

# The fecal microbiotas of women of Pacific and New Zealand European ethnicities are characterized by distinctive enterotypes that reflect dietary intakes and fecal water content

Nikki Renall<sup>a,b,c</sup>, Blair Lawley<sup>d</sup>, Tommi Vatanen<sup>e,f,g</sup>, Benedikt Merz<sup>h</sup>, Jeroen Douwes<sup>c</sup>, Marine Corbin<sup>c</sup>, Lisa Te Morenga<sup>b,c</sup>, Rozanne Kruger<sup>a</sup>, Bernhard H Breier<sup>a,b</sup>, and Gerald W Tannock<sup>d</sup>

<sup>a</sup>School of Sport, Exercise and Nutrition, College of Health, Massey University, Auckland, New Zealand; <sup>b</sup>Riddet Institute, Centre of Research Excellence, Massey University, Palmerston North, New Zealand; <sup>c</sup>Research Centre for Hauora and Health, Massey University, Wellington, New Zealand; <sup>d</sup>Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand; <sup>e</sup>Liggins Institute, University of Auckland, Auckland, New Zealand; <sup>f</sup>Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland; <sup>g</sup>The Broad Institute of MIT and Harvard, Cambridge, MA, USA; <sup>h</sup>Department of Physiology and Biochemistry of Nutrition, Max Rubner-Institut Karlsruhe, Germany

## ABSTRACT

Obesity is a complex, multifactorial condition that is an important risk factor for noncommunicable diseases including cardiovascular disease and type 2 diabetes. While prevention and management require a healthy and energy balanced diet and adequate physical activity, the taxonomic composition and functional attributes of the colonic microbiota may have a supplementary role in the development of obesity. The taxonomic composition and metabolic capacity of the fecal microbiota of 286 women, resident in Auckland New Zealand, was determined by metagenomic analysis. Associations with BMI (obese, nonobese), body fat composition, and ethnicity (Pacific, n = 125; NZ European women [NZE], n = 161) were assessed using regression analyses. The fecal microbiotas were characterized by the presence of three distinctive enterotypes, with enterotype 1 represented in both Pacific and NZE women (39 and 61%, respectively), enterotype 2 mainly in Pacific women (84 and 16%) and enterotype 3 mainly in NZE women (13 and 87%). Enterotype 1 was characterized mainly by the relative abundances of butyrate producing species, *Eubacterium rectale* and *Faecalibacterium prausnitzii*, enterotype 2 by the relative abundances of lactic acid producing species, *Bifidobacterium adolescentis*, *Bifidobacterium bifidum*, and *Lactobacillus ruminis*, and enterotype 3 by the relative abundances of *Subdoligranulum sp.*, *Akkermansia muciniphila*, *Ruminococcus bromii*, and *Methanobrevibacter smithii*. Enterotypes were also associated with BMI, visceral fat %, and blood cholesterol. Habitual food group intake was estimated using a 5 day nonconsecutive estimated food record and a 30 day, 220 item semi-quantitative Food Frequency Questionnaire. Higher intake of 'egg' and 'dairy' products was associated with enterotype 3, whereas 'non-starchy vegetables', 'nuts and seeds' and 'plant-based fats' were positively associated with enterotype 1. In contrast, these same food groups were inversely associated with enterotype 2. Fecal water content, as a proxy for stool consistency/colonic transit time, was associated with microbiota taxonomic composition and gene pools reflective of particular bacterial biochemical pathways. The fecal microbiotas of women of Pacific and New Zealand European ethnicities are characterized by distinctive enterotypes, most likely due to differential dietary intake and fecal consistency/colonic transit time. These parameters need to be considered in future analyses of human fecal microbiotas.

## ARTICLE HISTORY

Received 19 June 2022  
Revised 2 February 2023  
Accepted 7 February 2023



## KEYWORDS

Fecal microbiota;  
enterotypes; dietary intake;  
fecal water content; obesity


## Background

Obesity (adult body mass index, BMI  $\geq 30$  kg/m<sup>2</sup>) is a health problem in many countries as it is a risk factor for several non-communicable diseases such as cardiovascular disease and diabetes.<sup>1</sup> The worldwide prevalence of obesity has tripled since 1975.<sup>1</sup> Of the OECD countries, adult obesity rates are

highest in the USA, Mexico, New Zealand, and Hungary.<sup>2</sup> Moreover, the prevalence of adult obesity may differ between ethnic groups living in the same country. For example, in New Zealand, 71.3% of Pacific Island, 50.8% of Māori, 31.9% of European/other, and 18.5% of Asian adults are obese.<sup>3</sup> Obesity is 1.6 times more likely in people

**CONTACT** Gerald W Tannock  [gerald.tannock@otago.ac.nz](mailto:gerald.tannock@otago.ac.nz)  Department of Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2023.2178801>

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in socioeconomically deprived locations in New Zealand.<sup>3</sup> Clearly, primary disease prevention approaches (optimal nutrition, education, exercise) are required to solve the obesity problem, but ancillary knowledge of biological factors associated with the condition could be important in developing supportive prophylactic measures. One such factor is the gut microbiota that, at least in gnotobiotic animal models, has roles in energy harvest from the digesta, regulation of triglyceride uptake by fat cells and total body fat accumulation.<sup>4</sup> Obesity in genetically predisposed mice is associated with more bacterial species belonging to the phylum Firmicutes relative to phylum Bacteroidetes, with the opposite observed in lean mice.<sup>5–7</sup> Partial confirmation of this finding was obtained in a small human study in which the ratio of Firmicutes to Bacteroidetes (F:B) was shown to decrease as obese individuals lost weight while consuming a fat- or carbohydrate-restricted diet.<sup>8</sup> However, results from murine models have not always been confirmed in human studies, perhaps because the association of the composition of the gut microbiota with obesity is weak, confounded by inter-individual variation and insufficient sample sizes.<sup>9–11</sup> As concluded by Sze and Schloss,<sup>11</sup> the involvement of the microbiota in obesity may not be revealed by taxonomic investigations alone but may only be apparent if the metabolic capacity of different gene pools (structures) within the microbial community were investigated. The purpose of our study was to compare the taxonomic composition and metabolic capacity of the fecal microbiota of 286 women who varied principally in BMI (obese, non-obese), body fat composition, ethnicity (Pacific, NZ European), and socioeconomic deprivation level. Comprehensive details of habitual intake of food groups were collected and collated in relation to microbiota data.

## Materials and methods

### Study participants

Two hundred and eighty-six participants were recruited to the cross-sectional study 'Predictors linking Obesity and gut Microbiome' (PROMISE), which was conducted between July 2016 and

September 2017. Participants were recruited based on self-reported body mass index (BMI) so that half in each group either had normal BMI or were obese (BMI  $\geq 30$  kg/m<sup>2</sup>). Participants were stratified into Body Fat (BF) groups (low-BF% <35%; high-BF%  $\geq 35\%$ ) for all subsequent analysis because individuals with the same BMI can have different body compositions and metabolic disease risks.<sup>12–15</sup> Participants were of Pacific Island origins (n = 125) and NZ European women (n = 161), free from any chronic disease, aged 18–45 years, all resident in the Auckland region. Details of the study procedures and recruitment strategies have been published elsewhere.<sup>16</sup> Participants made two visits to the research unit and completed questionnaires at home between visits. The study was approved by the Southern Health Disability Ethics Committee (16/STH/32) and conducted according to the guidelines of the declaration of Helsinki. The trial was registered at anzctr.org.au (ACTRN12618000432213). All participants were informed in detail about the procedures and measurements and gave written informed consent.

### Anthropometric and demographic information

Fasting weight and height were used to calculate BMI as kg/m<sup>2</sup>. Body composition was assessed with a whole-body scan using Dual-energy X-ray Absorptiometry (DXA) (Hologic QDR Discovery A, Hologic Inc, Bedford, MA with APEX V. 3.2 software). Total body fat percentage (BF%), and visceral fat percentage were assessed with DXA. The New Zealand Deprivation 2013 index (NZDep2013), an area-based measure of socioeconomic deprivation, was used to assign a socioeconomic deprivation score ranging between decile 1 "least deprived" to decile 10 "most deprived" to each participant.<sup>17</sup> Blood chemistry assays (glucose, HbA1c, cholesterols, triglycerides, insulin) were conducted using standard diagnostic methods.

### Dietary assessment

Participants completed a 5-day nonconsecutive estimated food record (5DFR) at home. At study visit two, the 5DFR was reviewed in a face-to-face interview with a dietitian and participants completed a validated 220-item semi-quantitative Food Frequency Questionnaire (NZWFFQ)

regarding food intake during the previous 30 days.<sup>18</sup> Energy, macro- and micro-nutrient analyses of the 5DFR and NZWFFQ were completed using FoodWorks9 (Xyris Software [Australia] Pty Ltd, Queensland, Australia) nutrition analysis software, based on New Zealand's food composition database, FOODFiles 2016 (Plant & Food Research, NZ). All reported energy intakes were reviewed by dietitians and values between 2100 kJ/day and 27000 kJ/day were considered plausible for valid completion of the 5DFR and NZWFFQ; all others ( $n = 17$ ; Pacific  $n = 16$ , NZ European  $n = 1$ ) were excluded from further analyses. Foods from the 5DFR ( $n > 2850$ ) and the NZWFFQ were allocated to 55 food groups, based on similar food groupings used in previous studies.<sup>19</sup> Total energy (reported as kilojoules) includes the energy contribution from all the macronutrients as well as total dietary fiber.

The National Cancer Institute (NCI; USA)<sup>20,21</sup> method was used to calculate individual usual intakes of the 55 food groups (g/day) consumed for one month for each participant. The NCI method uses a two-part modeling approach to estimate the probability of consumption and the respective amount consumed, considering the effect of covariates which can influence the probability of consumption (e.g., seasonality) or the amount consumed (e.g., age). The individual habitual intake is then defined as the product of probability of consumption multiplied by the consumed amount (Usual intake = Probability x Amount). The 5DFR was used as the primary dietary data, and the covariates age, ethnicity, BMI, season (summer, autumn, winter, spring), weekend (weekday = Monday – Thursday, weekend = Friday – Sunday), and FFQ information (in standard units/day) were included. The 55 food groups were collapsed into 21 food groups, based on similar nutritional composition and characteristics, for all subsequent analyses.

### **Fecal microbiota analysis**

Fecal samples were collected and stored in the participants' home freezers 11 to 14 days prior to delivery to the research unit. Subsequent storage was at  $-80^{\circ}\text{C}$  until laboratory analysis. Fecal water content, a proxy for colonic transit time, was

determined by placing approximately 200 mg of feces in a pre-weighed microfuge tube; the weight was recorded, and the tube with cap open placed in a  $37^{\circ}\text{C}$  incubator. The tubes were subsequently dried until a constant dry weight was obtained, and percentage water content was then calculated.

DNA was extracted from 250 mg feces according to the protocol provided by the manufacturer (PowerSoil DNA isolation kit, Mo Bio, Carlsbad, CA, USA), with the following modification. Fecal samples were suspended in 1 mL of TN150 buffer (containing 10 mM TRIS-CL pH 8.0, 150 mM NaCl). The suspension was centrifuged at  $14,600 \times g$  (3 min,  $5^{\circ}\text{C}$ ) and the deposit was then resuspended in 700  $\mu\text{l}$  of PowerBead solution from the kit. The suspension was added back to the PowerBead Tubes and the standard protocol followed. DNA was eluted in 100  $\mu\text{l}$  of elution buffer (warmed to  $70^{\circ}\text{C}$ ) and then stored at  $-80^{\circ}\text{C}$ . Quality and quantity of genomic DNA was checked on a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and on a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) prior to sending the cleaned DNA to New Zealand Genomics Ltd. (NZGL) for shotgun metagenome sequencing using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). Libraries were sequenced across a minimum of six HiSeq lanes, and multiple libraries were prepared for several samples to test for library preparation and sequence run bias. An average of 13,150,561 (range 7,6940,894–17,081,755) reads were recovered for each sample. BBDuk (<https://sourceforge.net/projects/bbmap/>) was used to trim adapters, remove low quality reads and remove reads  $<100$  bp after trimming. KneadData (<http://huttenhower.sph.harvard.edu/kneaddata>) was used as quality control to remove human genome reads from bacterial reads, implementing the hg19 database. Sequence data from this study is deposited as NCBI Bioproject PRJNA657309.

Microbiota taxonomic profiles were created from DNA sequences using MetaPhlan 2.0 (version 2.6.0) according to default parameters.<sup>22</sup> Microbiota composition and diversity was further analyzed with QIIME2 (version 2018.8) using converted output tables from MetaPhlan 2.0. Beta diversity group significance for each

metric (Bray-Curtis Dissimilarity index, and Jaccard similarity matrix) was measured with PERMANOVA<sup>23</sup> and group dispersion was measured with PERMDISP.<sup>24</sup>

Enterotypes were predicted in R using the approach described in Arumugum et al.<sup>25,26</sup> and following the tutorial provided by EMBL (<http://enterotyping.embl.de>). Differential abundance testing to determine which species were driving enterotypes was carried out with Statistical Analysis of Taxonomic and Functional Profiles (STAMP).<sup>27</sup> Each enterotype was compared to all other samples using Welch's *t*-test using Benjamini-Hochberg for multiple testing correction. The association between the species that characterized the enterotypes and fecal water content was explored with Spearman's correlation.

### **Assessment of the metabolic capacity of the microbiota**

Metagenomic reads were assembled into contigs using MegaHIT,<sup>28</sup> individually for each sample, and open reading frames were predicted using Prodigal.<sup>29</sup> A non-redundant gene catalog was constructed by clustering genes based on sequence similarity at 95% identity and 90% coverage of the shorter sequence using CD-HIT.<sup>30</sup> Metagenomic gene abundances were estimated by mapping quality trimmed reads from each sample against the gene catalog with *k*-mer alignment in KMA.<sup>31</sup> Assembled genes were functionally annotated with eggNOG-mapper v2 based on orthology assignments using precomputed eggNOG v5.0 clusters.<sup>32</sup> Annotations by Gene Ontology (GO) terms, Enzyme Commission (EC) categories, Carbohydrate-Active Enzyme (CAZy) Database identifiers and KEGG Ontology (KO) terms. Data were tested for associations with CAZymes and microbial bile salt hydrolases (BSH) because of previously published associations between these metabolic features of the fecal microbiota and BMI in experimental or clinical situations.<sup>5,33,34</sup> Associations between metabolic functions and cohort metadata were tested using linear regression, where the abundance of a metabolic function (measured in copies-per-million) was used as the dependent variable and all cohort metadata were

used as independent variables simultaneously to control for any confounding.

Metabolic capacity of the microbiota was assessed with HUMAnN2 which quantifies MetaCyc metabolic pathways. Pathway abundances were normalized to copies-per-million units. Associations with subject metadata and MetaCyc pathway abundances were analyzed using linear regression with pathway abundance as a response variable and subject ethnicity, BMI, Shannon index, enterotype, habitual dietary fiber intake (g/day), fecal water content, and F:B ratio as fixed effects.

### **Statistical analysis**

Analysis of dietary data and associations with microbiota was conducted using SAS Enterprise Guide version 7.1 (SAS Institute, Cary, NC, USA). Normality of data was assessed using Kolmogorov-Smirnov tests and histograms, medians [25th, 75th] were used to present all non-normally distributed continuous data. Mann Whitney tests were used to test for differences between groups. Multiple logistic regression was used to assess associations between habitual food group intake and enterotypes. Univariate analyses were conducted followed by multivariate analyses adjusted for ethnicity, age, deprivation, and energy intake. Analyses were conducted separately for NZE and Pacific participants. Where effect estimates were similar for both groups, analyses for both groups combined were also conducted with analyses adjusted for ethnicity. Further adjustment for BF% group was conducted to assess the independent association of food group intake and enterotypes. The odds of the microbiota reflecting a particular enterotype was expressed per 1 serving size of change in food group intake. Collinearity between variables was assessed by computing the tolerance and the variance inflation factor (VIF); no collinearity was detected. *p*-values <0.05 were considered statistically significant.

## **Results**

### **Anthropometric and demographic information**

Complete data sets were obtained for 286 participants: 125 Pacific (44%) and 161 NZE (56%) women. Pacific women were younger (median age

23 y) than NZE women (median age 32 y), had higher BMI and visceral fat values, but lower total and high-density lipoprotein (HDL) cholesterol in comparison to NZE women. NZE women had lower HbA1c, lower fasting plasma concentrations of insulin, and lower Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) scores compared to Pacific women. NZE women resided in less deprived areas than Pacific women. There were no differences in BF% between Pacific and NZE women (Table 1).

### Comparison of microbiota compositions

The median number of predicted species per participant was 73 (Pacific median 75, interquartile range 67–81; NZE 72, 67–77). Stratifying the population by BF% groups did not reveal statistically significant differences in abundances of individual taxa following adjustment for multiple testing (Figure S1). Beta diversity analysis indicated that Pacific and NZE microbiotas were different (Figure 1) and this was confirmed by reference to the relative abundances of taxa present in the feces of the two groups (Figure 2). Searches for robust clusters of bacterial species in the microbiotas detected three enterotypes (Figure 3). Enterotype 1 was characterized mainly by the relative abundances of butyrate-producing species, *Eubacterium*

*rectale* and *Faecalibacterium prausnitzii*. Enterotype 2 was characterized by the relative abundances of lactic acid-producing species, *Bifidobacterium adolescentis*, *Bifidobacterium bifidum*, and *Lactobacillus ruminis*. Enterotype 3 was characterized by the relative abundances of *Subdoligranulum* sp., *Akkermansia muciniphila*, *Ruminococcus bromii*, and *Methanobrevibacter smithii* (Figure 4). Alpha diversity of the fecal microbiota was less in Enterotype 2 individuals relative to those characterized by Enterotypes 1 or 3 (Table 2). Enterotype 1 was found in 146 participants, including both Pacific and NZE women. Enterotype 2 (n = 70) was predominately found in Pacific women, and enterotype 3 (n = 70) predominately in NZE women (Table 2). Women with a microbiota characterized by enterotype 2 were younger and had a higher BMI and visceral fat %, higher fasting insulin, higher HbA1c concentrations and HOMA-IR scores in comparison to women with enterotypes 1 and 3. Women with a microbiota characterized by enterotype 3 were older, had a lower deprivation index, lower HbA1c, lower high density lipoprotein cholesterol ratio (TC:HDL), higher HDL cholesterol, and lower fecal water content compared to those of enterotypes 1 and 2 (Table 2).

**Table 1.** Demographic, metabolic, and body composition data of participants stratified by ethnicity.

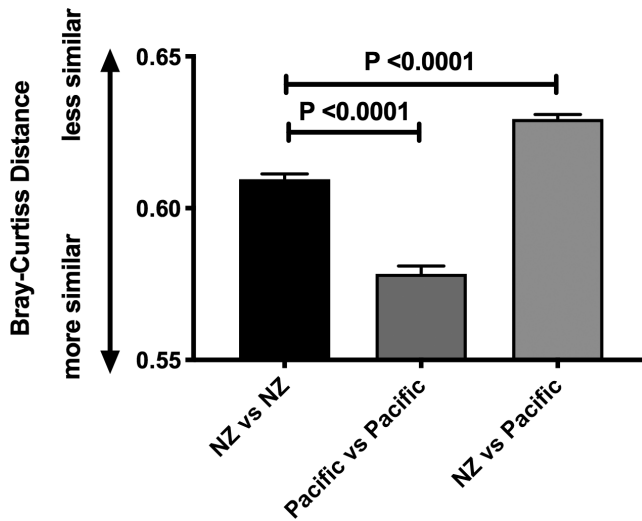
	Total population n = 286	Pacific n = 125	NZ European n = 161
Age (years)	28 [22, 35]	23 [20, 29]	32 [25, 37]*
Weight (kg)	77.6 [65.6, 96.0]	82.3 [72.3, 98.8]	70.7 [61.4, 92.5]*
Height (cm)	167.8 [163.7, 172.2]	168.6 [163.8, 174.1]	167.2 [163.7, 171.2]
BMI (kg/m <sup>2</sup> )	28.1 [23.0, 33.4]	29.5 [24.7, 34.8]	25.0 [22.2, 32.9]*
Total body fat (%)	34.5 [28.8, 39.6]	34.6 [29.5, 39.3]	34.3 [27.8, 39.9]
Visceral fat (%)	32.3 [23.7, 38.9]	34.3 [26, 39.8]	30.0 [21.4, 38.3]*
Body fat groups n (%)			
<35%	152 (53%)	65 (52%)	87 (54%)
>35%	134 (47%)	60 (48%)	74 (46%)
Total cholesterol (mmol/L) ∞	4.8 [4.3, 5.4]	4.5 [4.2, 5.1]	5.0 [4.5, 5.8]*
HDL cholesterol (mmol/L) ∞	1.5 [1.3, 1.9]	1.4 [1.3, 1.7]	1.7 [1.4, 2.0]*
LDL cholesterol (mmol/L) ∞	3.0 [2.4, 3.5]	2.9 [2.4, 3.3]	3.1 [2.4, 3.6]*
Triglycerides (mmol/L) ∞	0.9 [0.7, 1.2]	0.9 [0.7, 1.2]	0.9 [0.7, 1.2]
Total cholesterol:HDL ratio (mmol/L) ∞	3.1 [2.6, 3.7]	3.1 [2.6, 3.7]	3.0 [2.5, 3.7]
Glucose (mmol/L) ∞	5.3 [5.0, 5.6]	5.3 [5.1, 5.7]	5.3 [5.0, 5.6]
Insulin (U/ml) ∞	10.9 [7.3, 17.1]	14.6 [9.8, 23.2]	8.7 [6.4, 12.6]*
Homeostatic Model Assessment of Insulin Resistance	2.6 [1.7, 4.0]	3.4 [2.3, 5.9]	2.1 [1.5, 3.1]*
HbA1c (mmol/mol) #	31.7 [30.0, 33.6]	32.9 [31.4, 35.4]	30.7 [29.3, 32.6]*
Deprivation index #	6 [3, 8]	8 [6, 9]	4 [2, 6]*

All values are reported as medians [25th, 75th percentiles] or N (%).

\*Statistically significant difference between Pacific and NZ European women (Mann Whitney,  $p < 0.05$ ).

∞Pacific woman (n = 1) and #Pacific women (n = 2) have not been included in analyses due to missing data.

BMI: body mass index, HDL cholesterol: high density lipoprotein cholesterol, LDL cholesterol: low density lipoprotein cholesterol, HbA1c: glycosylated hemoglobin, Deprivation index: The 2013 New Zealand Deprivation Index.



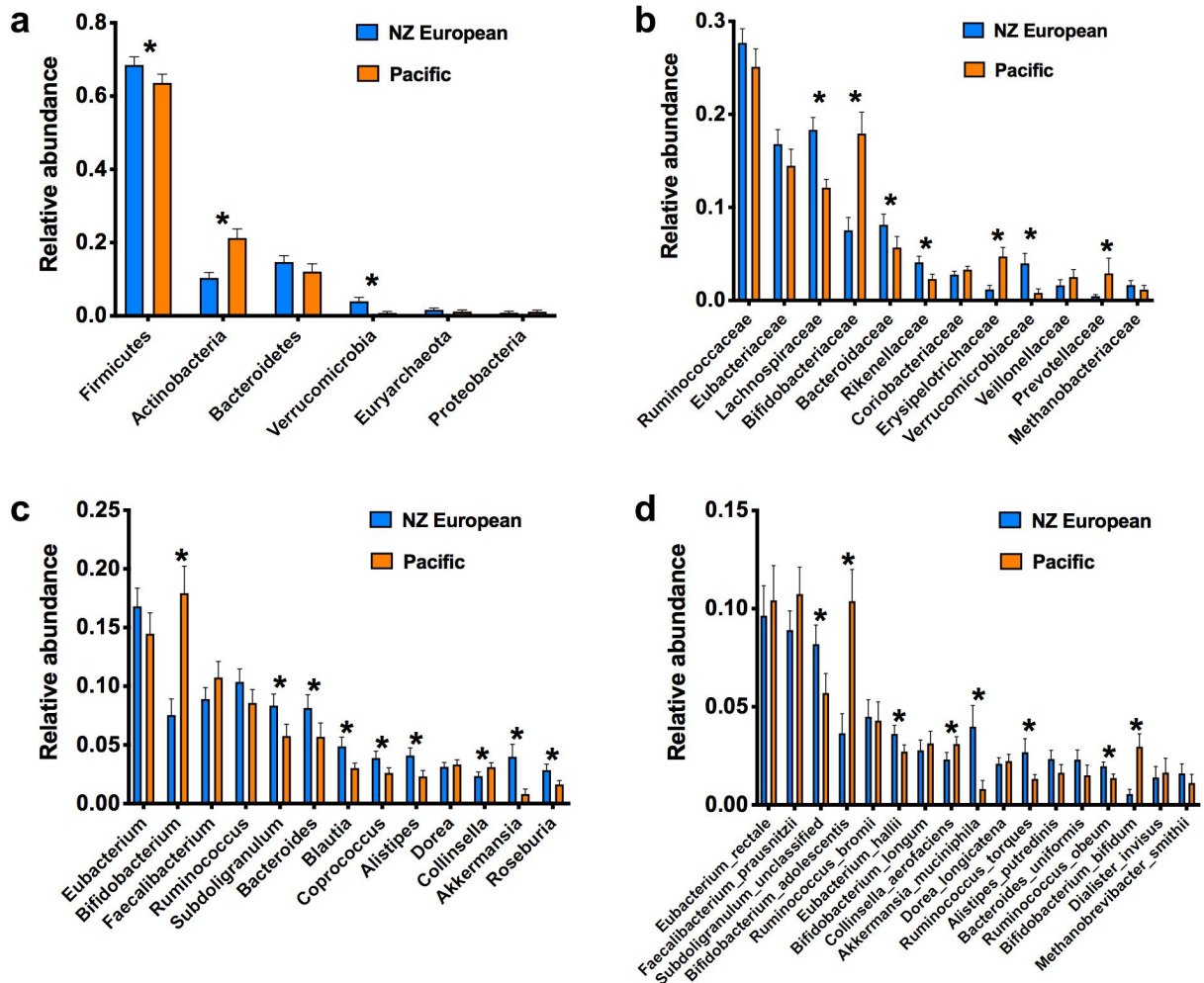
**Figure 1.** Beta diversity analysis (Bray-Curtis Distance) between fecal microbiotas of NZ European women and Pacific women, within and between ethnic groups.

### *Firmicutes to Bacteroidetes (F:B) ratio*

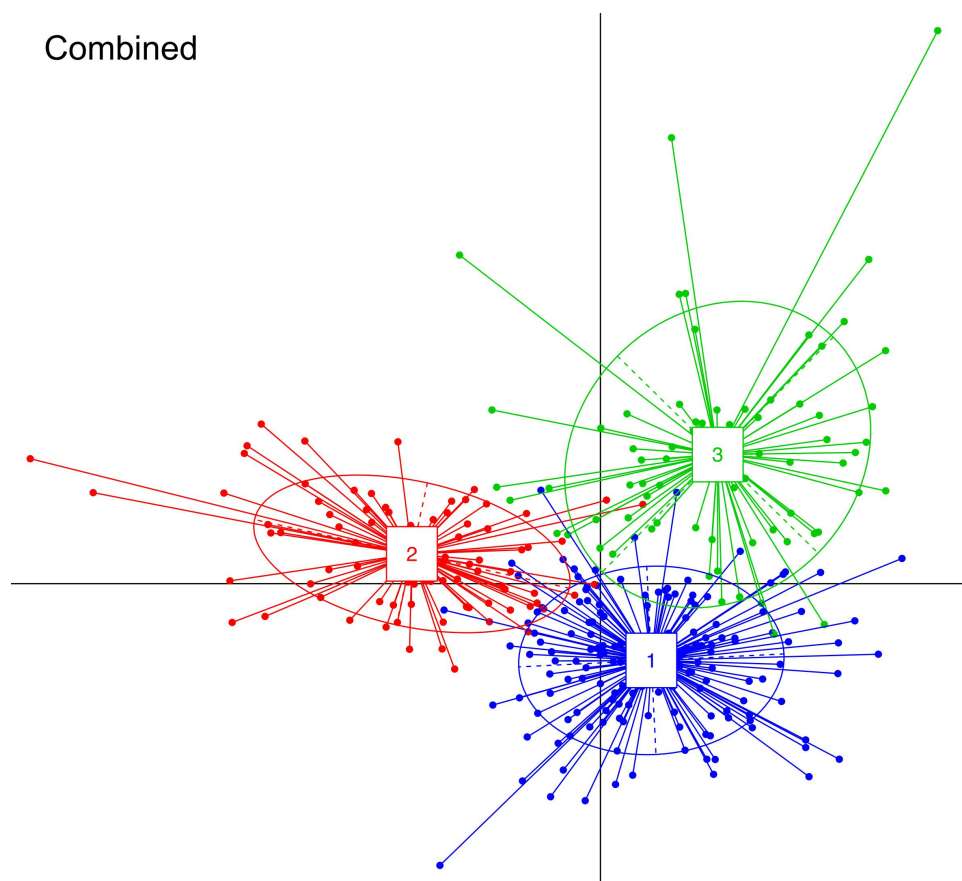
There were no significant differences in F:B ratio between Pacific and NZE women. However, participants with a low-BF% had a lower F:B ratio (5.2, 95%CL 3.1, 9.5) compared to those with high-BF% (9.8, 95%CL 4.4, 21.3). Enterotype 2 was associated with a greater F:B ratio in comparison to enterotypes 1 and 3. There was no difference in F:B ratio between enterotype 1 and 3 microbiotas (Table 2).

### *Fecal water content*

There were no significant differences in fecal water content between Pacific and NZE women. However, fecal water content was significantly lower in women with microbiotas



**Figure 2.** Relative abundances of the most commonly occurring taxa in the feces of NZE and Pacific women (\* $p < 0.05$ , Benjamini, Krieger, Yekutieli test).



**Figure 3.** BCA (between-class analysis) plots, showing clustering of samples from combined cohort data into three enterotypes.

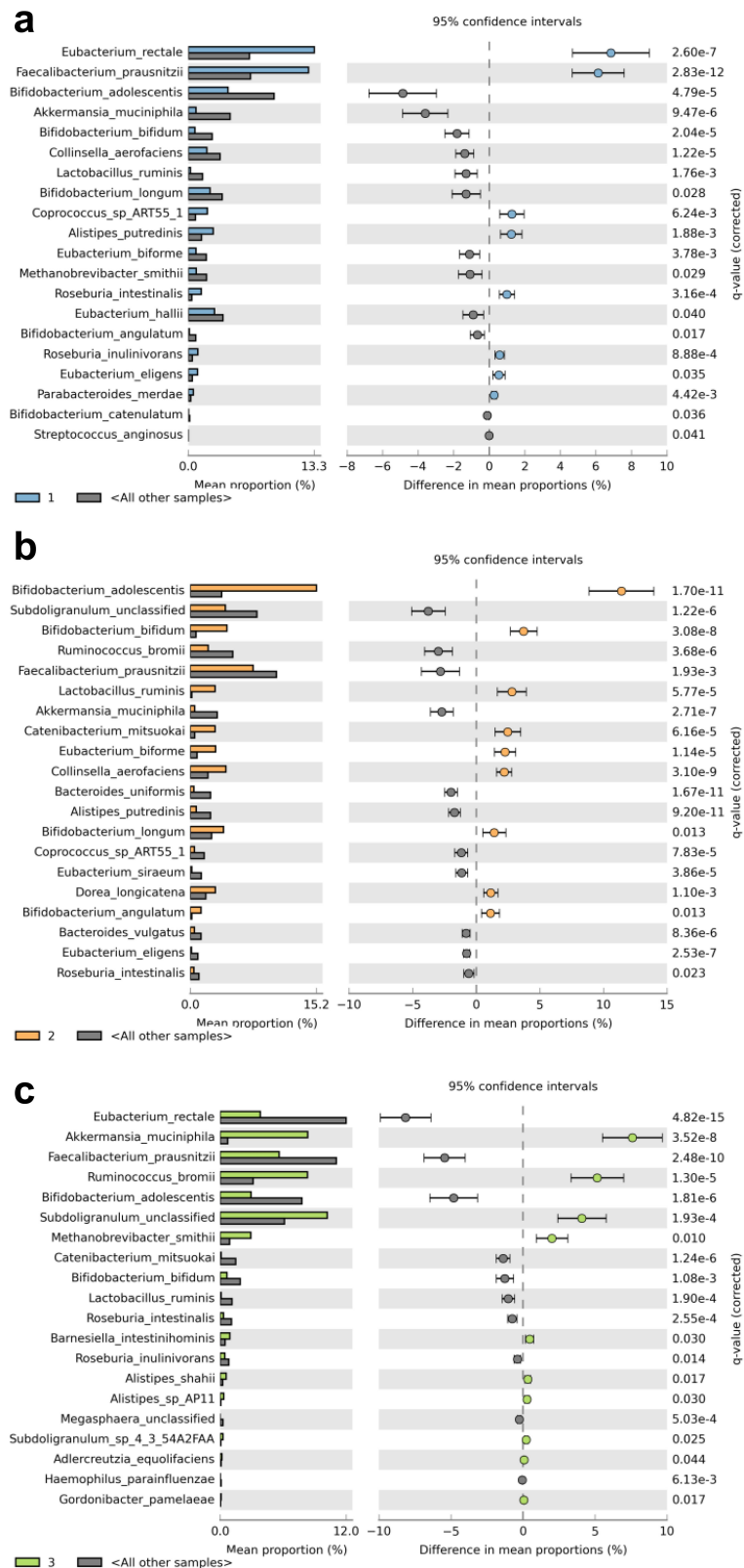
reflective of enterotype 3 (Table 2). In addition, the species that characterized enterotype 3 were all inversely correlated with fecal water content: *Akkermansia muciniphila* ( $r = -0.20$ ,  $p = 0.001$ ), *Ruminococcus bromii* ( $r = -0.19$ ,  $p = 0.001$ ), *Subdoligranulum spp.* ( $r = -0.23$ ,  $p < 0.001$ ), and *Methanobrevibacter smithii* ( $r = -0.14$ ,  $p = 0.022$ ).

#### **Association of diet with enterotypes**

Enterotypes were associated with habitual intake of particular food groups (Table 3). Participants whose microbiota was characterized by enterotype 1 or enterotype 3 had similar food group intakes but could be differentiated from each other on the basis of intakes of 'dairy products', 'cheese', 'non-starchy vegetables', and 'egg products' (Table 3). Participants with microbiotas

characterized by enterotype 2, consumed more 'discretionary savoury foods' and 'sugar sweetened beverages' compared to enterotypes 1 and 3, and less of the food groups that differentiated enterotype 1 from enterotype 3 (Table 3).

Multiple logistic regression showed significant associations between food groups and enterotype 1 and 3. In particular, for every serving size increase in 'dairy products', the likelihood of being characterized as enterotype 1 decreased by 54%, and conversely it increased the likelihood of being characterized as enterotype 3 by 180%. Moreover, for every serving size increase in 'starchy vegetables' the likelihood of being characterized as enterotype 1 increased by 145% and it decreased the likelihood of being characterized as enterotype 3 by 68%. Further, a strong positive association was observed between the intake of eggs and enterotype 3. This food group intake was



**Figure 4.** Description of predicted enterotypes. Each panel (a-c) contains the STAMP output (extended error bar plot) depicting species-level features within each enterotype with significant differential abundance (Welch's *t* test with Benjamini-Hochberg false discovery rate), compared to all other enterotypes.



**Table 2. Demographic, metabolic, and body composition data of participants stratified by enterotype.**

	Enterotype 1 n = 146	Enterotype 2 n = 70	Enterotype 3 n = 70
Ethnicity n (%) Pacific NZ European	57 (39%)	59 (84%)	9 (13%)
Age (years)	29 [23, 35] <sup>^</sup>	22 [20, 27] <sup>+</sup>	31 [24, 38] <sup>~</sup>
Weight (kg)	75.4 [64.5, 95.0] <sup>^</sup>	88.3 [68.5, 100.8] <sup>+</sup>	73.1 [63.9, 90.2]
Height (cm)	168.5 [163.9, 172.7]	168.4 [163.7, 172.6]	166.5 [164.0, 170.2]
BMI (kg/m <sup>2</sup> )	26.3 [22.8, 33.1] <sup>^</sup>	31.9 [24.6, 35.9] <sup>+</sup>	25.2 [22.7, 32.4]
Total body fat (%)	34.0 [28.5, 39.3] <sup>^</sup>	35.5 [31.9, 40.1]	32.5 [27.9, 40.2]
Visceral fat (%)	30.8 [23.1, 37.9] <sup>^</sup>	35.4 [29.1, 41.3] <sup>+</sup>	29.1 [20.4, 38.3]
Body fat groups n (%) <35% >35%	83 (57%) 63 (43%)	29 (41%) 41 (59%)	40 (57%) 30 (43%)
Total cholesterol (mmol/L) ∞	4.8 [4.4, 5.5] <sup>^</sup>	4.6 [4.2, 5.1]	4.8 [4.2, 5.5]
HDL-cholesterol (mmol/L) ∞	1.5 [1.3, 1.9] <sup>^</sup>	1.4 [1.2, 1.6] <sup>+</sup>	1.7 [1.4, 2.0] <sup>~</sup>
LDL - cholesterol (mmol/L) ∞	3.0 [2.5, 3.5]	2.9 [2.5, 3.4]	2.8 [2.4, 3.5]
Triglycerides (mmol/L) ∞	0.9 [0.7, 1.1] <sup>^</sup>	1.0 [0.8, 1.4] <sup>+</sup>	0.7 [0.6, 1.2]
Total cholesterol:HDL-cholesterol ratio (mmol/L) ∞	3.1 [2.7, 3.7] <sup>^</sup>	3.3 [2.7, 4.0] <sup>+</sup>	2.8 [2.4, 3.3] <sup>~</sup>
Glucose (mmol/L) ∞	5.3 [5.0, 5.6] <sup>^</sup>	5.4 [5.1, 5.8]	5.3 [4.9, 5.6]
Insulin (µU/ml) ∞	9.9 [6.9, 14.5] <sup>^</sup>	16.1 [10.7, 26.9] <sup>+</sup>	8.7 [6.4, 14.0]
Homeostatic Model Assessment of Insulin Resistance ∞	2.3 [1.5, 3.4] <sup>^</sup>	3.7 [2.4, 6.3] <sup>+</sup>	2.2 [1.6, 3.4]
HbA1c (mmol/mol) #	31.6 [29.6, 33.6] <sup>^</sup>	33.4 [31.7, 35.5] <sup>+</sup>	30.7 [29.3, 31.9] <sup>~</sup>
Deprivation index #	5 [3, 8] <sup>^</sup>	8 [6, 9] <sup>+</sup>	4 [3, 6] <sup>~</sup>
Predicted species abundance	72 [64, 78]	73 [67, 78]	76 [69, 81] <sup>~</sup>
Alpha diversity: Pielou's Evenness	0.71 [0.66, 0.74] <sup>^</sup>	0.68 [0.64, 0.71] <sup>+</sup>	0.70 [0.67, 0.74]
Shannon index	4.4 [4.1, 4.6] <sup>^</sup>	4.2 [3.9, 4.4] <sup>+</sup>	4.4 [4.2, 4.6]
Firmicutes:Bacteroidetes ratio	5.2 [2.6, 10.6] <sup>^</sup>	15.2 [7.0, 28.4] <sup>+</sup>	5.0 [2.8, 11.3]
Fecal water content (%) ≈	69 [62, 74]	70 [61, 77] <sup>+</sup>	65 [60, 69] <sup>~</sup>

All values are reported as medians [25th, 75th percentiles] or N (%).

<sup>^</sup>Statistically significant difference between enterotype 1 and 2, <sup>~</sup>Statistically significant difference between enterotype 1 and 3, <sup>+</sup>Statistically significant difference between enterotype 2 and 3 (Mann Whitney, p < 0.05).

<sup>∞</sup>Pacific woman (n = 1), <sup>≈</sup>NZE woman (n = 1), and #Pacific women (n = 2) have not been included in analyses due to missing data.

BMI: body mass index, HDL cholesterol: high density lipoprotein cholesterol, LDL cholesterol: low density lipoprotein cholesterol, HbA1c: glycosylated hemoglobin, Deprivation index: The 2013 New Zealand Deprivation Index.

**Table 3.** Habitual intake of food groups stratified by enterotype.

Food group g/day	Enterotype 1 n = 146	Enterotype 2 n = 70	Enterotype 3 n = 70
Dairy products	47.4 [21.8, 104.8]	48.0 [31.8, 92.4]~	94.7 [34.4, 159.8]+
Milk alternatives	0.0 [0.0, 30.0]^	0.0 [0.0, 0.0]~	0.0 [0.0, 31.0]
Cheese	13.2 [7.3, 19.7]^	6.8 [5.3, 9.5]~	17.4 [11.5, 24.7]+
Fruit	69.8 [44.5, 107.0]	61.0 [43.4, 106.1]	71.6 [49.4, 125.3]
Non starchy vegetables	83.0 [54.2, 120.8]^	46.0 [36.9, 59.7]~	93.8 [77.7, 126.1]+
Starchy vegetables	77.4 [66.4, 94.6]	73.1 [65.2, 85.1]	74.0 [64.0, 86.8]
Refined breads and cereals	58.5 [46.4, 76.3]^	70.3 [52.4, 84.5]	61.2 [44.1, 84.9]
Unrefined breads and cereals	48.4 [37.0, 66.3]^	40.3 [33.4, 52.8]	46.0 [34.0, 66.6]
Meat and seafood	70.5 [58.5, 81.8]	71.2 [55.3, 88.3]	70.0 [54.9, 81.8]
Processed meat	11.2 [7.4, 16.7]	9.4 [7.5, 14.5]	10.1 [6.8, 12.7]
Eggs	25.3 [18.4, 33.8]^	17.3 [14.7, 23.6]~	31.9 [23.1, 44.4]+
Legumes and meat alternatives	9.6 [5.6, 19.0]^	3.0 [0.0, 6.0]~	11.5 [7.4, 26.5]
Nuts and seeds	3.8 [2.1, 7.8]^	1.9 [0.9, 2.3]~	3.8 [2.2, 9.8]
Animal fat	4.6 [2.9, 7.0]	4.1 [3.0, 6.6]	3.8 [2.7, 7.0]
Plant based fat	9.2 [5.9, 13.9]^	6.1 [4.2, 7.7]~	8.9 [5.6, 13.1]
Savory sauces and condiments	43.7 [25.4, 64.9]^	33.7 [11.4, 47.5]~	54.4 [26.5, 74.2]
Discretionary sweet foods	62.8 [46.7, 88.8]	54.5 [44.1, 87.8]	61.0 [46.6, 84.5]
Discretionary savory foods	98.4 [73.5, 145.0]^	162.2 [119.9, 216.6]~	74.6 [61.0, 99.6]+
Sweetened beverages	251.7 [158.3, 320.9]^	351.1 [271.0, 411.6]~	193.8 [100.5, 250.7]+
Tea, coffee, water	992.4 [531.8, 1527.2]^	779.8 [370.3, 1173.5]~	1197.2 [725.8, 1718.5]
Alcoholic drinks	19.0 [8.0, 35.3]^	10.2 [0.0, 23.8]~	19.6 [11.2, 37.8]

All values are reported as medians [25th, 75th percentiles] Mann Whitney used to identify a significant difference between enterotypes ( $p < 0.05$ ) ^Statistically significant difference between enterotype 1 and 2 + Statistically significant difference between enterotype 1 and 3 ~ Statistically significant difference between enterotype 2 and 3 Enterotype 1: n = 146 (Pacific n = 57, NZ European n = 89); Enterotype 2: n = 70 (Pacific n = 59, NZ European n = 11); Enterotype 3: n = 70 (Pacific n = 9, NZ European n = 61)

negatively associated with enterotype 2. For every serving size increase in ‘cheese’, ‘non-starchy vegetables’, and ‘nuts and seeds’, the likelihood of being characterized as enterotype 2 decreased by 92%, 68% and 98% respectively (Table 4). Further adjustments for BF% groups did not alter results (data not shown).

### Metabolic capacity of the microbiota

The catalog of metagenomically assembled genes contained 3,019,279 non-redundant genes with, on average, 391,362 non-redundant genes per sample (standard deviation 80,183; minimum 128,516; maximum 553,887). We used linear regression to test for associations between EC enzymes, CAZymes, and other participant information, and the specific hypothesis that microbial bile salt hydrolases would be associated with BMI.<sup>33,34</sup> Associations were not found. However, fecal water content was positively associated with abundance of 117 MetaCyc pathways encoded by the gut microbiota (linear model, FDR corrected  $p < 0.1$ , top 50 associations summarized in Figure 5; complete data in Table S1). “NAD salvage pathway II”, “all trans farsenol biosynthesis (PWY-6859)”, and “taxadiene biosynthesis (PWY-7392)” were amongst the pathways associated with gut transit

time by others.<sup>35</sup> Shannon index of taxonomic profiles was associated with four MetaCyc pathways (adenine.and.adenosine.salvage.III [PWY.6609], superpathway.of.L.aspartate.and.L.asparagine.biosynthesis [ASPASN.PWY], aminoimidazole.ribonucleotide.biosynthesis.I [PWY.6121], peptidoglycan.biosynthesis.IV.Enterococcus.faecium [PWY.6471]), F:B ratio with two (formaldehyde.assimilation.III.dihydroxyacetone.cycle [P185.PWY], peptidoglycan.biosynthesis.IV.Enterococcus.faecium [PWY.6471]), and enterotype with one (dTDP.D.beta.fucofuranose.biosynthesis [PWY.7312]) (Table S1). Therefore, the main association between specific gene pools was with fecal water content.

### Discussion

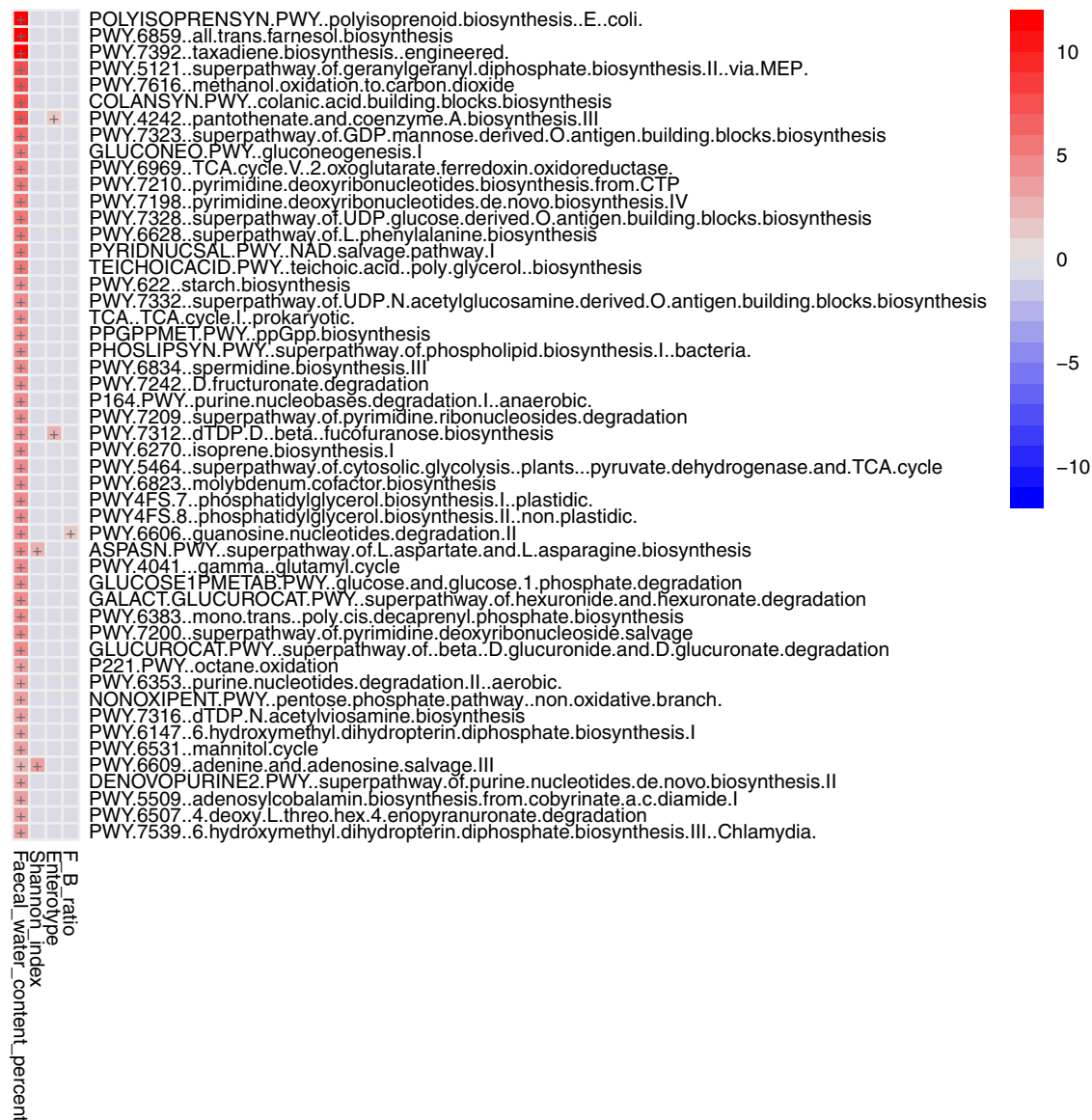
The complexity (number of bacterial species per microbiota) was similar between Pacific and NZE groups, but beta diversity analysis indicated differences in microbiota compositions. The microbiotas of the participants could be characterized by the definition of three enterotypes. Thus, enterotype 3 was mostly characteristic of NZE fecal microbiotas, whereas enterotype 2 was mostly characteristic of Pacific microbiotas, which indicated a possible ethnic

**Table 4.** The association of food group intake and the likelihood of being characterized as enterotype 1, 2 or 3.

Food group	Serving size	Enterotype 1 <sup>a</sup> n = 146		Enterotype 1 <sup>b</sup> n = 146		Enterotype 2 <sup>a</sup> n = 70		Enterotype 2 <sup>b</sup> n = 70		Enterotype 3 <sup>a</sup> n = 70		Enterotype 3 <sup>b</sup> n = 70		
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	
Dairy products	258 g	<b>0.46</b>	<b>[0.21,1.00]</b>	<b>0.46</b>	<b>[0.21,0.77]</b>	0.77	[0.36,1.67]		1.00	[0.46,2.80]	<b>2.80</b>	<b>[1.67,6.05]</b>	<b>2.80</b>	<b>[1.29,6.05]</b>
Milk alternatives	255 g	3.57	[0.77,12.65]	2.77	[0.77,12.65]	<b>0.01</b>	<b>[0.00,0.22]</b>		0.02	[0.00,1.00]	1.66	[0.46,5.92]	1.00	[0.22,3.57]
Cheese	40 g	1.68	[0.70,3.96]	1.38	[0.45,4.12]	<b>0.00</b>	<b>[0.00,0.02]</b>		<b>0.08</b>	<b>[0.01,0.92]</b>	<b>7.89</b>	<b>[3.02,21.72]</b>	1.61	[0.50,4.99]
Fruit	150 g	1.00	[0.47,1.82]	0.86	[0.41,1.57]	0.74	[0.30,1.57]		1.82	[0.74,4.45]	1.57	[0.74,2.85]	1.00	[0.47,2.11]
Non starchy vegetables	75 g	<b>1.82</b>	<b>[1.25,2.84]</b>	<b>1.96</b>	<b>[1.16,3.54]</b>	<b>0.06</b>	<b>[0.03,0.15]</b>		<b>0.32</b>	<b>[0.11,0.86]</b>	<b>1.96</b>	<b>[1.25,2.84]</b>	0.74	[0.44,1.35]
Starchy vegetables	75 g	<b>2.27</b>	<b>[1.00,5.11]</b>	<b>2.45</b>	<b>[1.00,5.50]</b>	0.59	[0.22,1.45]		0.93	[0.30,2.84]	0.55	[0.22,1.45]	<b>0.32</b>	<b>[0.11,0.93]</b>
Refined breads and cereals	40 g	0.79	[0.57,1.08]	0.82	[0.59,1.17]	<b>1.49</b>	<b>[1.08,2.12]</b>		1.32	[0.85,1.96]	0.89	[0.59,1.27]	1.08	[0.73,1.55]
Unrefined breads and cereals	40 g	1.04	[0.79,1.32]	1.00	[0.79,1.27]	0.67	[0.41,1.13]		0.85	[0.50,1.49]	1.17	[0.89,1.55]	1.08	[0.82,1.43]
Meat and seafood	65 g	1.07	[0.63,1.79]	1.07	[0.63,1.91]	1.14	[0.63,2.04]		1.07	[0.52,2.04]	0.82	[0.43,1.48]	0.88	[0.40,1.91]
Processed meat	50 g	1.28	[0.35,4.60]	1.35	[0.35,5.32]	0.70	[0.15,3.44]		1.49	[0.23,9.94]	1.00	[0.23,4.38]	0.45	[0.08,2.44]
Eggs	60 g	0.84	[0.34,2.05]	0.66	[0.23,1.82]	<b>0.02</b>	<b>[0.00,0.10]</b>		<b>0.11</b>	<b>[0.00,0.58]</b>	<b>14.03</b>	<b>[4.66,43.75]</b>	<b>6.62</b>	<b>[1.82,22.17]</b>
Legumes and meat alternatives	150 g	8.05	[1.00,72.83]	5.16	[0.55,54.40]	<b>0.00</b>	<b>[0.00,0.00]</b>		<b>0.00</b>	<b>[0.00,0.41]</b>	5.99	[0.86,35.07]	0.74	[0.09,6.94]
Nuts and seeds	30 g	<b>3.54</b>	<b>[1.35,9.00]</b>	<b>3.54</b>	<b>[1.27,9.79]</b>	<b>0.00</b>	<b>[0.00,0.01]</b>		<b>0.02</b>	<b>[0.00,0.55]</b>	1.61	[0.70,3.64]	0.56	[0.21,1.52]
Animal fat	20 g	0.90	[0.32,2.60]	0.72	[0.24,2.19]	0.42	[0.10,1.74]		1.00	[0.18,5.50]	2.28	[0.72,6.98]	1.37	[0.41,4.66]
Plant based fat	20 g	<b>1.91</b>	<b>[1.13,3.21]</b>	<b>1.70</b>	<b>[1.00,2.92]</b>	<b>0.10</b>	<b>[0.03,0.32]</b>		<b>0.21</b>	<b>[0.05,0.83]</b>	1.17	[0.74,1.88]	0.80	[0.48,1.35]
Savory sauces and condiments	30 g	1.23	[0.97,1.56]	1.20	[0.94,1.56]	<b>0.51</b>	<b>[0.35,0.72]</b>		0.76	[0.50,1.16]	1.23	[0.97,1.61]	0.94	[0.70,1.23]
Discretionary sweet foods	40 g	0.96	[0.82,1.17]	1.00	[0.82,1.22]	0.89	[0.70,1.13]		0.89	[0.64,1.22]	1.13	[0.92,1.38]	1.08	[0.89,1.38]
Discretionary savory foods	60 g	0.84	[0.66,1.00]	0.94	[0.70,1.27]	<b>2.17</b>	<b>[1.71,2.75]</b>		1.27	[0.89,1.71]	<b>0.46</b>	<b>[0.32,0.66]</b>	0.74	[0.48,1.20]
Sweetened beverages	340 g	0.71	[0.36,1.00]	1.00	[0.51,1.97]	<b>5.45</b>	<b>[2.77,10.72]</b>		1.40	[0.71,3.89]	<b>0.26</b>	<b>[0.13,0.51]</b>	0.71	[0.36,1.97]
Tea, coffee, water	255 g	1.00	[1.00,1.00]	1.00	[1.00,1.00]	<b>0.77</b>	<b>[0.77,1.00]</b>		1.00	[0.77,1.00]	<b>1.00</b>	<b>[1.00,1.29]</b>	1.00	[1.00,1.29]
Alcoholic drinks	150 g	1.35	[0.86,2.11]	1.16	[0.74,1.82]	0.47	[0.16,1.16]		1.35	[0.64,2.45]	1.00	[0.64,1.57]	0.74	[0.41,1.16]

Total n = 286; Pacific n = 125, NZ European n = 161; Enterotype 1: n = 146 (Pacific n = 57, NZ European n = 89); Enterotype 2: n = 70 (Pacific n = 59, NZ European n = 11); Enterotype 3: n = 70 (Pacific n = 9, NZ European n = 61). <sup>a</sup>Unadjusted models <sup>b</sup>Models adjusted for NZDep2013, Age, ethnicity, and energy intake Odds ratio (bolded when statistically significant) represent the change in the outcome per 1 serving size of change in food group intake (presented as the g/serve equivalent e.g., 250 mL of standard cow milk = 258 g). Comparisons focus on the likelihood of being in 1 enterotype compared to both other enterotypes

### Top 50 features with significant associations ( $-\log(\text{qval}) \cdot \text{sign}(\text{coeff})$ )



**Figure 5.** The top 50 most represented biochemical features in the fecal metagenome positively associated with fecal water content. A full list of features is given in Table S1.

influence on fecal microbiota composition. Further comparisons revealed that the different categories of microbiotas were also associated with some anthropometric measurements, notably BMI, visceral fat %, and cholesterol values. Thus, enterotyping of fecal microbiotas helped delineate human groups with different ethnic and metabolic characteristics. The enterotypes that we describe here are different from those previously described in relation to stratification of microbiota composition.<sup>25</sup> Commonly, three enterotypes have been observed, usually characterized by the

abundance of the genera *Bacteroides* (E1), *Prevotella* (E2), or *Ruminococcus* (E3). We used the same bioinformatics approach for detection of enterotypes as described by others<sup>25</sup> based on the Calinski-Harabasz index, but our taxonomic data is derived from MetaPhlan analysis of shot-gun DNA sequencing which provides accurate species discrimination. Together with STAMP, we find that robust clusters of bacterial species whose relative abundances differ between individuals characterize and differentiate microbiotas.

The composition of the microbiota could be confounded by features of human diet, environment, and innate physiology.<sup>36–38</sup> Indeed, we found that fecal water content was different between subjects whose microbiotas belonged to enterotype 3 compared to enterotypes 1 and 2. Fecal water content was used as a surrogate measurement of stool consistency which has been considered to reflect gut transit time.<sup>37,39</sup> The Bristol Stool Form Scale is often used as a tool for subjective scoring of stool consistency but has only moderate correlation when used by untrained participants.<sup>40</sup> Fecal dry weight (fecal water content) has been recommended as a reliable method to assess stool consistency because it is an objective measurement and has high reproducibility.<sup>40</sup> Fecal water content is also influenced by the water-holding capacity of insoluble solids, and the water-absorbing ability of the participant.<sup>41</sup> Enterotype 3 participants had feces of lower water content, indicating slower colonic transit time. This enterotype was characterized by the relative abundances of *Subdoligranulum* sp., *Akkermansia muciniphila*, *Ruminococcus bromii*, and *Methanobacter smithii* and is consistent with the observations of Vandeputte and colleagues.<sup>42</sup> They found an increased abundance of methane-producing bacteria (*Methanobrevibacter*), *Akkermansia*, and *Ruminococaceae* in association with stool consistency indicative of slower gut transit time. An association of Archaea with firmer stool consistency was also reported by Tigchelaar et al.<sup>43</sup> Both these studies used Bristol Stool Scale assessment of fecal properties. Gut transit time can be measured accurately as the duration of time from ingestion of blue dye within a standardized food to its first excretion of blue color within a stool. Asnicar and colleagues<sup>35</sup> used this method to measure gut transit time and reported that longer transit time was associated with increased abundances of *Akkermansia muciniphila*, *Bacteroides*, and *Alistipes* spp. Steenackers et al.<sup>44</sup> linked alterations in fecal microbiota composition specifically to colonic transit time. This topic has been reviewed recently by Prochazkova and colleagues<sup>45</sup> who concluded that disease-related microbiota

compositions may be confounded by changes in gut transit time. There is, therefore, considerable agreement between studies on the influence of colonic transit time on the taxonomic composition of fecal microbiotas. As in our study, functional pathways and gene families represented in fecal metagenomes were also assessed by Asnicar et al.<sup>35</sup> “NAD salvage pathway II”, “all trans farsenol biosynthesis (PWY-6859)”, and “taxadiene biosynthesis (PWY-7392)” were amongst the pathways that had greater representation in fecal metagenomes associated with slower gut transit. We also detected different abundances of genes associated with these pathways, but there was a positive rather than a negative correlation with water content of feces. Overall, knowledge of metabolic capacity of microbiotas did not aid in the discrimination of groups with different BMI or body fat composition.

*Bifidobacterium* abundance has been inversely correlated with the fat content of the diet rather than stool consistency/gut transit time.<sup>39</sup> Higher bifidobacterial abundances characterized enterotype 2 and significant negative associations between the intake of the food groups ‘plant-based fat’ and ‘nuts and seeds’, which are characteristically high in unsaturated fatty acids, were observed.

Limitations of our study include the use of self-reported dietary data that may be prone to misreporting but which nevertheless provide insight into eating habits unavailable by other means.<sup>46</sup> Lifestyle factors other than diet (for example, amount and type of physical activity) could influence metabolic health profiles and are under investigation with the study participants.<sup>47</sup>

The results of our study pointed to marked associations with ethnicity on microbiota composition because 84% of Pacific women had microbiotas characterized by enterotype 2. Of particular note, the Firmicutes to Bacteroidetes (F:B) ratio was greater in participants with enterotype 2 microbiotas. As reported by others, higher F:B ratio was associated with higher BMI values. Ethnic differences in microbiota composition point to the potential impact of human genetics, and/or fecal microbiotas of family members, and/or general

household environment (including habitually ingested foods) on the composition of the gut microbiota.<sup>26,48</sup> Comparison of ethnicities in previous studies sometimes shows differences in fecal microbiota composition.<sup>49–53</sup> However, the differences are likely to be due to lifestyle differences, and environmental factors that tend to be characteristic of ethnic groups including dietary patterns, and living and working in similar neighborhoods, rather than ethnicity per se.<sup>26,48,54</sup> All things considered, the genotype of the host probably has little impact on the composition of the adult microbiota which is mainly influenced by environmental factors such as diet and shared environment.<sup>55</sup> Indeed, we observed significant associations between food group intake and enterotypes. Higher intakes of egg and dairy products increased the likelihood of being characterized as enterotype 3, which primarily consisted of older NZE women. Moreover, ‘non-starchy vegetables’, ‘nuts and seeds’ and ‘plant-based fats’ were positively associated with the likelihood of being characterized as enterotype 1, detected in both Pacific and NZE microbiotas. These food groups were inversely associated with enterotype 2, primarily detected in Pacific women. It is unlikely that the differences we observed were driven by study design (i.e., selectively recruiting low- and high-BF% groups) because the observations were independent of further adjustment for BF% groups. Clearly, the causative role of specific food groups identified in our study on the selection of enterotypes detected in feces needs to be further investigated through dietary intervention studies.<sup>56</sup>

## Conclusions

Sze and Schloss<sup>11</sup> proposed that evaluation of metabolic capacity represented by different gene pools within the microbiota might assist in understanding the contribution of the microbial community to obesity. Although we observed differences in the representation of biochemical pathways among participants, this was associated with fecal water content (stool consistency/gut transit time) rather than BMI or body fat percentage. Importantly, our study of NZ women showed differences in the taxonomic compositions and metabolic capacities of the fecal microbiota that were associated with dietary intakes and fecal water content. Clearly, future studies in

which fecal microbiotas are characterized must be conducted with cognizance of the ethnicity, habitual dietary intake, and colonic transit times of the participants.

## Acknowledgments

The PROMISE study was funded by the Health Research Council (HRC) of New Zealand. NR was supported by a PhD scholarship provided by the Riddet Institute, Massey University, Palmerston North, New Zealand. The authors would like to thank all the volunteers who participated in the PROMISE study, including staff and students who participated in recruitment and data collection. Joanne Slater, Niamh Brennan, Sophie Kindleysides and Moana Manukia played key roles in participant recruitment, coordination, and data acquisition of the PROMISE study. Elizabeth Cullen, Bronte Anscombe, Ashleigh Jackson, Owen Mugridge, Shakeela Jayasinghe, Sherina Holland, and Ridvan Firestone provided support during data acquisition and tabulation.

## Disclosure statement

There are no competing interests.

## Funding

Authors' contributions: NR, PhD student; BL, GWT, supervision of microbiota analysis; TV, metabolic capacity analysis; BM, JD, MC, supervision of statistical analysis; LTM, BM, RK, BB supervision of nutritional analysis; NR and GWT wrote the manuscript with input from other authors.

## References

1. Obesity and overweight. WHO. 2021. [accessed 2022 April 5]. [www.who.int/news-room/fact-sheets/detail/obesity-and-overweight](http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight).
2. OECD obesity update. 2017. [accessed 2022 April 5]. [www.oecd.org/health/obesity-update.htm](http://www.oecd.org/health/obesity-update.htm)
3. Obesity statistics. NZ Ministry of Health. 2021. [accessed 2022 April 5]. <https://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/obesity-statistics>.
4. Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A*. 2004 Nov 2;101(44):15718–15723. doi:10.1073/pnas.0407076101.
5. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy

- harvest. *Nature*. 2006 Dec 21;444(7122):1027–1031. doi:10.1038/nature05414.
6. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*. 2005 Aug 2;102(31):11070–11075. doi:10.1073/pnas.0504978102.
  7. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med*. 2009 Nov 11;1(6):6ra14. doi:10.1126/scitranslmed.3000322.
  8. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006 Dec 21;444(7122):1022–1023. doi:10.1038/4441022a.
  9. Duncan SH, Lobley GE, Holtrop G, Ince J, Johnstone AM, Louis P, Flint HJ. Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obes (Lond)*. 2008 Nov;32(11):1720–1724. doi:10.1038/ijo.2008.155.
  10. Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett*. 2014 Nov 17;588(22):4223–4233. doi:10.1016/j.febslet.2014.09.039.
  11. Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio*. 2016 Aug 23;7(4):e01018–16. doi:10.1128/mBio.01018-16.
  12. Dickey RA. AACE/ACE obesity task force. AACE/ACE position statement on the prevention, diagnosis, and treatment of obesity (1998 revision). *Endocr Pract*. 1998;4:297–350.
  13. Jo A, Mainous AG. Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: a nationally representative cross-sectional study. *BMJ Open*. 2018 Apr 13;8(4):e019200. doi:10.1136/bmjopen-2017-019200.
  14. Oliveros E, Somers VK, Sochor O, Goel K, Lopez-Jimenez F. The concept of normal weight obesity. *Prog Cardiovasc Dis*. 2014 Jan-Feb;56(4):426–433. doi:10.1016/j.pcad.2013.10.003.
  15. Kramer CK, Zinman B, Retnakaran R. Are metabolically healthy overweight and obesity benign conditions?: a systematic review and meta-analysis. *Ann Intern Med*. 2013 Dec 3;159(11):758–769. doi:10.7326/0003-4819-159-11-201312030-00008.
  16. Kindleysides S, Kruger R, Douwes J, Tannock GW, Renall N, Slater J, Lawley B, McGill A-T, Brennan N, Manukia M. Predictors Linking Obesity and the Gut Microbiome (the PROMISE Study): protocol and Recruitment Strategy for a Cross-Sectional Study on Pathways That Affect the Gut Microbiome and Its Impact on Obesity. *JMIR Res Protoc*. 2019;8(8):e14529. doi:10.2196/14529.
  17. Atkinson J, Salmond C, Crampton P. NZDep2013 index of deprivation. Wellington, New Zealand: University of Otago; 2014.
  18. Beck KL, Houston ZL, McNaughton SA, Kruger R. Development and evaluation of a food frequency questionnaire to assess nutrient intakes of adult women in New Zealand. *Nutr Diet*. 2020 Apr;77(2):253–259. doi:10.1111/1747-0080.12472.
  19. Beck KL, Jones B, Ullah I, McNaughton SA, Haslett SJ, Stonehouse W. Associations between dietary patterns, socio-demographic factors and anthropometric measurements in adult New Zealanders: an analysis of data from the 2008/09 New Zealand Adult Nutrition Survey. *Eur J Nutr*. 2018 Jun;57(4):1421–1433. doi:10.1007/s00394-017-1421-3.
  20. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman LS. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*. 2009 Dec;65(4):1003–1010. doi:10.1111/j.1541-0420.2009.01223.x.
  21. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc*. 2006 Oct;106(10):1575–1587. doi:10.1016/j.jada.2006.07.003.
  22. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods*. 2015 Oct;12(10):902–903. doi:10.1038/nmeth.3589.
  23. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol*. 2001;26:32–46.
  24. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*. 2006 Mar;62(1):245–253. doi:10.1111/j.1541-0420.2005.00440.x.
  25. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, et al. Enterotypes of the human gut microbiome. *Nature*. 2011 12;473(7346):174–180. doi:10.1038/nature09944.
  26. Lawley B, Otal A, Moloney-Geany K, Diana A, Houghton L, Heath AM, Taylor RW, Tannock GW. Fecal microbiotas of Indonesian and New Zealand children differ in complexity and bifidobacterial taxa during the first year of life. *Appl Environ Microbiol*. 2019 Sep 17;85(19):e01105–19. doi:10.1128/AEM.01105-19.
  27. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014 Nov 1;30(21):3123–3124. doi:10.1093/bioinformatics/btu494.
  28. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015 May 15;31(10):1674–1676. doi:10.1093/bioinformatics/btv033.

29. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 2010 Mar 8;11:119. doi:10.1186/1471-2105-11-119.
30. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012 Dec 1;28(23):3150–3152. doi:10.1093/bioinformatics/bts565.
31. Clausen PTL, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinform.* 2018 Aug 29;19(1):307. doi:10.1186/s12859-018-2336-6.
32. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019 Jan;47(D1):D309–D314. doi:10.1093/nar/gky1085.
33. Jones BV, Begley M, Hill C, Gahan CG, Marchesi JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A.* 2008 Sep 9;105(36):13580–13585. doi:10.1073/pnas.0804437105.
34. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, Shanahan F, Hill C, Gahan CG. Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *Proc Natl Acad Sci U S A.* 2014 May 20;111(20):7421–7426. doi:10.1073/pnas.1323599111.
35. Asnicar F, Leeming ER, Dimidi E, Mazidi M, Franks PW, Al Khatib H, Valdes AM, Davies R, Bakker E, Francis L, et al. Blue poo: impact of gut transit time on the gut microbiome using a novel marker. *Gut.* 2021 Sep;70(9):1665–1674. doi:10.1136/gutjnl-2020-323877.
36. Zhu L, Liu W, Alkhoury R, Baker RD, Bard JE, Quigley EM, Baker SS. Structural changes in the gut microbiome of constipated patients. *Physiol Genomics.* 2014 Sep 15;46(18):679–686. doi:10.1152/physiolgenomics.00082.2014.
37. Roager HM, Hansen LB, Bahl MI, Frandsen HL, Carvalho V, Gøbel RJ, Dalgaard MD, Plichta DR, Sparholt MH, Vestergaard H, et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat Microbiol.* 2016 Jun 27;1(9):16093. doi:10.1038/nmicrobiol.2016.93.
38. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. *Nature.* 2020 Nov;587(7834):448–454. doi:10.1038/s41586-020-2881-9.
39. Müller M, Hermes GDA, Canfora EE, Smidt H, Masclee AAM, Zoetendal EG, Blaak EE. Distal colonic transit is linked to gut microbiota diversity and microbial fermentation in humans with slow colonic transit. *Am J Physiol Gastrointest Liver Physiol.* 2020 Feb 1;318(2):G361–G369. doi:10.1152/ajpgi.00283.2019.
40. Vork L, Wilms E, Penders J, Jonkers DMAE. Stool consistency: looking beyond the Bristol Stool Form Scale. *J Neurogastroenterol Motil.* 2019 Oct 30;25(4):625. doi:10.5056/jnm19086.
41. Wenzl HH, Fine KD, Schiller LR, Fordtran JS. Determinants of decreased fecal consistency in patients with diarrhea. *Gastroenterology.* 1995;108(6):1729–1738. doi:10.1016/0016-5085(95)90134-5.
42. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut.* 2016 Jan;65(1):57–62. doi:10.1136/gutjnl-2015-309618.
43. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. Gut microbiota composition associated with stool consistency. *Gut.* 2016 Mar;65(3):540–542. doi:10.1136/gutjnl-2015-310328.
44. Steenackers N, Falony G, Augustijns P, Van der Schueren B, Vanuytsel T, Vieira-Silva S, Wauters L, Raes J, Matthys C. Specific contributions of segmental transit times to gut microbiota composition. *Gut.* 2022 Jul;71(7):1443–1444. doi:10.1136/gutjnl-2021-325916.
45. Procházková N, Falony G, Dragsted LO, Licht TR, Raes J, Roager HM. Advancing human gut microbiota research by considering gut transit time. *Gut.* 2023 Jan;72(1):180–191. doi:10.1136/gutjnl-2022-328166.
46. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhauser ML, Thompson FE, Potischman N, Guenther PM, Tarasuk V, et al. Addressing current criticism regarding the value of self-report dietary data. *J Nutr.* 2015;145(12):2639–2645. doi:10.3945/jn.115.219634.
47. Slater J, Kruger R, Douwes J, O'Brien WJ, Corbin M, Miles-Chan JL, Breier BH. Objectively measured physical activity is associated with body composition and metabolic profiles of Pacific and New Zealand European women with different metabolic disease risks. *Front Physiol.* 2021 May 12 DOI: 10.3389/fphys.2021.684782
48. Xu J, Lawley B, Wong G, Otal A, Chen L, Ying TJ, Lin X, Pang WW, Yap F, Chong YS, et al. Ethnic diversity in infant gut microbiota is apparent before the introduction of complementary diets. *Gut Microbes.* 2020 Sep 2;11(5):1362–1373. doi:10.1080/19490976.2020.1756150.
49. Brooks AW, Priya S, Blekhan R, Bordenstein SR. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol.* 2018 Dec 4;16(12):e2006842. doi:10.1371/journal.pbio.2006842.
50. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, Tremaroli V, Bakker GJ, Attaye I, Pinto-Sietsma SJ, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat Med.* 2018 Oct;24(10):1526–1531. doi:10.1038/s41591-018-0160-1.



51. Brewster R, Tamburini FB, Asimwe E, Oduaran O, Hazelhurst S, Bhatt AS. Surveying gut microbiome research in Africans: toward improved diversity and representation. *Trends Microbiol.* 2019 Oct;27(10):824–835. doi:10.1016/j.tim.2019.05.006.
52. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature.* 2019 Apr;568(7753):505–510. doi:10.1038/s41586-019-1058-x.
53. Dwiyanto J, Ayub Q, Lee SM, Foo SC, Chong CW, Rahman S. Geographical separation and ethnic origin influence the human gut microbial composition: a meta-analysis from a Malaysian perspective. *Microb Genom.* 2021;7(8):000619. doi:10.1099/mgen.0.000619.
54. Lu J, Zhang L, Zhai Q, Zhao J, Zhang H, Lee Y-K, Lu W, Li M, Chen W. Chinese gut microbiota and its associations with staple food type, ethnicity, and urbanization. *npj Biofilms Microbiomes.* 2021;7(71). doi:10.1038/s41522-021-00245-0.
55. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature.* 2018 Mar 8;555(7695):210–215. doi:10.1038/nature25973.
56. Tannock GW. Modulating the gut microbiota of humans by dietary intervention with plant glycans. *Appl Environ Microbiol.* 2021 Feb 26;87(6):e02757–20. doi:10.1128/AEM.02757-20.