

Abundance of *Prevotella copri* in gut microbiota is inversely related to a healthy diet in patients with type 2 diabetes

Chih-Yiu Tsai^{a,b}, Po-Yu Liu^c, Meng-Chuan Huang^{d,e}, Chiao-I Chang^e, Hsin-Yun Chen^f, Yu-Hsien Chou^a, Chi-Neu Tsai^{b,g,**}, Chia-Hung Lin^{a,h,*}

^a Division of Endocrinology and Metabolism, Chang Gung Memorial Hospital, Taoyuan, Taiwan

^b Graduate Institute of Clinical Medical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan

^c School of Medicine, College of Medicine, National Sun Yat-sen University, Kaohsiung, Taiwan

^d Department of Nutrition and Dietetics, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

^e Department of Public Health and Environmental Medicine, School of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

^f Department of Nutrition Therapy, Chang Gung Memorial Hospital, Taoyuan, Taiwan

^g Department of Surgery, New Taipei Municipal Tucheng Hospital, New Taipei City, Taiwan

^h Department of Chinese Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan

Abstract

While the gut microbiota is known to be influenced by habitual food intake, this relationship is seldom explored in type 2 diabetes patients. This study aims to investigate the relationship between dietary patterns and gut microbial species abundance in 113 type 2 diabetes patients (mean age, 58 years; body mass index, 29.1; glycohemoglobin [HbA1c], 8.1%). We analyzed the gut microbiota using 16S amplicon sequencing, and all patients were categorized into either the *Bacteroides* enterotype (57.5%, n = 65) or the *Prevotella* enterotype (42.5%, n = 48) using the partitioning around medoids clustering algorithm, based on the most representative genera. Patients with the *Bacteroides* enterotype showed better glycemic control with a 2.71 odds of HbA1c \leq 7.0% compared to the *Prevotella* enterotype (95% confidence interval, 1.02–7.87; *P*, 0.034). Dietary habits and the nutrient composition of all patients were assessed using a validated food frequency questionnaire. It was observed that the amounts of dietary fiber consumed were suboptimal, with an average intake of 16 g per day. Additionally, we extracted four dietary patterns through factor analysis: eating-out, high-sugar foods, fish–vegetable, and fermented foods patterns. Patients with the *Bacteroides* enterotype had higher scores for the fish–vegetable pattern compared to the *Prevotella* enterotype (0.17 \pm 0.13 versus -0.23 ± 0.09 ; *P*, 0.010). We further investigated the relationship between the microbiota and the four dietary patterns and found that only the fish–vegetable dietary pattern scores were correlated with principal coordinate values. A lower pattern score was associated with the accumulated abundance of the 31 significant microbial features. Among these features, *Prevotella copri* was identified as the most significant by using a random forest model, with an area under the receiver operating characteristic of 0.93 (95% confidence interval, 0.88–0.98). To validate these results, we conducted a custom quantitative polymerase chain reaction assay. This assay confirmed the presence of *P. copri* (sensitivity, 0.96; specificity, 0.97) in our cohort, with a prevalence of 47.8%, and a mean relative abundance of 21.0% in subjects harboring *P. copri*. In summary, type 2 diabetes patients with the *Prevotella* enterotype demonstrated poorer glycemic control and deviations from a healthy dietary pattern. The abundance of *P. copri*, as a major contributing microbial feature, was associated with the severity in the deficiency in dietary fish and vegetables. Emphasis should be placed on promoting a healthy dietary pattern and understanding the microbial correlations.

Keywords: Dietary pattern, Gut microbiota, *Prevotella copri*, Type 2 diabetes

Received 11 June 2023; revised 24 November 2023; accepted 13 November 2023.
Available online 15 December 2023

* Corresponding author at: No.5, Fuxing Street, Guishan Dist., Taoyuan City 33305, Taiwan.

** Corresponding author at: No.259, Wenhua 1st Rd., Guishan Dist., Taoyuan City 33302, Taiwan.

E-mail addresses: pink7@mail.cgu.edu.tw (C.-N. Tsai), adronlin@cgmh.org.tw (C.-H. Lin).

<https://doi.org/10.38212/2224-6614.3484>

2224-6614/© 2023 Taiwan Food and Drug Administration. This is an open access article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The prevalence of type 2 diabetes is increasing worldwide, with an estimated 700 million patients predicted by 2045 [1]. This increase in type 2 diabetes is partially attributed to lifestyle changes, particularly the Westernized diet [2]. Medical nutrition therapy should be re-emphasized as the center of diabetes management. A healthful eating pattern such as the Mediterranean diet should be widely adopted because of its metabolic and cardiovascular benefits [3]. Nutrition counseling should take into account factors such as cultural background, socioeconomic setting, and personal preference [4]. In addition, the therapeutic effect of a dietary plan may vary between patients according to their microbiota [5]. Consideration of the characteristics of individual microbiota is important for personalized nutrition but is often neglected.

The gut microbiota is closely associated with the pathogenesis of type 2 diabetes. Bacterial fermentation of dietary fibers produces short-chain fatty acids that modulate human energy homeostasis. Microbial dysbiosis may contribute to leaky gut syndrome, endotoxemia, and insulin resistance [6]. The gut microbiota is the mediator between food and the host metabolism. Moreover, the long-term dietary pattern is the major force that shapes the microbial composition of the gut [7]. Understanding the relationship between the microbiota and dietary patterns would elucidate key dietary factors important for diabetes control.

The concept of enterotypes provides a diagnostic tool in searching for biomarkers among complex microbiome compositions [8]. The prevalence of *Prevotella* and *Bacteroides* in the microbiota tends to differ according to diet [9]. Enterotype *Prevotella* is highly dominant in rural populations with a pre-industrial diet, while *Bacteroides* is prevalent among modern westernized populations. Because the Westernized diet may predispose individuals to the onset of diabetes, *Bacteroides* has been proposed as marker for populations at risk for type 2 diabetes. However, studies investigating this theory report conflicting results [10,11]. *Prevotella copri*, the major species in this enterotype, plays a role in human carbohydrate homeostasis, but its effects on glycemic control are also controversial [12,13]. Possible reasons for the conflicting results are that population-based studies did not include consistent dietary evaluation and that confounding effects of nutrients were sometimes omitted.

To address these previous shortcomings, the present study aims to investigate the microbiota together

with dietary information to identify associations with type 2 diabetes. Because the Western-Pacific is the largest region with a rapid increase in type 2 diabetes, a cohort study of populations in this area is needed to address this issue. The cohort of this study is a population of Han Chinese with type 2 diabetes who underwent team care in a diabetes treatment program. To assess the severity of diabetes, we collected clinical measurements related to glycemic control and comorbidities such as obesity, dyslipidemia, and diabetic kidney disease. These characteristics were compared with the enterotyping results to determine their clinical significance in relation to the microbiota. Habitual food intake was evaluated in relation to the therapeutic results of nutrition therapy. A series of analyses were performed to identify dietary patterns associated with gut microbiota. Furthermore, a quantitative polymerase chain reaction (PCR) assay was designed for confirming the identity of the most abundant microbial species. The relationships between clinical characteristics, enterotypes, dietary patterns, and significant microbial features in type 2 diabetes patients were discussed in this study.

2. Methods

2.1. Study participants and clinical data

We recruited 113 participants with type 2 diabetes. These participants were consecutively enrolled from all patients visiting the outpatient departments of endocrinology and metabolism at Linkou, Taipei and Taoyuan Chang Gung Memorial hospital in northern Taiwan. The participants were undergoing standardized team care provided by certificated diabetes health care institutes. Inclusion criteria included a diagnosis of type 2 diabetes and age ≥ 20 years old. Exclusion criteria included active inflammatory or infectious diseases, advanced chronic kidney disease (estimated glomerular filtration rate [eGFR] < 30 mL/min/1.73 m²), recent abdominal pain or diarrhea within the previous month, and recent use of antibiotics, probiotics, or prebiotics within the previous month.

Each participant's age, sex, duration of type 2 diabetes, and concurrent antidiabetic medication use were recorded. Body mass index (BMI), serum creatinine, alanine aminotransferase (ALT), fasting plasma glucose (FPG), glycohemoglobin (HbA1c), lipid profiles and urinary albumin-creatinine ratio (UACR) were measured by standard protocols. Kidney function was assessed using the isotope dilution mass spectrometry (IDMS) traceable Modification of Diet in Renal Disease (MDRD) study

equation. The Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) index was calculated for participants with paired FPG and fasting serum insulin levels. An informed consent form approved by the Institutional Review Board of Chang Gung Memorial Hospital was signed by each participant (certificate numbers: 201900467B0 and 202002572B0).

2.2. Dietary habit evaluation and dietary pattern analysis

A validated food frequency questionnaire designed for type 2 diabetes patients in Taiwan was used to evaluate consumption frequency of foods, food servings, and eating habits [14]. Our subjects were interviewed by registered dietitians working in our hospital and were asked about the frequency of consumption of a list of foods (groups). Nine frequency options ranging from “almost never” to “4–6 times/day” were listed for all food items. The consumption of 15 food items (groups) was assessed as the number of portions per day, week, or month. These data were converted to portion equivalents per week to determine the overall energy intake (kcal/day), macronutrient intake (carbohydrate, fat, and protein [g/day]), and fiber intake (g/day) [15]. The consumption frequency or portions of 38 specific food items or dietary habits on a per-week basis were extracted into four dietary major dietary patterns using factor analysis. The factors were rotated using Varimax rotation to achieve a simpler structure with greater interpretability. Each participant obtained four pattern scores. Four dietary patterns were identified: fish and vegetables, fermented foods, eating out, and high-sugar foods. Food items with factor loading ≥ 0.2 were considered to contribute significantly to the dietary pattern (Fig. 1). The modeling process of factor analysis was conducted using SPSS Version 22.0 (SPSS Inc., Chicago, IL, USA). The association between each dietary factor score for each food pattern and clinical parameters was analyzed using Pearson's correlation.

2.3. 16S ribosomal RNA amplicon sequencing and raw reads processing

The collection and processing of fecal samples, extraction and storage of non-human DNA, as well as the subsequent library preparation, adhered closely to the procedures outlined in our prior study [16]. Additionally, 16S ribosomal RNA (rRNA) amplicon sequencing was performed by the Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou. Following the processing

and denoising of raw sequence data, high-quality reads of amplicon sequence variants (ASVs) were obtained, ranging from 17,354 to 191,764 per sample (with an average of $81,425 \pm 27,365$). Subsequent to this step, the classification of taxonomic assignment was carried out. These procedures were conducted in accordance with the same established protocols. To optimize the biological relevance to clinical factors, a sequence alignment-free algorithm was applied to cluster 5309 ASVs into 1684 k-mer taxonomic units (KTUs) [17].

2.4. Microbial community analysis

Microbial community analysis was performed using R Software Version 4.2.3 and the *vegan* package [18]. Enterotyping for all samples was based on relative genus abundances using Jensen–Shannon distance and the Partitioning Around Medoids clustering algorithm. The Calinski–Harabasz Index was used to assess the optimal number of clustering. To evaluate the alpha diversity in groups of different enterotypes, we calculated the observed KTU number, Shannon index, and Simpson's index. Wilcoxon rank sum tests were used to detect significant differences (α at 0.05). Principal coordinate analysis (PCoA) with Bray–Curtis dissimilarity was applied to visualize the results of enterotyping. Beta diversity was compared between groups using the permutational multivariate analysis of variance using distance matrices (PERMANOVA) method. Differences in clinical parameters between different enterotype groups were determined using the independent-samples t-test, Fisher's exact test, or Kruskal–Wallis test in accordance with data types and with a significance level α at 0.05.

To investigate the association between a specific dietary pattern and gut microbiota, Pearson's correlation of the pattern scores and principal coordinate values (the projection on PCoA axis 1 or PCoA axis 2) were calculated for each pattern. Pearson's correlation of all KTUs and the related axis in PCoA were calculated to identify the significant bacterial features that were candidates for association with the dietary pattern. The significantly correlated KTUs were categorized into two communities with positive or negative coefficients. To further evaluate the significance of the positive and negative microbial communities, Pearson's correlation of the sum of KTUs in log-transformed abundances and the microbiota-associated dietary pattern scores, respectively, were calculated.

A random forest model was used for selecting the most important KTUs that could distinguish participants with different enterotypes. Receiver

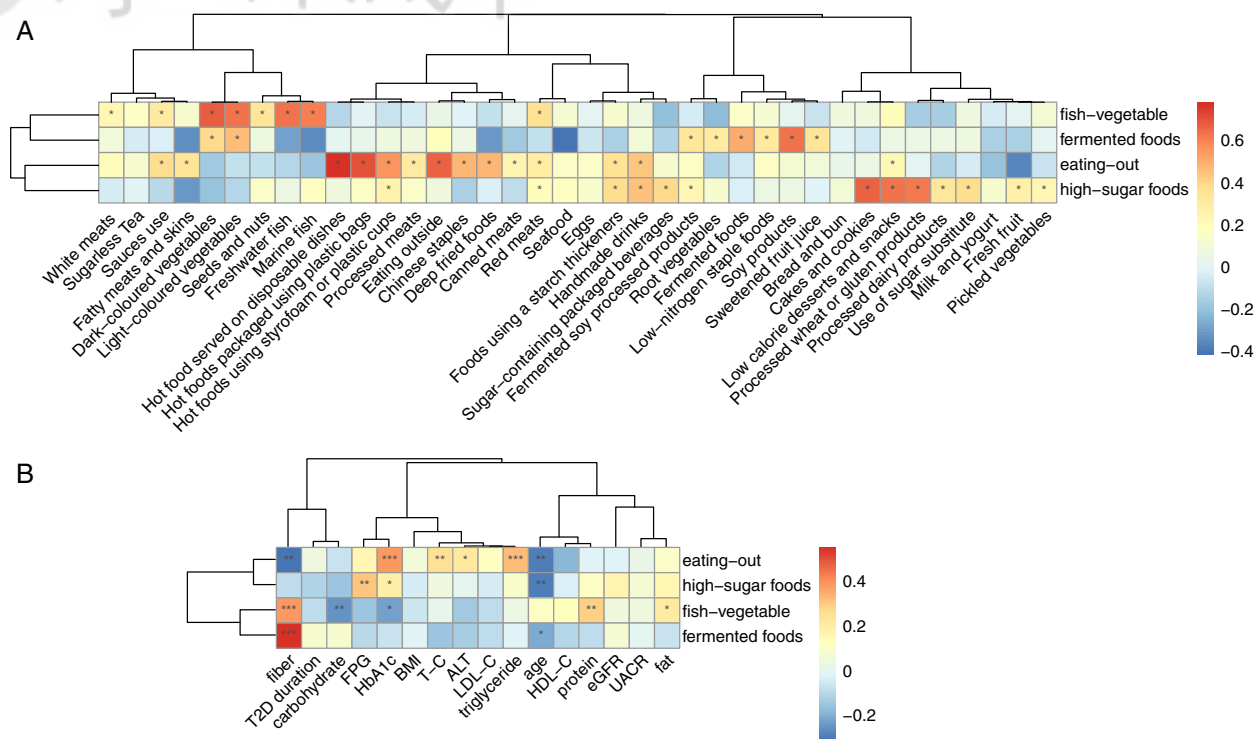


Fig. 1. A) The loading values of 38 dietary factors are presented by the color gradient in the heatmap. Dietary patterns are named according to representative foods components or behaviors. *Cells with loading values ≥ 0.2 B) Pearson's correlation coefficients of dietary patterns versus macronutrients and clinical parameters are shown by a color gradient. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. ALT, alanine aminotransferase; BMI, body mass index; eGFR, estimated glomerular filtration rate; FPG, fasting plasma glucose; HbA1c, glycohemoglobin; HDL-C, HDL-cholesterol; LDL-C, LDL-cholesterol; T-C, total-cholesterol; T2D, type 2 diabetes; UACR, urinary albumin-creatinine ratio.

operating characteristic (ROC) analysis was applied and the area under the receiver operating characteristic (AUROC) was used to evaluate the performance of selected features at a 95% confidence level. The R codes for plots of alpha and beta diversity, heatmap, random forest, and ROC analysis were modified from the source code of the MARco package [19].

2.5. Real-time quantitative PCR validation for *P. copri*

A custom TaqMan assay (Assay ID: APH6D2Y, Thermo Fisher Scientific Inc., Waltham, MA, USA) was designed in accordance with 9 sequences (AB244770.1, AB244773.1, AB244771.1, AB244772.1, MG592701.1, MG592702.1, MN537545.1, NR_113411.1, and NR_040877.1) obtained from the National Center for Biotechnology Information (NCBI) database (access date: Nov 21, 2019) for *P. copri* strains (taxid: 165179). The consensus sequence was identified after alignment using Vector NTI software (Thermo Fisher Scientific Inc., Waltham, MA, USA) (Supplementary Table). This assay was utilized for the detection of *P. copri*. Another TaqMan assay (Assay ID:

Ba04930791_S1, Thermo Fisher Scientific Inc., Waltham, MA, USA) was used for detecting universal bacterial 16S rRNA as an internal control.

Quantitative PCR was performed using the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific Inc., Waltham, MA, USA). The reaction mixture contained 9 μL sample DNA, 10 μL of TaqMan Gene Expression Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA), and 1 μL of TaqMan Gene Expression assay (Total volume, 20 μL /well in 96-well optical reader plates). DNA samples were diluted to 1 ng/ μL as the final concentration. Each sample underwent a duplicate run with the two expression assays. The amplification program consisted of one cycle at 95 $^{\circ}\text{C}$ for 10 min, followed by 30 cycles at 95 $^{\circ}\text{C}$ for 30 s, and 60 $^{\circ}\text{C}$ for 1 min. The cycle threshold (Ct) values for each sample were determined using QuantStudio 3 & 5 Real-time PCR Data Analysis Software (Thermo Fisher Scientific Inc., Waltham, MA, USA). The mean ΔCt value between the two assays indicated the relative abundance of *P. copri* in each sample. An increased ΔCt value ($\text{Ct}_{\text{universal bacterial control}} - \text{Ct}_{\text{Prevotella assay}}$) indicated a higher abundance.

To compare the results of quantitative PCR with 16S sequencing data, Pearson's correlation of each KTU in log-abundance versus the ΔCt of samples was calculated. The significantly correlated KTUs were filtered by p-value adjustment using the False Discovery Rate (FDR) method. To verify the scientific names of the KTUs correlated to the *P. copri* assay, the Basic Local Alignment Search Tool (BLAST) was used to search against NCBI non-redundant (nr) and rRNA databases using the following parameters: query coverage, 100%; percent identity > 99%; E value, 0 (access date: Oct 23, 2022). The optimal ΔCt value was obtained as the maximal sum of sensitivity and specificity for enterotype classification.

3. Results

3.1. The enterotype *Bacteroides* associated with better glycemic control

Our cohort of 113 patients with type 2 diabetes was middle-aged (58 ± 1 years) and overweight (BMI 29.1 ± 0.4), with a mean HbA1c of $8.1 \pm 0.2\%$. About half of the estimated calorie intake was from carbohydrates (51%), while 17% and 32% were from protein and fat, respectively. Metformin-based therapy was predominant among the patients.

The optimal number of clusters for enterotyping, determined by the highest Calinski-Harabasz Index, was two. The predominant clusters were enterotype *Bacteroides* (57.5%) and enterotype *Prevotella* (42.5%). The on-target HbA1c ($\leq 7.0\%$) rate was higher in the *Bacteroides* group than in the *Prevotella* group (35.4% versus 16.7%; $P, 0.034$). Patients with *Bacteroides* enterotype had 2.71 odds of HbA1c $\leq 7.0\%$ than did those with the *Prevotella* enterotype (95% confidence interval [CI], 1.02–7.87). Additionally, patients with *Bacteroides* enterotype showed a higher fish–vegetable pattern score compared to those with *Prevotella* enterotype (0.17 ± 0.13 versus -0.23 ± 0.09 ; $P, 0.010$). No significant difference was observed between the two groups with respect to all other clinical parameters (age, sex, BMI, ALT, lipid profiles, eGFR, UACR, macronutrients proportions, estimated daily calorie intake, and antidiabetic treatment) (Table 1).

3.2. The healthy fish–vegetable dietary pattern

Four distinct dietary patterns were extracted from the 38 dietary factors: 1) The eating-out pattern is featured with foods from a restaurant or takeaway; these food contents are characterized by high oil, high salt, or refined carbohydrates. 2) The high-

sugar foods pattern shows increased intake of sweetened beverages, desserts, and snacks that consist mainly of high carbohydrate and fats. 3) The fish–vegetable pattern includes vegetables and whole food sources of protein and fats, including fish, meat, nuts, and seeds. 4) The fermented foods pattern is characterized by processed foods, including soy-based or fermented products with high salt content (Fig. 1A).

The fish–vegetable pattern had the following macronutrient proportions: positive for protein ($r, 0.29$; $P, 0.002$) and fat ($r, 0.21$; $P, 0.026$) and negative for carbohydrate ($r, -0.25$; $P, 0.008$). A strong correlation with fiber intake was also noted ($r, 0.38$; $P < 0.001$). The eating-out pattern and high-sugar pattern were associated with unhealthy metabolic parameters. The HbA1c level was the only clinical parameter that correlated with the fish–vegetable pattern ($r, -0.22$; $P, 0.017$) (Fig. 1B).

3.3. Enterotype *Bacteroides* had mildly decreased alpha diversity

The indexes of alpha diversity differed slightly between the two enterotype groups. The *Prevotella* group had higher values than the *Bacteroides* group in observed KTU numbers (167 [129, 212] vs. 146 [115, 178]; $P, 0.022$) and the Shannon index (3.59 [3.37, 3.79] versus 3.35 [3.21, 3.65]; $P, 0.030$) (Fig. 2A–C). The heterogeneity between the groups in beta diversity analysis was significant by classification with enterotypes ($P, 0.001$) (Fig. 2D).

3.4. *Prevotella copri* inversely associated with the fish–vegetable pattern

The pattern scores of the fish–vegetable pattern correlated negatively with PCoA 2 ($r = -0.21$; $P, 0.024$). The fish–vegetable pattern was the only dietary pattern that showed a gut–microbiota association (Fig. 2E). Of the 128 KTUs that had significance to the scores on PCoA 2, 31 correlated positively and 97 correlated negatively. The accumulated log-abundance of the 31 positively correlated KTUs correlated with the fish–vegetable pattern score ($r, -0.246$; $P, 0.009$), while no other KTUs did (Fig. 2F). Moreover, the 31 key KTUs differed significantly in their accumulated log-abundances between participants with enterotype *Prevotella* and enterotype *Bacteroides* ($P < 0.001$).

Analysis of random forest models with the 31 microbial features revealed that the top 10 KTUs showed good discriminative power in classification of enterotypes (AUROC, 0.96; 95% CI, 0.92–0.99). The top 3 KTUs, all classified as *Prevotella* 9, had a

Table 1. Clinical features of type 2 diabetic subjects with *Bacteroides* and *Prevotella* enterotypes.

	All (n = 113)	<i>Prevotella</i> (n = 48)	<i>Bacteroides</i> (n = 65)	P
Sex (male/female)	68/45	33/15	35/30	0.124
Age (years)	58 ± 1	59 ± 2	57 ± 1	0.401
Duration of diabetes (years)	10 ± 1	11 ± 1	10 ± 1	0.772
BMI	29.1 ± 0.4	29.3 ± 0.6	28.9 ± 0.5	0.645
FPG (mg/dL)	150 ± 5	151 ± 8	148 ± 6	0.751
HbA1c (%)	8.1 ± 0.2	8.2 ± 0.2	8.0 ± 0.2	0.491
Ratio of HbA1c ≤ 7.0% (n/%)	31/27.4	8/16.7	23/35.4	0.034*
HOMA-IR	3.80 ± 0.32	4.56 ± 0.66	3.26 ± 0.26	0.428
eGFR (mL/min/1.73 m ²)	84 ± 3	80 ± 4	86 ± 5	0.354
UACR	491 ± 128	285 ± 124	643 ± 202	0.134
ALT (U/L)	32 ± 2	34 ± 3	30 ± 3	0.379
Total cholesterol (mg/dL)	166 ± 4	160 ± 5	171 ± 5	0.130
LDL-cholesterol (mg/dL)	94 ± 3	93 ± 4	94 ± 4	0.845
HDL-cholesterol (mg/dL)	42 ± 1	42 ± 1	42 ± 1	0.913
Triglycerides (mg/dL)	200 ± 15	175 ± 14	219 ± 24	0.119
Estimated calorie intake (Kcal/day)	1636 ± 56	1678 ± 93	1605 ± 70	0.535
Protein (%)	17 ± 0	16 ± 0	17 ± 0	0.073
Fat (%)	32 ± 1	31 ± 1	33 ± 1	0.265
Carbohydrate (%)	51 ± 1	53 ± 2	50 ± 1	0.161
Fiber (g/day)	16 ± 1	15 ± 1	16 ± 1	0.309
Fish–vegetable pattern	NA	−0.23 ± 0.09	0.17 ± 0.13	0.010*
Fermented foods pattern	NA	−0.01 ± 0.16	0.01 ± 0.12	0.924
Eating-out pattern	NA	−0.03 ± 0.13	0.03 ± 0.13	0.737
High-sugar foods pattern	NA	0 ± 0.15	0 ± 0.12	0.996
Antidiabetic agents (n/%)				0.812
Metformin-based oral therapy	43/38.1	18/37.5	25/38.5	
Metformin with SGLT-2i or GLP-1 RA	32/28.3	12/25.0	20/30.8	
Insulin-based therapy	35/31.0	17/35.4	18/27.7	
Other	3/2.7	1/2.1	2/3.1	

Continuous data are shown as the mean ± standard error; independent-samples t-test, Fisher's exact test, and Kruskal–Wallis test. The HOMA-IR index is calculated using data from 36 subjects in the *Prevotella* group and 50 subjects in the *Bacteroides* group. Four types of dietary pattern scores are dimensionless values extracted through factor analysis. * $P < 0.05$. ALT, alanine aminotransferase; BMI, body mass index; eGFR, estimated glomerular filtration rate; FPG, fasting plasma glucose; GLP-1 RA, glucagon-like peptide 1 receptor agonist; HbA1c, glycohemoglobin; HOMA-IR, Homeostatic Model Assessment for Insulin Resistance; NA, not applicable; SGLT-2i, sodium–glucose cotransporter 2 inhibitor; UACR, urinary albumin-creatinine ratio.

diagnostic power of 0.93 for AUROC [95% CI, 0.88–0.98]. The top KTU was *P. copri* DSM 18205 strain (accession number NR_040877.1) in the NCBI database (Fig. 3A–C).

3.5. Performance of the *P. copri* assay in detecting and enterotyping

Of the 54 KTUs belonging to the *Prevotella* genus according to 16S sequencing results, 4 were identified as *P. copri* DSM 18205 strain and 31 were *Prevotella* 9 spp. with closest phylogeny to the same strain (accession number as NR_040877.1 or NR_113411.1; identity score < 99%). These 35 KTUs accounted for 82.9% of the total *Prevotella* genus according to the reference database. The remaining 19 KTUs were of other *Prevotella* species.

The ΔC_t values obtained from the *P. copri* assay correlated well with the accumulated log-abundance of the 35 KTUs that were associated with *P. copri* DSM 18205 (r , 0.801; $P < 0.001$); the other 19

KTUs did not (r , 0.055; P , 0.565). Of these KTUs, 7 remained significant after adjustment and accounted for 91.9% of the 35 KTUs. Applying -2.71 as the cut-off ΔC_t value gave the optimal diagnostic power (sensitivity, 0.96; specificity, 0.97) for detecting the presence of *P. copri* (defined as relative abundance > 0.1%). The average prevalence was 47.8% and the average relative abundance was 10.2% for the total cohort (21.0% in subjects harboring *P. copri*) (Fig. 3D).

In summary, *P. copri* might emerge as a prominent microbial feature, frequently observed and associated with suboptimal glycemic control and an unhealthy diet.

4. Discussion

Our study demonstrated that patients with the *Prevotella* enterotype were more likely to have hyperglycemia, and the abundance of *P. copri* was inversely associated with a healthy fish-vegetable dietary pattern. To our knowledge, this is the first

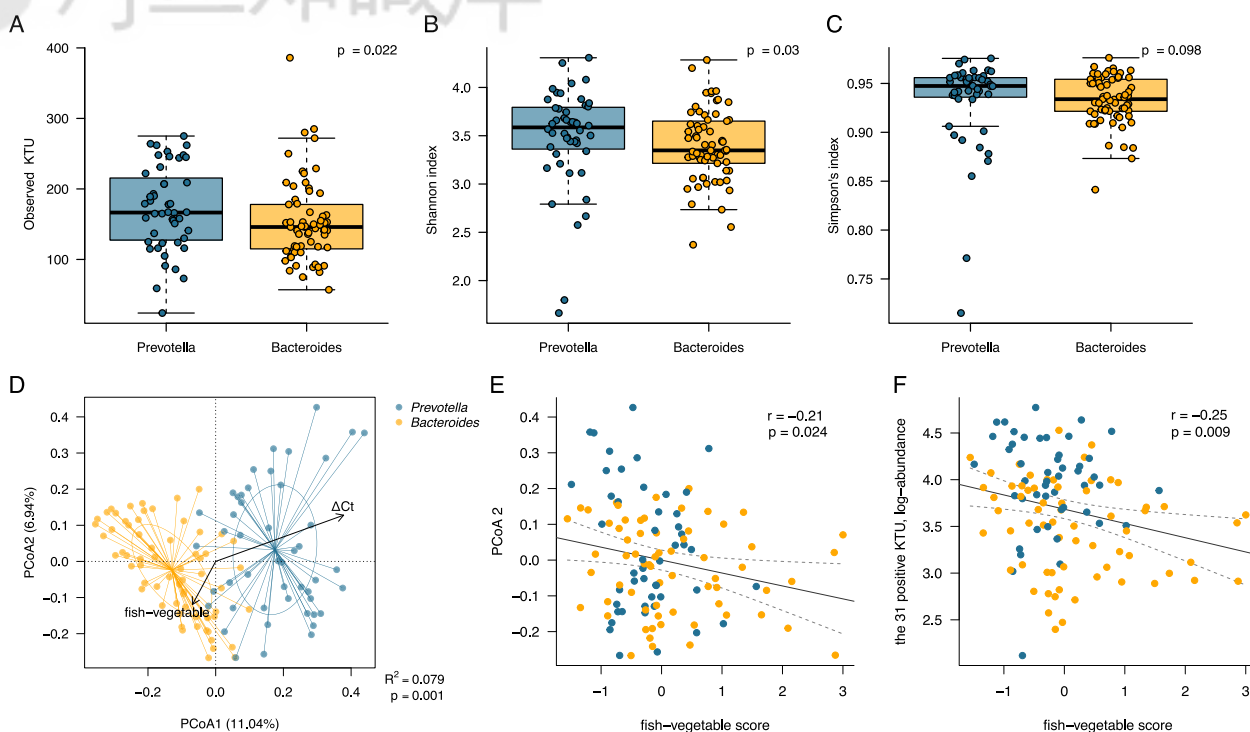


Fig. 2. A–C) The alpha diversity indexes of microbiota differed weakly between patients with diverse enterotypes. D) The dissimilarity between the two enterotype groups in beta diversity is shown. The fish–vegetable pattern score is shown as an arrow towards *Bacteroides*; ΔCt for the *Prevotella copri* assay is shown as an opposite arrow. E, F) The fish–vegetable pattern scores correlated with PCoA 2 and accumulated log-abundance of the 31 positive PCoA 2-associated KTUs. ΔCt , delta cycle threshold; KTU, k-mer taxonomic unit; PCoA, principal coordinate analysis.

study to investigate the relationship between enterotypes and dietary patterns in type 2 diabetes patients.

The type 2 diabetes patients in this cohort came from urbanized communities in northern Taiwan. A high disease severity was observed because the participants were enrolled from the diabetes reference centers. Under regular surveillance, their habitual diet showed moderate proportions in macronutrients but a low daily fiber intake. Either healthy or unhealthy dietary patterns were observed. One possible reason for inadequate diet control in real-world practice is patient or clinical inertia in adherence to a healthy diet.

The concept of enterotypes offers a structured approach to classify distinct patterns of gut microbial communities, offering insights into the dietary patterns and glycemic control for individuals with type 2 diabetes. However, due to factors such as choices for taxonomic levels, distance metrics, and clustering algorithms, the process of enterotyping can lead to varying numbers of clusters [20]. Nevertheless, a consistent observation emerges: either a three-cluster model revealing *Bacteroides*, *Prevotella*, and Firmicutes-dominated clusters, or a two-cluster model distinguishing *Prevotella*-driven

samples from the rest [9]. In our study, a two-cluster model distinguishing *Prevotella* from *Bacteroides* demonstrated the most optimal classification.

The on-target HbA1c rate was 27.4% on average and differed between the two enterotype groups (*Bacteroides*, 35.4%; *Prevotella*, 16.7%). Nonetheless, we observed no apparent difference in fiber intake between the two enterotype groups. We observed a weak tendency toward more protein and fat in the *Bacteroides* group and more carbohydrate in the *Prevotella* group. The *Bacteroides* enterotype was more prevalent than the *Prevotella* enterotype (57.5% vs. 42.5%, respectively). This lower abundance of *Prevotella* is consistent with data from other westernized cohorts (*Prevotella*, 29.6% on average), which contrasts the extremely high prevalence (95.4%) in non-westernized cohorts [21]. *Bacteroides* has been linked to a Western diet characterized as high consumption of animal proteins and fat, whereas *Prevotella* has been linking to a diet of staple foods and many vegetables in isolated hunter-gatherer cohorts [22]. Decreased richness in the gut microbiota was also found in the *Bacteroides* group. However, an important finding is that the fiber-rich dietary pattern, fish–vegetable, correlated with *Bacteroides* rather than *Prevotella* in our cohort. One possible

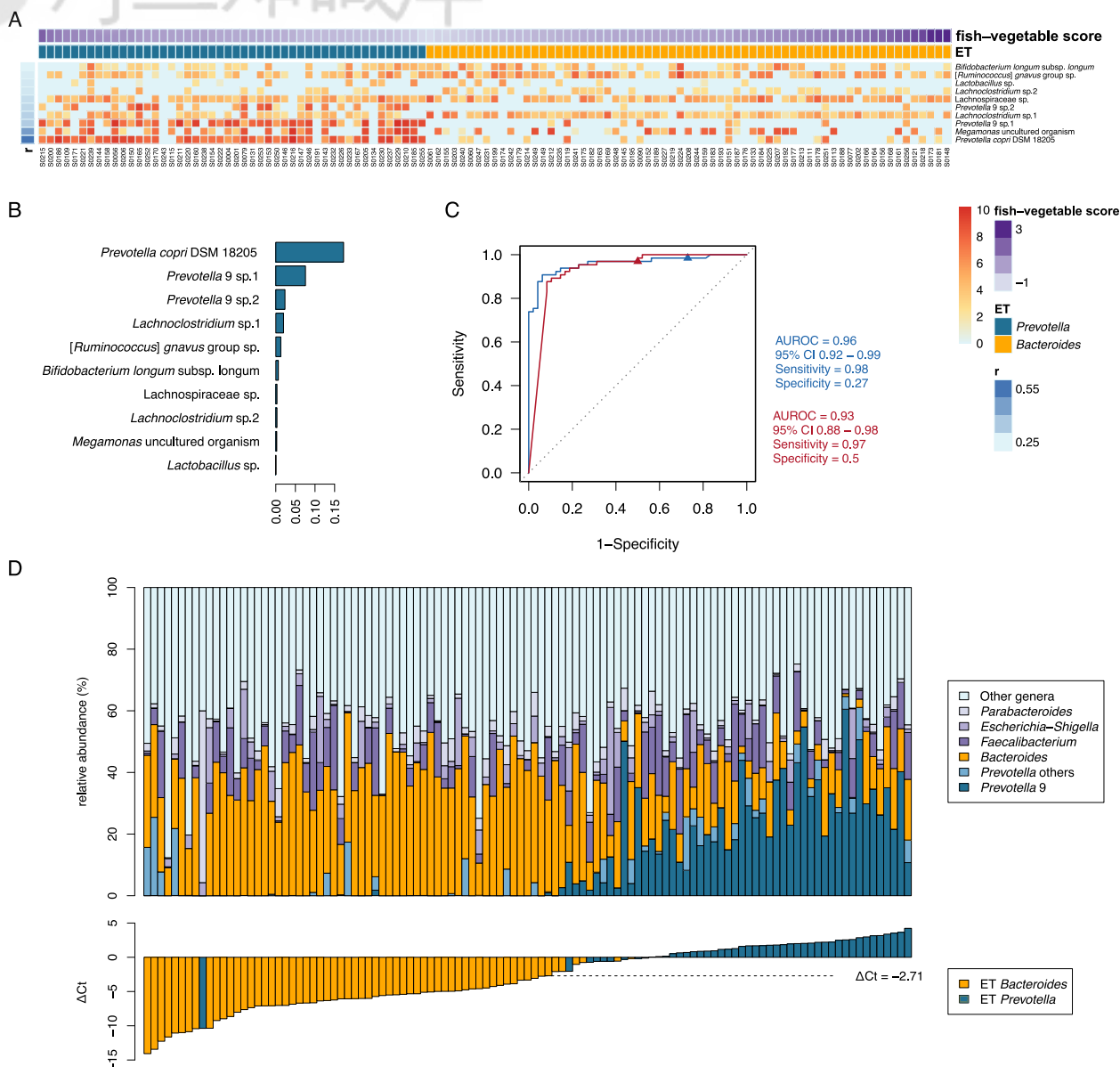


Fig. 3. A) The top 10 KTUs are plotted in the order of their correlation coefficient values. Samples are grouped by enterotypes in sequences of the fish-vegetable pattern scores. The color gradient in cells represents the log-transformed abundance of microbial features. B) Top 10 KTUs arranged by their mean decreasing accuracy scores in a random forest model. C) The ROC curves illustrate the diagnostic power of the top microbial features (blue, top 10; red, top 3). D) Comparison of 16S ribosomal RNA amplicon sequencing, enterotyping, and the *Prevotella copri* assay. All samples are arranged in accordance with ΔCt values on the x-axis. The top 5 genera are shown by relative abundance on the y-axis. ΔCt of -2.71 indicates the presence of *Prevotella copri* ($> 0.1\%$), whereas the balance between *Bacteroides* and *Prevotella* contributes to the determination of enterotypes. AUROC, area under the receiver operating characteristic; r , correlation coefficient value; ΔCt , delta cycle threshold; ET, enterotype; KTU, k-mer taxonomic unit; ROC, Receiver operating characteristic.

explanation is the differences in study cohorts. Most evidence of the association between *Prevotella* and high fiber intake was observed in rural populations. The microbiota among type 2 diabetes patients on a Western diet might differ [23]. Furthermore, *Bacteroides* have distinctive mechanisms to degrade complex polysaccharides, including dietary fiber, present in the intestinal tract [9]. While both

Bacteroides and *Prevotella* are capable of fermenting dietary fiber [24], it has been observed that augmenting the intake of fiber in a low-fiber diet leads to an increase in specific *Bacteroides* species [25]. Notably, in the context of a Westernized diet, the preference for a fish-vegetable dietary pattern was linked to the *Bacteroides* enterotype, rather than the *Prevotella* enterotype.

An unhealthy dietary pattern such as eating out or consuming high-sugar foods correlates significantly with hyperglycemia or dyslipidemia, but no gut–microbiota association was found (Fig. 2B). In contrast, the fish–vegetable diet includes whole food sources including a variety of colored vegetables and freshwater or marine fish. These foods are enriched in good protein and fat, as well as fiber. The fish–vegetable is associated with healthier metabolic profiles in patients with type 2 diabetes [26]. Remarkably, this diet also is also linked to a lower abundance of *P. copri*, a major contributor to the *Prevotella* enterotype. This finding further confirms that macronutrients and fiber are instrumental in the shaping of the gut microbiota. A healthy dietary pattern, particularly in relation to enterotypes, plays a crucial role in understanding and addressing the association between gut microbiota and human diseases [27].

P. copri was found to be the most representative microbial species that was inversely associated with the healthy fish–vegetable pattern. One possible reason was the high relative abundance of these bacteria (10.2% in our cohort). The question of their clinical relevance has attracted great interest [21]. *P. copri* might be a microbial feature associated with non-response to glucagon-like peptide 1 (GLP-1) receptor agonists and may be associated with increased inflammatory cytokines in populations with type 2 diabetes [6,16]. Although the causal relationship remains uncertain, it might be explained by the finding that *P. copri* induces insulin resistance and increases circulating levels of branched-chain amino acids [12]. In addition, glucose metabolism was modulated by *P. copri* in the presence of dietary fiber supplements; this pathway might not be adopted in individuals with a lower intake of vegetables [13]. A recent study reports that *P. copri* is a complex comprising four genetically distinct clades [28]. An omnivorous diet may result in the selection of *P. copri* with a higher prevalence of the *leuB* gene, which is involved in branched-chain amino acid biosynthesis and a risk factor for glucose intolerance [29]. In a westernized cohort of type 2 diabetes patients, the *P. copri* complex might favor the predominance of sub-strains that predispose to hyperglycemia. A randomized control trial showed that a high-fiber and vegetable-protein–based diet induced a decrease in *P. copri* and improved glycemic control, dyslipidemia, and inflammation [23]. Based on this evidence, we inferred that the predominance of *P. copri* in a westernized cohort of type 2 diabetes patients may indicate a lack of dietary fiber and a status predisposing to hyperglycemia.

P. copri is typically abundant, although its presence is not universal. Variations in distribution, whether due to geographic or dietary factors, enable the simplified detection of this bacterium [21]. Quantitative PCR has become a cost-effective tool for such detection. Recently studies have validated the use of PCR for detecting *P. copri* and highlighted the importance of these [30]. As an alternative, the assay described in the present study detected the presence of the bacteria as well as the enterotype *Prevotella*.

The main limitation of this study is the small number of participants. Parameters could not be perfectly randomized, and the significance of our findings may be affected. The causal relationship between an unhealthy diet and *P. copri* could not be proven in a cross-section design. This study merely demonstrates survey results in this specific local population. The ethnicity, region of residence, or disease severity in diabetes are crucial confounding factors and should be adjusted in our further nationwide study. Future metagenomic studies should be considered to identify the sub-strains and biodiversity of the *P. copri* complex, as well as the functional metabolites including short-chain fatty acids, in order to understand the mechanism underlying the interaction between *P. copri* and human energy homeostasis. In conclusion, our results suggest that *P. copri* might serve as a representative microbial feature among type 2 diabetes patients on an unhealthy Westernized diet with insufficient vegetables and fish-based protein and fat.

Data availability

The datasets supporting the conclusions of this article are available in the NCBI repository, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA936824>.

Funding

This work was supported by the Chang Gung Memorial Hospital, Taoyuan City, Taiwan [grant numbers CMRPG3K1171 and CMRPG3K1172].

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Chih-Yi Li, Hsin-Yi Liu, Pei-Yu Lai, Wan-Ting Hsiao, and Yi-Ching Wang, the

dietitians in the Department of Medical Nutrition Therapy, Chang Gung Memorial Hospital, Linkou, Taipei, and Taoyuan branches for their help and expertise in dietitian interviews. They received no financial support for their participation. The authors thank all members of the Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou for their help in library preparation and sequencing.

References

- [1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract* 2022;183:109119.
- [2] Oggioni C, Lara J, Wells JC, Soroka K, Siervo M. Shifts in population dietary patterns and physical inactivity as determinants of global trends in the prevalence of diabetes: an ecological analysis. *Nutr Metabol Cardiovasc Dis* 2014;24:1105–11.
- [3] Delgado-Lista J, Alcalá-Díaz JF, Torres-Pena JD, Quintana-Navarro GM, Fuentes F, García-Ríos A, et al. CORDIOPREV Investigators. Long-term secondary prevention of cardiovascular disease with a Mediterranean diet and a low-fat diet (CORDIOPREV): a randomised controlled trial. *Lancet* 2022;399:1876–85.
- [4] Evert AB, Dennison M, Gardner CD, Garvey WT, Lau KHK, MacLeod J, et al. Nutrition therapy for adults with diabetes or prediabetes: a consensus report. *Diabetes Care* 2019;42:731–54.
- [5] Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. *Cell* 2015;163:1079–94.
- [6] Iljazovic A, Roy U, Galvez EJC, Lesker TR, Zhao B, Gronow A, et al. Perturbation of the gut microbiome by *Prevotella* spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol* 2021;14:113–24.
- [7] de Moraes AC, Fernandes GR, da Silva IT, Almeida-Pititto B, Gomes EP, Pereira AD, et al. Enterotype may drive the dietary-associated cardiometabolic risk factors. *Front Cell Infect Microbiol* 2017;7:47.
- [8] Gorvitovskaia A, Holmes SP, Huse SM. Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle. *Microbiome* 2016;4:15.
- [9] Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* 2018;3:8–16.
- [10] He Y, Wu W, Wu S, Zheng HM, Li P, Sheng HF, et al. Linking gut microbiota, metabolic syndrome and economic status based on a population-level analysis. *Microbiome* 2018;6:172.
- [11] Wang J, Li W, Wang C, Wang L, He T, Hu H, et al. Enterotype *Bacteroides* is associated with a high risk in patients with diabetes: a pilot study. *J Diabetes Res* 2020;2020:1–11.
- [12] Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyötyläinen T, Nielsen S, Jensen BA, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 2016;535:376–81.
- [13] Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee YS, De Vadder F, Arora T, et al. Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metabol* 2015;22:971–82.
- [14] Huang MC, Lin KD, Chen HJ, Wu YJ, Chang CI, Shin SJ, et al. Validity of a short food frequency questionnaire assessing macronutrient and fiber intakes in patients of Han Chinese descent with type 2 diabetes. *Int J Environ Res Publ Health* 2018;15:1142.
- [15] Chung HF, Hsu CC, Mamun AA, Long KZ, Huang YF, Shin SJ, et al. Dietary patterns, dietary biomarkers, and kidney disease in patients with type 2 diabetes: a repeated-measure study in Taiwan. *Asia Pac J Clin Nutr* 2018;27:366–74.
- [16] Tsai CY, Lu HC, Chou YH, Liu PY, Chen HY, Huang MC, et al. Gut microbial signatures for glycemic responses of glp-1 receptor agonists in type 2 diabetic patients: a pilot study. *Front Endocrinol* 2022:12.
- [17] Liu PY, Yang SH, Yang SY. KTU: K-mer taxonomic units improve the biological relevance of amplicon sequence variant microbiota data. *Methods Ecol Evol* 2021:1–9.
- [18] Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGinn D, et al. *vegan*: community Ecology Package. R package version 2.5-7. 2020.
- [19] Liu PY. MARCO: microbiome analysis RcodeDB. 2021. <https://doi.org/10.5281/zenodo.4589897>. [Accessed 9 March 2021].
- [20] Liang C, Tseng HC, Chen HM, Wang WC, Chiu CM, Chang JY, et al. Diversity and enterotype in gut bacterial community of adults in Taiwan. *BMC Genom* 2017;18:932.
- [21] Tett A, Pasolli E, Masetti G, Ercolini D, Segata N. *Prevotella* diversity, niches and interactions with the human host. *Nat Rev Microbiol* 2021;19:585–99.
- [22] Hansen MEB, Rubel MA, Bailey AG, Ranciaro A, Thompson SR, Campbell MC, et al. Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana. *Genome Biol* 2019;20:16.
- [23] Medina-Vera I, Sanchez-Tapia M, Noriega-Lopez L, Grandos-Portillo O, Guevara-Cruz M, Flores-Lopez A, et al. A dietary intervention with functional foods reduces metabolic endotoxaemia and attenuates biochemical abnormalities by modifying faecal microbiota in people with type 2 diabetes. *Diabetes Metab* 2019;45:122–31.
- [24] Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. Fiber-utilizing capacity varies in *Prevotella*- versus *Bacteroides*-dominated gut microbiota. *Sci Rep* 2017;7:2594.
- [25] Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell* 2019;179:59–73 e13.
- [26] Huang MC, Chang CI, Chang WT, Liao YL, Chung HF, Hsu CC, et al. Blood biomarkers of various dietary patterns correlated with metabolic indicators in Taiwanese type 2 diabetes. *Food Nutr Res* 2019;63:3592.
- [27] Cai JA, Zhang YZ, Yu ED, Ding WQ, Jiang QW, Cai QC, et al. Gut microbiota enterotypes mediate the effects of dietary patterns on colorectal neoplasm risk in a Chinese population. *Nutrients* 2023;15.
- [28] Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, et al. The *Prevotella* copri complex comprises four distinct clades underrepresented in Westernized populations. *Cell Host Microbe* 2019;26:666–679 e7.
- [29] De Filippis F, Pasolli E, Tett A, Tarallo S, Naccarati A, De Angelis M, et al. Distinct genetic and functional traits of human intestinal *Prevotella* copri strains are associated with different habitual diets. *Cell Host Microbe* 2019;25:444–453 e3.
- [30] Verbrugge P, Van Aken O, Hallenius F, Nilsson A. Development of a real-time quantitative PCR method for detection and quantification of *Prevotella* copri. *BMC Microbiol* 2021;21:23.