

Gut Microbiome-Based Diagnostic Model to Predict Diabetes Mellitus

Hai-Tao Yang

First Affiliated Hospital of Xinjiang Medical University

Wen-Juan Xiu

First Affiliated Hospital of Xinjiang Medical University

Jing-Kun Liu

First Affiliated Hospital of Xinjiang Medical University

Yi Yang

First Affiliated Hospital of Xinjiang Medical University

Ying-Ying Zheng

First Affiliated Hospital of Xinjiang Medical University

Xian-Geng Hou

First Affiliated Hospital of Xinjiang Medical University

Ting-Ting Wu

First Affiliated Hospital of Xinjiang Medical University

chengxin wu

First Affiliated Hospital of Xinjiang Medical University

Xiang Xie (✉ xiangxie999@sina.com)

First Affiliated Hospital of Xinjiang Medical University

Research Article

Keywords: intestinal microorganisms, diabetes, diagnostic model, fasting blood glucose

Posted Date: June 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-650888/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Bioengineered on December 19th, 2021.
See the published version at <https://doi.org/10.1080/21655979.2021.2009752>.

Abstract

Introduction: In this study, our aim is to describe the intestinal microbiome structure of diabetes, and to establish a diagnostic model of diabetes mellitus based on intestinal microorganisms.

Methods: The experimental group included 44 patients with diabetes, while the control group included 47 healthy patients without statistical differences in age, gender, smoking, drinking alcohol and other basic diseases after matching. Firstly, their population information, biochemical indicators, and stool samples were collected. Subsequently, the stool samples were conducted the next-generation sequencing of the 16S rRNA genes of fecal bacteria.

Results: In this study, there were significant differences in alpha-diversity and beta-diversity between the two groups. Moreover, the differences within the group were analyzed by Adonis, finding that $R^2=0.031$ and $P=0.001$. Through the LEfSe differential bacteria analysis, it was found that the distribution of differential bacteria between the two groups was significantly abnormal. Among them, the abundances of the genus *faecalibacterium*, the genus *Prevotella*, and the genus *Roseburia* were higher in the diabetes group, while the abundances of the genus *Shigella* and the genus *Bifidobacterium* were lower. In the correlation analysis between bacteria and clinical indicators, it was found that the genus *Veillonella* and the genus *unclassified_Enterobacteriaceae* were negatively related to blood glucose, while the genus *Phascolarctobacterium*, the genus *unidentified_Bacteroidales* and the genus *Prevotella* were significantly positively correlated with fasting blood glucose. Through 10-fold cross-validation of the random forest model, a total of 12 microbial markers were detected, and the area under the curve (AUC) was 84.1%. Notably, we conducted the combined analysis on the intestinal microbial markers and clinical variables, and found that the AUC was 90.8 %.

Conclusions: In this study, it was found that the host intestinal microorganisms could predict the occurrence of diabetes, and its predictive ability was no less than that of fasting blood sugar. It can be seen that the intestinal microorganisms are closely related to the occurrence and development of diabetes.

Introduction

According to the report from the World Health Organization (WHO), diabetes mellitus has become an epidemic in the 21st century, which is the third most serious chronic lifelong disease that threatens human health after tumors and cardiovascular diseases. According to the statistics from the International Diabetes Federation, in 2011, there were 370 million diabetic patients in the world, 80% of which were in developing countries. It is estimated that by 2030, the number of diabetic patients in the world will reach 550 million [1]. Among the complications of diabetes mellitus, especially the complications of chronic diseases, they can involve multiple organs, disability and high mortality rate, affecting the physical and mental health of patients seriously, and bringing a heavy burden on individuals, families and the society. The main complication of diabetes is autonomic neuropathy, which can induce adverse events such as

cardiovascular death, nonfatal myocardial infarction and nonfatal stroke. It was reported that in a 6-year randomized controlled study, the total incidence of the adverse cardiovascular events was 23.2%, and the total mortality was 9.5% [2]. Although the diagnosis and treatment methods of diabetes mellitus have been constantly updated in recent years, it generally does not fundamentally prevent the occurrence and development of diabetes mellitus. Therefore, it is crucial to further explore the new mechanisms.

For a long time, the research on intestinal microorganisms has been unlimited to digestive tract diseases such as difficile *Clostridium* infection. The continuous research has revealed that the relationship between the host and the intestinal microorganisms was not a simple parasitic relationship, but a mutually beneficial symbiosis relationship. As early as the early twentieth century, some scholars proposed after research that there were about 100 trillion bacteria in the intestinal flora of a normal person, the total number of which was equivalent to the total number of human cells. The enormous genome contained in these microorganisms is also called human "Second Genome" [3–5]. The subsequent studies have continuously confirmed that this huge genome is closely related to the host's birth, aging, sickness and death. It is reported that the host's long-term diet [6], smoking [7, 8], drinking [9] and even stress [10, 11] and obesity [12] will change the structure of the intestinal flora, and the changes in the structure of the flora will further lead to changes in its microbial metabolites. These specific compounds can affect and promote the occurrence and development of diseases. For instance, MAO is closely related to cardiovascular diseases, diabetes and other chronic diseases [13].

At present, there are many research reports on diabetes and intestinal flora, while the results are different. In a recent clinical study of 180 people, researchers compared the intestinal microorganisms of the diabetic group and the healthy people. On the genus level, the relative abundance of *Prevotella* and *Alloprevotella* was significantly higher in the T2DM group [14]. However, in a prospective study in the same year, 341 samples were followed up and found that in contrast to the matched controls, individuals who went on to develop T2D had lower abundances of *Bifidobacterium longum*, *Coprobacillus unclassified*, and *Veillonella dispar* and higher abundances of *Roseburia hominis*, *Porphyromonas bennonis*, and *Paraprevotella unclassified* [15]. Currently, as to which bacteria are probiotics for diabetic patients, and which bacteria play a vital role in the occurrence and development of diabetes, this important conclusion has not yet been uniformly answered. Therefore, in this study, the age, gender, smoking, drinking, and other underlying diseases of the diabetic population were matched and included in the healthy population, and the two groups were analyzed and compared by bioinformatics to study this important research subject.

Methods

Research Design and Population

We included 44 patients with diabetes in the First Affiliated Hospital of Xinjiang Medical University and 47 healthy people. Patients with diabetes recruited in this study were diagnosed according to the *China's Guidelines for the Prevention and Treatment of Type 2 Diabetes Mellitus (2020 Edition)*. The diagnosis of

hypertension in the population was based on the *Hypertension Guidelines* in China in 2020, and the number of people diagnosed with hypertension was 67. The 91 patients included were all underwent coronary angiography. Moreover, coronary heart disease was diagnosed in patients when at least one main coronary artery stenosis of them were larger than 50%, and there were totally 72 cases of coronary heart disease.

We collected patient clinical data containing basic population information and biochemical indicators. We excluded the following patients: 1. Patients diagnosed with heart failure, structural heart disease, and pulmonary heart disease 2. Patients with a history of using antibiotic or probiotic within 3 months. 3. Patients with severe liver and kidney dysfunction, such as patients whose creatinines were not less than 2-fold of the normal upper limit, patients whose aspartate transaminases or alanine transaminases were not less than 3-fold of the normal upper limit, etc. 4. Patients with abnormal stool morphology such as diarrhea and dry stools. The study design complied with the Declaration of Helsinki, and it was approved by the Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University. Before the recruitment, the informed consent of eligible patients was obtained.

Past History and Clinical Data

For alcohol intake, subjects who drank more than eight standard drinks per week were classified as heavy drinkers [7]. For smoking, we adopted the WHO definition. Those who had smoked continuously or cumulatively for 6 months or more were defined as smokers (Current Smoking status). Smokers had no longer smoked at the time of the survey and persisted for more than 6 months become Past Smoking status [7]. 5ml peripheral venous blood was taken from each patient after 12 hours' fasting. The testing laboratory data included blood routine testing parameters, blood biochemical analysis results, renal function parameters and liver function parameters, and blood lipid analysis results.

Fecal Specimen Collection, DNA Extraction, and Sequencing

We provided a stool sampler for each participant for sample collection. All the participants underwent a formal training on how to collect the sample before recruitment. The stool sample freshly collected from each participant was divided into five aliquots of 200mg and immediately transported to the laboratory and frozen at -80°C. The bead-beating method was utilized to isolate the bacterial DNA from fecal samples as described previously [16].

Polymerase chain reaction was applied to amplify the V3-V4 region of 16S rRNA genes through applying the extracted DNA from each sample as the template. All the DNA extraction and sequencing were performed by Shanghai Personal Biotechnology Co., Ltd. (<http://www.personalbio.cn>, Shanghai, China). The sequencing data were processed through using the Quantitative Insights Into Microbial Ecology (v1.8.0) pipeline, as described previously [17].

Microbiome Data Analysis

When exploring the microbial diversity of both groups, we applied dimensionality reduction in bioinformatics. For the analysis of alpha-diversity, assessing the richness and diversity of the sample was evaluated by the Chao1, Shannon, Observed-species and Faith's PD indexes. The Chao1 and Observed-species indices are utilized for the community's abundance; the Shannon index is applied to assess the community's diversity, and the Faith's PD index is calculated in combination with the diversity of the flora to assess the kinship between the floras of species. For the analysis of beta-diversity, we applied Bray Curtis distance and unweighted UniFrac distance to calculate. In the commonly-applied microbial beta-diversity calculation distance, we believed that Bray Curtis distance paid more attention to the change of microbial abundance, while unweighted UniFrac distance paid more attention to the evolutionary relationship of microorganisms. We demonstrated beta-diversity through these two aspects. For the demonstration method, we employed Principal coordinates analysis (PCoA) and Nonmetric Multidimensional scaling (NMDS) analysis. For the statistical tests between the two groups, we applied Analysis of similarities (Anosim) analysis. For which species are different, we applied the non-parametric Kruskal-Wallis and Wilcoxon rank sum test, combined with the linear discriminant analysis (LDA) LDA Effect Size (LEfSe) analysis [18], [19]. For the correlation between bacterial abundance and biochemical indicators, we adopted the spearman correlation coefficient. In addition, for the screening of marker bacteria, we employed a mechanical learning model to screen for important bacteria. For the calculation of the probability of disease (POD) index, we performed logarithmic transformation on the bacterial abundance, and combined logistics regression to calculate the probability as the POD index [20].

Statistical Analysis

The SPSS version 22 (SPSS Inc., Chicago, IL, USA) and R version 3.2.4 (R Foundation for Statistical Computing, Vienna, Austria) was used for the statistical analysis of clinical data. For LEfSe (LDA Effect Size) analysis, we applied the Galaxy online analysis platform (<http://huttenhower.sph.harvard.edu/galaxy/>), and continuous variables were analyzed through applying Student's t-test. The chi-square test or Fisher exact test was used to analyze categorical variables. All results were considered statistically significant when p-values were less than 0.05.

Results

1. Basic Population Information and Intestinal Microbial Composition

We included a total of 91 cases in the study population, among which 44 cases were patients with a clear diagnosis of type 2 diabetes. When we included the control group, we matched the age, gender, BMI, living habits (smoking, drinking), and underlying diseases. Consequently, 47 cases were finally included. The baseline characteristics of the total sample are shown in Table 1. The average age of the enrolled patients was from 58 years old to 59 years old. The average BMI in the two groups was from 24 to 26. The two groups of smoking population were 38.64% and 34.04%; the drinking populations were 27.27% and 27.66%, respectively. In the included experimental group and the control group, the coronary heart

disease population accounted for 79% and 78% of the total population in each group, and the hypertensive population accounted for 70% and 76% of the total population in each group. Since the basic information of the population has been matched, in our long-term oral drug status records and biochemical examination index statistics of patients, we found that there were no statistical differences except for the differences in the treatment of diabetes drugs and fasting blood glucose.

Table 1
Baseline characteristics of study population

Variable	2 diabetes(n = 44)	Control(n = 47)	P value
Age, yr	58 ± 8	59 ± 12	0.556
Male	24(54.55%)	28(59.57%)	0.393
Body mass index, kg/m ²	26.06 ± 3.639	24.55 ± 3.159	0.61
Smoking	17(38.64%)	16(34.04%)	0.406
Drinking	12(27.27%)	13(27.66%)	0.577
Basic diseases			
Coronary heart disease	35(79.55%)	37(78.72%)	0.565
Essential hypertension	31(70.45%)	36(76.60%)	0.335
Long term oral/ subcutaneous injection medication			
Metformin	40(90.9%)	/	
Glucosidase inhibitors	11(25%)	/	
DDP-4 inhibitors	2(4.55%)	/	
Insulin	16(36.36%)	/	
Aspirin	33(75%)	34(72.34%)	0.481
Clopidogrel	30(68.18%)	25(53.19%)	0.106
Statin	32(72.72%)	26(55.32%)	0.065
β-block	27(61.36%)	28(59.57%)	0.516
Calcium Channel Blockers	12(27.27%)	7(14.89%)	0.116
ACEI or ARB ^a	18(40.91%)	18(38.30%)	0.484
Laboratory results			
White blood cells, ×10 ⁹ /L	6.76 ± 1.9	6.67 ± 1.98	0.838
Neutrophilicgranulocyte,%	59.91 ± 11.36	57.49 ± 13.39	0.354

Data are presented as median (interquartile range), mean ± SD, or number (%).

ACEI or ARB^a:angiotensin converting enzyme inhibitors or angiotensin receptor blocker

FBG^b:Fasting blood glucose;LDL^c:Low Density Lipoprotein;HDL^d:High Density Lipoprotein

Variable	2 diabetes(n = 44)	Control(n = 47)	P value
Hemoglobin, $\times 10^{12}/L$	136.50 \pm 21.34	138.55 \pm 13.47	0.587
Platelets, $\times 10^9/L$	235.34 \pm 65.98	221.85 \pm 53.13	0.284
blood urea nitrogen,mmol/L	5.37 \pm 1.91	5.05 \pm 1.32	0.366
Creatinine,umol/L	62.00(46.34-72.00)	64.00(44.20-72.50)	0.948
Uric acid,umol/L	314.45 \pm 91.63	320.28 \pm 77.47	0.746
FBG ^b ,mmol/L	6.59(5.77-9.43)	4.7(4.27-5.395)	< 0.001
Total cholesterol, mmol/L	3.64 \pm 0.90	3.69 \pm 1.13	0.849
Triglyceride, mmol/L	1.72(1.11-2.39)	1.38(1.01-2.02)	0.184
LDL ^c ,mmol/L	2.20 \pm 0.77	2.20 \pm 0.92	0.972
HDL ^d ,mmol/L	1.01 \pm 0.27	1.15 \pm 0.27	0.13
Total bilirubin,umol/L	10.1(7.4-14.4)	10.6(8.3-14.8)	0.553
Data are presented as median (interquartile range), mean \pm SD, or number (%).			
ACEI or ARB ^a :angiotensin converting enzyme inhibitors or angiotensin receptor blocker			
FBG ^b :Fasting blood glucose;LDL ^c :Low Density Lipoprotein;HDL ^d :High Density Lipoprotein			

We collected stools from the included population for the intestinal microbial analysis. We performed 16s rRNA sequencing on 91 patients. In order to determine whether the sample size is sufficient and to estimate the abundance of the community, we applied the species dilution curve and the abundance grade curve (See Fig. 1A and Fig. 1B). After the curve, it can be seen that they are all stable, indicating that the included sample size is sufficient for statistical analysis. Through the sequencing data analysis, we identified approximately 20,000 OTUs from the discovery cohort. We displayed two sets of Venn diagrams for the identified OTUs (See Fig. 1C). The results indicated that the OTU level in the diabetes group was significantly higher than the control group. We performed annotated species display for the top 100 OTU levels in abundance, and annotated Clostridia, Gammaproteobacteria, Actinobacteria, and Bacteroidia bacteria respectively (See Fig. 1D). We displayed the top 10 bacterial abundances at the phylum level, family level, genus level, and species level for the annotated species. The comparison of the histogram revealed that the species composition between the two groups was quite different (See Fig. 2A-2D). Among the bacteria at the phylum level, the specific top three bacteria proportions in the experimental group are as follows: Firmicutes (57.91%), Bacteroidetes (15.08%), and Actinobacteria (13.3%); the specific top three bacteria proportions in the control group are as follows: Firmicutes (51.32%), Proteobacteria (21.09%), and Actinobacteria (17.38%).

2. Specific Gut Microbial Signature in Type 2 Diabetes Mellitus Patients

The alpha diversity analysis on the two groups was presented (See Fig. 3A-3B). The results indicated that there were significant differences between the diabetes group and the control group. Furthermore, the Chao1, Observed-species and Shannon indexes were high in the diabetes group, suggesting a greater abundance of bacterial communities in the diabetes group than in the control group. The Faith's PD index was higher in the diabetic group, suggesting greater affinity between bacteria in the diabetic group. We performed a beta diversity analysis between the two groups based on unweighted_unifrac distance and Bray-Curtis distance, The beta diversity analysis results were presented in the form of Principle coordinates analysis (PCoA) and Nonmetric Multidimensional scaling (NMDS) dimension reduction (See Fig. 4A-4D). In the NMDS Analysis, Stress = 0.143 < 0.2 on the basis of the unweighted_unifrac distance, and Stress = 0.192 < 0.2 on the basis of the Bray-Curtis distance. In addition, we performed an Adonis analysis of the intra-group differences, and the results suggested R² = 0.031 and PR² = 0.001. All the species diversity analysis indicated significant differences between the two groups. Moreover, in the diabetes group, the species richness is higher than that in the control group, and the genetic distance between bacteria is larger than that in the control group. These results suggested that during the development of diabetes, the composition of the gut microbiota might change.

We performed a Lefse analysis on the different species between the two groups, where the threshold value of the LDA is more than 3. We found 27 bacteria with distinct differences in genus levels between the two groups. In the diabetes group, the abundances of the genus faecalibacterium, the genus Prevotella and the genus Roseburia are very high, while the abundances of the genus Shigella and the genus Bifidobacterium are very low (See Fig. 5A). Subsequently, we analyzed the correlation between these different bacteria and clinical biochemical indicators, and displayed them in the form of heat maps. Except for the genus Veillonella and genus unclassified_Enterobacteriaceae which are negatively correlated with blood sugar, the genus Phascolarctobacterium, the genus unidentified_Bacteroidales, and the genus Prevotella have an obvious positive correlation with the fasting blood glucose. In addition, among these indicators, in addition to the greater correlation with blood sugar, the correlation with high-density lipoprotein is also obvious. For instance, the abundance of the genus unidentified_Bacteroidales has a significant negative correlation with the high-density lipoprotein (See Fig. 5B).

3. Identification of Microbial Genus-Based Markers of 2DM

In this study, a 10 x cross-examination (10- fold cross-validations) of genus-level bacteria was performed to further find important signature species through random forests in mechanical learning models. We identified 12 bacteria with the highest importance at the genus level, including Shigella and Parabacteroides, as shown in Fig. 6A. Subsequently, we applied these 12 bacteria to diagnose diabetes and drew the ROC curve (Fig. 6B). The cut-off value of the ROC curve is 0.78; the 95% confidence curve is 0.386–0.936, and the AUC area is 0.841. We calculated the probability of disease (POD) index through

applying the identified optimal set of 12 genus. As is shown in Fig. 6C results, the biomarker can complete 84.1% of the diagnosis between the diabetes group and the control group, indicating that this indicator has a good diagnostic effect on the disease.

In the included population, the clinical indicators were screened for diagnostic markers. As shown in Table 2, the clinical indicators with p value less than 0.05 and AUC area greater than 0.5 are only fasting blood glucose and hypertension. The diagnostic curve and AUC area of a patient's random fasting blood glucose for diabetes is 0.839 ($p < 0.001$, 95% CI: 0.742–0.937). There are more people diagnosed with diabetes in people with high blood pressure, namely, the AUC area of the diagnosis of hypertension for diabetes is 0.658 ($p < 0.019$, 95% CI: 0.533–0.783).

Table 2
Candidate Variables for Clinical Model Development^a

Variables	AUC	P values	95% CI
FBG	0.839	0.000	0.742–0.937
HDL	0.337	0.016	0.212–0.462
EH	0.658	0.019	0.533–0.783
BMI	0.613	0.095	0.481–0.744
CAD	0.599	0.140	0.471–0.728
Ne	0.589	0.189	0.458–0.720
SMOKING	0.552	0.437	0.421–0.684
LDL	0.451	0.469	0.318–0.584
TG	0.546	0.496	0.414–0.678
TC	0.457	0.527	0.326–0.589
WBC	0.463	0.580	0.330–0.595
Cr	0.471	0.663	0.336–0.605
UA	0.474	0.698	0.340–0.608
SEX	0.475	0.706	0.342–0.607
TBIL	0.477	0.737	0.343–0.612
AGE	0.480	0.766	0.345–0.614
PLT	0.519	0.774	0.383–0.656
HBL	0.491	0.898	0.356–0.627
DRINKING	0.503	0.962	0.371–0.635
^a Note:			
WBC: white blood cells; Cr: Creatinine; UA: uric acid; PLT: platelets; HBL: High density lipoprotein;			
FBG: Fasting blood glucose; HDL: High density lipoprotein; BMI: Body mass index; CAD: coronary heart disease; Ne: Neutrophilicgranulocyte. EH: Essential hypertension; TC: Total cholesterol;			
TBIL: total bilirubin; LDL: Low density lipoprotein; TG: Triglyceride.			

4. Nomogram Prediction of DM

Subsequently, we performed logarithmic transformation on the microbial abundance of these 12 genera microorganisms. We calculated the POD value in combination with logistics regression [20], and

established the nomogram diagnostic model in combination with age, gender (if female), and clinical variable indicators such as the denial from the diagnosis of hypertension and fasting blood glucose (See Fig. 7A). A patient was randomly selected from our population. This patient was a 45-year-old woman who had been diagnosed with high blood pressure. Her biochemical index fasting blood glucose was 6 mmol/L. The stool was tested for landmark microorganisms and the POD value was calculated to be 0.7. Combining the above information into this diagnostic model, we concluded that the probability of diabetes in the patient was 95%. The results are shown in Fig. 7B. Then, we verified the efficacy of this diagnostic model and drew the C-index curve. As shown in Fig. 7C, the C-index of this model is 0.924, which has an excellent effect on the diagnosis of diabetes. We drew the ROC curve of the clinical indicators, microbial POD indicators and the combined model in the model, and found that the microbial POD indicators independently diagnosed diabetes and the clinical indicator models including fasting blood glucose have similar diagnostic performance. If we combine these two aspects for diagnosis, the AUC area can reach to 0.908 (See Fig. 7D). This result suggests that microbial indicators can optimize the diagnosis of diseases by clinical indicators. In addition, we have also drawn a DCA curve, and we have reached similar conclusions in the DCA curve (See Fig. 7E).

Discussion

In this study, for the first time, we successfully established a diabetic diagnosis model through applying intestinal microorganisms combined with clinical indicators, and the verification of the diagnosis model revealed that the diagnosis effect of this model was very good. For the differences in the intestinal flora between diabetes, it has recently been reported in several studies. For instance, Awgichew et al. attributed differences between diabetes and health groups to the role of short-chain fatty acid bacteria [21]. In the research, Zhang et al. reported patients at the genus level of bacteria in the diabetes group. They found that the relative abundance of *Prevotella* and *Alloprevotella* was significantly higher [14]. In our study, 91 patients were included, and there was no significant statistical difference between the two groups in basic information such as age and gender. In the diversity analysis, significant differences between the two groups were found, which is consistent with existing research reports. The specific differences in bacterial levels are not consistent with recent studies, and it is considered that most of the patients in this study have underlying diseases. Actually, in the common process of clinical diagnosis and treatment, such similar patients are more common, which are more in line with the complexity of patients in the process of clinical practice, and more suitable for clinical diagnosis and treatment. We employed the random forest model to find 12 distinctly different genus-level bacteria in the included sample population, in combination with POD indicators to establish a powerful diagnostic model (UC = 90.8%, $P < 0.001$). It is important that the C index is close to 1 when this diagnostic model is validated. These results indicate that the intestinal flora is closely related to the occurrence and development of diabetes.

In our findings, the abnormal abundance of the genus *parabacteroides* (the intestinal microorganism) in diabetic patients seems to have a significant effect on the diagnosis of diabetes. For the relationship of the genus *parabacteroides* is not the first reported with diabetes. In a recent prospective randomized controlled study in Spain, the relationship between diet, gut microbes and diabetes was explored. The

same similar results were obtained in the study results. In the study results, the role of the genus *parabacteroides* in affecting the occurrence and development of diabetes is confirmed. This effect seems to be related to the enhancement of the metabolic pathways including terpenoid-quinone, lipopolysaccharides and N-glycan biosynthesis [22]. The genus *Alistipes* is also very important in the diabetic group, and the effect of the genus *Alistipes* on diabetes has also been reported recently [23]. The genus *Alistipes* is currently considered to be a bacterium that produces short-chain fatty acids (SCFA), which have a number of potential roles in modulating metabolic health and DM risk factors, such as blood glucose regulation and metabolic regulation, and maintaining the integrity of the intestinal barrier [24]. There are also effects on the genus *Faecalibacterium* in diabetic patients, and the different bacterial distribution is more obvious in the diabetic group. Currently, more studies have been reported on the genus *Faecalibacterium*, and the abnormal abundance of the genus *Faecalibacterium* was also reported by different institutes in experimental groups [25] and [26]. In our study, the abundance of the genus *Faecalibacterium* was positively associated with the presence of triglycerides, presuming that it was likely to be related to the immune effects of the genus *Faecalibacterium*. This similar report has also been mentioned in previous studies [27]. In this study, the genus *streptococcus* was also selected in randomized forest importance in the diabetes group. The genus *streptococcus* was widely found in nature, human and animal stools, and healthy people's nasopharynx and intestines, which can mainly cause suppurative inflammation, toxin diseases and hypersensitivity reactive diseases. It was also reported in diabetes in previous studies [23]. In our study, there were two *prevotella* genera in bacterial abundance and clinical indicators, which is the result of the current Greengenes database. Both *prevotella* were annotated, which also emphasize the relationship between *prevotella* and blood glucose. In our study, the abundance of the genus *prevotella* was high in the diabetes group, and it was positively associated with the clinical biochemical index fasting blood glucose. This conclusion was also mentioned in a 2020 study in the United States. In this study, the researchers attributed this effect to lipopolysaccharide (LPS), which is a component of the Gram-negative bacteria wall. It can activate the local immune response and may cause low-grade systemic inflammation, leading to insulin resistance and affecting the occurrence and development of diabetes [28]. In addition to the genus mentioned above, the genus *Dialister*, the genus *Butyricimonas* and the genus *Gemmiger* have been shown in the random forest model diabetes group. This conclusion has been confirmed in different studies [29], [30], and [31]. In our heat map of differential bacteria and biochemical indicators, it is suggested that the genus *Gemmiger* is surely negatively correlated with the percentage of neutrophils, suggesting that genus *Gemmiger* is likely to play a role through immune regulation.

Among the 12 types of bacteria included in the model in this study, the genus *Shigella*, the genus *Ruminococcus*, the genus *Actinomyces*, and the genus *Bifidobacterium* were relatively important in the control group. It is not the first time that *Shigella* has been included in diabetes-related diagnostic models. In a recent study on diabetic nephropathy, the study model included 25 genus bacterial to diagnose diabetic nephropathy, and the AUC area of the diagnostic model after drawing the ROC curve was 0.972. In such a model, the genus *Shigella* was also emphasized. However, the action mechanism of the genus *Shigella* and diabetes or diabetes-related diseases is currently unclear. The genus *Ruminococcus* is

closely related to the diet structure of the host. This bacterium has also been reported in diabetes. This conclusion is similar to the structure of this study [31]. The genus *Actinomyces* is widely distributed in nature and has a wide variety of species. It is a member of the normal flora of the human body and can cause endogenous infections. There are few reports of the genus *Actinomyces* in the diagnosis model of diabetes. In 2019, a Chinese study reported the microbial structure of patients with hyperlipidemia and gestational diabetes. The report revealed that the abundance of this genus was abnormal between the two groups, which can also verify the results in this study [32]. Among these genera with high control abundance, the genus *Bifidobacterium* should be the most reported microbial bacteria, which has been clearly defined as probiotics. It has been added to dairy products for consumption and also plays a role in the pharmaceutical industry. In our Lefse analysis and research results, the genus *Bifidobacterium* has the highest LDA in the control group. This conclusion does not seem to be a surprise. The improvement of blood sugar by *Bifidobacterium* was also mentioned in a study. The elderly patients with type 2 diabetes took 200ml of a compound drink containing *Bifidobacterium bifidum* every day. After taking it for 1 month, the fasting blood glucose level was significantly reduced [33]. However, at present, there are rare reports on the mechanism of *Bifidobacterium* to improve blood sugar, which should be the future research direction of probiotics to improve disease. In addition, there are also research reports on the negative correlation between the genus *Bifidobacterium* and creatinine. In a randomized controlled study in Brazil in 2020, the subjects were given drugs containing *Bifidobacterium* genus regularly. After follow-up, it was found that the blood creatinine level was different between the two groups, and the creatinine level of the experimental group taking the drug decreased significantly [34].

In our study, there are several limitations that may require close attention. First, although we developed a precise diagnostic model through the intestinal flora, the function of these microbiomes is unknown and we have not done the detection of bacterial metabolites. In this way, we have insufficient research on the mechanism. Second, we included less sample sizes, which were not verified in different regions and different time periods. Third, the method we tested is 16S gene sequencing. In the database, there are still many tentative names, unidentified, unclassified and unclassified bacteria. It is suggested to further explore the multiple group methods such as macrogene. In conclusion, for the first time, we have successfully developed a diagnostic model with high diagnostic effect through utilizing intestinal microbes combined with the clinical biochemical indicators. This model reveals that the intestinal flora can significantly improve the diagnostic ability of diabetes, and also shows that the intestinal flora is involved in the occurrence and development of the disease, providing new mentality for exploring metabolic diseases such as diabetes.

Declarations

Due to technical limitations, Declarations section is not available for this version.

References

1. Chinese Medical Association Diabetes Society. Guidelines for the prevention and treatment of type 2 diabetes in China (2013 edition) [J]. Chinese Journal of Diabetes, 2014, 6(7): 447–498. DOI: 10.3760/cma.j.issn.1674-5809.2014.07.004.
2. Lai, Y. R. *et al.* Severity of Cardiovascular Autonomic Neuropathy Is a Predictor Associated With Major Adverse Cardiovascular Events in Adults With Type 2 Diabetes Mellitus: A 6-Year Follow-up Study. *Can J Diabetes*, **45** (2), 155–161 (2021 Mar).
3. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome [J]. *Science*, 2006, 312(5778): 1355–1359.
4. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing [J]. *Nature*, 2010, 464(7285): 59–65.
5. Nicholson, J. K. & Holmes E, Wilson, I. D. Gut microorganisms, mammalian metabolism and personalized health care [J]. *Nature Reviews Microbiology*, 2005, 3(5): 431–438.
6. Valle, M. *et al.* Cholecalciferol Supplementation Does Not Prevent the Development of Metabolic Syndrome or Enhance the Beneficial Effects of Omega-3 Fatty Acids in Obese Mice. *J Nutr*, **151** (5), 1175–1189 (2021).
7. Yoon, H. *et al.* Characteristics of the Gut Microbiome of Healthy Young Male Soldiers in South Korea: The Effects of Smoking. *Gut Liver*, **15** (2), 243–252 (2021).
8. Yang, Y. *et al.* Cigarette smoking exposure breaks the homeostasis of cholesterol and bile acid metabolism and induces gut microbiota dysbiosis in mice with different diets., **450**, 152678 (2021).
9. Ballway, J. W. & Song, B. J. Translational Approaches with Antioxidant Phytochemicals against Alcohol-Mediated Oxidative Stress, Gut Dysbiosis, Intestinal Barrier Dysfunction, and Fatty Liver Disease. *Antioxidants (Basel)*, **10** (3), 384 Published 2021 Mar 4. (2021).
10. Stevens, B. R., Pepine, C. J., Richards, E. M., Kim, S. & Raizada, M. K. Depressive-hypertension: A proposed human endotype of brain/gut microbiome dysbiosis [published online ahead of print, 2021 May 10]. *Am Heart J*. 2021;100540.
11. Yano, Y. & Niiranen, T. J. Gut Microbiome over a Lifetime and the Association with Hypertension. *Curr Hypertens Rep*, **23** (3), 15 Published 2021 Mar 8. (2021).
12. Minichino, A. *et al.* Endocannabinoid system mediates the association between gut-microbial diversity and anhedonia/amotivation in a general population cohort [published online ahead of print, 2021 May 17]. *Mol Psychiatry*, <https://doi.org/10.1038/s41380-021-01147-5> (2021).
13. He, M., Tan, C. P., Xu, Y. J. & Liu, Y. Gut microbiota-derived trimethylamine-N-oxide: A bridge between dietary fatty acid and cardiovascular disease? *Food Res Int*, **138** (Pt B), 109812 <https://doi.org/10.1016/j.foodres.2020.109812> (2020).
14. Zhang, Z. *et al.* Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. *PeerJ*, **9**, e10952 Published 2021 Mar 24 <https://doi.org/10.7717/peerj.10952> (2021).
15. Wang, L. *et al.* The Fecal Microbiota Is Already Altered in Normoglycemic Individuals Who Go on to Have Type 2 Diabetes. *Front Cell Infect Microbiol*, **11**, 598672 Published 2021 Feb 18 <https://doi.org/10.3389/fcimb.2021.598672> (2021).

16. Liu, H. *et al.* Alterations in the gut microbiome and metabolism with coronary artery disease severity. *Microbiome*, **7**, 68 (2019).
17. Li, P. *et al.* Zearalenone changes the diversity and composition of caecum microbiota in weaned rabbit. *BioMed Res. Int* 2018, 2018, 1 – 10.
18. Harakeh, S. *et al.* Impact of smoking cessation, coffee and bread consumption on the intestinal microbial composition among Saudis: A cross-sectional study. *PLoS One*, **15** (4), e0230895 Published 2020 Apr 29 <https://doi.org/10.1371/journal.pone.0230895> (2020).
19. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol*, **12**, R60 (2011).
20. Zheng, Y. Y. *et al.* Gut Microbiome-Based Diagnostic Model to Predict Coronary Artery Disease. *J Agric Food Chem*, **68** (11), 3548–3557 (2020).
21. Chambers, E. S., Preston, T., Frost, G. & Morrison, D. J. Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. *Curr. Nutr. Rep*, **7**, 198–206 (2018).
22. Mateo-Gallego, R. *et al.* An alcohol-free beer enriched with isomaltulose and a resistant dextrin modulates gut microbiome in subjects with type 2 diabetes mellitus and overweight or obesity: a pilot study. *Food Funct*, **12** (8), 3635–3646 (2021).
23. Pinna, N. K. *et al.* Trans-ethnic gut microbial signatures of prediabetic subjects from India and Denmark. *Genome Med*, **13** (1), 36 Published 2021 Mar 3. (2021).
24. Chambers, E. S., Preston, T., Frost, G. & Morrison, D. J. Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. *Curr. Nutr. Rep*, **7**, 198–206 (2018).
25. Almugadam, B. S. *et al.* Alterations of Gut Microbiota in Type 2 Diabetes Individuals and the Confounding Effect of Antidiabetic Agents. *J Diabetes Res*. 2020;2020:7253978. Published 2020 Sep 28. doi:10.1155/2020/7253978
26. Alvarez-Silva, C. *et al.* Trans-ethnic gut microbiota signatures of type 2 diabetes in Denmark and India. *Genome Med*. 2021;13(1):37. Published 2021 Mar 3. doi:10.1186/s13073-021-00856-4
27. Munukka, E. *et al.* Gut-adipose tissue axis in hepatic fat accumulation in humans. *J Hepatol*, **61** (1), 132–138 (2014).
28. Radwan, S. *et al.* A comparative study of the gut microbiome in Egyptian patients with Type I and Type II diabetes. *PLoS One*, **15** (9), e0238764 Published 2020 Sep 9. (2020).
29. Park, J. U. *et al.* Influence of Microbiota on Diabetic Foot Wound in Comparison with Adjacent Normal Skin Based on the Clinical Features. *Biomed Res Int*, **2019**, 7459236 Published 2019 Aug 19 <https://doi.org/10.1155/2019/7459236> (2019).
30. Harbison, J. E. *et al.* Gut microbiome dysbiosis and increased intestinal permeability in children with islet autoimmunity and type 1 diabetes: A prospective cohort study. *Pediatr Diabetes*, **20** (5), 574–583 (2019).
31. Breuninger, T. A. *et al.* Associations between habitual diet, metabolic disease, and the gut microbiota using latent Dirichlet allocation. *Microbiome*, **9** (1), 61 Published 2021 Mar 16. (2021).

32. Liu, H. *et al.* Alterations of Gut Microbiota and Blood Lipidome in Gestational Diabetes Mellitus With Hyperlipidemia. *Front Physiol*, **10**, 1015 Published 2019 Aug 6. (2019).
33. Moroti, C. *et al.* Effect of the consumption of a new symbiotic shake on glycemia and cholesterol levels in elderly people with type 2 diabetes mellitus[J]. *Lipids Health Dis*, **11**, 29 (2012).
34. Romão da Silva, L. F. *et al.* Effects of probiotic therapy on cardio-metabolic parameters and autonomic modulation in hypertensive women: a randomized, triple-blind, placebo-controlled trial. *Food Funct*, **11** (8), 7152–7163 (2020).

Figures

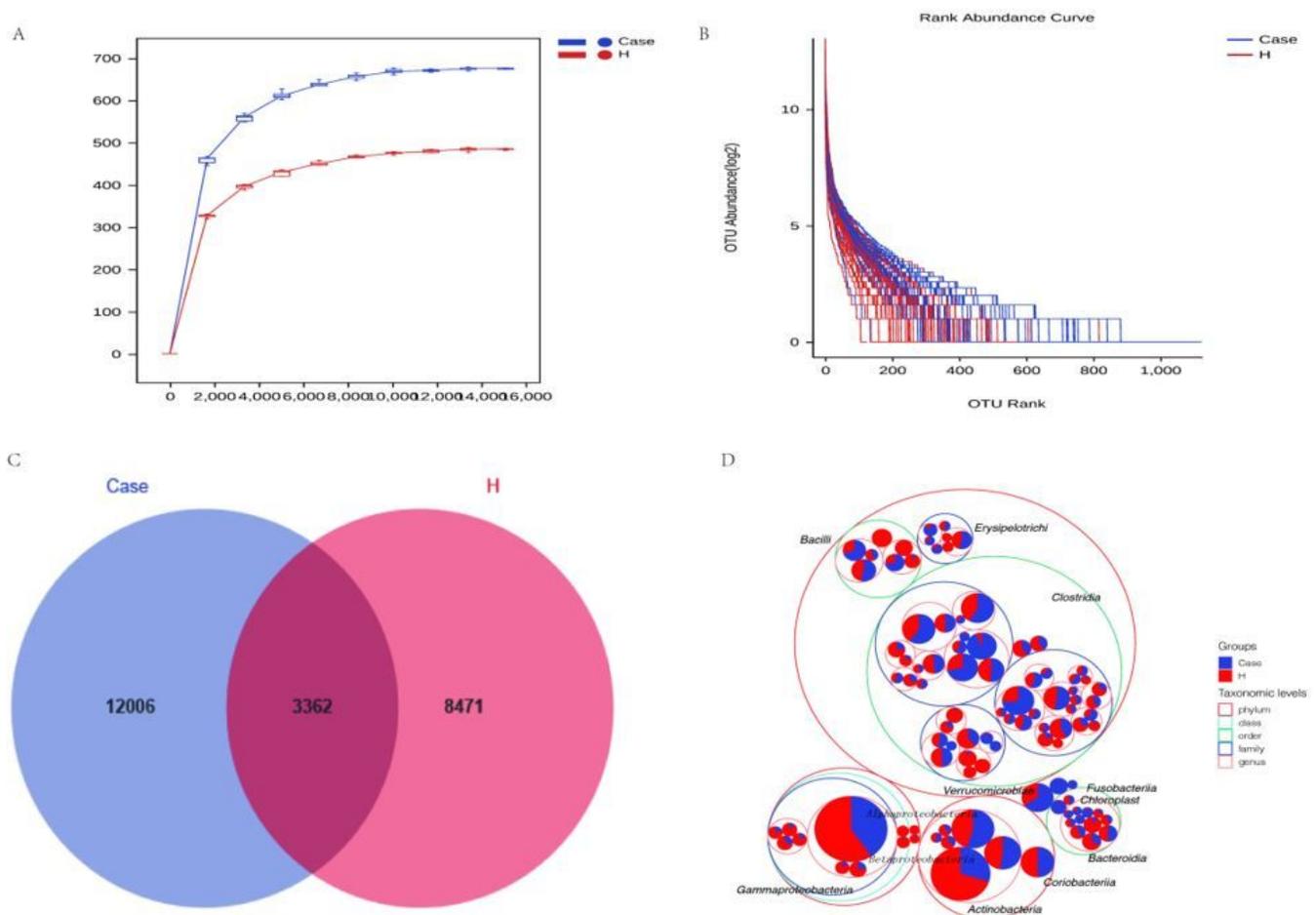


Figure 1

Increased fecal microbial diversity in patients with DM vs healthy controls. [A] Rarefaction Curve. [B] Rank abundance curve. [C] Venn diagram displaying the overlaps between groups showing that 3362 of the total richness of 23839 OTUs were shared between the CAD patients and the healthy controls. OTUs, operational taxonomy units. [D] circle packing chart.

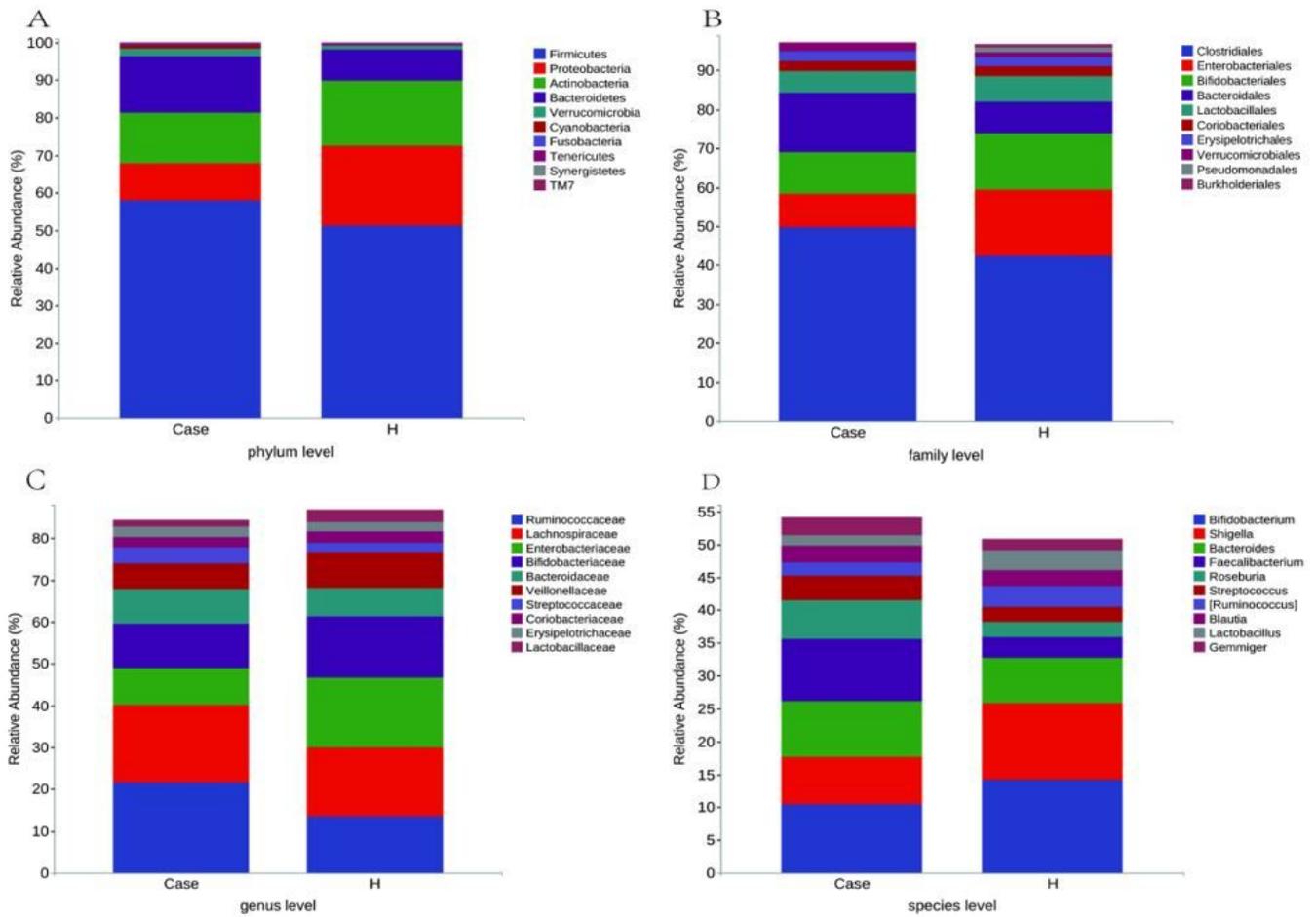


Figure 2

Stacked histogram of species composition in the top 10 of each group. [A] phylum level. [B] family level. [C] genus level. [D] species level.

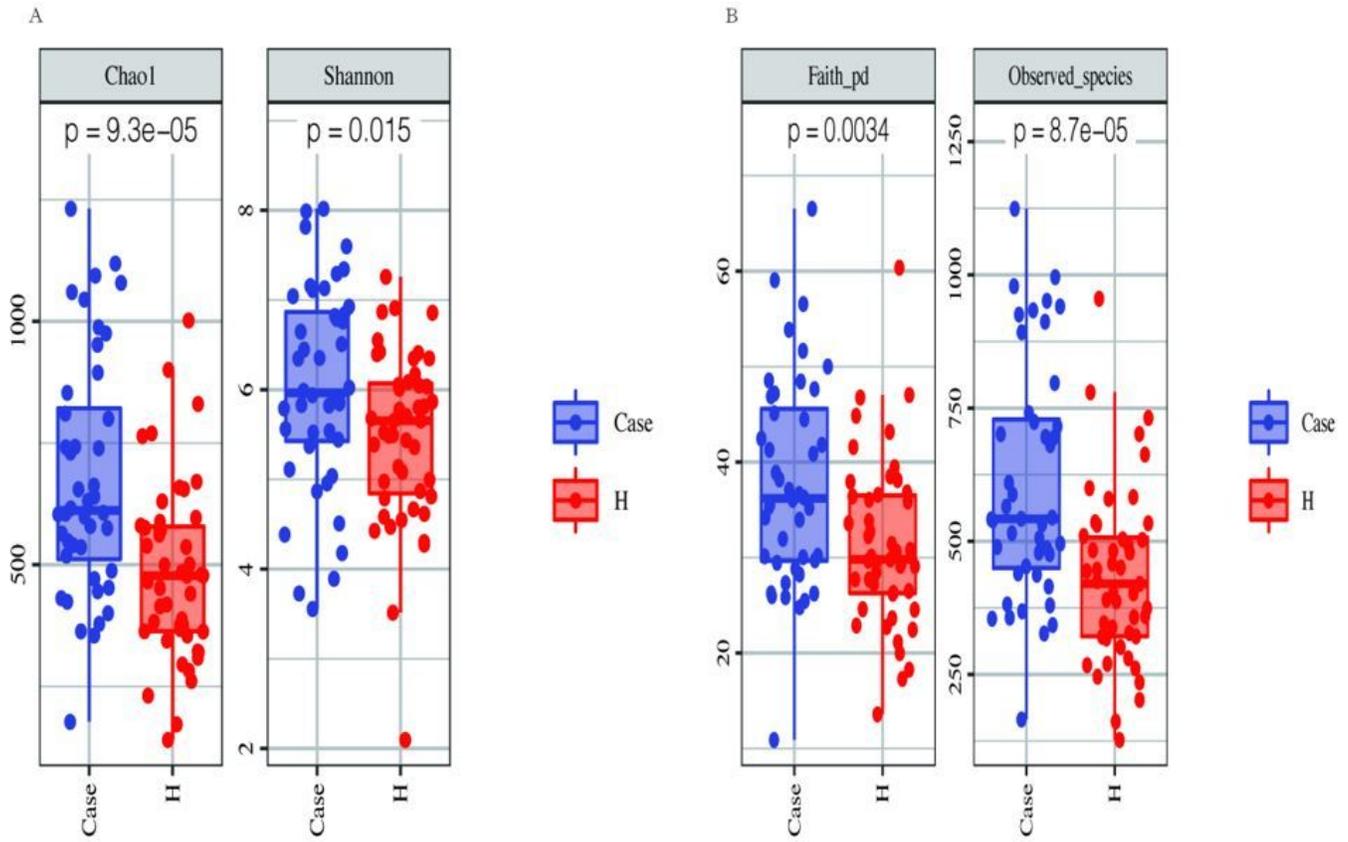


Figure 3

Comparison of fecal microbial diversity, as estimated by the Chao1 index and Shannon index(A),Faith-pd index and Observed_species index(B).

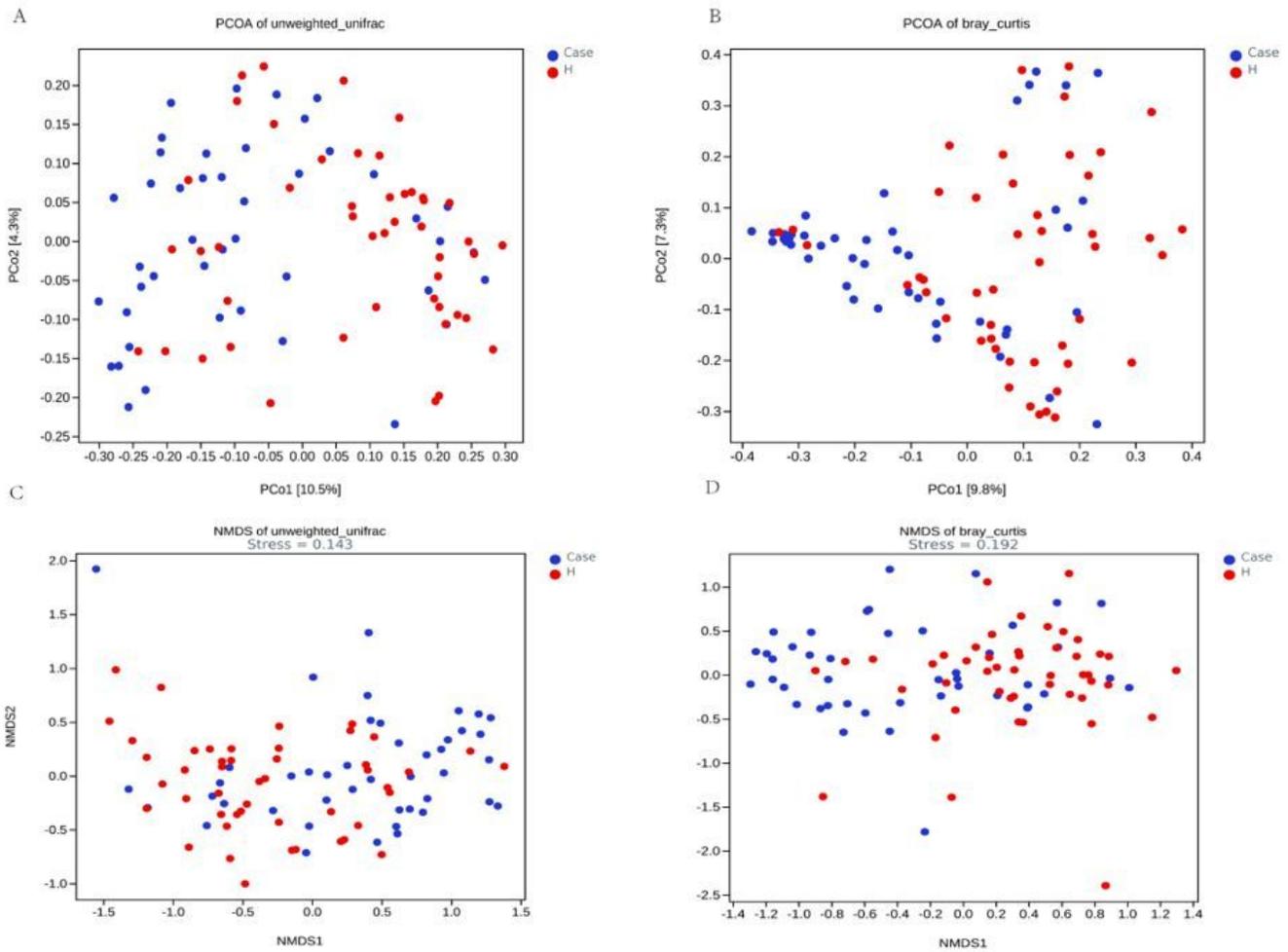
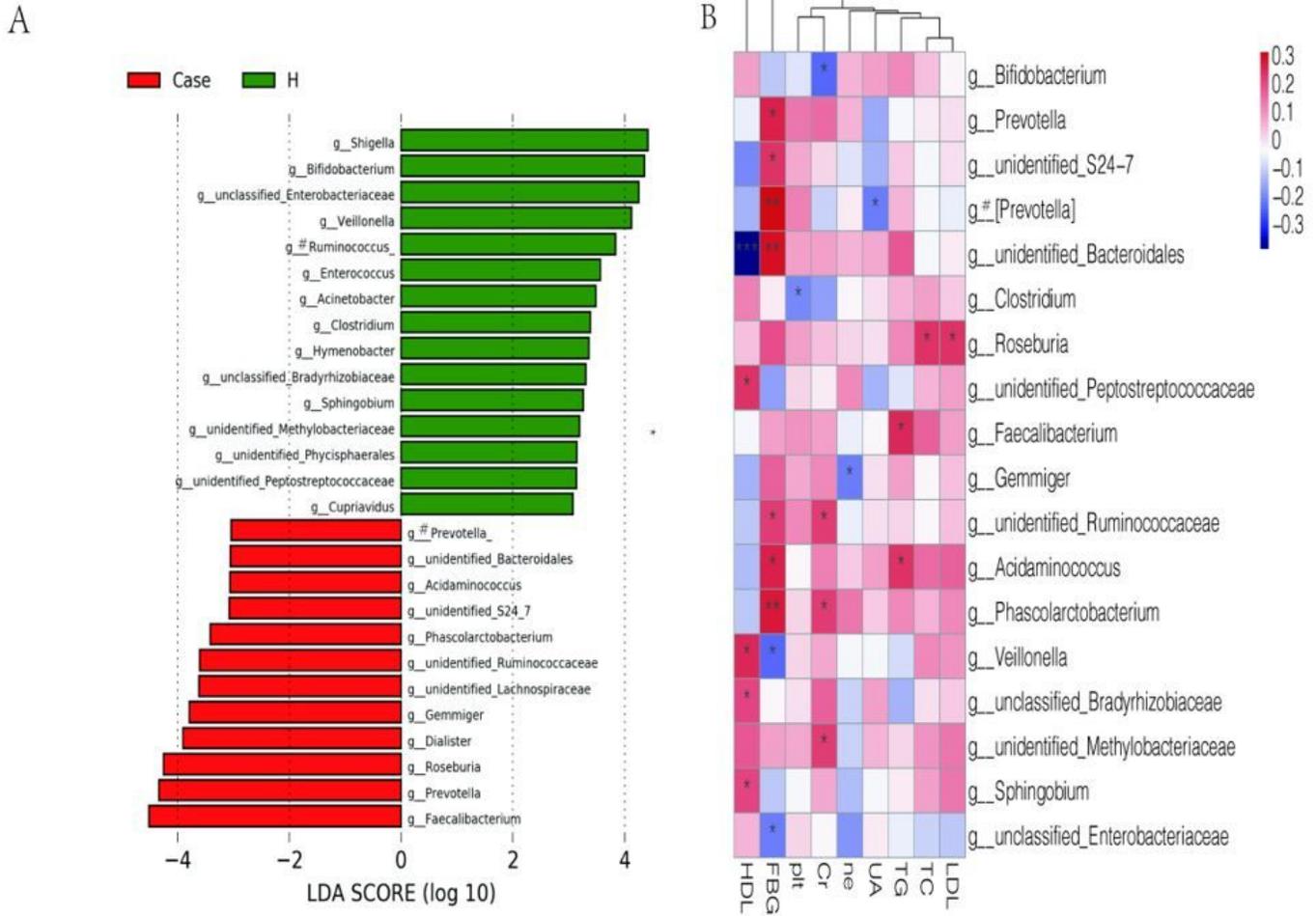


Figure 4

Beta diversity was calculated using unweighted UniFrac [left] or brau_curtis [right] by PCoA and NMDS, indicating a symmetrical distribution of fecal microbial community among all the samples.



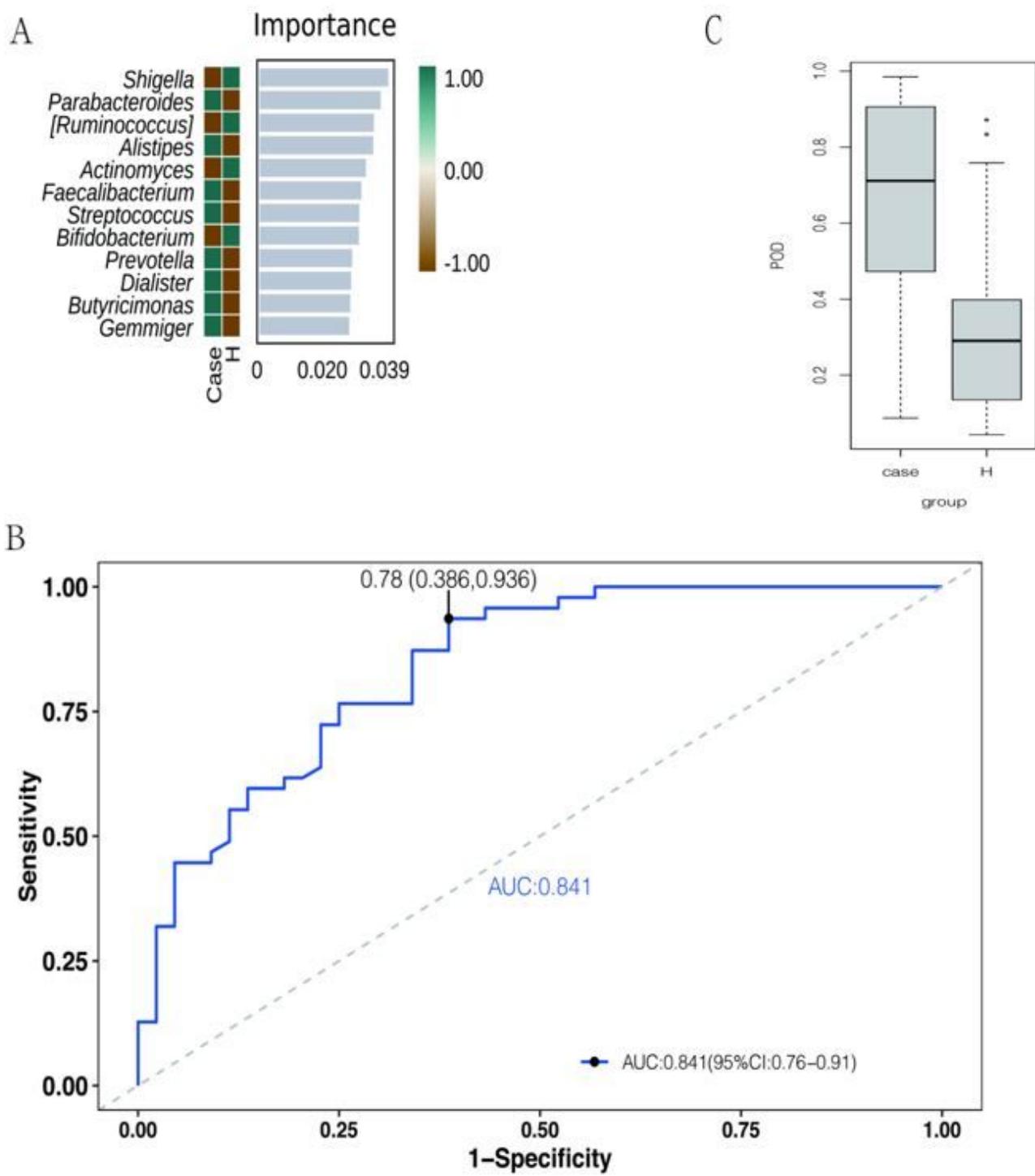


Figure 6

Important biomarkers.[A] The top 12 bacteria belong to the genus level.[B] ROCs curve with AUC for the diagnostic performance of the gut microbial model. [C] Comparison of the POD of gut microbiome. [*Ruminococcus*] Temporary name.

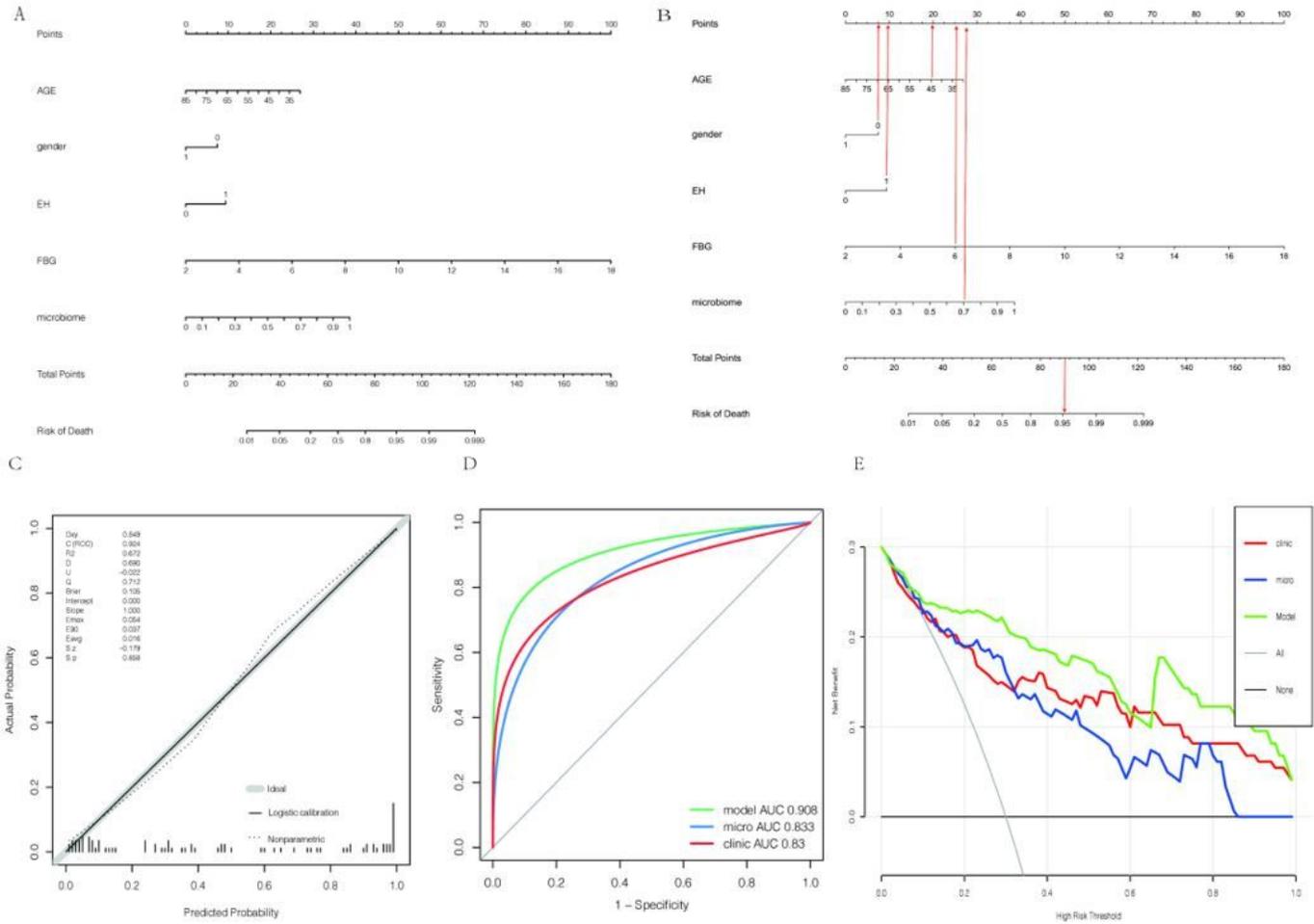


Figure 7

Nomogram and its performance. [A] Instructions for using the nomogram.[B] Use the nomogram specifically.[C] calibration_curve.[D] The AUCs for the diagnostic performances of the clinical model ,microbiome model and the combined model .[E] Decision Curve Analysis of the clinical model ,microbiome model and the combined model.