Microbiome Features Differentiating Unsupervised-Stratification-Based Clusters of Patients with Abnormal Glycometabolism

Ting Xu,[a] Xuejiao Wang,[b] Yu Chen,[a] Hui Li,[a] Liping Zhao,[a,c] Xiaoying Ding,[b] Chenhong Zhang[a]

aState Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

bDepartment of Endocrinology and Metabolism, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

cDepartment of Biochemistry and Microbiology and New Jersey Institute for Food, Nutrition and Health, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, New Jersey, USA

Ting Xu and Xuejiao Wang contributed equally to this article. Author order was determined in order of decreasing seniority.

ABSTRACT The alteration of gut microbiota structure plays a pivotal role in the pathogenesis of abnormal glycometabolism. However, the microbiome features identified in patient groups stratified solely based on glucose levels remain controversial among different studies. In this study, we stratified 258 participants (discovery cohort) into three clusters according to an unsupervised method based on 16 clinical parameters involving the levels of blood glucose, insulin, and lipid. We found 67 cluster-specific microbiome features (i.e., amplicon sequence variants [ASVs]) based on 16S rRNA gene V3-V4 region sequencing. Specifically, ASVs belonging to *Barnesville* and *Alistipes* were enriched in cluster 1, in which participants had the lowest blood glucose levels, high insulin sensitivity, and a high-fecal short-chain fatty acid concentration. ASVs belonging to *Prevotella copri* and *Ruminococcus gnavus* were enriched in cluster 2, which was characterized by a moderate level of blood glucose, serious insulin resistance, and high levels of cholesterol and triglyceride. Cluster 3 was characterized by a high level of blood glucose and insulin deficiency, enriched with ASVs in *P. copri* and *Bacteroides vulgatus*. In addition, machine learning classifiers using the 67 cluster-specific ASVs were used to distinguish individuals in one cluster from those in the other two clusters both in discovery and testing cohorts (*n* = 83). Therefore, microbiome features identified based on the unsupervised stratification of patients with more inclusive clinical parameters may better reflect microbiota alterations associated with the progression of abnormal glycometabolism.

IMPORTANCE The gut microbiota is altered in patients with type 2 diabetes (T2D) and prediabetes. The association of particular bacteria with T2D, however, varied among studies, which has made it challenging to develop precision medicine approaches for the prevention and alleviation of T2D. Blood glucose level is the only parameter in clustering patients when identifying the T2D-related bacteria in previous studies. This stratification ignores the fact that patients within the same blood glucose range differ in their insulin resistance and dyslipidemia, which also may be related to disordered gut microbiota. In addition to parameters of blood glucose levels, we also used additional parameters involving insulin and lipid levels to stratify participants into three clusters and further identified cluster-specific microbiome features. We further validated the association between these microbiome features and glycometabolism with an independent cohort. This study highlights the importance of stratification of patients with blood glucose, insulin, and lipid levels when identifying the microbiome features associated with the progression of abnormal glycometabolism.

KEYWORDS glycometabolism, unsupervised stratification, insulin resistance, microbiome features

A linkage between gut microbiota dysbiosis and type 2 diabetes (T2D) has been extensively explored (1–3). The alterations of the gut microbiota of patients with T2D are characterized by a decrease in the abundance of beneficial bacteria (e.g., *Akkermansia muciniphila*, the genera of *Bifidobacterium* and *Roseburia* spp.) and an increase in opportunistic pathogens, including *Clostridium bolteae* and *Desulfovibrio* sp. (1, 4–6). The loss of potential butyrate-producing bacteria, such as *Faecalibacterium prausnitzii*, and the deficiency in butyrate production are the most common findings in patients with T2D, as well as in patients with prediabetes (1, 3, 5–7). Nonetheless, the association of particular bacteria with T2D has varied among studies. For instance, Zhang et al. found that *A. muciniphila* was less abundant, whereas *Clostridiales* sp. strain SS3/4 was more abundant in patients with T2D (4), but the abundances of these two taxa showed the opposite result from those reported in the study by Qin (1). This confusion has led to difficulty in using bacterial features to elucidate the mechanism of gut microbiota in the development of T2D for developing new biomarkers and therapies.

One possible reason for the challenge in identifying gut microbial characteristics in patients with T2D and prediabetes is that all of the previous studies have solely used blood glucose levels as the criteria to stratify their cohorts. Clinical investigation, however, has shown that people within the same blood glucose range, as defined by the American Diabetes Association (ADA) and World Health Organization (WHO) criteria, are heterogeneous in insulin sensitivity and islet β-cell function (8, 9). Insulin resistance and islet β-cell dysfunction are the two pathogeneses of abnormal glycometabolism and occur much early than the onset of hyperglycemia (10). In addition, the condition of dyslipidemia, a risk factor of T2D and diabetic complications, is varied in people within the same blood glucose range. For example, in the Framingham Heart Study, 19% of the men and 17% of the women with diabetes had increased total plasma triglyceride levels, and the prevalence of increased total plasma triglyceride levels in men and women with normal blood glucose levels were 9 and 8% (11). Recently, apart from blood glucose levels, additional variables related to insulin resistance and sensitivity, as well as dyslipidemia, have been proposed for use in glycometabolism classifications (12–14). In addition to blood glucose levels, a study classified individuals, including healthy people and people with prediabetes, into six distinct clusters by including parameters such as insulin secretion, insulin resistance, and high-density lipoprotein-cholesterol (HDL), and monitored them for 4.1 years and 16.3 years (mean follow-up years in two cohorts, respectively) (12). Longitudinal follow-up revealed that different clusters had different risks of diabetes and diabetic complications. Individuals in one of the clusters experienced slower progression to overt T2D than those in other clusters but had a higher risk of nephropathy. In addition to the hyperglycemia, the disorder of gut microbiota has been shown to have a causative role in the progression of obesity and insulin resistance both in mouse models and in human studies by gut microbiota transplantation. For example, the *Enterobacter cloacae* B29 strain isolated from a morbidly obese human's gut induced obesity and insulin resistance in germfree mice (15). A gut microbiota transplantation study on humans revealed that gut microbiota from lean donors increased the insulin sensitivity in obese recipients with metabolic syndrome (16). In addition, the gut microbiota could regulate insulin resistance by producing metabolites, such as imidazole propionate, which could directly impair insulin signaling at the level of insulin receptor substrate (17). Furthermore, a study in germfree and conventionally raised mice showed that the gut microbiota had an effect on the host's serum lipidomes, especially the triglyceride levels (18). In obese mice fed a high-fat diet, supplementation with probiotics, such as *Lactobacillus curvatus*, alone or together with *L. plantarum*, reduced cholesterol in plasma; supplementation with *Bifidobacterium* spp. decreased the levels of circulating triglycerides and low-density lipoprotein-cholesterol (LDL) and increased the HDL level (19–21). Therefore, it is essential to consider insulin and lipid levels in the criteria for cohort stratification when studying the association of glycometabolism and gut microbiota.

In this study, to identify the microbiome features that correlate with progression of abnormal glycometabolism, we classified 258 individuals into three glycometabolism

clusters according to variables related to blood glucose level, insulin resistance, and dyslipidemia. We found that the gut microbiota structure was different among the three glycometabolism clusters, and the characteristics of gut microbiota were associated with metabolic phenotypes. Moreover, we identified cluster-specific microbiome features and further validated their association with glycometabolism with a testing cohort.

RESULTS

Unsupervised-stratification-based clusters of patients with abnormal glycometabolism. During a survey of patients with T2D conducted at Sijing Community Health Service Center of Shanghai Songjiang District, we recruited 267 participants as a discovery cohort and examined 33 bioclinical parameters of these participants. According to ADA criteria (8), 84 participants had normal glucose tolerance (NGT), 52 had isolated impaired fasting glucose (IFG), 34 had isolated impaired glucose tolerance (IGT), 38 had combined glucose intolerance (CGI), and 59 had T2D. We observed that the coefficients of variation (CV) of insulin levels for the oral glucose tolerance test (OGTT), homeostatic model assessment for insulin resistance (HOMA-IR), β-cell function (HOMA-β), serum triglyceride level, and cholesterol levels (total cholesterol, HDL, and LDL) were high in participants with the same classification based on ADA criteria (Fig. 1A). We evaluated insulin resistance and insulin secretion using the HOMA-IR and HOMA-β index, which also showed significant variations in participants within the same glucose range (Fig. 1B; see also Fig. S1 in the supplemental material). These results confirmed that the subjects with same glucose level had high heterogeneity of metabolic state.

Then, we reclassified the same group of participants in the discovery cohort according to variables, including HbA1c, OGTT-derived glucose levels (five-time point blood glucose levels during OGTT), and insulin levels (five-time point blood insulin levels during OGTT), and anthropometric variables (body mass index [BMI], waist circumference, hip circumference), as well as variables related to dyslipidemia and insulin resistance (fasting triglyceride, HDL cholesterol). We standardized the variables before the clustering procedure. We excluded nine participants (NGT = 1, IFG = 2, IGT = 2, CGI = 1, T2D = 3) with at least one outlier variable. By using the K-Mediods clustering algorithm (22), we classified 258 participants into three clusters that were determined according to the maximum silhouette coefficient (silhouette coefficient = 0.21) (Fig. 1C and D). The participants with NGT and prediabetes were clustered only into clusters 1 and 2. Specifically, 80.7% participants with NGT were clustered into cluster 1 and the rest with NGT were clustered into cluster 2. In addition, 68% with IFG, 37.5% with IGT, and 45.9% with CGI were clustered into cluster 1 and the remainder were clustered into cluster 2. In contrast, most of T2D (53.6%) were clustered into cluster 3, only 3.6 and 42.9% were clustered into clusters 1 and 2, respectively. Finally, the Jaccard coefficient means for the three clusters were greater than 0.7 (0.78, 0.71, and 0.83, respectively), which indicated the three clusters were stable.

Compared to cluster 1, clusters 2 and 3 showed more serious disruptions in glycometabolism and lipometabolism and an increased inflammatory state (Table 1). The levels of HbA1c, FBG (fasting blood glucose), 0.5-h PBG (postprandial blood glucose), 1-h PBG, 2-h PBG, and 3-h PBG were lowest in cluster 1 and highest in cluster 3 (Fig. 1E and H and Table 1). The insulin levels during OGTT except 3-h insulin level were significantly higher in cluster 2 than in clusters 1 and 3 (Fig. 1I and Table 1), and the HOMA-IR index was significantly higher in clusters 2 and 3 than in cluster 1 (Fig. 1F), indicating that cluster 1 had the highest insulin sensitivity, and cluster 2 had compensatory secretion insulin, whereas the cluster 3 was insulin deficiency (Fig. 1G). In addition, BMI, waist-to-hip ratio (WHR), and systolic blood pressure (SBP) were significantly higher in cluster 2 than those in the other two clusters (Fig. 1J and Table 1). The levels of total cholesterol, triglyceride, and leptin were significantly higher in clusters 2 and 3 than those in cluster 1 (Fig. 1J and Table 1). The lipopolysaccharide (LPS)-binding protein (LBP) level, an indicator of chronic inflammation, was lowest in cluster 1 and highest in cluster 3 (Table 1).

FIG 1 Metabolic characteristics of the unsupervised-stratification-based clusters. (A, top panel) Coefficients of variation (CV) of clinical variables in each ADA group. The size of the circle indicates the CV value. (Bottom panel) Heatmap of clinical variables. The values were scale transformed by column. (B)

Taken together, participants in cluster 1 showed high insulin sensitivity and the lowest glucose levels and serum lipids. Participants in cluster 2 had serious insulin resistance and high serum lipids levels. Participants in cluster 3 had insulin deficiency, with the highest levels of blood glucose and chronic inflammation.

Differences in gut microbiota in unsupervised-stratification-based clusters. To investigate the differences in the gut microbiota among three unsupervised-stratification-based clusters, we performed 16S rRNA gene V3-V4 region sequencing on fecal samples collected from all of the participants. Although there were no significant differences in diversity and richness of the gut microbiota (see Fig. S2), the PCA of phylogenetic-ILR (PhILR; phylogenetic-isometric log ratio transformation)-transformed Euclidean distances and score plots of the linear discriminant analysis (LDA) showed that the structure and composition of the gut microbiota differed significantly among the three clusters (Fig. 2A and B). Moreover, the distance between cluster 1 and cluster 3 was larger than it was between cluster 1 and cluster 2 (Fig. 2C).

Then we constructed a coabundance network of prevalent ASVs (shared by more than 20% of the samples) in each cluster based on Pearson correlation coefficient to explore the ecological relationship of the members in gut microbial community (Fig. 2D). We calculated the topological parameters of networks in three clusters to explore whether any differences existed in complexity among the microbial networks. The total numbers of nodes were similar (168, 153, and 158) among the three networks, whereas the total numbers of edges varied relatively (638, 420, and 332); in particular, the network in cluster 3 had the lowest number of edges. The network density, which is defined as the ratio of the number of actual edges and the number of possible edges, decreased progressively in the three clusters (the density values in clusters 1, 2, and 3 were 0.045, 0.036, and 0.027, respectively). Moreover, the network degree centrality, a measure of the relative connectivity of each node in a network, decreased from cluster 1 to cluster 3 (Fig. 2E). These results suggested that cluster 1 had more microbial interactions than the other two clusters. Since bacteria act as functional groups (guilds) in the gut ecosystem (23), we next clustered the 188 nodes (i.e., ASVs) of the three networks into 29 coabundance groups (CAGs). Correlation analysis between CAGs and clinical parameters showed significant correlation between gut microbiota and glycometabolism, insulin secretion, and lipid levels (see Fig. S3).

Moreover, we measured the content of short-chain fatty acids (SCFAs) in the fecal samples of all participants. SCFAs are important microbiota-derived metabolites and have been proved to be associated with glycometabolism. We found that cluster 1 had the highest concentration of fecal butyric acid, whereas cluster 3 had the lowest (Fig. 2F). The difference of acetic acid concentration among three clusters was similar with butyric acid (Fig. 2G). We further examined the genes involved in the production of butyric and acetic acid, e.g., *buk* for butyrate and *fhs* for acetate production. The gene abundances showed a pattern similar to that of the fecal butyric and acetic acid concentration among three clusters (see Fig. S4).

Taken together, the β-diversity of gut microbiota, the gut microbial network topology, and the capacity for producing butyrate and acetate were significantly different among these three clusters.

Features of gut microbiota in unsupervised-stratification-based clusters. To identify the cluster-specific microbiome features of three clusters, we compared the abundance of ASVs among the three clusters using Wilcoxon rank sum test (false discovery rate

FIG 1 Legend (Continued)

Variations of HOMA-IR among members within the same blood glucose range. (C) Silhouette coefficient corresponding to the number of clusters from 2 to 20. (D) The number of participants in each cluster, with colors indicating glycemic categories (NGT, normal glucose tolerance; IFG, impaired fasting glycemia; IGT, impaired glucose tolerance; CGI, combined impaired fasting glycemia and impaired glucose tolerance; T2D, type 2 diabetes). (E to G) Comparisons of HbA1c (E), HOMA-IR (F), and HOMA-β (G) among clusters. Boxes show the medians and the interquartile ranges (IQRs), the whiskers denote the lowest and highest values that were within 1.5 times the IQR from the first and third quartiles, and the outliers are shown as individual points. The Kruskal-Wallis test P value is shown at the bottom of each plot. A Wilcoxon rank sum test was used for comparisons between two clusters (adjusted by FDR). Clusters with common characters were not significantly different (FDR > 0.05). (H to J) Radar charts show the median values of clinical parameters related to blood glucose levels (H), blood insulin levels (I), and lipometabolism (J). Each spoke in the chart represents one cluster.

TABLE 1 Comparisons of clinical parameters among unsupervised-stratification-based clusters in a discovery cohort[a]

Category	Mean ± SEM			P
	Cluster 1 ($n = 132$)	Cluster 2 ($n = 96$)	Cluster 3 ($n = 30$)	
Glycemic category (ADA criteria)				
NGT (n)	67	16	0	/
IFG (n)	34	16	0	/
IGT (n)	12	20	0	/
CGI (n)	17	20	0	/
T2D (n)	2	24	30	/
General information				
Age (yrs)	57.42 ± 0.57[A] (131)	57.59 ± 0.62[A] (95)	60.83 ± 1.01[B] (30)	9.2E−03
Gender (no. of females/no. of males)	68/64	53/43	11/19	NS
Glycometabolism indicators				
HbA1c (%)	5.49 ± 0.03[A] (132)	5.76 ± 0.05[B] (96)	7.1 ± 0.19[C] (30)	4.6E−16
FBG (mmol/L)	5.47 ± 0.05[A] (132)	5.7 ± 0.08[B] (96)	8.45 ± 0.28[C] (30)	4.5E−18
0.5-h PBG (mmol/L)	9.22 ± 0.15[A] (129)	10.1 ± 0.23[B] (91)	13.06 ± 0.5[C] (26)	3.9E−11
1-h PBG (mmol/L)	8.77 ± 0.2[A] (129)	10.84 ± 0.32[B] (91)	16.29 ± 0.57[C] (26)	6.4E−18
2-h PBG (mmol/L)	6.66 ± 0.14[A] (132)	8.76 ± 0.27[B] (96)	15.21 ± 0.52[C] (30)	6.6E−24
3-h PBG (mmol/L)	4.41 ± 0.11[A] (125)	5.44 ± 0.21[B] (87)	10.12 ± 0.5[C] (26)	2.4E−16
Glucose AUC (mmol/L/min)	1253.99 ± 23.93[A] (132)	1480.7 ± 43.71[B] (96)	2213.3 ± 131.75[C] (30)	3.7E−14
HOMA-IR	0.8 ± 0.04[A] (130)	1.6 ± 0.11[B] (91)	1.33 ± 0.23[B] (28)	6.7E−09
HOMA-IS	1.69 ± 0.09[A] (130)	1.02 ± 0.08[B] (91)	1.28 ± 0.15[B] (28)	6.7E−09
HOMA-β	38.09 ± 2.66[A] (130)	63.18 ± 4.77[B] (91)	14.71 ± 2.19[C] (28)	4.1E−14
Fasting insulin (μU/mL)	3.32 ± 0.16[A] (130)	6.24 ± 0.41[B] (91)	3.45 ± 0.53[A] (28)	3.6E−09
0.5-h insulin (μU/mL)	20.88 ± 1.3[A] (127)	31.8 ± 2.35[B] (87)	6.62 ± 0.84[C] (28)	5.6E−15
1-h insulin (μU/mL)	24.51 ± 1.29[A] (127)	42.33 ± 2.95[B] (87)	11.18 ± 1.54[C] (28)	7.6E−13
2-h insulin (μU/mL)	19.64 ± 1.38[A] (131)	43.44 ± 3.77[B] (91)	13.15 ± 1.74[A] (28)	1.6E−12
3-h insulin (μU/mL)	5.07 ± 0.44[A] (127)	14.44 ± 1.51[B] (87)	8.67 ± 1.19[B] (28)	7.3E−12
Insulin AUC (μU/mL/min)	3028.32 ± 156.09[A] (130)	5745.88 ± 377.77[B] (91)	1802.82 ± 197.32[C] (28)	1.2E−15
Fasting C-peptide (nmol/L)	1.19 ± 0.05[A] (131)	1.96 ± 0.11[B] (91)	1.32 ± 0.17[A] (28)	2.6E−09
0.5-h C-peptide (nmol/L)	3.67 ± 0.17[A] (127)	5.24 ± 0.32[B] (87)	1.86 ± 0.22[C] (28)	1.2E−11
1-h C-peptide (nmol/L)	4.91 ± 0.23[A] (127)	7.28 ± 0.44[B] (87)	2.67 ± 0.3[C] (28)	7.6E−12
2-h C-peptide (nmol/L)	4.81 ± 0.24[A] (131)	8.09 ± 0.42[B] (91)	4.08 ± 0.54[A] (28)	3.3E−12
3-h C-peptide (nmol/L)	2.41 ± 0.13[A] (127)	4.63 ± 0.29[B] (87)	3.47 ± 0.37[C] (28)	1.7E−11
Anthropometric markers				
BMI (kg/m^2)	23.75 ± 0.18[A] (132)	27.86 ± 0.28[B] (96)	25.17 ± 0.58[A] (30)	6.4E−21
Ht (cm)	160.62 ± 0.67 (132)	161.81 ± 0.86 (96)	163.83 ± 1.06 (30)	9.9E−02
Wt (kg)	61.39 ± 0.68[A] (132)	72.88 ± 0.87[B] (96)	67.71 ± 1.89[C] (30)	5.7E−17
Waist circumference (cm)	80.79 ± 0.51[A] (132)	91.57 ± 0.67[B] (96)	84.82 ± 1.63[A] (30)	9.9E−23
Hip circumference (cm)	92.54 ± 0.34[A] (132)	99.5 ± 0.57[B] (96)	96.04 ± 1.03[C] (30)	1.5E−17
Waist-to-hip ratio	0.87 ± 0[A] (132)	0.92 ± 0[B] (96)	0.88 ± 0.01[A] (30)	8.9E−11
SBP (mm Hg)	128.62 ± 1.1[A] (132)	133.58 ± 1.41[B] (96)	127.13 ± 2.19[A] (30)	3.2E−03
DBP (mm Hg)	81.58 ± 0.68 (132)	83.21 ± 0.74 (96)	82 ± 1.55 (30)	9.4E−02
Lipometabolism indicators				
Total cholesterol (mmol/L)	1.84 ± 0.1[A] (127)	2.41 ± 0.16[B] (91)	1.8 ± 0.25[AB] (27)	1.1E−02
Triglyceride (mmol/L)	0.66 ± 0.04[A] (127)	1.3 ± 0.16[B] (91)	1.05 ± 0.25[AB] (27)	8.5E−04
HDL (mmol/L)	0.63 ± 0.04 (127)	0.7 ± 0.04 (91)	0.56 ± 0.08 (27)	1.2E−01
LDL (mmol/L)	1.06 ± 0.1 (127)	1.11 ± 0.1 (91)	0.8 ± 0.12 (27)	2.4E−01
Leptin (ng/mL)	0.34 ± 0.03[A] (95)	0.51 ± 0.04[B] (78)	0.36 ± 0.05[AB] (27)	1.1E−02
Inflammatory indicators				
LBP (μg/mL)	12.55 ± 0.53[A] (77)	14.12 ± 0.51[B] (64)	16.51 ± 0.92[C] (24)	1.2E−04

[a]A Kruskal-Wallis test was used to compare the differences among the three clusters. For comparisons between two clusters, we used the Wilcoxon rank sum test (adjusted by FDR) for continuous variables and the chi-squared test for categorical variables. Data indicated by the same superscript letters were not significantly different (FDR > 0.05). n, number of samples; FBG, fasting blood glucose; PBG, postprandial blood glucose; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; SBP, systolic blood pressure; DBP, diastolic blood pressure. LBP, lipopolysaccharide-binding protein. HOMA-IR = [FBG (mmol/L) × fasting insulin (μU/mL)]/22.5; HOMA-IS = 22.5/(FBG × fasting insulin); HOMA-β = (20 × fasting insulin)/(FBG −3.5).
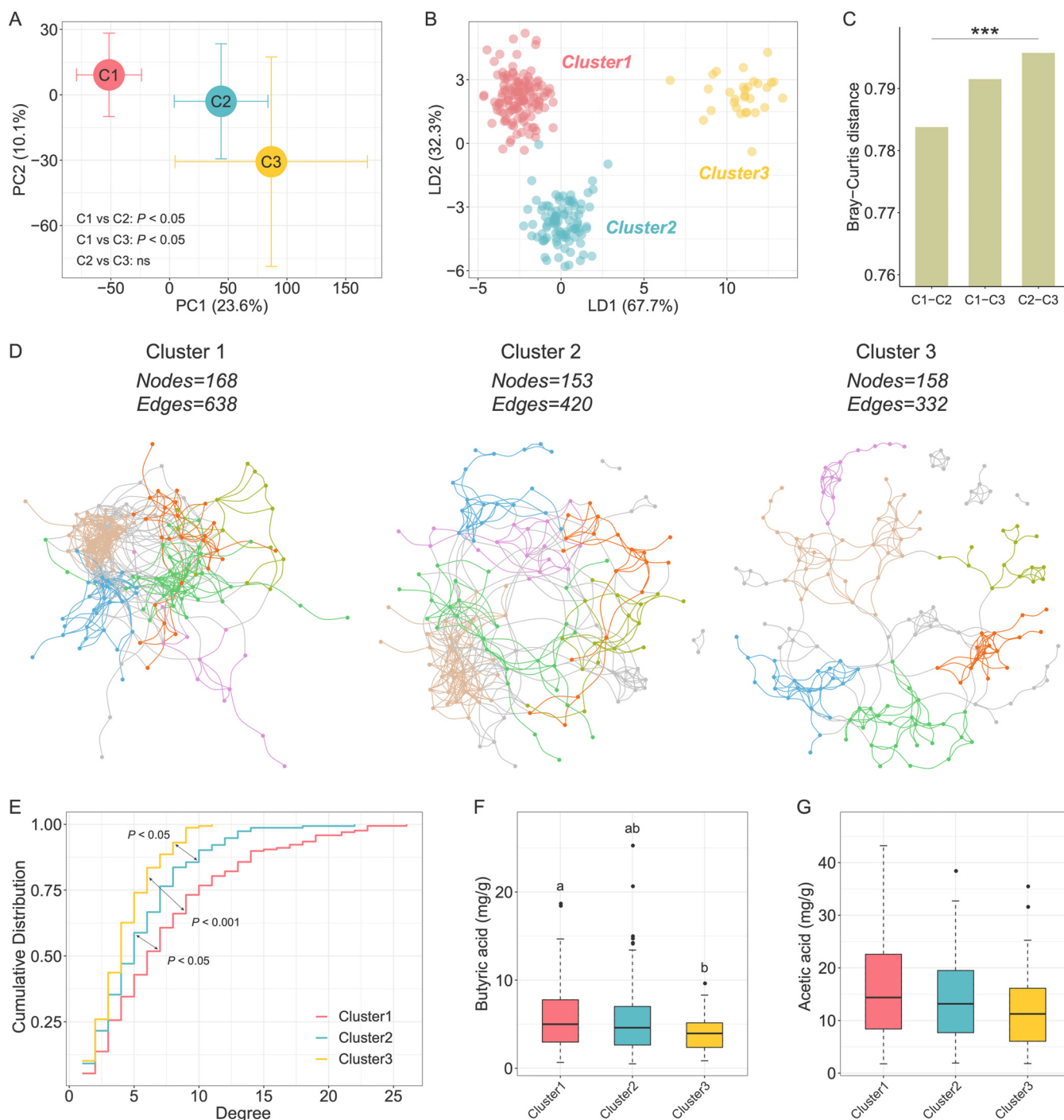
FIG 2 Characterization of the gut microbiota in unsupervised-stratification-based clusters. (A) PCA of PhILR-transformed Euclidean-distance based on the abundance of ASVs. The circles and error bars indicate the means and standard errors of the mean (SEM). Comparisons of gut microbiota structures among three clusters were tested by permutational multivariate analysis of variances (PERMANOVA; permutations = 9,999). (B) LDA score plot of the gut microbiota structure of the three clusters. (C) Between-sample Bray-Curtis distances of the gut microbiota of three clusters. Kruskal-Wallis test (***, $P < 0.001$). (D) Visualization of constructed networks based on Pearson correlation coefficient. The first six modules with a large number of nodes are shown in different colors, and the other modules are shown in gray. (E) Degree centralities of networks from three clusters. Kolmogorov-Smirnov tests were used to test the differences in cumulative distributions. (F and G) Comparisons of fecal butyric acid (F) and fecal acetic acid (G) concentrations among clusters. Boxes, whiskers, and outliers denote values as described for Fig. 1E. A Wilcoxon rank sum test was used for comparisons between two clusters (adjusted by FDR). Clusters with common characters were not significantly different (FDR > 0.05).
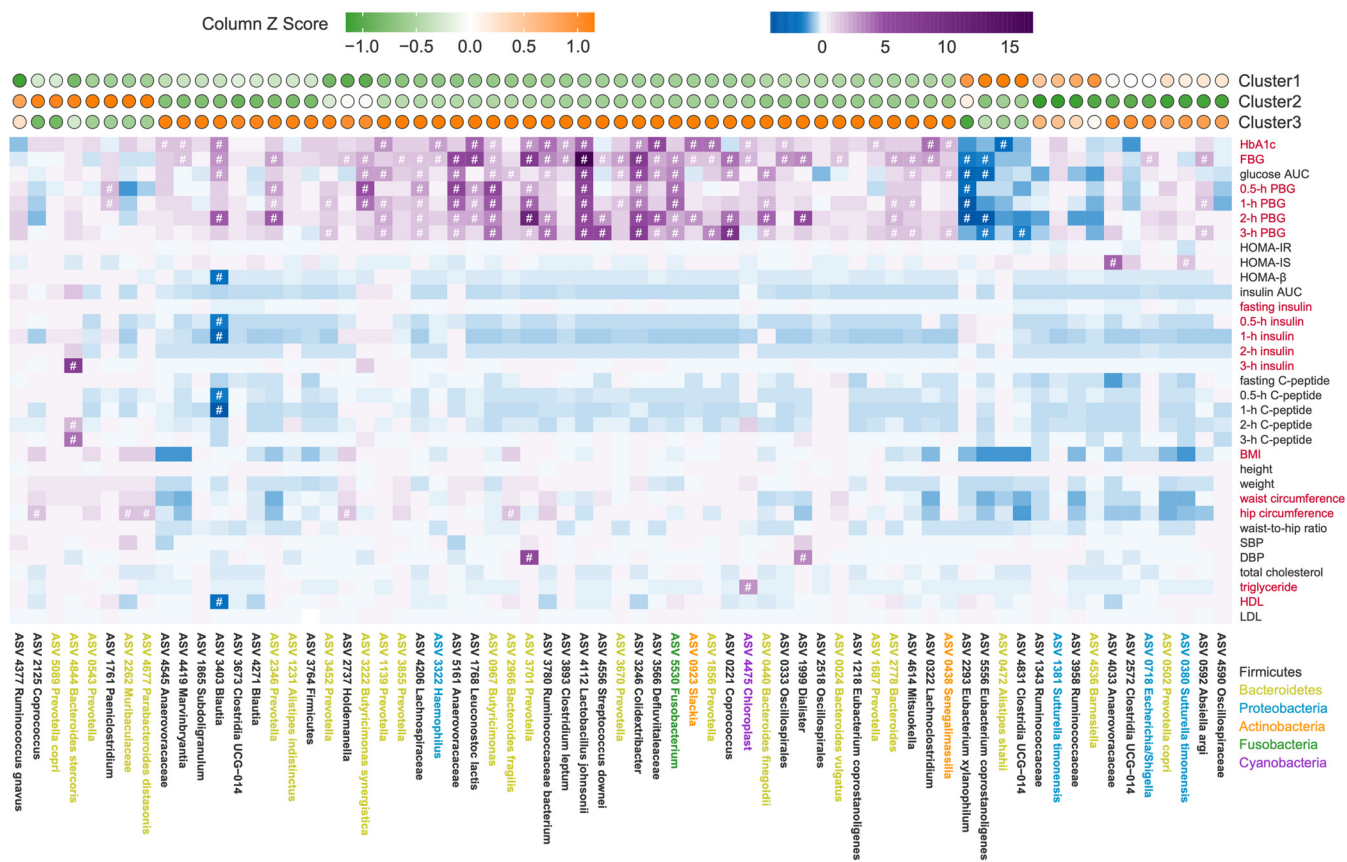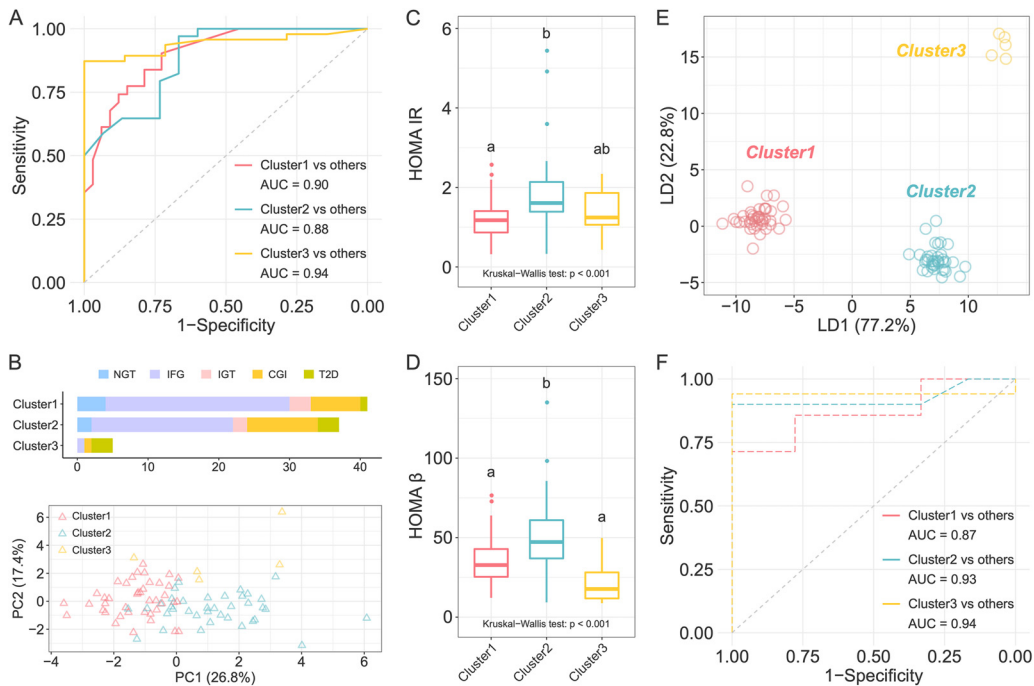
FIG 3 Cluster-specific microbiome features identified based on unsupervised-stratification-based clusters. (Top panel) The colors of circles indicate the scale-transformed mean abundance of the 67 ASVs in each cluster. These ASVs were clustered with a Spearman correlation coefficient and ward linkage based on their scale-transformed abundance values. (Middle panel) Association between ASVs and clinical variables. The colors denote the correlation coefficients. P values were adjusted by Benjamini-Hochberg procedure. #, Adjusted $P < 0.25$ was considered to be statistically significant based on the instruction of MaAslin2. Age and gender were considered to be covariates. Red text on the right indicates the variables used for classification. (Bottom panel) Taxonomy of ASVs. Colors represent the phyla.

[FDR] < 0.05, |log$_2$-fold change| > 1). We found that 67 ASVs were significantly different between at least two clusters (Fig. 3, top panel; see Table S1 in the supplemental material). Only 14 of the 67 ASVs showed significantly altered abundance between clusters 1 and 2, and 36 ASVs were significantly altered between clusters 1 and 3, whereas the abundance of 43 ASVs changed significantly between clusters 2 and 3. β-Diversity analysis based on the Bray-Curtis distance showed significant correlations between the profiles of 67 cluster-specific ASVs and all ASVs (Mantel test, $R = 0.4$, $P = 0.001$; Fig. S5, Procrustes analysis, $P < 0.001$). In addition, the Mantel statistic based on Euclidean distance of 67 cluster-specific ASVs and Euclidean distance of 16 clinical variables used for unsupervised clustering, showed that the alteration pattern of microbiome features was significantly associated with the clinical phenotype (Mantel test, $R = 0.14$, $P = 0.003$).

We subsequently assessed the correlation between the cluster-specific ASVs and all of the host clinical variables based on a modified general linear model (Fig. 3). Among the 67 cluster-specific ASVs, we found that 50 ASVs were correlated with at least one clinical variable. Four ASVs (three belonged to *Clostridia*: *Eubacterium xylanophilum* ASV2293, *Eubacterium coprostanoligenes* ASV5556, and *Clostridia* UCG-014 ASV4831; one belonged to *Bacteroidia*: *Alistipes shahii* ASV0472), which were significantly higher in cluster 1, were negatively correlated with the parameters related to glucose intolerance. Most of the ASVs (39 ASVs) enriched in cluster 3 were significantly positively correlated with parameters related to glucose intolerance, and five of them were positively correlated with parameters related to lipid metabolism. Among them, two ASVs showed significant correlation with insulin-related variables. ASV3403, belonging to *Blautia*, was

FIG 4 Classification based on cluster-specific microbiome features in the testing cohort. (A) Receiver-operating characteristic (ROC) curves for classification of individuals in one cluster from the other two clusters in the discovery cohort. The random forest classifier was constructed based on leave-one-out cross-validation using the 67 cluster-specific ASVs. (B) The number of participants in each cluster in the testing cohort, with colors indicating glycemic categories according to ADA criteria, and PCA plot showing the different clinical phenotype of three clusters. (C and D) Comparisons of HOMA-IR (C) and HOMA-β (D) among clusters. The Kruskal-Wallis test P value is shown at the bottom of each plot. Boxes, whiskers, and outliers denote values as described for Fig. 1E. A Wilcoxon rank sum test was used for comparisons between two clusters (adjusted by FDR). Clusters with common characters were not significantly different (FDR > 0.05). (E) LDA score plot of the three clusters based on the abundance of ASVs. (F) ROC curves for classification of individuals in one cluster from the other two clusters in the testing cohort using the 67 cluster-specific microbial features identified in the discovery cohort.

negatively correlated with HOMA-β, but ASV4033 enriched in cluster 3, was positively correlated with HOMA-IS. One of the ASVs enriched in cluster 2, *Bacteroides stercoris* ASV4844, was positively correlated with insulin and C-peptide levels. Other ASVs (*Paeniclostridium* ASV1761, *Coprococcus* ASV2125, *Muribaculaceae* ASV2262, and *Parabacteroides distasonis* ASV4677) enriched in cluster 2 were positively correlated with parameters related to glucose intolerance or hip circumference. Taken together, the results indicated a significant correlation between cluster-specific microbiome features and clinical phenotypes.

Stratification in testing cohort based on cluster-specific microbiome features. We next developed machine-learning classifiers based on a random forest algorithm by a leave-one-out cross-validation to distinguish individuals in one cluster from those in other two clusters using the 67 cluster-specific ASVs. Receiver operating characteristic curve analysis suggested that the models had high prediction power with area under the curve (AUC) ranging from 0.88 to 0.94 (Fig. 4A). We then tested whether these features of gut microbiota could distinguish the subjects from different glycometabolism clusters in the testing cohort, who were the survey participants at the same community 2 years ago. We assigned the participants in the testing cohort to the nearest one of the three clusters based on Euclidean distance of the clinical variables that had been used for clustering in the discovery cohort except for the 3-h glucose level and 3-h insulin level (because these two variables were not available for the testing cohort) (Fig. 4B). The principal-component analysis (PCA) based on clinical variables, which we used for classification, revealed a separation among the three clusters of testing cohort (Fig. 4B). HOMA-IR and HOMA-β were significantly higher in cluster 2 than in the other two clusters (Fig. 4C and D), which was similar to the differences identified among the clusters in the discovery cohort. In addition, cluster 1 showed the

lowest levels of glucose AUC and cluster 3 showed the highest levels (see Table S2), which suggested that cluster 1 had the best glucose metabolism. The gut microbiota structure of the three clusters in the testing cohort were clearly separated from each other (Fig. 4E), and the score plot of the LDA showed that the gut microbiota structure was different among clusters and not cohorts (see Fig. S6). To explore whether the participants in the different clusters in the testing cohort could be distinguished by the microbial profile, we developed machine-learning classifiers again using the abundance of the 67 cluster-specific ASVs in the testing cohort. Receiver operating characteristic curve analysis suggested that the models had a moderate prediction power with an AUC ranging from 0.87 to 0.94 (Fig. 4F). The classification for the testing cohort suggested that the cluster-specific ASVs reflected the microbiota alterations associated with abnormal glycometabolism.

DISCUSSION

In the current population-based cross-sectional study, we showed that unsupervised stratification of patients with abnormal glycometabolism based on more inclusive clinical parameters could help identify microbiome features more robustly associated with glycometabolism. We confirmed the association between identified microbiome features and glycometabolism in a validation cohort.

The microbiome composition of individuals with T2D has been controversial among studies (24). *Bifidobacterium* and *Bacteroides* were the most reported genera containing microbes related with T2D. *Bifidobacterium* has been reported to be potentially protective against T2D in most studies, whereas only one study has reported conflicting result (2, 25–28). *Bacteroides* has been reported to be negatively correlated with T2D in five cross-sectional studies and positively correlated with T2D in three studies that had involved some type of treatment (4, 5, 24, 29). One of the reasons for this unreliable relationship between gut microbiota and T2D is that individuals within the same blood glucose range are heterogeneous in insulin sensitivity and secretion as well as lipid metabolism, which are also associated with gut microbiota (21). In our study, based on 16 variables that combined blood glucose levels and parameters related to insulin resistance and dyslipidemia, we classified 258 individuals into three clusters with unique metabolic characteristics: cluster 1 was characterized by the lowest blood glucose levels, insulin sensitivity, and lowest lipid levels; cluster 2 was characterized by a moderate level of blood glucose, serious insulin resistance, and high levels of cholesterol and triglyceride; and cluster 3 was characterized by the highest blood glucose levels and insulin deficiency. The Swedish All New Diabetics study reported that clusters identified based on more inclusive indexes and an unsupervised method showed different risk of diabetic complications (13). This result suggested that the stratification based on more inclusive clinical parameters was better than that based only on blood glucose levels, because it not only could separate people within different glucose levels and insulin levels but also could predict the risk of diabetic complications. Based on the unsupervised-stratification-based clusters, we identified cluster-specific microbiome features that not only were related to glycometabolism but also were available for population classification in another general cohort. This finding implied an association between these cluster-specific microbiome features and glycometabolism. Thus, our research findings suggested that the stratification combining the blood glucose levels and indicators related to insulin resistance and dyslipidemia could make it possible to identify the microbiome features associated with abnormal glycometabolism.

The identified cluster-specific microbiome features may contribute to the progression of abnormal glycometabolism. For example, we found that two ASVs belonging to *Prevotella copri* were enriched in clusters 2 and 3, respectively, and one ASV belonging to *Bacteroides vulgatus* was enriched in cluster 3. Both cluster 2 and cluster 3 were characterized by the most resistance to insulin. One study found that *P. copri* and *B. vulgatus* were the strongest driver species for the positive association between HOMA-IR and microbial branched-chain amino acids (BCAAs) biosynthesis in Danish individuals without diabetes and further found

that *P. copri* caused insulin resistance and impaired glucose intolerance by changing the circulating serum levels of BCAAs in mice (30). Transplantation of *B. vulgatus* resulted in insulin resistance in recipient mice (31). Therefore, these two bacteria may have contributed to a high level of insulin resistance in clusters 2 and 3 in our work. Furthermore, we also found a high abundance of *Ruminococcus gnavus* ASV4377 in clusters 2 and 3. Studies reported that *R. gnavus* is a mucin-degrading bacterium that may directly break the integrity of gut barrier and is associated with inflammatory bowel diseases (32–35). The disruption of gut barrier may lead to the translocation of endotoxins produced by gut bacteria to the host. In our study, the level of LBP, a load marker of gut-derived antigens, was higher in clusters 2 and 3, which suggested an increased level of plasma endotoxin load produced by gut bacteria. The increased circulating endotoxin load would induce chronic inflammation, which is a driving factor for insulin resistance and dyslipidemia (36). Thus, the disordered gut microbiome, such as increased levels of *R. gnavus*, may contribute to the insulin resistance and dyslipidemia by disrupting gut barrier, elevating circulated endotoxin load, and inducing chronic inflammation. Moreover, cluster 1 enriched the ASVs belonging to *Barnesiella*. Some species of *Barnesiella* have been reported to produce acetate (37). In addition, high levels of acetate and butyrate concentration, as well as the functional genes involved in the production of these metabolites were observed in cluster 1. Studies have revealed that acetate suppresses body fat accumulation and inflammation in obese or diabetic rodents through multiple mechanisms (38–41). Butyrate also has been shown to improve gut integrity by increasing the tight junction assembly (42), inducing mucin synthesis (43), and decreasing gut bacterial transport across the epithelium (44). Thus, the individuals in cluster 1 may have benefited from the integrity of the intestinal barrier, which was protected by higher acetate/butyrate concentration, and may have avoided chronic inflammation that can be induced by elevated endotoxin load. Taken together, the cluster-specific microbiome features found in the present study may have contributed to the distinct glycometabolism phenotype of the three clusters. The contribution and mechanism of these bacteria in the progression of glycometabolism disorder need to be further experimentally verified.

In this study, we showed that unsupervised stratification based on blood glucose, insulin, and lipid levels led to the identification of cluster-specific gut microbiome features associated with glycometabolism. A well-stratified cohort is a prerequisite for the identification of bacteria associated with glycometabolism. Upon further validation in larger cohorts and follow-up with a longer duration, the elucidation of the mechanism of these identified cluster-specific microbiome features in the progression of abnormal glycometabolism may lead to the future development of biomarkers for early diagnosis and therapeutic treatment.

MATERIALS AND METHODS

Ethical approval. The protocols for both studies were approved by the Human Research Ethics Committee of Shanghai General Hospital (2009KY037, 2013KY083) before the procedure of enrollment. This clinical trial was registered in the Chinese Clinical Trial Registry under number ChiCTR-IPC-14005346. All of the participants signed an informed consent form before sample collection.

Overview of the cohorts. Patients at the Sijing Community Health Service Center of Songjiang District participated in a survey about type 2 diabetes (T2D). We recruited 267 individuals from a diabetes survey taken in 2014, which was considered to be the discovery cohort in the analysis. To test our findings in the discovery cohort, we recruited 86 individuals from a diabetes survey taken in 2012 as the testing cohort.

The participants were asked to fast overnight (more than 10 h) to collect the fasting venous blood. After physical examination and fasting venous blood collection, we performed a 3-h OGTT (75 g glucose) and collected venous blood samples at 30, 60, 120, and 180 min. The blood samples were set at room temperature for 30 min and then centrifuged to obtain the serum. The serum of fasting venous blood was divided into two parts: one was used to evaluate the fasting blood glucose, blood lipid, and inflammation, and the other was immediately stored at −80°C for quantification of LBP and leptin. Stool samples were collected on the day of the physical examination and stored at −80°C quickly until fecal DNA extraction.

Biochemical assays. The levels of HbA1c, serum glucose, serum insulin, serum C-peptide, triglyceride, total cholesterol, HDL cholesterol, and LDL cholesterol were determined at Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine. Enzyme-linked immunosorbent assays (ELISAs) were used to quantify the levels of LBP (Hycult Biotech, PA, USA) and leptin (DL Develop, Wuhan, China) in the lab of Shanghai Jiao Tong University.

Unsupervised stratification. We used 16 clinical variables (HbA1c, five-time point blood glucose levels and insulin levels during OGTT, BMI, waist circumference, hip circumference, triglyceride, HDL) to stratify the discover cohort. We performed unsupervised stratification in an R environment (version 3.6.1). Euclidean distances were calculated (the *vegdist* function in the "vegan" package) for the standardized clinical variables (scaled to a mean of 0 and a standard deviation of 1) to complete the clustering analysis. Individuals with outlier variables (absolute standardized levels of ≥5) were excluded from the clustering analysis. We performed the K-Mediods clustering algorithm using the *pam* function in the "cluster" package to complete the clustering procedure. The silhouette-width was calculated using the *silhouette* function in the "cluster" package. Clusterboot algorithm from the "fpc" package was used to assess the stability of clusters. Finally, other than 9 participants who had at least one outlier variable, 258 participants were clustered into three clusters.

We used the median values of the 14 selected clinical variables in the discovery cohort, except for the 3-h glucose and 3-h insulin levels (because these two variables were not available for the testing cohort), to assign participants to clusters in the testing cohort. We took the nearest neighbors of the three cluster centers based on Euclidean distances. After we removed three participants with outlier variables, 83 participants were used in the testing cohort.

Statistical analysis of clinical data. Statistical analyses of clinical data were performed in an R environment (version 3.6.1). The difference in the clinical variables among clusters was tested by Kruskal-Wallis test. For differences between two clusters, a Wilcoxon rank sum test was used (adjusted by FDR). FDR values were converted into a character-based display in which common characters represented clusters that were not significantly different (the *multcompLetters* function in the "multcompView" package). The Pearson chi-square test was performed to compare the differences in categorical data. $P < 0.05$ (for Kruskal-Wallis test) and FDR < 0.05 (for Wilcoxon rank sum test) were considered to have a significant difference.

Fecal DNA extraction and 16S rRNA gene V3-V4 region sequencing. We extracted fecal microbial DNA based on the previously published method (45). A total of 353 samples were sequenced in four batches on Illumina Miseq system (Illumina Inc., USA). The sequencing library of 16S rRNA gene V3-V4 regions was prepared as previously described (46), according to a modified version of the manufacturer's instructions.

We used QIIME2 software (v2018.11) (47) to process and analyze the 16S rRNA gene read pairs. The raw sequence data were demultiplexed, denoised, and filtered for chimeric reads with the DADA2 plugin (48) to obtain the frequency table and representative sequence file of amplicon sequence variants (ASVs). After we removed the ASVs considered to be contaminants (49), the decontamination table composed of 353 sample and 5,448 ASVs was downsized to 10,000,000 to standardize sequence depth. We used the representative sequence file for taxonomic annotation using the SILVA database (version 138).

Functional gene prediction. The functional genes for producing butyrate (*but*, butyrate kinase) and acetate (*fhs*, formate-tetrahydrofolate ligase) were predicted based on 16S rRNA gene information by using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) (50).

Bioinformatics and statistical analysis of microbiota data. The following analyses of microbiota were performed by R (version 3.6.1). The α-diversity of each sample was calculated with Shannon index, Simpson index, Observed ASVs, and Faith's phylogenetic diversity (PD whole tree) (R-packages "picante," "phyloseq," and "ape"). Structural differences in gut microbiota were assessed by β-diversity based on PhILR-transformed Euclidean distance using the R-packages "phyloseq" and "philr." A random forest model was trained to distinguish the individuals in one cluster from those in other two clusters using the *train* function in the R-package "caret."

Wilcoxon rank sum test (adjusted by FDR) was used to analyze the difference of α-diversity index and identify the significantly different ASVs between two clusters. ASVs were considered to be significantly different between two clusters when the FDR was <0.05 and the absolute value of the logarithmic (base 2) fold change ($|\log_2$-fold change$|$) in relative abundance was >1. Permutational multivariate analysis of variance (PERMANOVA; permutations = 9,999) was used to assess the structural difference between different clusters.

Network construction. In each cluster, prevalent ASVs shared by more than 20% of the samples were used to construct the microbial association network. Networks of clusters 1, 2, and 3 were generated based on Pearson correlations of the prevalent ASVs using the R-package "WGCNA." The correlations with $P < 0.05$ (adjusted by Benjamini-Hochberg procedure) were retained for further analysis. The layout of nodes and edges was determined by the Fruchterman-Reingold layout algorithm using the correlation efficient as weight. The topological characteristics calculation and visualization of the networks were performed using the R-package "igraph." Kolmogorov-Smirnov test (the *ks.test* function in the "stats" package) was used to compare the network's topological characteristics between clusters. $P < 0.05$ was considered to have a significant difference.

Next, the ASVs from the three networks were clustered using the "ward.D2" (the *hclust* function in the "vegan" package) based on the correlation distance which were converted from correlation values. Permutational MANOVA (permutations = 9,999, $P < 0.01$) was used to determine whether the two clades of the cluster tree were not significantly different and clustered into one CAG.

Association of microbiome features and clinical phenotypes. To calculate the associations between microbiome features, as well as CAGs and clinical phenotypes, we used a modified general linear model as implemented by MaAsLin2 (Multivariate microbial Association with Linear Models), which combines an arcsine square root transformed analysis of relative abundances in a standard multivariable linear model while adjusting for gender and age. The P values were adjusted by using the Benjamini-Hochberg procedure. According to the instructions of the "Maaslin2" package, adjusted P value lower than 0.25 was considered to be significant.

Quantification of fecal short-chain fatty acids. Fecal SCFAs were quantified by gas chromatography/mass spectrometry as previously described (2). A Wilcoxon rank sum test was used to analyze the difference of SCFAs between clusters.

Data availability. The raw sequence data reported have been deposited (PRJCA013291) in the Genome Sequence Archive (GSA) database under accession number CRA008952.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.02 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.3 MB.
FIG S4, PDF file, 0.2 MB.
FIG S5, PDF file, 0.03 MB.
FIG S6, PDF file, 0.3 MB.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, DOCX file, 0.02 MB.

REFERENCES

1. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55–60. https://doi.org/10.1038/nature11450.

2. Zhao L, Zhang F, Ding X, Wu G, Lam YY, Wang X, Fu H, Xue X, Lu C, Ma J, Yu L, Xu C, Ren Z, Xu Y, Xu S, Shen H, Zhu X, Shi Y, Shen Q, Dong W, Liu R, Ling Y, Zeng Y, Wang X, Zhang Q, Wang J, Wang L, Wu Y, Zeng B, Wei H, Zhang M, Peng Y, Zhang C. 2018. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science 359:1151–1156. https://doi.org/10.1126/science.aao5774.

3. Wu H, Tremaroli V, Schmidt C, Lundqvist A, Olsson LM, Kramer M, Gummesson A, Perkins R, Bergstrom G, Backhed F. 2020. The gut microbiota in prediabetes and diabetes: a population-based cross-sectional study. Cell Metab 32:379–390. https://doi.org/10.1016/j.cmet.2020.06.011.

4. Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, Chen Y, Ji L. 2013. Human gut microbiota changes reveal the progression of glucose intolerance. PLoS One 8:e71108. https://doi.org/10.1371/journal.pone.0071108.

5. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F. 2013. Gut metagenome in European women with normal, impaired, and diabetic glucose control. Nature 498:99–103. https://doi.org/10.1038/nature12198.

6. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Costea PI, Kultima JR, Li J, Jørgensen T, Levenez F, Dore J, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O, MetaHIT Consortium. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528:262–266. https://doi.org/10.1038/nature15766.

7. Allin KH, Tremaroli V, Caesar R, Jensen BAH, Damgaard MTF, Bahl MI, Licht TR, Hansen TH, Nielsen T, Dantoft TM, Linneberg A, Jorgensen T, Vestergaard H, Kristiansen K, Franks PW, Consortium I-D, Hansen T, Backhed F, Pedersen O, IMI-DIRECT Consortium. 2018. Aberrant intestinal microbiota in individuals with prediabetes. Diabetologia 61:810–820. https://doi.org/10.1007/s00125-018-4550-1.

8. American Diabetes Association. 2010. Diagnosis and classification of diabetes mellitus. Diabetes Care 33(Suppl 1):S62–S69. https://doi.org/10.2337/dc10-S062.

9. Alberti KG, Zimmet PZ. 1998. Definition, diagnosis, and classification of diabetes mellitus and its complications. 1. Diagnosis and classification of diabetes mellitus: provisional report of a WHO consultation. Diabet Med 15: 539–553. https://doi.org/10.1002/(SICI)1096-9136(199807)15:7%3C539::AID-DIA668%3E3.0.CO;2-S.

10. Kahn SE. 2003. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. Diabetologia 46:3–19. https://doi.org/10.1007/s00125-002-1009-0.

11. Kannel WB. 1985. Lipids, diabetes, and coronary heart disease: insights from the Framingham Study. Am Heart J 110:1100–1107. https://doi.org/10.1016/0002-8703(85)90224-8.

12. Wagner R, Heni M, Tabák AG, Machann J, Schick F, Randrianarisoa E, Hrabě de Angelis M, Birkenfeld AL, Stefan N, Peter A, Häring H-U, Fritsche A. 2021. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 27:49–57. https://doi.org/10.1038/s41591-020-1116-9.

13. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, Vikman P, Prasad RB, Aly DM, Almgren P, Wessman Y, Shaat N, Spégel P, Mulder H, Lindholm E, Melander O, Hansson O, Malmqvist U, Lernmark Å, Lahti K, Forsén T, Tuomi T, Rosengren AH, Groop L. 2018. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 6: 361–369. https://doi.org/10.1016/S2213-8587(18)30051-2.

14. Stidsen JV, Henriksen JE, Olsen MH, Thomsen RW, Nielsen JS, Rungby J, Ulrichsen SP, Berencsi K, Kahlert JA, Friborg SG, Brandslund I, Nielsen AA, Christiansen JS, Sorensen HT, Olesen TB, Beck-Nielsen H. 2018. Pathophysiology-based phenotyping in type 2 diabetes: a clinical classification tool. Diabetes Metab Res Rev 34:e3005. https://doi.org/10.1002/dmrr.3005.

15. Fei N, Zhao L. 2013. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J 7:880–884. https://doi.org/10.1038/ismej.2012.153.

16. Vrieze A, Van Nood E, Holleman F, Salojärvi J, Kootte RS, Bartelsman JFWM, Dallinga–Thie GM, Ackermans MT, Serlie MJ, Oozeer R, Derrien M, Druesne A, Van Hylckama Vlieg JET, Bloks VW, Groen AK, Heilig HGHJ, Zoetendal EG, Stroes ES, de Vos WM, Hoekstra JBL, Nieuwdorp M. 2012. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology 143: 913–916. https://doi.org/10.1053/j.gastro.2012.06.031.

17. Koh A, Molinaro A, Stahlman M, Khan MT, Schmidt C, Manneras-Holm L, Wu H, Carreras A, Jeong H, Olofsson LE, Bergh PO, Gerdes V, Hartstra A, de Brauw M, Perkins R, Nieuwdorp M, Bergstrom G, Backhed F. 2018. Microbially produced imidazole propionate impairs insulin signaling through mTORC1. Cell 175:947–961. https://doi.org/10.1016/j.cell.2018.09.055.

18. Velagapudi VR, Hezaveh R, Reigstad CS, Gopalacharyulu P, Yetukuri L, Islam S, Felin J, Perkins R, Borén J, Oresic M, Bäckhed F. 2010. The gut microbiota modulates host energy and lipid metabolism in mice. J Lipid Res 51:1101–1112. https://doi.org/10.1194/jlr.M002774.

19. An HM, Park SY, Lee DK, Kim JR, Cha MK, Lee SW, Lim HT, Kim KJ, Ha NJ. 2011. Antiobesity and lipid-lowering effects of *Bifidobacterium* spp. in high fat diet-induced obese rats. Lipids Health Dis 10:116. https://doi.org/10.1186/1476-511X-10-116.

20. Yoo SR, Kim YJ, Park DY, Jung UJ, Jeon SM, Ahn YT, Huh CS, McGregor R, Choi MS. 2013. Probiotics L. plantarum and L. curvatus in combination alter hepatic lipid metabolism and suppress diet-Induced obesity. Obesity (Silver Spring) 21:2571–2578. https://doi.org/10.1002/oby.20428.

21. Aron-Wisnewsky J, Warmbrunn MV, Nieuwdorp M, Clément K. 2021. Metabolism and metabolic disorders and the microbiome: the intestinal microbiota associated with obesity, lipid metabolism, and metabolic health—pathophysiology and therapeutic strategies. Gastroenterology 160:573–599. https://doi.org/10.1053/j.gastro.2020.10.057.

22. Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. 2006. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J Mathemat Model Algorithms 5:475–504. https://doi.org/10.1007/s10852-005-9022-1.

23. Zhang C, Yin A, Li H, Wang R, Wu G, Shen J, Zhang M, Wang L, Hou Y, Ouyang H, Zhang Y, Zheng Y, Wang J, Lv X, Wang Y, Zhang F, Zeng B, Li W, Yan F, Zhao Y, Pang X, Zhang X, Fu H, Chen F, Zhao N, Hamaker BR, Bridgewater LC, Weinkove D, Clement K, Dore J, Holmes E, Xiao H, Zhao G, Yang S, Bork P, Nicholson JK, Wei H, Tang H, Zhang X, Zhao L. 2015. Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. EBioMedicine 2:968–984. https://doi.org/10.1016/j.ebiom.2015.07.007.

24. Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, Shulzhenko N. 2020. Role of gut microbiota in type 2 diabetes pathophysiology. EBioMedicine 51:102590. https://doi.org/10.1016/j.ebiom.2019.11.051.

25. Sedighi M, Razavi S, Navab-Moghadam F, Khamseh ME, Alaei-Shahmiri F, Mehrtash A, Amirmozafari N. 2017. Comparison of gut microbiota in adult patients with type 2 diabetes and healthy individuals. Microb Pathog 111:362–369. https://doi.org/10.1016/j.micpath.2017.08.038.

26. Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Manneras-Holm L, Stahlman M, Olsson LM, Serino M, Planas-Felix M, Xifra G, Mercader JM, Torrents D, Burcelin R, Ricart W, Perkins R, Fernandez-Real JM, Bäckhed F. 2017. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med 23:850–858. https://doi.org/10.1038/nm.4345.

27. Xu J, Lian F, Zhao L, Zhao Y, Chen X, Zhang X, Guo Y, Zhang C, Zhou Q, Xue Z, Pang X, Zhao L, Tong X. 2015. Structural modulation of gut microbiota during alleviation of type 2 diabetes with a Chinese herbal formula. ISME J 9:552–562. https://doi.org/10.1038/ismej.2014.177.

28. Sasaki M, Ogasawara N, Funaki Y, Mizuno M, Iida A, Goto C, Koikeda S, Kasugai K, Joh T. 2013. Transglucosidase improves the gut microbiota profile of type 2 diabetes mellitus patients: a randomized double-blind, placebo-controlled study. BMC Gastroenterol 13:81. https://doi.org/10.1186/1471-230X-13-81.

29. Sun L, Xie C, Wang G, Wu Y, Wu Q, Wang X, Liu J, Deng Y, Xia J, Chen B, Zhang S, Yun C, Lian G, Zhang X, Zhang H, Bisson WH, Shi J, Gao X, Ge P, Liu C, Krausz KW, Nichols RG, Cai J, Rimal B, Patterson AD, Wang X, Gonzalez FJ, Jiang C. 2018. Gut microbiota and intestinal FXR mediate the clinical benefits of metformin. Nat Med 24:1919–1929. https://doi.org/10.1038/s41591-018-0222-4.

30. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, Forslund K, Hildebrand F, Prifti E, Falony G. 2016. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature 535:376–381. https://doi.org/10.1038/nature18646.

31. Qi X, Yun C, Sun L, Xia J, Wu Q, Wang Y, Wang L, Zhang Y, Liang X, Wang L, Gonzalez FJ, Patterson AD, Liu H, Mu L, Zhou Z, Zhao Y, Li R, Liu P, Zhong C, Pang Y, Jiang C, Qiao J. 2019. Gut microbiota–bile acid–interleukin-22 axis orchestrates polycystic ovary syndrome. Nat Med 25:1225–1233. https://doi.org/10.1038/s41591-019-0509-0.

32. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin TH. 2010. Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. Am J Gastroenterol 105:2420–2428. https://doi.org/10.1038/ajg.2010.281.

33. Tailford LE, Owen CD, Walshaw J, Crost EH, Hardy-Goddard J, Le Gall G, de Vos WM, Taylor GL, Juge N. 2015. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. Nat Commun 6:7624. https://doi.org/10.1038/ncomms8624.

34. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N. 2013. Utilization of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain dependent. PLoS One 8:e76341. https://doi.org/10.1371/journal.pone.0076341.

35. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. 2019. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. Proc Natl Acad Sci U S A 116:12672–12677. https://doi.org/10.1073/pnas.1904099116.

36. Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget A, Delmee E, Cousin B, Sulpice T, Chamontin B, Ferrieres J, Tanti JF, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin R. 2007. Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes 56:1761–1772. https://doi.org/10.2337/db06-1491.

37. Sakamoto M, Lan PTN, Benno Y. 2007. *Barnesiella viscericola* gen. nov., sp. nov., a novel member of the family *Porphyromonadaceae* isolated from chicken caecum. Int J Syst Evol Microbiol 57(Part 2):342–346. https://doi.org/10.1099/ijs.0.64709-0.

38. Sakakibara S, Yamauchi T, Oshima Y, Tsukamoto Y, Kadowaki T. 2006. Acetic acid activates hepatic AMPK and reduces hyperglycemia in diabetic KK-A(y) mice. Biochem Biophys Res Commun 344:597–604. https://doi.org/10.1016/j.bbrc.2006.03.176.

39. Yamashita H, Fujisawa K, Ito E, Idei S, Kawaguchi N, Kimoto M, Hiemori M, Tsuji H. 2007. Improvement of obesity and glucose tolerance by acetate in type 2 diabetic Otsuka Long-Evans Tokushima Fatty (OLETF) rats. Biosci Biotechnol Biochem 71:1236–1243. https://doi.org/10.1271/bbb.60668.

40. Kondo T, Kishi M, Fushimi T, Kaga T. 2009. Acetic acid upregulates the expression of genes for fatty acid oxidation enzymes in liver to suppress body fat accumulation. J Agric Food Chem 57:5982–5986. https://doi.org/10.1021/jf900470c.

41. Carvalho BM, Guadagnini D, Tsukumo DML, Schenka AA, Latuf-Filho P, Vassallo J, Dias JC, Kubota LT, Carvalheira JBC, Saad MJA. 2012. Modulation of gut microbiota by antibiotics improves insulin signalling in high-fat fed mice. Diabetologia 55:2823–2834. https://doi.org/10.1007/s00125-012-2648-4.

42. Peng L, Li ZR, Green RS, Holzman IR, Lin J. 2009. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. J Nutr 139:1619–1625. https://doi.org/10.3945/jn.109.104638.

43. Burger-van Paassen N, Vincent A, Puiman PJ, van der Sluis M, Bouma J, Boehm G, van Goudoever JB, van Seuningen I, Renes IB. 2009. The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: implications for epithelial protection. Biochem J 420:211–219. https://doi.org/10.1042/BJ20082222.

44. Lewis K, Lutgendorff F, Phan V, Soderholm JD, Sherman PM, McKay DM. 2010. Enhanced translocation of bacteria across metabolically stressed epithelia is reduced by butyrate. Inflamm Bowel Dis 16:1138–1148. https://doi.org/10.1002/ibd.21177.

45. Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. 1997. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. Appl Environ Microbiol 63:2802–2813. https://doi.org/10.1128/aem.63.7.2802-2813.1997.

46. Zhang Q, Wu Y, Wang J, Wu G, Long W, Xue Z, Wang L, Zhang X, Pang X, Zhao Y, Zhao L, Zhang C. 2016. Accelerated dysbiosis of gut microbiota during aggravation of DSS-induced colitis by a butyrate-producing bacterium. Sci Rep 6:27572–27511. https://doi.org/10.1038/srep27572.

47. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9.

48. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

49. Moossavi S, Fehr K, Khafipour E, Azad MB. 2021. Repeatability and reproducibility assessment in a large-scale population-based microbiota study: case study on human milk microbiota. Microbiome 9:41. https://doi.org/10.1186/s40168-020-00998-4.

50. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. 2020. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38:685–688. https://doi.org/10.1038/s41587-020-0548-6.