British Journal of Nutrition (2022), 128, 93–102 doi:10.1017/S0007114521002968

Long-term association between diet quality and characteristics of the gut microbiome in the multiethnic cohort study

Erica Ma[1], Gertraud Maskarinec[1]*, Unhee Lim[1], Carol J. Boushey[1], Lynne R. Wilkens[1], V. Wendy Setiawan[2], Loïc Le Marchand[1], Timothy W. Randolph[3], Isaac C. Jenkins[3], Keith R. Curtis[3], Johanna W. Lampe[3] and Meredith A.J. Hullar[3]

[1]*University of Hawai'i Cancer Center, Honolulu, HI, USA*
[2]*Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA*
[3]*Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

(Submitted 10 April 2021 – Final revision received 21 July 2021 – Accepted 2 August 2021 – First published online 9 August 2021)

Abstract

As past usual diet quality may affect gut microbiome (GM) composition, we examined the association of the Healthy Eating Index (HEI)-2015 assessed 21 and 9 years before stool collection with measures of fecal microbial composition in a subset of the Multiethnic Cohort. A total of 5936 participants completed a validated quantitative FFQ (QFFQ) at cohort entry (Q1, 1993–1996), 5280 at follow-up (Q3, 2003–2008) and 1685 also at a second follow-up (Adiposity Phenotype Study (APS), 2013–2016). All participants provided a stool sample in 2013–2016. Fecal microbial composition was obtained from 16S rRNA gene sequencing (V1–V3 regions). HEI-2015 scores were computed based on each QFFQ. Using linear regression adjusted for relevant covariates, we calculated associations of HEI-2015 scores with gut microbial diversity and 152 individual genera. The mean HEI-2015 scores increased from Q1 (67 (SD 10)) to Q3 (71 (SD 11)) and APS (72 (SD 10)). Alpha diversity assessed by the Shannon Index was significantly higher with increasing tertiles of HEI-2015. Of the 152 bacterial genera tested, seven (*Anaerostipes*, *Coprococcus_2*, *Eubacterium eligens*, *Lachnospira*, *Lachnospiraceae_ND3007*, *Ruminococcaceae_UCG-013* and *Ruminococcus_1*) were positively and five (*Collinsella*, *Parabacteroides*, *Ruminiclostridium_5*, *Ruminococcus gnavus* and *Tyzzerella*) were inversely associated with HEI-2015 assessed in Q1, Q3 and APS. The estimates of change per unit of the HEI-2015 score associated with the abundance of these twelve genera were consistent across the three questionnaires. The quality of past diet, assessed as far as ~20 years before stool collection, is equally predictive of GM composition as concurrently assessed diet, indicative of the long-term consistency of this relation.

Key words: **Gut microbiome: Diet: Healthy Eating Index: 16S rRNA gene sequencin**

Over the past decade, interest in gut microbiome (GM) composition in relation to health status has increased tremendously due to technical advances in genomic analyses[1]. Many determinants of microbiome patterns, e.g. ageing, use of antibiotics, other medications, environmental exposures and genetics, have been identified[2,3]. Nutritional factors influence the presence and activity of specific gut bacteria and are linked to conditions such as the metabolic syndrome, colorectal cancer, inflammatory bowel disease and CVD have emerged[2,4]. Diet quality indices are commonly utilised to analyse the impact of overall diet composition as a single measure[5]. The Healthy-Eating Index (HEI) describes compliance with the USA federal dietary guidelines[6,7].

Our previous report based on the Adiposity Phenotype Study (APS), a subset of the Multiethnic Cohort (MEC) with close to 2000 participants, found that higher diet quality was associated with greater α diversity and the relative abundance of the phylum Actinobacteria and twenty-one genera[8], including seven members of the *Lachnospiraceae* family from the phylum Firmicutes. Similarly, in a Swedish study, higher scores for a "health-conscious" food pattern were related to the abundance of six genera of the phylum Firmicutes[9]: loading positively in relation to *Roseburia*, *Lachnospira* and unclassified in order RF39 and negatively in relation to *Blautia*, *Eubacterium* and *Anaerotruncus*. Lower HEI-2005 scores were associated with a lower relative abundance of *Roseburia* in a recent report[10]. Beneficial effects of the Mediterranean Diet on the diversity and GM composition have also been detected in several investigations[8,11–13]. For example, adherence to the MD was associated with significantly higher levels of total SCFA in fecal samples and linked to a significant reduction in overall mortality and

Abbreviations: DXA, Dual-energy X-ray Absorptiometry; GM, gut microbiome; HEI, Healthy-Eating Index; mGWAS, Microbiome Genome-Wide Association Study; MEC, Multiethnic Cohort; MRI, Magnetic Resonance Imaging; QFFQ, quantitative FFQ.

* **Corresponding author:** Gertraud Maskarinec, email gertraud@cc.hawaii.edu

morbidity[11–13]. Also, a greater presence of Bacteroidetes was associated with a lower animal protein intake[12]. A recent report identified a panel of intestinal species that were associated with healthy dietary habits and overlapped with those associated with favourable cardiometabolic and postprandial markers[14].

The current analysis examines the association of the HEI-2015 with gut microbial diversity and the structure of the gut microbial community, expanding our previous MEC cross-sectional analysis[8] of 1735 participants and 103 individual bacteria to a larger study population of more than 5000 cohort members and an updated reference library with 152 genera. By analysing diet quality assessed at cohort entry (Q1) and follow-up (Q3), 21 and 9 years before stool collection, respectively, long-term exposure and changes in diet quality over time can be evaluated. Significant findings for past diet as assessed in Q1 and Q3 were confirmed in the data set that assessed diet quality (APS) and the GM at the same time[8].

Materials and methods

Study population

The MEC is an ongoing study of more than 215 000 participants aged 45–75 years from five ethnic groups (Japanese American, white, Latino, African American and Native Hawaiian) recruited in Hawaii and Los Angeles, California (online Supplemental Fig. S1). As previously published[15], all participants completed a self-administered questionnaire (Q1) at cohort entry (1993–1996), which collected information on anthropometrics, socio-demographic factors, medical history, physical activity and a quantitative FFQ (QFFQ). A follow-up questionnaire (Q3) during 2003–2008 was completed by 98 214 (46 %) participants[16]. As part of a Microbiome Genome-Wide Association Study (mGWAS) study and the APS conducted in 2013–2016[17], 6094 participants were recruited from the MEC study using a uniform stool collection protocol[18]. The study protocol was approved by the Biomedical Institutional Review Boards at the University of Hawaii (CHS#17200) and the University of Southern California (#HS-12-00623); all participants signed an informed consent form. After excluding individuals with invalid dietary information as well as missing BMI and/or microbiome data (*n* 158), the final data set consisted of 5936 mGWAS and APS participants. As 656 did not complete the follow-up questionnaire, 5280 cohort members were part of the Q3 data set.

Microbiome genome-wide association study

The mGWAS includes a MEC subset of 4502 participants aged 60–90 years who had been previously genotyped for different GWAS projects[18]. Exclusion criteria due to their known effects on the gut microbiota[19,20] included ileostomy/colectomy and dialysis. Participation was deferred for recent treatment with chemotherapy, radiation therapy, corticosteroid hormones, prescription weight-loss drugs, insulin or thyroid medications, antibiotics and flu shot or other vaccination; recent endoscopy or irrigation and/or cleansing of the large intestine and recent substantial weight change[18]. Participants were asked to mail their stool sample with a completed questionnaire addressing exposures related to the fecal microbiome, such as recent antibiotics use. The overall participation rate was 43 % among eligible individuals.

Adiposity phenotype study

The APS recruited a subset of 1861 MEC participants aiming to study body fat distribution and its determinants[17]. Participants aged 60–77 years were stratified by sex, ethnicity and six BMI categories (18·5–21·9, 22–24·9, 25–26·9, 27–29·9, 30–34·9 and 35–40 kg/m^2). Exclusion criteria included current BMI outside 18·5–40 kg/m^2, current/recent (<2 years) smoking, soft or metal implants or serious health conditions, with similar deferment criteria as in the mGWAS for recent relevant exposures. The overall participation rate was 23 % of eligible individuals. During a clinic visit, blood and stool samples, anthropometric measures, questionnaire data and Magnetic Resonance Imaging (MRI) and Dual-energy X-ray Absorptiometry (DXA) scans were obtained[17].

Dietary assessment and healthy eating index-2015

A detailed description of the QFFQ development and calibration has been reported previously[21]. Briefly, the QFFQ was designed to include food groups that contributed >85 % (for each ethnic group) of the intake of nutrients of interest during the past year[21]. The frequency of over 180 food items was assessed using eight categories for food groups and nine for beverages and the amount of each item was recorded in three serving sizes[15]. Although the substance of the QFFQ remained the same, the original QFFQ was updated in 2003 for Q3 and the APS to modify the food lists, amounts and examples or names given for the food items without substantial change.

Using an extensive food composition database[22] and the same approach, estimated daily intakes were computed to calculate the HEI-2015, which is an updated version of the original HEI-2005 and HEI-2010[23,24] reflecting the Dietary Guidelines for American 2015–2020[6]. The HEI-2015 is composed of thirteen components with a maximum of 100 points[7] and includes nine adequacy components (total fruits, whole fruits, total vegetables, greens and beans, whole grains, dairy, total protein foods, seafood and plant proteins and fatty acids) as well as four moderation components (refined grains, Sodium, added sugars and saturated fats). Most components are scored as amounts per 1000 kcal of intake. The key differences between the 2010 and 2015 index include the allocation of legumes, which was modified to total vegetables, greens and beans, total protein foods and seafood/plant proteins and the empty energy component, which was replaced with saturated fat and added sugars.

Stool collection

During mGWAS and APS studies in 2013–2016, stool samples were collected in RNA*later* (Ambion) and stored at −80 °C until bulk shipment on dry ice from Honolulu or Los Angeles to Seattle for processing. Stool sampling was delayed for at least 3 months for participants who first reported use of antibiotics. As previously described[25], DNA was extracted and amplified for the V1–V3 region of the 16S rRNA genes, and amplicons were sequenced on the MiSeq platform (Illumina, San Diego, CA). To classify bacterial taxonomy, sequences were processed using

QIIME v.1.8[26] and SILVA 1.32, as previously described but updated in comparison with the previous study[25]. The filtering strategy for operational taxonomic units included parameters in QIIME to exclude low abundant sequences, singletons and chimeras and final filtering at the genera level, in which we removed genera that appeared in <10 % of the samples[27]. Diversity indices were calculated on ComBat-adjusted[28] operational taxonomic unit counts and then exported for statistical analysis.

Statistical analysis

All statistical modelling was conducted with SAS version 9.4 software (SAS Institute Inc., Cary, NC). All phylum and genus variables had undergone ComBat-adjustment[28] to correct values across laboratory batches and centered log-ratio transformation to account for their compositional nature[29]. The HEI-2015 scores were categorised into tertiles with tertile 1 indicating the lowest and tertile 3 the highest diet quality. To examine the association of diet quality with α diversity (Shannon, Chao1 and PD Whole Tree) and relative abundance of phyla and genera, general linear models were applied to estimate covariate-adjusted least square means of gut microbial measures by HEI-2015 tertiles; trend tests using the continuous variables assessed possible dose–response relations. The regression coefficients in these models indicate the change in relative abundance of a GM component per one unit of the HEI-2015 score. All models were adjusted for ethnicity, sex, age at stool collection, BMI, alcohol intake, physical activity (<0·5 *v.* ≥0·5 h moderate/vigorous activity), total energy intake (log-transformed), smoking status and antibiotic use within the last year, but not other medications as only information on use of antifungals (1 %), laxatives (19 %) and probiotics (8 %) was requested at the time of stool collection.

Assuming a nominal type-I error of 0·05, Bonferroni corrections were applied to the analysis of the 10 phyla (0·05/10 = 0·005) and the 152 genera (0·05/152 = 0·00033) to lower the likelihood of chance associations. To analyse diet quality as assessed 20·8 (SD 1·2) and 9·3 (SD 1·4) years before stool sample collection, separate models were performed for nutritional information obtained at Q1 and Q3. In addition, the APS data set with 1685 participants who completed the QFFQ at the same time as stool sample collection were analysed. Participants who had completed Q1 and Q3 were assigned change variables (T1T1, T1T2, T1T3, T2T1, T2T2, T2T3, T3T1, T3T2 and T3T3). For instance, T1T2 indicates that the individual was in tertile 1 at Q1 and in tertile 2 at Q3. Because of the small sample size, change from Q1 to APS (*n* 1685) was not computed. To evaluate the influence of change in diet quality between Q1 and Q3, these nine categories were modelled as independent variables using the same approach as above and represented in a heat map. The 13 HEI-2015 components were plotted in a spider plot to identify changes in diet quality responsible for a decrease or increase in tertiles from Q1 to Q3.

Results

The 5936 participants at Q1 (Table 1) were equally distributed between women (52·7 %) and men (48·3 %). A total of 5280 participants completed Q3 and 1685 at APS. The respective mean ages at Q1, Q3 and APS were 54, 65 and 69 years, while the mean time periods to stool collection were 21, 9 and 0 years. Mean HEI-2015 scores increased from Q1 (67 (SD 10)) to Q3 (71 (SD 11)) and APS (72 (SD 10)). Japanese Americans comprised the majority of the study population. A larger proportion of women and normal weight individuals were classified as HEI-2015 tertile 3, whereas men, as well as overweight and obese individuals, were more likely to belong to tertile 1. The Spearman correlation coefficient between Q1 and Q3 was 0·49 ($P < 0·0001$). Mean total energy intake and alcohol consumption were lower in participants in tertile 3 than 1.

Alpha diversity

Alpha diversity as assessed by the Shannon Index (Table 2) was significantly greater across tertiles for diet assessed in Q1 ($P < 0·0001$), Q3 ($P < 0·0001$) and APS ($P < 0·02$). The difference between tertiles 1 and 3 was approximately 1 % for both Q1 and Q3. This trend was also seen for PD Whole Tree, but was only significant for diet assessed at Q1 ($P = 0·002$). There were no significant associations of the Chao1 index with diet quality at any point in time.

Phyla

Of the 10 phyla, Firmicutes had the highest mean relative abundance (54 %), followed by Bacteroidetes (40 %) and Proteobacteria (3 %) (data not shown). Although only the HEI-2015 assessed at the time of stool collection showed a significant inverse relation with Actinobacteria, all regression coefficients describing the difference in relative abundance per one unit of the HEI-2015 score were in the same direction (Table 3). Diet quality at Q1 was positively associated with the presence of Firmicutes but not with diet assessed at stool collection. A significant inverse association of the HEI-2015 with Proteobacteria was only seen for diet at stool collection.

Genera

Of the 152 genera, the following had a relative abundance of at least 1 %: *Bacteroides* (28 %), *Faecalibacterium* (9·4 %), *Parabacteroides* (1·5 %), *Ruminococcus 1* (1·5 %), *Lachnoclostridium* (1·1 %) and *Escherichia-Shigella* (1·0 %). The genera *Eubacterium xylanophilum* and *Tyzzerella* had the highest zero prevalence (26·7 % and 17·8 %). Among the 152 bacterial genera, twenty eight were significantly associated with diet as assessed in at least one of the questionnaires (online Supplemental Table S1). At the phylum (family) level, they were twenty-three families in Firmicutes (12 *Lachnospiraceae*, 9 *Ruminococcaceae*, 1 *Christensenellaceae* and 1 *Erysipelotrichaceae*), three in Bacteroidetes (*Bacteroidaceae*, *Marinifilaceae* and *Tannerellaceae*), one in Actinobacteria (*Coriobacteriaceae*) and one in Proteobacteria (*Enterobacteriaceae*).

Specifically, twelve genera were significantly associated with diet quality assessed at all three time points as shown by the regression coefficients for relative abundance per one unit of the HEI-2015 score (Table 4), while six genera were significantly

Table 1. Characteristics of the study population by tertiles of the HEI-2015 at overtime*
(Mean values and standard deviations)

	Q1 (1993–1996)				Q3 (2003–2008)				APS (2013–2016)			
Category	All	T1†	T2	T3	All	T1†	T2	T3	All	T1†	T2	T3
Range of index scores		31·7–62·3	62·3–72·0	72·0–96·5		33·4–66·7	66·7–76·7	76·7–99·9		35·4–67·7	67·7–76·4	76·4–98·9
N		1964	1975	1997		1743	1773	1764		561	562	562
Age at Qx, y	5936				5280							
Mean		52·2	53·7	55·3		64·3	65·4	66·2				
SD		6·2	6·8	7·1		6·7	6·9	7·0				
Age at stool sample, y	5936				5280				1685			
Mean		73·0	74·5	76·2		73·7	74·8	75·5		69·2	69·5	69·1
SD		6·3	6·8	7·1		6·7	7·0	7·0		2·8	2·7	2·7
Time to stool collection, y	5936				5280							
Mean		20·8	20·8	20·9		9·4	9·3	9·3				
SD		1·3	1·2	1·2		1·4	1·4	1·4				
Sex, %	5936				5280				1685			
M	2866	39	34	27	2566	38	34	28	831	39	32	30
F	3070	27	33	40	2714	28	33	39	854	28	35	37
Ethnicity, %												
Whites	1004	23	33	44	920	27	36	37	380	28	33	38
African American	935	23	32	46	753	26	32	42	272	28	36	36
Native Hawaiian	793	37	32	31	684	38	30	31	278	38	30	33
Japanese American	2087	39	32	29	1973	33	33	34	413	33	35	32
Latino	1117	38	38	25	950	40	36	24	342	40	32	27
BMI (kg/m^2), %												
Normal (18·5–24·9)	2601	29	33	38	2025	26	32	42	500	23	30	46
Overweight (25–29·9)	2350	36	33	31	2021	35	35	30	675	32	37	31
Obese (30+)	985	37	36	28	1074	42	33	25	510	45	32	24
Moderate/vigorous active, %												
<0·5 h/d	1626	37	33	30	1003	41	33	26	335	47	27	26
0·5+ h/d	4140	31	33	36	4146	31	34	36	1350	30	35	35
Smoking Status, %												
Never	2845	28	34	37	2484	28	33	38	1030	30	34	36
Ever‡	3045	38	32	30	2689	38	33	29	655	38	33	29
Antibiotic use, %	1119	34	33	34	986	31	38	31	366	30	36	34
Total energy, kcal/d												
Mean		2321	2251	2097		1890	1892	1827		1931	1851	1778
SD		1086	1017	919		841	840	744		941	862	790
Alcohol intake, g/d												
Mean		7·9	8·0	8·1		6·4	7·8	6·6		7·1	8·3	7·2
SD		19·1	1017	19·0		14·0	17·5	13·9		17·7	16·1	14·3

* Healthy Eating Index-2015 (HEI-2015).
† T1 = lowest or first tertile, T2 = intermediate or second tertile, T3 = highest or third tertile.
‡ Individuals who were either past-smokers or current smokers.

British Journal of Nutrition

Table 2. Mean (95 % CI) for diversity measures by tertiles of HEI-2015 over time*
(Mean values and 95 % confidence intervals)

Measure	Tertile†	Q1 (1993–1996)			Q3 (2003–2008)			APS (2013–2016)		
		Mean	95 % CI	P_{trend}‡	Mean	95 % CI	P_{trend}‡	Mean	95 % CI	P_{trend}‡
Shannon Index	T1	6·44	6·27, 6·61	<0·0001	6·46	6·29, 6·63	0·0001	6·36	6·22, 6·49	0·02
	T2	6·51	6·34, 6·68		6·53	6·36, 6·70		6·43	6·29, 6·57	
	T3	6·53	6·36, 6·71		6·55	6·38, 6·73		6·44	6·30, 6·57	
Chao 1	T1	1061	1027, 1095	0·08	1044	1009, 1078	0·85	1043	1013, 1072	0·78
	T2	1072	1038, 1106		1047	1013, 1082		1046	1017, 1075	
	T3	1068	1034, 1103		1040	1005, 1075		1037	1007, 1067	
PD Whole Tree	T1	24·7	23·8, 25·6	0·002	24·5	23·6, 25·4	0·38	23·7	22·9, 24·5	0·60
	T2	24·9	24·0, 25·7		24·6	23·7, 25·5		23·8	23·1, 24·6	
	T3	24·9	24·0, 25·8		24·5	23·7, 25·5		23·4	22·6, 24·1	

* Healthy Eating Index-2015 (HEI-2015) as independent variable; adjusted means obtained by linear regression for diversity indices; values shown are as means (95% CI) adjusted for sex, age at stool collection in 2013–2016, ethnicity, BMI (kg/m^2), total energy intake (log-transformed), physical activity (hours/day), smoking status (never/ever), antibiotic use in past year (yes/no) and alcohol intake
† T1=lowest or first tertile, T2=intermediate or second tertile, T3=highest or third tertile
‡ P_{trend} obtained with HEI-2015 as continuous variable and adjusted for sex, age at stool collection, ethnicity, BMI, total energy intake (log-transformed), physical activity, smoking status, antibiotic use in past year and alcohol intake

related to diet from two questionnaires and ten with one (online Supplemental Table S1). However, the regression coefficients were in the same direction and of similar magnitude across time even when statistical significance was not reached. For ten out of the twelve significant genera, the regression coefficients were greatest at APS with 12–88 % lower values at Q1 and intermediate estimates at Q3. The exception was *Escherichia-Shigella:* the inverse association was only significant for diet as assessed at stool collection, and the coefficient was greater at that time than with diet at Q1 and Q3 (−0·023 *v.* −0·008 and −0·004). Of the twelve consistent bacterial genera (Table 4), seven (*Anaerostipes*, *Coprococcus 2*, *Eubacterium eligens* group, *Lachnospira*, *Lachnospiraceae ND3007* group, *Ruminococcaceae UCG-013* and *Ruminococcus 1*) showed positive associations and five (*Collinsella*, *Parabacteroides*, *Ruminiclostridium 5*, *Ruminococcus gnavus* group and *Tyzzerella*) showed inverse associations with diet quality.

Change in diet quality

Of the 5280 participants who completed Q1 and Q3, 2849 (54 %) remained in the same tertile of diet quality, 1225 (23 %) shifted to a lower and 1206 (23 %) to a higher tertile. These proportions were very similar across ethnic groups with 53–55 % of cohort members remaining in the same tertiles, 19–28 % decreasing and 19–26 % increasing. Looking at change in diet quality between Q1 and Q3 (Fig. 1), the associations across the nine categories from T1T1 to T3T3 with T1T1 as reference became stronger compared with the cross-sectional associations as evaluated by the regression estimates. The magnitude of changes became smaller for the five inverse associations (*Collinsella, Lachnospira, Parabacteroides, Ruminiclostridium 5, Ruminococcus gnavus, Tyzzerella*) and gradually greater for the seven positive associations (*Anaerostipes*, *Coprococcus 2*, *Eubacterium eligens*, *Lachnospira*, *Lachnospiraceae ND3007*, *Ruminococcaceae UCG-013*, *Ruminococcus 1*). As indicated in the heat map (Fig. 1), being in the highest tertile for Q3 predicted strong positive associations for *Coprococcus 2*, *Eubacterium eligens* group and *Lachnospira* and inverse associations for *Ruminococcus gnavus* group and *Tyzzerella* even for participants with lower diet quality at Q1.

As to the HEI-2015 components responsible for a change in tertiles, participants who moved to a higher tertile from Q1 to Q3 (Fig. 2A) showed stronger adherence to the whole grains and refined grains recommendations but also improved their scores for the total vegetables, total fruits, greens and beans and dairy categories. On the other hand, participants with a lower tertile at Q3 than Q1 (Fig. 2B) showed worse adherence to the saturated fat, fatty acids and whole grain components.

Discussion

In the current analysis, higher diet quality measured at three points in time over approximately 30 years was associated with greater α diversity as indicated by the Shannon index although the absolute differences across tertiles of the HEI-2015 were small (around 1 %). After Bonferroni adjustment, the abundance

Table 3. Phyla associated with HEI-2015 among participants at 3 points in time*

	Q1 (1993–1996)		Q3 (2003–2008)		APS (2013–2016)	
Phylum	β†	P_{trend}‡	β†	P_{trend}‡	β†	P_{trend}‡
Actinobacteria	–0·0006	0·72	–0·004	0·009	–0·013	0·0001
Bacteroidetes	0·0000	0·97	–0·002	0·05	–0·006	0·01
Cyanobacteria	0·001	0·61	0·0007	0·77	0·005	0·29
Firmicutes	0·004	<0·0001	0·001	0·36	–0·001	0·57
Fusobacteria	–0·002	0·50	–0·001	0·69	0·004	0·45
Lentispaerae	–0·004	0·10	–0·001	0·67	0·003	0·54
Proteobacteria	–0·0007	0·72	0·0001	0·95	–0·01	0·003
Synergistetes	0·0001	0·96	0·002	0·45	0·009	0·06
Tenericutes	0·002	0·51	0·004	0·15	0·008	0·14
Verrucomicrobia	–0·0002	0·94	0·0009	0·75	0·003	0·55

* Healthy Eating Index-2015 (HEI-2015) as independent variable.
† Regression parameter obtained by linear regression for 10 phyla with Bonferroni adjusted significance ($P < 0·005$); values shown are as means (95 % CI) adjusted for sex, age at stool collection in 2013–2016, ethnicity, BMI (kg/m^2), total energy intake (log-transformed), physical activity (hours/day), smoking status (never/ever), antibiotic use in past year (yes/no) and alcohol intake.
‡ P_{trend} obtained with HEI-2015 as continuous variable and adjusted for sex, age at stool collection, ethnicity, BMI, total energy intake (log-transformed), physical activity, smoking status, antibiotic use in past year and alcohol intake.

Table 4. Genera associated with HEI-2015 across all three questionnaires*

			Q1 (1993–1996)		Q3 (2003–2008)		APS (2013–2016)	
Genus	*R*† (%)	*M*‡ (%)	β§	P_{trend}	β§	P_{trend}	β§	P_{trend}
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Anaerostipes	0·8	0·02	0·0098	<0·0001	0·0069	<0·0001	0·0084	0·0001
Actinobacteria; Coriobacteriia; Coriobacteriales; Coriobacteriaceae; Collinsella	0·6	0·7	–0·0080	<0·0001	–0·0084	<0·0001	–0·0151	<0·0001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Coprococcus_2	0·4	0·4	0·0099	<0·0001	0·0112	<0·0001	0·0175	<0·0001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; [Eubacterium]_eligens_group	0·7	0·3	0·0191	<0·0001	0·0225	<0·0001	0·0259	<0·0001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Lachnospira	0·3	2·0	0·0178	<0·0001	0·0175	<0·0001	0·0162	<0·0001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Lachnospiraceae_ND3007_group	0·3	2·3	0·0102	<0·0001	0·0111	<0·0001	0·0127	0·0001
Bacteroidetes; Bacteroidia; Bacteroidales; Tannerellaceae; Parabacteroides	1·5	0·02	–0·0074	<0·0001	–0·0097	<0·0001	–0·0140	<0·0001
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminiclostridium_5	0·5	0·02	–0·0058	<0·0001	–0·0084	<0·0001	–0·0108	<0·0001
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminococcaceae_UCG-013	0·2	2·3	0·0084	<0·0001	0·0077	<0·0001	0·0123	0·0001
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminococcus_1	1·5	0·0	0·0095	<0·0001	0·0084	<0·0001	0·0108	<0·0001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; [Ruminococcus] gnavus_group	0·1	9·1	–0·0124	<0·0001	–0·0133	<0·0001	–0·0143	0·0002
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Tyzzerella	0·1	17·8	–0·0103	<0·0001	–0·0117	<0·0001	–0·0166	<0·0001

* Healthy Eating Index-2015 (HEI-2015) as independent variable; regression parameters obtained by linear regression for 152 genera with Bonferroni adjusted significance ($P < 0·00033$).
† Relative abundance of the genera based on the total sum of the bacteria.
‡ Percent zero prevalence of genera.
§ β = regression parameter and P_{trend} from trend tests with HEI-2015 as continuous variable and adjusted for sex, age at stool collection, ethnicity, BMI, total energy intake (log-transformed), physical activity, smoking status, antibiotic use in past year and alcohol intake.

of twenty-eight genera was significantly associated with diet assessed in at least one questionnaire; the majority belonged to the *Lachnospiraceae* or *Ruminococcaceae* families of the *Clostridiales* order (online Supplemental Table S1). Of these, twelve genera were associated with diet quality as assessed at all three time points (Table 4): *Anaerostipes*, *Coprococcus 2*, *Eubacterium eligens* group, *Lachnospira*, *Lachnospiraceae ND3007* group, *Ruminococcaceae UCG-013*, *Ruminococcus*, *Collinsella*, *Parabacteroides*, *Ruminiclostridium 5*, *Ruminococcus gnavus* group and *Tyzzerella*. The fact the strength of the association as assessed by the regression coefficient for ten of the significant twelve genera were remained

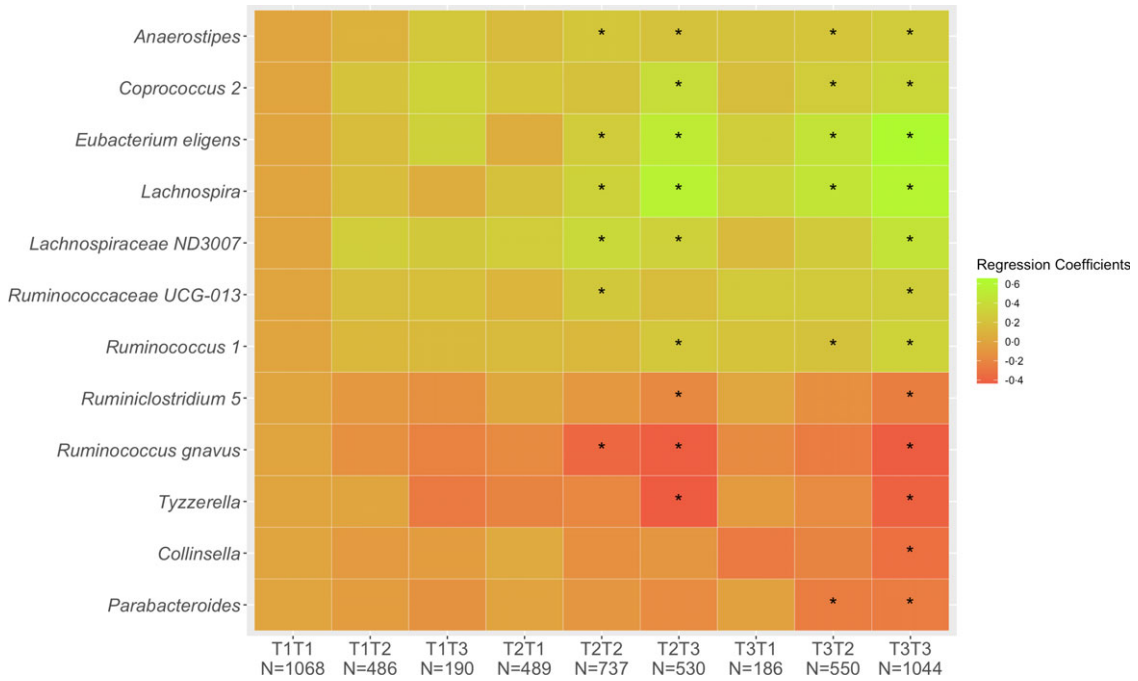

Fig. 1. Associations of 12 genera with tertiles of HEI-2015 at Q1 and Q3. Values shown are regression coefficients obtained by multiple linear regression, which indicate the difference in relative abundance of each genus per category of combined HEI-2015 (T1T1, etc.) with T1T1 as reference. The β estimates were adjusted for sex, age at stool collection, ethnicity, BMI, total energy intake (log-transformed), physical activity, smoking status, antibiotic use in past year and alcohol intake. The asterisk symbol (*) indicates significance ($P < 0{\cdot}05$).

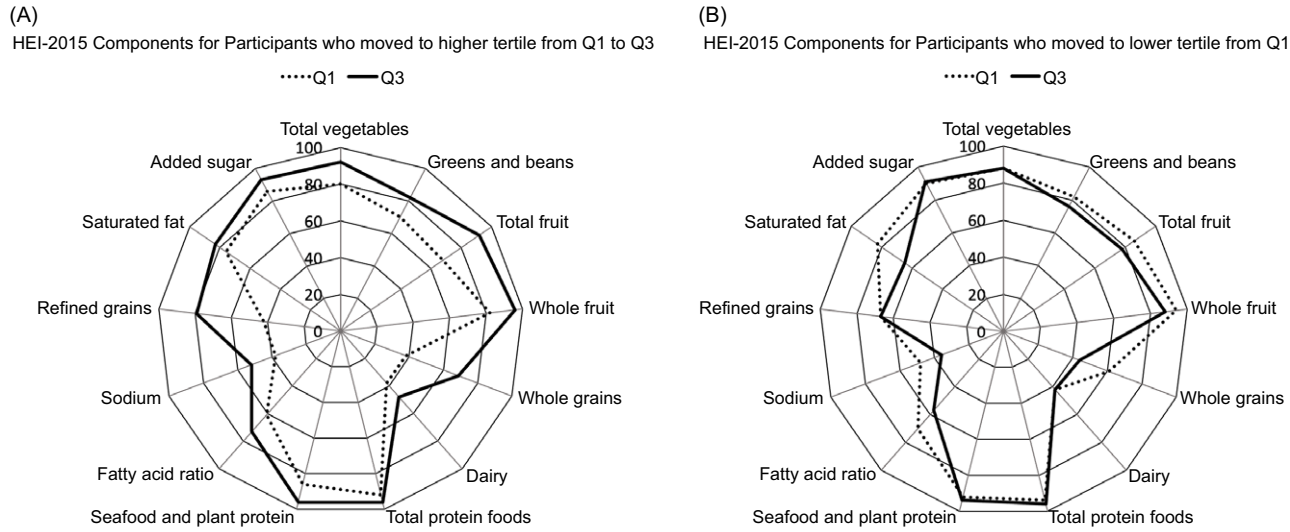

Fig. 2. Radar graphs representing the 13 component scores of the Healthy Eating Index-2015 (HEI-2015) in Q1 and Q3 among participants who increased (A) and decreased (B) diet quality from Q1 to Q3. The percentages represent the mean HEI-2015 score divided by the maximum score per category.

quite stable supports the importance of long-term dietary exposure in shaping gut microbial composition although diet close to stool collection shows a stronger association.

Combined diet quality as assessed in Q1 and Q3 (Fig. 1) indicated trends, positive for some and inverse for other bacteria, in the abundance of the twelve genera for participants reporting higher diet quality at both time points and in particular when closer to stool collection as compared with those with low HEI-2015 scores at Q1 and Q3. The availability of repeated dietary information and the analysis of change in diet can be considered novel in comparison with previous publications reporting on diet quality and the gut microbiome[8,9,11–13].

The current findings are consistent with a previous MEC report that found α diversity to be significantly associated with better diet quality[8]. However, the importance of the very small change or lack of change in α diversity seen here and in previous reports[8,9,30,31] is not clear. A review of dietary intervention studies also detected a limited effect on microbial diversity across

studies and suggested that an increase in microbial diversity in already healthy people may not be as important as among diseased individuals[32].

The fibre-fermenting bacteria of the phylum Firmicutes, such as *Lachnospira* and *Ruminococcus*, identified here and in the previous report[8], are able to metabolise dietary plant polysaccharides to produce SCFA, such as butyrate, propionate and acetate, which act as a major energy source for intestinal epithelial cells and strengthen the mucosal barrier[33], possibly with beneficial effects on glucose metabolism and inflammation[9]. *Lachnospira* is one of the core genera of the GM and the only Firmicutes besides *Faecalibacterium* and *Eubacterium* with the ability to degrade pectins[11].

The greater abundance of members from the *Lachnospira/Lachnospiraceae ND3007* group is also consistent with previous reports showing a positive relation with a health-conscious diet characterised by fibre-rich and plant-based food[9], a vegetarian pattern[11], a prudent dietary pattern[34] and a validated index of diet quality[30]. Similarly, the bacteria *Ruminococcaceae UCG-013/Ruminococcus* were more abundant among individuals with a fibre-rich diet[35] and on a high-quality diet[30]. However, in two other reports, their abundance was higher for consumers of a Western pattern, which is characterised by high consumption of processed meats, refined grains, potatoes, eggs, sweets and salty snacks[34], omnivores[11], an animal-based diet[36] and a pro-inflammatory diet[37]. The differences across studies may be due to the fact that particular species within a particular genus have different functions and that specific bacteria use different types of dietary fibres as substrates for SCFA production[38,39]. This suggests that variation in fibre intake can influence the selection of bacteria with similar functions[9] and may explain why within the *Lachnospiraceae* group, *E. eligens* and *Lachnospira* were both positively associated with diet quality, but *Roseburia* was not seen as one of the significant genera as in other reports[9].

The positive association of diet quality with the *Eubacterium eligens* group agrees with its higher abundance among those with a vegetarian diet in Thailand[40]. Although we are not aware of any connections with diet for *Ruminococcus gnavus*, it is known that this organism is involved in non-alcoholic fatty liver disease[41]. Also diets rich in animal products have been associated with *Collinsella* and *Ruminococcus gnavus*, two pro-inflammatory genera[42]. For *Ruminiclostridium 5*, which was inversely associated with diet quality, the only report in humans we identified suggested that the relative abundance of *Ruminiclostridium_9* increased during a whole wheat diet as compared with a refined wheat diet[43].

In addition, *Anaerostipes*, a fermenting organism, was positively associated with the aMED and DASH in our previous report[8] and with different diet indices in a recent report from the Personalized Responses to Dietary Composition Trial (PREDICT 1) study[14]. *Coprococcus 2*, also a fermenting organism, was positively associated with the DASH in our previous report[8], a validated index of diet quality among pregnant women with excess weight[30] and a vegetarian diet in Thailand[40]. The inverse relation to a Western diet among older community-dwelling men[34] is also consistent, but the inverse association with a plant-based pattern in a dietary modification trial contradicts the idea of its beneficial influence[36]. Similar to in this report, a higher abundance of *Tyzzerella*, which has been associated with cardiovascular disease[44], was seen for lower scores of the HEI[10].

The only significant member of the phylum Actinobacteria, *Collinsella* was inversely associated with diet across time periods in the current analysis and with higher scores of four indices (HEI-2010, AHEI, DASH and aMED) among APS participants[8]. A prudent pattern, characterised by high consumption of fruits, vegetables, nuts, fish and chicken and turkey, among older community-dwelling men showed similar findings[34]. *Collinsella* is known to be associated with low fibre diets[38] and result in inflammatory status[42] , fatty liver[45] and related diseases[46]. *Escherichia-Shigella* was associated with diet quality at the time of stool collection only (online Supplemental Table S1). This may be due to the higher mean age of participants who participated in the APS as it is well known that ageing increases the likelihood to acquire these bacteria due to lower immunity or medication use[47–49].

Parabacteroides, the only significant Bacteroidetes, is involved in bile acid metabolism, fat degradation and protein digestion and showed contradictory findings. Our inverse relation agrees with a report that *Parabacteroides* were part of a cluster that increased on a diet rich in animal foods[36] and an inverse association with a vegetarian diet in Thailand[40]. However, their abundance was greater with a higher HEI-2005 score in colonic mucosa-associated GM among healthy individuals[10]. *Bacteroides* are known to be associated with consumption of animal protein, a variety of amino acids and saturated fat, which represent a more Westernised diet[3].

An increasing number of studies have indicated that dietary components may influence or shift the composition of the GM with the potential of beneficial health effects, such as prevention of CVD and obesity, potentially through alteration in host immunity and metabolic activity[33]. As shown in the current analysis, a high-quality diet, as assessed by the HEI-2015, maintained over many years predicted a GM composition that included several bacteria that have been associated with favourable health effects. As diet quality closer to the time of stool collection had a stronger association with the GM, changes in dietary quality even at an older age may be beneficial[50,51]. However, the current results suggest that a consistently high-quality diet over time predicts the most favourable pattern of gut bacteria. Although the number of years between diet quality assessment and stool collection varied, the strength of the associations remained similar across the three questionnaires, supporting the long-term effects of diet on the characteristics of the GM and the stability of dietary habits over time. Alternatively, a certain diet quality at a younger age may be formative in the selection of bacteria that will remain part of a person's GM[52,53]. This idea is supported by reports that sudden changes in diet temporarily modify the composition and diversity of the gut microbiota before reverting to its original state[2,54]. Although the microbial composition changed in ten subjects randomly controlled for either a high-fat/low-fibre or low-fat/high-fibre diet for 10 days, the dominant genera remained stable over time indicating its strong association with long-term diet[3].

The novelty of the current report lies in the long-term observation with repeated assessments of diet quality and the ability to examine changes in diet quality. The larger sample size and the

updated reference libraries partially confirm the previous findings[8] but also make comparisons between the two reports challenging as the different taxonomic classifications of individual bacteria are not analogous. The choice of the HEI-2015 provides a valid and widely used measure of overall diet quality shown to be related to mortality[7,55], type 2 diabetes[56] and colorectal cancer risk[57]. However, several limitations have to be considered in the interpretation of the current findings. Foremost, the one-time collection of stool samples that may have been affected by concurrent disease[58] or other events, the known measurement errors of FFQ[59] and the different sample sizes at each point in time. As the important influence of different medications has been documented[14,60], our lack of detailed use at stool collection may have confounded the current findings. Applying 16S sequencing instead of whole genome shotgun sequencing restricted our ability to identify bacteria at the species level. For a better understanding of metabolic mechanism involved in the relation between diet quality and the GM, the imputation of functional pathways with PICRUSt[61] or similar tools will provide an assessment of community function and insights into predictive metagenomics in future studies.

The most remarkable finding of the current analysis is the stability of the associations of diet quality as assessed many years before stool collection with characteristics of the GM. Participants who improved their diet quality scores between Q1 and Q3 showed microbial patterns that were comparable to individuals with high HEI-2015 scores at Q3. The quality of past diet, assessed as far as 21 years before stool collection, was equally predictive of GM composition as concurrently assessed diet, indicative of the long-term nature of this relation.

Acknowledgements

We thank the participants for their time to participate in the study.

This work was supported by US National Cancer Institute (NCI) grants P01 CA168530, U01 CA164973, P30 CA071789, P30 CA015704 and UL1TR000130.

Author contributions were as follows: conceptualisation: U. L., L. L. M., J. W. L., M. A. J. H., C. J. B., and G. M.; data curation: L. R. W., K. R. C., I. C. J., and T. W. R.; formal analysis: E. M., G. M.; funding acquisition: L. L. M., L. R. W., M. A. J. H., and J. W. L.; investigation: E. M. and G. M.; methodology: U. L., C. J. B, and T. W. R.; project administration: U. L., L. L. M.; supervision: G. M.; visualisation: E. M.; writing – original draft: E. M.; and writing – review and editing: C. J. B., U. L., L. R. W., V. W. S., M. A. J. H., J. W. L., T. W. R., and L. L. M.

All authors reviewed and approved the final version of the manuscript.

None of the authors reported a conflict of interest.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S0007114521002968

References

1. Morgan XC & Huttenhower C (2012) Chapter 12: human microbiome analysis. *PLoS Comput Biol* **8**, e1002808.
2. Fu BC, Randolph TW, Lim U, *et al.* (2019) Temporal variability and stability of the fecal microbiome: the multiethnic cohort study. *Cancer Epidemiol Biomarkers Prev* **28**, 154–162.
3. Wu GD, Chen J, Hoffmann C, *et al.* (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108.
4. Fabiani R, Naldini G , Chiavarini M (2019) Dietary patterns and metabolic syndrome in adult subjects: a systematic review and meta-analysis. *Nutrients* **11**(9), 2056.
5. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
6. US Department of Health Human Services (2015) 2015–2020 Dietary guidelines for Americans. Washington (DC): USDA.
7. Panizza CE, Shvetsov YB, Harmon BE, *et al.* (2018) Testing the predictive validity of the healthy eating index-2015 in the multiethnic cohort: is the score associated with a reduced risk of all-cause and cause-specific mortality? *Nutrients* 10.
8. Maskarinec G, Hullar MAJ, Monroe KR, *et al.* (2019) Fecal microbial diversity and structure are associated with diet quality in the multiethnic cohort adiposity phenotype study. *J Nutr* **149**, 1575–1584.
9. Ericson U, Brunkwall L, Hellstrand S, *et al.* (2020) A health-conscious food pattern is associated with prediabetes and gut microbiota in the Malmo offspring study. *J Nutr* **150**, 861–872.
10. Liu Y, Ajami NJ, El-Serag HB, *et al.* (2019) Dietary quality and the colonic mucosa-associated gut microbiome in humans. *Am J Clin Nutr* **110**, 701–712.
11. De Filippis F, Pellegrini N, Vannini L, *et al.* (2016) High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut* **65**, 1812–1821.
12. Garcia-Mantrana I, Selma-Royo M, Alcantara C, *et al.* (2018) Shifts on gut microbiota associated to Mediterranean diet adherence and specific dietary intakes on general adult population. *Front Microbiol* **9**, 890.
13. Mitsou EK, Kakali A, Antonopoulou S, *et al.* (2017) Adherence to the Mediterranean diet is associated with the gut microbiota pattern and gastrointestinal characteristics in an adult population. *Br J Nutr* **117**, 1645–1655.
14. Asnicar F, Berry SE, Valdes AM, *et al.* (2021) Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat Med* **27**, 321–332.
15. Kolonel LN, Henderson BE, Hankin JH, *et al.* (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* **151**, 346–357.
16. Park SY, Shvetsov YB, Kang M, *et al.* (2020) Changes in diet quality over 10 years are associated with baseline sociodemographic and lifestyle factors in the multiethnic cohort study. *J Nutr* **150**, 1880–1888.
17. Lim U, Monroe KR, Buchthal S, *et al.* (2019) Propensity for intra-abdominal and hepatic adiposity varies among ethnic groups. *Gastroenterology* **156**, 966–975.
18. Fu BC, Randolph TW, Lim U, *et al.* (2016) Characterization of the gut microbiome in epidemiologic studies: the multiethnic cohort experience. *Ann Epidemiol* **26**, 373–379.
19. Jakobsson HE, Jernberg C, Andersson AF, *et al.* (2010) Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS One* **5**, e9836.
20. O'Brien CL, Allison GE, Grimpen F, *et al.* (2013) Impact of colonoscopy bowel preparation on intestinal microbiota. *PLoS One* **8**, e62815.

21. Stram DO, Hankin JH, Wilkens LR, *et al.* (2000) Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol* **151**, 358–370.
22. Murphy SP (2002) Unique nutrition support for research at the Cancer Research Center of Hawaii. *Hawaii Med J* **61**, 17.
23. Guenther PM, Casavale KO, Reedy J, *et al.* (2013) Update of the healthy eating index: HEI-2010. *J Acad Nutr Diet* **113**, 569–580.
24. Guenther PM, Reedy J , Krebs-Smith SM (2008) Development of the healthy eating index-2005. *J Am Diet Assoc* **108**, 1896–1901.
25. Fu BC, Randolph TW, Lim U, *et al.* (2019) Temporal variability and stability of the fecal microbiome: the Multiethnic Cohort Study. *Cancer Epidemiol Biomarkers Prev* **28**, 154–162.
26. Caporaso JG, Kuczynski J, Stombaugh J, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336.
27. Langille MGI, Zaneveld J, Caporaso JG, *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**, 814–821.
28. Zhang Y, Parmigiani G , Johnson WE (2020) ComBat-Seq: batch effect adjustment for RNA-Seq count data. *bioRxiv*, 202020012013904730.
29. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, *et al.* (2017) Microbiome datasets are compositional: and this is not optional. *Front Microbiol* **8**, 2224.
30. Laitinen K & Mokkala K (2019) Overall dietary quality relates to gut microbiota diversity and abundance. *Int J Mol Sci* **20**, 1835.
31. Yu D, Nguyen SM, Yang Y, *et al.* (2021) Long-term diet quality is associated with gut microbiome diversity and composition among urban Chinese adults. *Am J Clin Nutr* **113**, 684–694.
32. Willis HJ & Slavin JL (2020) The influence of diet interventions using whole, plant food on the gut microbiome: a narrative review. *J Acad Nutr Diet* **120**, 608–623.
33. Singh RK, Chang HW, Yan D, *et al.* (2017) Influence of diet on the gut microbiome and implications for human health. *J Transl Med* **15**, 73.
34. Shikany JM, Demmer RT, Johnson AJ, *et al.* (2019) Association of dietary patterns with the gut microbiota in older, community-dwelling men. *Am J Clin Nutr* **110**, 1003–1014.
35. So D, Whelan K, Rossi M, *et al.* (2018) Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. *Am J Clin Nutr* **107**, 965–983.
36. David LA, Maurice CF, Carmody RN, *et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563.
37. Zheng J, Hoffman KL, Chen JS, *et al.* (2020) Dietary inflammatory potential in relation to the gut microbiome: results from a cross-sectional study. *Br J Nutr* **124**, 931–942.
38. Adamberg K, Jaagura M, Aaspollu A, *et al.* (2020) The composition of faecal microbiota is related to the amount and variety of dietary fibres. *Int J Food Sci Nutr* **71**, 845–855.
39. Makki K, Deehan EC, Walter J, *et al.* (2018) The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* **23**, 705–715.
40. Ruengsomwong S, La-Ongkham O, Jiang J, *et al.* (2016) Microbial community of healthy Thai vegetarians and non-vegetarians, their core gut microbiota, and pathogen risk. *J Microbiol Biotechnol* **26**, 1723–1735.
41. Alferink LJ, Radjabzadeh D, Erler NS, *et al.* (2020) Microbiomics, metabolomics, predicted metagenomics and hepatic steatosis in a population-based study of 1355 adults. *Hepatology*.
42. van Soest APM, Hermes GDA, Berendsen AAM, *et al.* (2020) Associations between pro- and anti-inflammatory gastro-intestinal microbiota, diet, and cognitive functioning in Dutch healthy older adults: the NU-AGE study. *Nutrients* 12.
43. van Trijp MPH, Schutte S, Esser D, *et al.* (2020) Minor changes in the composition and function of the gut microbiota during a 12-week whole grain wheat or refined wheat intervention correlate with liver fat in overweight and obese adults. *J Nutr*.
44. Kelly TN, Bazzano LA, Ajami NJ, *et al.* (2016) Gut microbiome associates with lifetime cardiovascular disease risk profile among Bogalusa heart study participants. *Circ Res* **119**, 956–964.
45. Astbury S, Atallah E, Vijay A, *et al.* (2020) Lower gut microbiome diversity and higher abundance of proinflammatory genus Collinsella are associated with biopsy-proven nonalcoholic steatohepatitis. *Gut Microbes* **11**, 569–580.
46. Menni C, Louca P, Berry SE, *et al.* (2021) High intake of vegetables is linked to lower white blood cell profile and the effect is mediated by the gut microbiome. *BMC Med* **19**, 37.
47. Brussow H (2013) Microbiota and healthy ageing: observational and nutritional intervention studies. *Microb Biotechnol* **6**, 326–334.
48. Lakshminarayanan B, Harris HMB, Coakley M, *et al.* (2013) Prevalence and characterization of *Clostridium perfringens* from the faecal microbiota of elderly Irish subjects. *J Med Microbiol* **62**, 457–466.
49. O'Sullivan O, Coakley M, Lakshminarayanan B, *et al.* (2013) Alterations in intestinal microbiota of elderly Irish subjects post-antibiotic therapy. *J Antimicrob Chemother* **68**, 214–221.
50. Wilmanski T, Diener C, Rappaport N, *et al.* (2021) Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat Metab* **3**, 274–286.
51. Jeffery IB, Lynch DB , O'Toole PW (2016) Composition and temporal stability of the gut microbiota in older persons. *ISME J* **10**, 170–182.
52. Tamburini S, Shen N, Wu HC, *et al.* (2016) The microbiome in early life: implications for health outcomes. *Nat Med* **22**, 713–722.
53. Sarkar A, Yoo JY, Valeria Ozorio Dutra S, *et al.* (2021) The association between early-life gut microbiota and long-term health and diseases. *J Clin Med* 10.
54. Valdes AM, Walter J, Segal E, *et al.* (2018) Role of the gut microbiota in nutrition and health. *BMJ* **361**, k2179.
55. Harmon BE, Boushey CJ, Shvetsov YB, *et al.* (2015) Associations of key diet-quality indexes with mortality in the Multiethnic Cohort: the Dietary Patterns Methods Project. *Am J Clin Nutr* **101**, 587–597.
56. Jacobs S, Harmon BE, Boushey CJ, *et al.* (2015) A priori-defined diet quality indexes and risk of type 2 diabetes: the Multiethnic Cohort. *Diabetologia* **58**, 98–112.
57. Park SY, Boushey CJ, Wilkens LR, *et al.* (2017) High-quality diets associate with reduced risk of colorectal cancer: analyses of diet quality indexes in the multiethnic cohort. *Gastroenterology* **153**, 386–394.e382.
58. Allaband C, McDonald D, Vazquez-Baeza Y, *et al.* (2019) Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. *Clin Gastroenterol Hepatol* **17**, 218–230.
59. Freedman LS, Schatzkin A, Midthune D, *et al.* (2011) Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst* **103**, 1086–1092.
60. Weersma RK, Zhernakova A , Fu J (2020) Interaction between drugs and the gut microbiome. *Gut* **69**, 1510–1519.
61. Langille MG, Zaneveld J, Caporaso JG, *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**, 814–821.