number of genes whose abundance are nonzero.

(4) The exhibition of generation situation of relative abundance, the exhibition of abundance cluster heat map, and NMDS (R vegan package, Version 2.15.3) decrease-dimension analysis are based on the abundance table of each taxonomic hierarchy. LEfSe analysis is used to look for the different species between groups. Permutation test between groups is used in Metastats analysis for each taxonomy and get the P value, then use Benjamini and Hochberg False Discovery Rate to correct P value and acquire q value. LEfSe analysis is conducted by LEfSe software (the default LDA score is 4). Finally, random forest (RandoForest) (R pROC and randomForest packages, Version 2.15.3) was used to construct a random forest model. Screen out important species by MeanDecreaseAccuracy and MeanDecreaseGin, then cross-validate each model (default 10 times) and plot the ROC curve.

5. Common functional database annotations

(1) Adopt DIAMOND software (V0.9.9) to blast Unigenes to KEGG functional database (Version 2018-01-01, http://www.kegg.jp/kegg/)with the parameter setting of blastp, -e 1e-5 [20,24]. For each sequence's blast result, the best Blast Hit is used for subsequent analysis.

(2) Statistic of the relative abundance of different functional hierarchy, the relative abundance of each functional hierarchy equal the sum of relative abundance annotated to that functional level.

(3) Based on the function annotation result and gene abundance table, the gene number table of each sample in each taxonomy hierarchy is obtained. The gene number of a function in a sample equal the gene number that annotated to this function and the abundance is nonzero.

(4) Based on the abundance table of each taxonomy hierarchy, not only the counting of annotated gene numbers, the exhibition of the general relative abundance situation, the exhibition of abundance cluster heat map are conducted, but also the Anosim analysis of the difference between groups (inside) based on functional abundance, comparative analysis of metabolic pathways between groups are performed.

Statistical analysis

SPSS26 and R software were used for statistical analysis. The measurement data of normal distribution is expressed by mean ± standard deviation, which meets the conditions of parameter test. The data are compared between groups by one-way ANOVA. The data that did not meet the conditions of parameter test were expressed by median (interquartile interval), and the comparison between groups was analyzed by rank sum test and multi classification logistic regression. Chi square test was used for counting data. Spearman nonparametric inertia analysis is used for correlation analysis. When $*p < 0.05$ $**p < 0.01$ and $***p < 0.001$ the differences were statistically significant. Same figures were completed using the Wekemo Bioincloud (https:// www.bioincloud.tech).

Results

Characteristics of study participants

This study enrolled 24 qualified volunteers, who were divided into diabetes group (T2D, 6 cases), prediabetes group (PRET2D, 6 cases) and normal glucose control group (Control, 12 cases) according to the FPG values. The sex ratio of men to women of the three groups was 1:1. We found statistically significant differences in waistline of the participants in the three groups, and the waistline is positively correlated with FPG (r=0.434, p=0.034) table 1 .

Table 1 Characteristics of participants in the three groups

Group	FPG, mmol/L	Age, years	BMI	Waistlin, cm	Hipline, cm	WHR	SBP, mmHg	DBP, mmHg
T2D	10.8±5.5	61.3±7.7	27.4±2.5	101.3±6.8	104.2±6.1	1.0±0.0	132.3±17.1	82.2±12.4
PRET2D	6.4±0.3	59.2±8.5	26.8±1.3	95.8±6.2	102.3±5.6	0.9±0.1	134.8±15.7	88.5±6.4
Control	5.6±0.3	56.3±8.3	24.9±3.4	89.8±8.3	98.9±6.6	0.9±0.1	129.8±16.9	87.3±10.3
F value	7.738	0.816	1.941	4.922	1.547	2.648	0.192	0.700
p value	0.003	0.456	0.168	0.018	0.236	0.094	0.826	0.508

Metagenomic analysis

1.Species analysis

(1) Number of Genes

We measured the number of genes in T2D, PRET2D and Control group, analyzed the common and unique characteristic of the genes among the different groups, and drew a flower map showed as below. At the same time, the Control group was divided into Control 1(5.4±0.2 mmol/L) and Control 2(5.8±0.1 mmol/L) group according to the mean value of FPG, and the genetic differences of the four groups were shown by venn map(figure 1). When the FPG value rises or reaches the value of pre T2D state, the number of specific genes decreases, but that increases again in T2D state. These difference of genes in different groups may be the result of changes in the structure of gut microbiota.

(2) Analysis of the relative abundance of microbial species

The relative abundance of kingdom, phylum, class, order, family, genus and species were analyzed, and the top 35 species with the highest relative abundance (top 35) were shown. We found that there was no significant difference among the three groups in seven classification hierarchies except *clostridiaceae* family *k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae* (p=0.045) and *clostridium* genus *k__Bacteria;p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium* (p=0.049) (figure 2). The relative abundance of *clostridiaceae* family and *clostridium* genus all decreased in PRET2D group, but increased in T2D group, this difference was similar to that of above referenced unique genes.

(3) LEfSe analysis of different species among groups

In order to screen bacterial biomarkers with significant difference among groups, we measured different species among groups by rank sum test, reduced the dimension by LDA (linear discriminant analysis) and evaluated the influence of different species through LDA score. The analysis shows that *clostridium* genus has the highest LDA score in T2D group, and it can be used as a potential bacterial biomarker for T2D (figure 3).

2.Analysis of function

(1) Statistics on the number of genes annotated in the KEGG database

Unigenes were blasted against the KEGG database. Results showed most of the genes annotated in the first level of KEGG database were metabolism related genes (figure 4). In this study, we focused on the difference of the metabolic function of gut microbiotic among T2D, PRET2D and Control groups.

(2)Functional analysis

Functional diffrence among the three groups in level 1,level 2 and level 3 of KEGG database was analyzed. There was no significant functional diffrence in level 1,while digestive system in level 2 belonging to organismal systems was significantly diffrent among groups (PRET2D>Control>T2D, p=0.034). These results may suggest that the digestive function of gut microbiota increases first and then decreases in the development of T2D. Seven significant functional differences in level 3 were observed, including five differeces in metabolic function (ko00600, ko00361, ko00791, ko00941 and ko00945) (table 2). The trend of difference of ko00600 among the three groups is consistent with that of digestive system,and other four metabolic functions are the highest in T2D group (figure 5).

Table 2 Seven significant functional differences in level 3, annotation and *p* value

Name	Level1	Level2	Level3	*p* value
ko00600	Metabolism	Lipid metabolism	Sphingolipid metabolism	0.045
ko00361	Metabolism	Xenobiotics biodegradation and metabolism;	Chlorocyclohexane and chlorobenzene degradation	0.043
ko00791	Metabolism	Xenobiotics biodegradation and metabolism	Atrazine degradation	0.022
ko00941	Metabolism	Biosynthesis of other secondary metabolites	Flavonoid biosynthesis	0.026
ko00945	Metabolism	Biosynthesis of other secondary metabolites;	Stilbenoid, diarylheptanoid and gingerol biosynthesis	0.026
ko04213	Organismal Systems;	Aging	Longevity regulating pathway - multiple species	0.046
ko05150	Human Diseases	Infectious diseases: Bacterial	Staphylococcus aureus infection	0.019

(3) Difference analysis of the relative abundance of metabolic enzymes

In gut microbiota, 35 metabolic enzymes show significant difference in relative abundance among the 3 groups, and most of that are involved in carbohydrate metabolism, amino acid metabolism and metabolism of cofactors and vitamins. Some of the 35 enzymes are involved in multiple metabolic pathways (table 3). The relative abundance of most enzymes is either the highest in T2D group or the highest in PRET2D group. The relative abundance of tagaturonate reductase catalyzing the interconversions of pentose and glucuronate (1.1.1.58) is gradually declined in Control group, PRET2D group and T2D group sequentially. Therefore, we may come to a conclusion that the metabolic function of gut microbiota changed significantly in the T2D stage and even in the early stage of T2D (figure 6).

Table 3 Name, function and *p* value of the 35 gut microbiotic enzymes that show significant difference between T2D, PRET2D and Control group

Level2	EC ID	Name	Function	*p* value
Carbohydrate metabolism	1.1.1.58	tagaturonate reductase	pentose and glucuronate interconversions.	0.049
	2.3.1.182	Transferases	C5-Branched dibasic acid metabolism; Valine, leucine and isoleucine biosynthesis.	0.047
	2.4.2.53	undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	Amino sugar and nucleotide sugar metabolism.	0.015
	2.6.1.19	4-aminobutyrate-2-oxoglutarate transaminase	Butanoate and Propanoate metabolism; Alanine, aspartate, glutamate, beta-Alanine metabolism.	0.015
	2.7.1.1	Hexokinase	Starch, Sucrose, Glycolysis, Gluconeogenesis, Galactose, Fructose and mannose, Amino sugar and nucleotide sugar metabolism; Streptomycin, Neomycin, kanamycin, gentamicin biosynthesis.	0.022
	2.7.1.55	allose kinase	Fructose and mannose metabolism.	0.027
	2.8.3.12	glutaconate CoA-transferase	Butanoate metabolism; Styrene degradation.	
	3.5.99.6	glucosamine-6-phosphate deaminase	Amino sugar and nucleotide sugar metabolism.	0.027
	4.1.2.17	L-fuculose-phosphate aldolase	Fructose and mannose metabolism.	0.017
Amino acid metabolism	1.1.1.103	L-threonine 3-dehydrogenase	Glycine, serine and threonine metabolism.	
	2.1.1.104	caffeoyl-CoA O-methyltransferase	Phenylalanine metabolism; Flavonoid biosynthesis; Phenylpropanoid biosynthesis; Stilbenoid, diarylheptanoid, gingerol biosynthesis.	0.026
	2.8.1.2	3-mercaptopyruvate sulfurtransferase	Cysteine and methionine metabolism.	0.045
	3.3.2.12	oxepin-CoA hydrolase	Phenylalanine metabolism	0.033
	3.5.1.5	Urease	Arginine biosynthesis; Nucleotide metabolism; Purine metabolism; Atrazine degradation.	0.017
	3.5.2.14	N-methylhydantoinase (ATP-hydrolysing)	Arginine and proline metabolism.	0.045
	4.4.1.1	cystathionine gamma-lyase	Cysteine and methionine metabolism; Glycine, serine and threonine metabolism; Selenocompound metabolism.	0.007

	5.1.1.4	proline racemase	Arginine and proline metabolism	0.020
	6.1.1.13	D-alanine-poly(phosphoribitol) ligase	D-Alanine metabolism.	0.042
	6.3.2.1	pantoate-beta-alanine ligase (AMP-forming)	beta-Alanine metabolism; Pantothenate and CoA biosynthesis.	0.018
Metabolism of cofactors and vitamins	2.1.2.3	phosphoribosylaminoimidazolecarboxamide formyltransferase	One carbon pool by folate.	0.046
	2.1.2.11	3-methyl-2-oxobutanoate hydroxymethyltransferase	Pantothenate and CoA biosynthesis.	0.026
	3.5.4.25	GTP cyclohydrolase II	Riboflavin metabolism; Folate biosynthesis;	0.038
	3.6.1.67	dihydroneopterin triphosphate diphosphatase	Folate biosynthesis.	0.043
	4.1.99.12	3,4-dihydroxy-2-butanone-4-phosphate synthase	Riboflavin metabolism.	0.036
	4.1.99.19	2-iminoacetate synthase	Thiamine metabolism.	0.030
	4.3.99.3	7-carboxy-7-deazaguanine synthase	Folate biosynthesis.	0.027
Lipid metabolism	2.4.1.337	1,2-diacylglycerol 3-alpha-glucosyltransferase	Glycerolipid metabolism.	0.017
	3.1.4.46	glycerophosphodiester phosphodiesterase	Glycerophospholipid metabolism.	0.013
	4.2.1.59	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	Fatty acid biosynthesis.	0.020
Nucleotide metabolism	3.5.4.10	IMP cyclohydrolase	Purine metabolism.	0.047
Energy metabolism	1.8.5.3	respiratory dimethylsulfoxide reductase	Sulfur metabolism.	0.011
	2.1.1.246	[methyl-Co (III) methanol-specific corrinoid protein]-coenzyme M methyltransferase	Methane metabolism.	0.023
Metabolism of terpenoids and polyketides	2.6.1.33	dTDP-4-amino-4,6-dideoxy-D-glucose transaminase	Polyketide sugar unit biosynthesis; Biosynthesis of other secondary metabolites Acarbose and validamycin biosynthesis.	0.007
	5.3.3.2	sopentenyl-diphosphate Delta-isomerase	Terpenoid backbone biosynthesis.	0.013
Xenobiotics metabolism	3.5.4.43	hydroxydechloroatrazine ethylaminohydrolase	Atrazine degradation.	0.032

3.Correlation analysis of dietary factors, *clostridium* genus and metabolic enzymes

(1) Correlation analysis of dietary factors and *clostridium* genus

We calculated the average daily intake of food or nutrients for each participant. The differences of carotene intake in 3 groups were statistically significant(481.3±216.9μg *vs* 444.0±170.3μg *vs* 666.0±183.8μg) p=0.048 , the carotene intake and seafood intake showed a negative correlation with FPG (r=-0.503, p=0.012; r=-0.570, p=0.004). Meanwhile, the carotene intake and the relative abundance of *clostridium* genus was negatively correlated (r=-0.423, p=0.039) (figure 7).

(2) **Correlation analysis of dietary factors, *clostridium* genus and metabolic enzymes**

Among the above 35 metabolic enzymes, 19 may be produced by *clostridium* genus. We analyzed correlations of the relative abundance of the19 metabolic enzymes with *clostridium* genus and other factors that may be relevant, including age, FPG, BMI, waistline, hipline, WHR, DBP, SBP, carotene intake, beans intake and fruits intake (figure 8). There are ten metabolic enzymes significantly correlated to *clostridium* genus, among which three have negative and seven have positive correlations. The three enzymes that are negatively correlated with *clostridium* genus are tagaturonate reductase (1.1.1.58), glucosamine-6-phosphate deaminase (3.5.99.6) and transferases (2.3.1.182) respectively, which are all involved in carbohydrate metabolism. The seven enzymes that have positive correlation with *clostridium* genus are glutaconate CoA-transferase (2.8.3.12), 1,2-diacylglycerol 3-alpha-glucosyltransferase (2.4.1.337), [methyl-Co(III) methanol-specific corrinoid protein]-coenzyme M methyltransferase (2.1.1.246), Urease (3.5.1.5), L-threonine 3-dehydrogenase (1.1.1.103), undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase (2.4.2.53) and L-fuculose-phosphate aldolase (4.1.2.17) respectively, some of the seven enzymes involved in carbohydrate metabolism, amino acid metabolism, some participating in lipid metabolism and energy metabolism. Carotene intake is positively correlated with tagaturonate reductase (1.1.1.58) that catalyzes the interconversions of pentose and glucuronate. Fruits intake show a negative correlation with respiratory dimethylsulfoxide reductase (1.8.5.3), dihydroneopterin triphosphate diphosphatase (3.6.1.67) and 3-hydroxyacyl-[acyl-carrier-protein] dehydratase (4.2.1.59), that participate in the sulfur metabolism, folate biosynthesis and fatty acid biosynthesis respectively. Beans intake is related to amino acid metabolic enzymes (3.5.2.14, 6.3.2.1). Blood pressure is negatively correlated with glycerolipid metabolic enzymes (2.4.1.337). Waistline, hipline and BMI are also have correlation with metabolic enzymes.

Disscussion

In order to identify intestinal microbial species which are closely associated with the T2D, we paid more attention to species and genus difference among the three groups. We found *clostridium* genus difference exits in the three groups, which is consistent with other studies[2, 25]. Therefore, *clostridium* genus plays an important role in the development of T2D. Specifically, compared with Control group, the abundance of *clostridium* genus decreased in PRET2D group and increased in T2D group, which indicated that some species of *clostridium* genus undergone complex changes in the development of T2D, for example some pathogenic species increased and some beneficial species decreased in abundance. Similar phenomena have been found in Allin's and Zhong's studies [26, 27]. Besides, we found that the changes of the metabolic function of gut microbiota were significantly related to T2D, and the changes of the metabolic function of carbohydrate played an important role, which was consistent with other studies [28]. In addition, we also analyzed the association between carotene intake, fruits intake, beans intake and gut microbiota. The carotene intake may affect the reproduction and metabolic function of *clostridium* genus. The T2D patients and PreT2D participants are all new diagnosis, excluding the interference of drug treatment and other diseases. Limited by the sample size, there may be sampling errors in this study, and the research results need to be further confirmed by more population experiments.

Conclusions

There is a significant difference in the relative abundance and metabolic function of gut microbiota among T2D, PRET2D and Control groups. And *clostridium* genus can be a potential biomarker for T2D in Mongolian population. Meanwhile, compared with Control group, PRT2D and T2D group have an increased metabolic function of gut microbiota, and the changes of the metabolic function of carbohydrate, amino acid, lipid and energy of *clostridium* genus may play an important role. In addition, the carotene intake may negatively regulate the relative abundance of *clostridium* genus but positively regulate tagaturonate reductase of *clostridium* genus.

Declarations

Acknowledgements

We thank all participants involved in the study, and thank Wuhan metware Biotechnology Co., Ltd for samples detection.

Authors' contributions

Yanchao Liu and Mingxiao Wang performed the bioinformatics analysis and drafted the manuscript; Yanchao Liu,Wuyun-tana Li, Yumin Gao and Hailing Li performed the dietary survey, physical examination and fecal samples collection; Ning Cao and Wenli Hao performed data management and preparation of samples; Lingyan Zhao designed and supervised the study . All authors reviewed the final manuscript.

Founding

This study was founded by "Science and Technology Million Project" of Inner Mongolia Medical University (YKD2020KJBW005) and National Natural Science Foundation of China (81660564).

Availability of data and materials

The detail data and materials available please see http://www.ebi.ac.uk/arrayexpress/help/FAQ.html#cite.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of the Inner Mongolia Medical University (reference number: YKD2016066, signed on 07/03/2016. All participants provided written informed consent. This work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for Experiments in Humans

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

References

1. International Diabetes Federation. IDF Diabetes Atlas.10th ed. Brussels International Diabetes Federation,2021

2. Zhou Z, Sun B, Yu D, Zhu C. Gut Microbiota: An Important Player in Type 2 Diabetes Mellitus. Front Cell Infect Microbiol.2022;12:834485.
3. Wang L, Yu X, Xu X, Ming J, Wang Z, Gao B, Xing Y, Zhou J, Fu J, Liu T et al. The Fecal Microbiota Is Already Altered in Normoglycemic Individuals Who Go on to Have Type 2 Diabetes. Front Cell Infect Microbiol.2021;11:598672.
4. Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature.2019;569(7758):663–671.
5. Liu Y, Lou X. Type 2 diabetes mellitus-related environmental factors and the gut microbiota: emerging evidence and challenges. Clinics (Sao Paulo). 2020;75: e1277.
6. Jiang Z, Sun TY, He Y, Gou W, Zuo LS, Fu Y, Miao Z, Shuai M, Xu F, Xiao C et al. Dietary fruit and vegetable intake, gut microbiota, and type 2 diabetes: results from two large human cohort studies. BMC Med.2020;18(1):371.
7. Sharma S, Tripathi P. Gut microbiome and type 2 diabetes: where we are and where to go? J Nutr Biochem. 2019;63:101-108.
8. Zhu T, Goodarzi MO. Metabolites Linking the Gut Microbiome with Risk for Type 2 Diabetes. Curr Nutr Rep. 2020;9(2):83-93.
9. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol.1998;5(10): R245-249.
10. Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet. 2005;6(11):805-814.
11. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, et al. Comparative metagenomics of microbial communities. Science. 2005;308(5721):554-557.
12. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, Bäckhed F, Nielsen J. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun. 2012;3:1245.
13. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J, Bäckhed F. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99-103.
14. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18.
15. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L et al. Alterations of the human gut microbiome in liver cirrhosis. Nature. 2014;513(7516):59-64.
16. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014;10(11):766.
17. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65.
18. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-3152.
19. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A et al. Ocean plankton. Structure and function of the global ocean microbiome. Science. 2015;348(6237):1261359.
20. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32(8):834-841.
21. Villar E, Farrant GK, Follows M, Garczarek L, Speich S, Audic S, Bittner L, Blanke B, Brum JR, Brunet C et al. Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. Science. 2015;348(6237):1261447.

22. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60.
23. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 2011;21(9):1552-1560.
24. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199-205.
25. Chen PC, Chien YW, Yang SC. The alteration of gut microbiota in newly diagnosed type 2 diabetic patients. Nutrition. 2019;63-64:51-56.
26. Allin KH, Tremaroli V, Caesar R, Jensen BAH, Damgaard MTF, Bahl MI, Licht TR, Hansen TH, Nielsen T, Dantoft TM et al. Aberrant intestinal microbiota in individuals with prediabetes. Diabetologia. 2018;61(4):810-820.
27. Zhong H, Ren H, Lu Y, Fang C, Hou G, Yang Z, Chen B, Yang F, Zhao Y, Shi Z, et al. Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. EBioMedicine. 2019; 47:373-383.
28. Wang TY, Zhang XQ, Chen AL, Zhang J, Lv BH, Ma MH, Lian J, Wu YX, Zhou YT, Ma CC et al. A comparative study of microbial community and functions of type 2 diabetes mellitus patients with obesity and healthy people. Appl Microbiol Biotechnol. 2020;104(16):7143-7153.

Figures

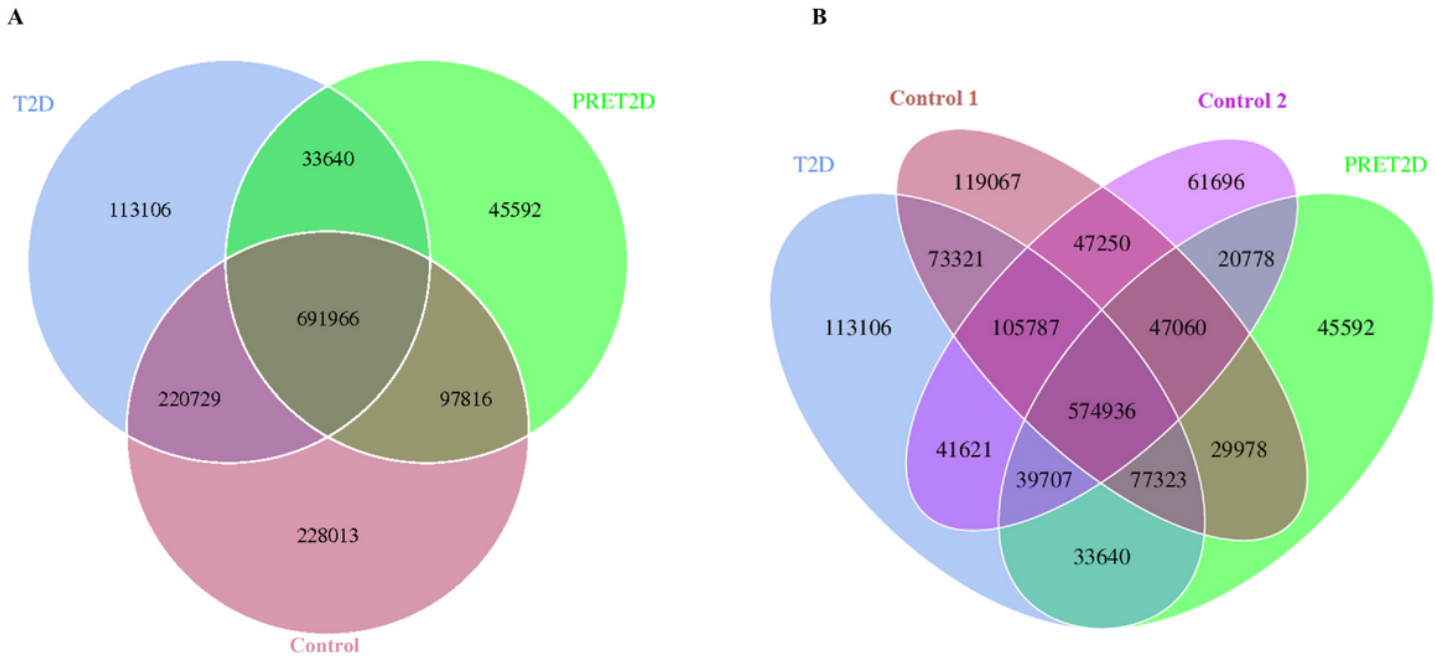

Figure 1

Common and unique characteristic of genes among the different groups. A Comparison of T2D, PRET2D and Control groups; B Comparison of T2D, PRET2D, Control 1 and Control2 groups.

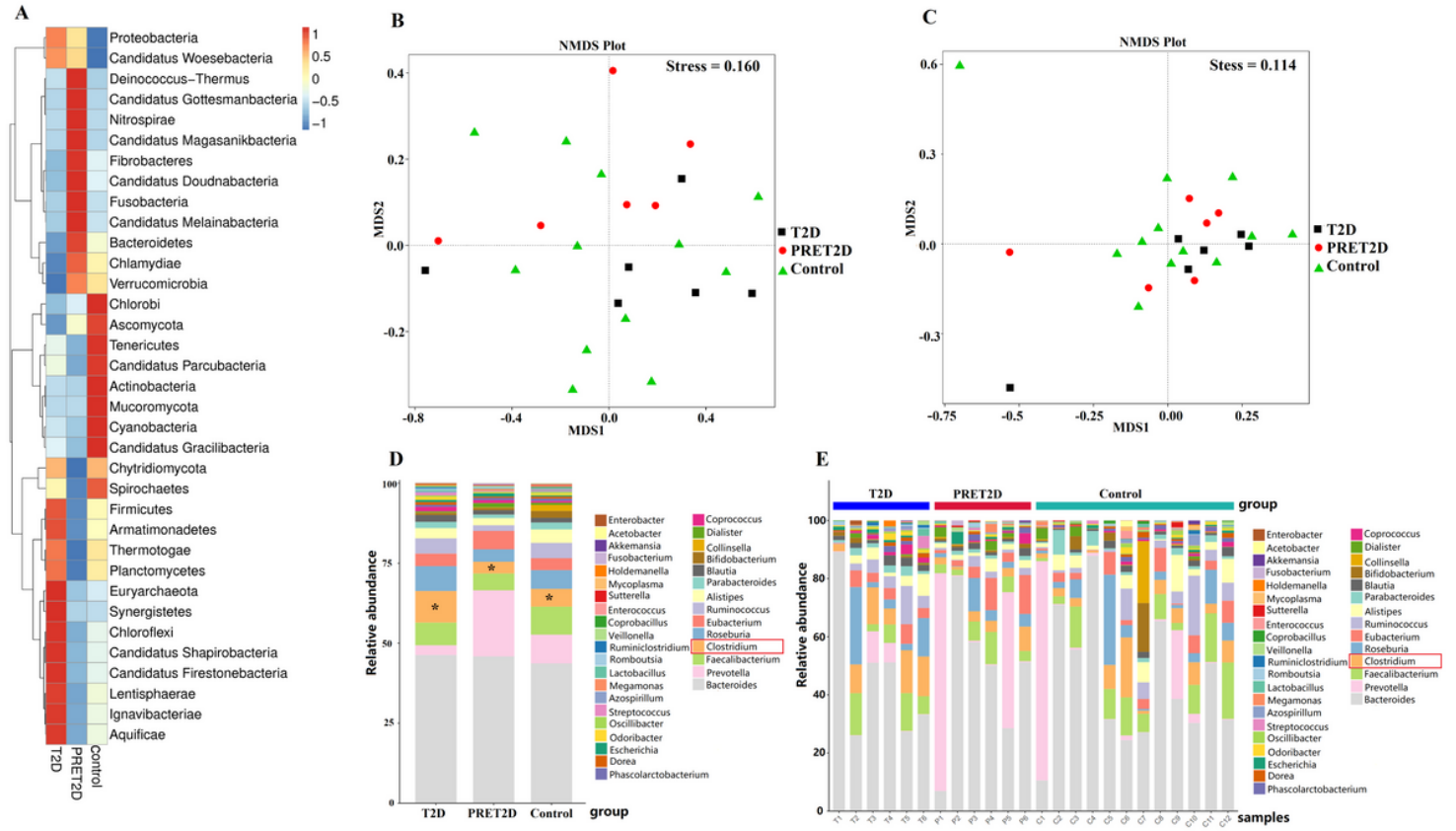

Figure 2

Results of species abundance analysis in phylum and genus hierarchies. **A** Cluster heat map of abundance in phylum in the three groups. **B** NMDS decrease-dimension analysis in phylum in the three groups. **C** NMDS decrease-dimension analysis in genus in the three groups. **D** Abundance comparison in genus, and the abundance of *clostridium* genus differed significantly in the three groups. **E** Abundance comparison of each sample in genus.

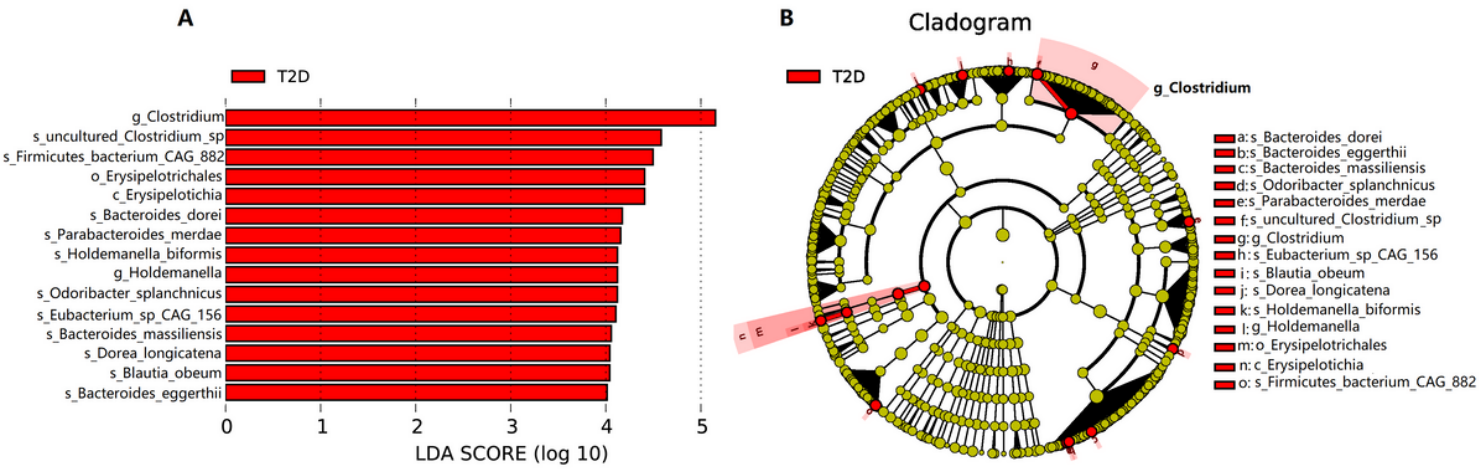

Figure 3

Results of the LEfSe analysis. **A** The distribution diagram of LDA score of different species, the length of the histogram represents the influence of different species, and the species whose LDA score is higher than the set value of 4 can be regarded as biomarkers with statistical difference among groups. **B** The evolutionary branch diagram of different

species, and the circle radiating from inside to outside represents the classification hierarchy from phylum to species. Each small circle in different classification hierarchies represents a classification at this hierarchy. The diameter of the small circle is directly proportional to the relative abundance. The species with no significant difference are uniformly colored as yellow , and the species with significant difference which can be used as a biomarker for T2D are colored as red . The red node indicates the microbial flora that play an important role in the red group.

Figure 4

Number of genes that were annotated in the first level of KEGG database, most of which were metabolism related genes.

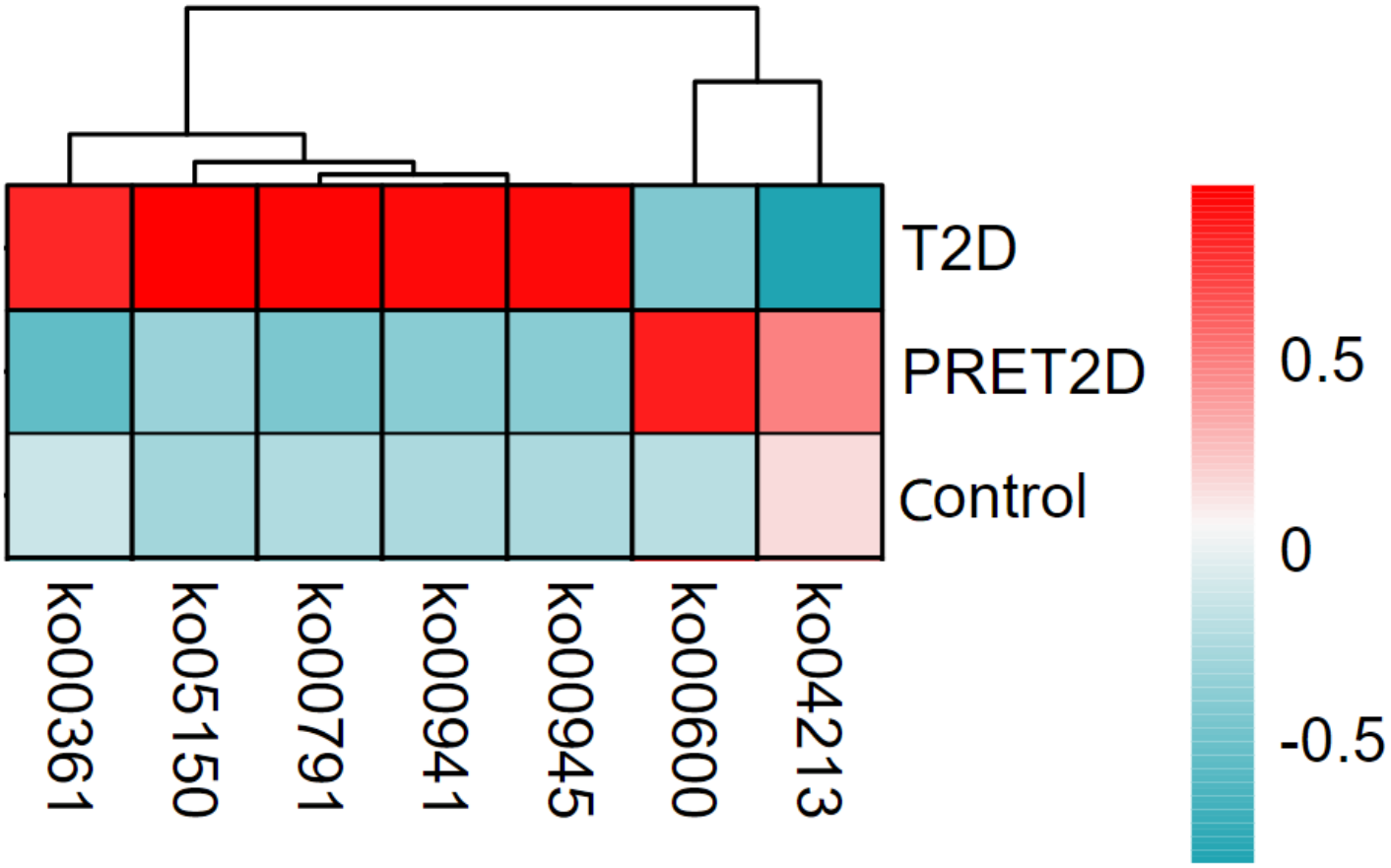

Figure 5

The significant functionnal differences in level 3. All of the seven different functions are higher in T2D or PRET2D group than that in Control group.

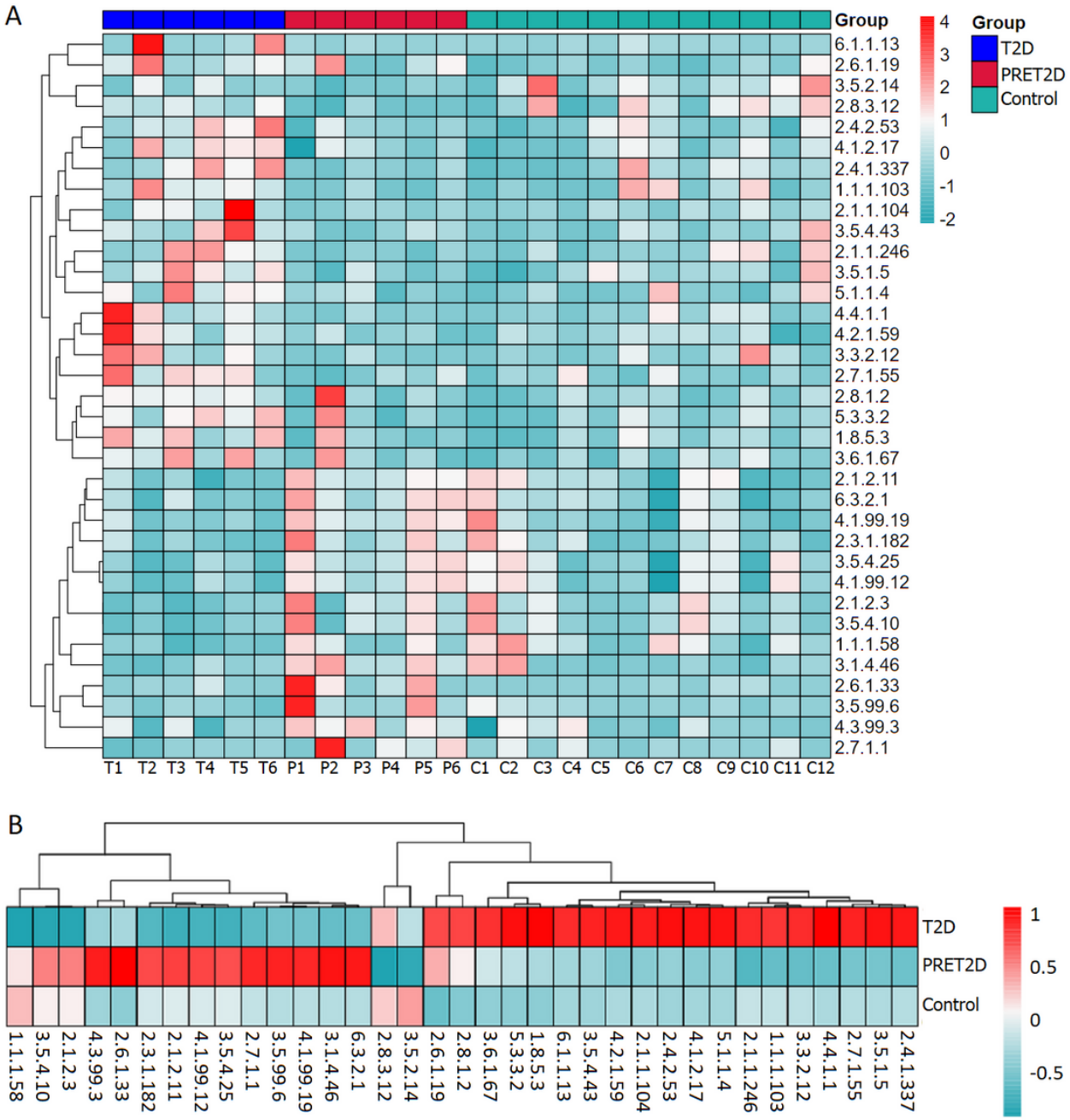

Figure 6

Relative abundances of 35 metabolic enzymes in gut microbiota show significant difference between T2D, PRET2D and Control group. **A** Clustering heat map of the 35 metabolic enzymes in each sample. **B** Clustering heat map of the 35 metabolic enzymes in the three groups. Almost all the enzymes with the highest relative abundance were fromT2D group or PRET2D group.

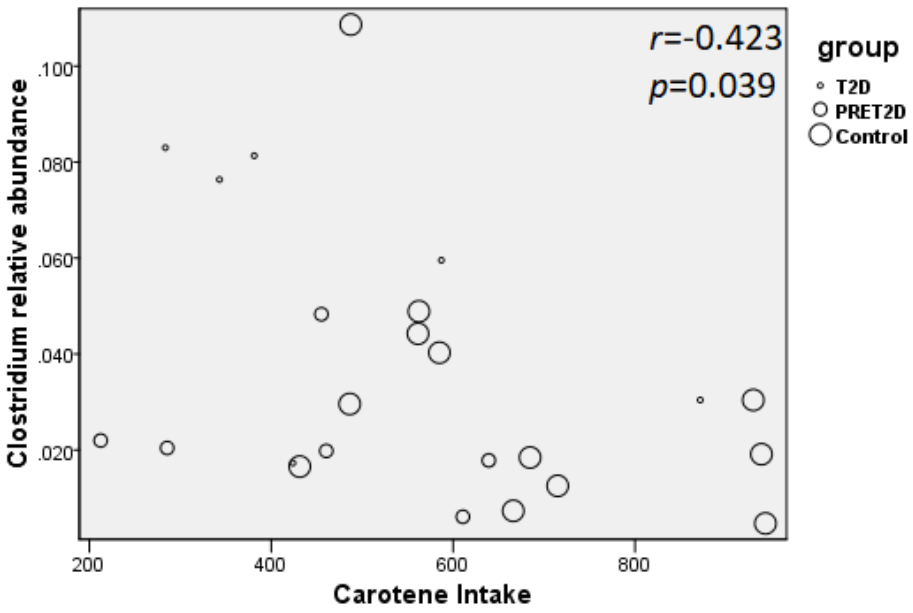

Figure 7

Correlation between the carotene intake and the relative abundance of *clostridium* genus, the distribution trend of the samples from the three groups of showed negative correlation.

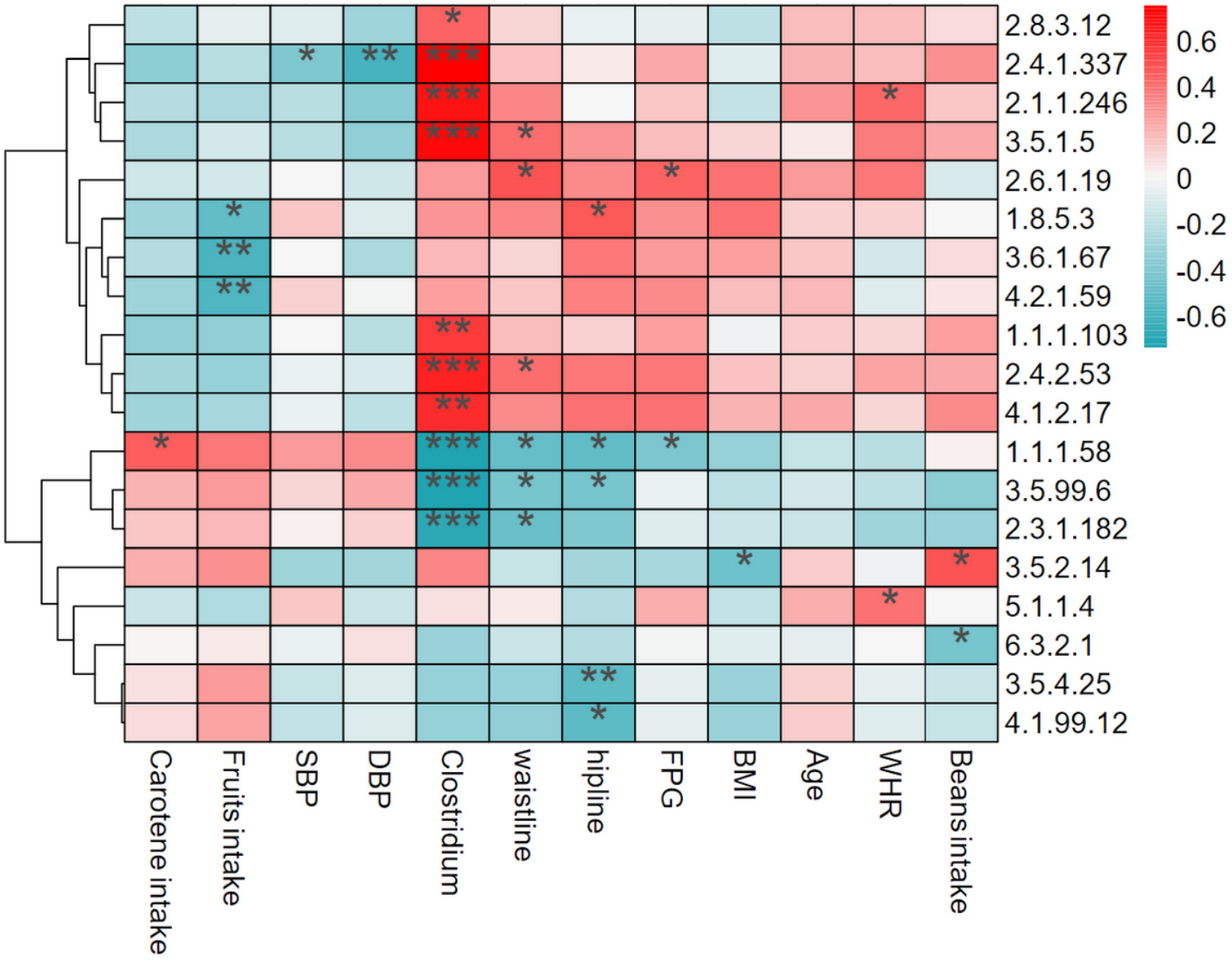

Figure 8

Correlations of 19 metabolic enzymes with *clostridium* genus and other factors.