Taxonomic composition and diversity of the gut microbiota in relation to habitual diet in Korean adults

Hwayoung Noh
International Agency for Research on Cancer
Hwan-Hee Jang
Rural Development Administration
Gichang Kim
Rural Development Administration
Semi Zouiouich
International Agency for Research on Cancer
Su-Yeon Cho
Rural Development Administration
Hyeon-Jeong Kim
Rural Development Administration
Jeongseon Kim
National Cancer Center
Jeong-Sook Choe
Rural Development Administration
Marc J. Gunter
International Agency for Research on Cancer
Pietro Ferrari
International Agency for Research on Cancer
Augustin Scalbert
International Agency for Research on Cancer
Heinz Freisling (freislingh@iarc.fr)
International Agency for Research on Cancer https://orcid.org/0000-0001-8648-4998

Research

Keywords: Human gut microbiota, Enterotypes, Habitual diet, Dietary pattern, Korean adults

Posted Date: February 8th, 2020

DOI: https://doi.org/10.21203/rs.2.22869/v1

Abstract

Background: Little is known of the relationship between the Korean habitual diet and gut microbiota composition. We investigated associations of habitual dietary intake of foods and nutrients with the taxonomic composition and diversity of gut microbiota in 222 Korean adults aged 18-58 years in a cross-sectional study. Gut microbial taxonomic composition and diversity data were obtained by 16S rRNA gene sequencing of bacterial DNA extracted from fecal samples. Habitual diet for the previous year was collected by a validated food frequency questionnaire. Correlations between intakes of food and nutrients and gut microbial taxonomic composition were examined with adjustment for sex, age, body mass index, dietary supplement, smoking status, and sample batch. Specific dietary patterns associated with α-diversity were identified by reduced rank regression. Enterotypes of gut microbiota were explored by principal coordinate analysis based on β-diversity.

Results: The intakes of vegetables, fermented legumes, and potatoes were positively associated with the *Firmicutes-to-Bacteroidetes* (F/B) ratio, while the intakes of noodle products and non-alcoholic beverages were inversely related to the *F/B* ratio (all $P<0.05$). A dietary pattern associated with higher α-diversity (HiαDP) was characterized by greater intakes of fermented legumes, vegetables, potatoes, tea, and fruit/fruit juice and lower intakes of non-alcoholic beverages. Among three different enterotypes identified based on the β-diversity, the *Ruminococcus* enterotype had higher scores of the HiαDP and was more strongly associated with intakes of vegetables and nuts/seeds, compared to the two other enterotypes.

Conclusions: We conclude that the habitual diet in Korean adults was associated with gut microbial taxonomic composition and diversity. A higher intake of plant-based and fermented foods was associated with distinct gut microbial enterotypes in Korean adults.

Background

The human gut microbiome is a complex community consisting of $10^{13} \sim 10^{14}$ microorganisms, dominated by bacteria, which inhabit the human gastrointestinal tract [1]. The volume of the collective microbial genome is over 100 times larger than the human genome [1, 2]. In a symbiotic relationship with the host, the gut microbiota contributes to numerous physiological processes, such as modulating the intestinal barrier [3], regulating energy metabolism [4, 5], protection against pathogens [6], and regulating the immune system [7].

Host dietary intake is one of the main factors that can modulate the taxonomic composition and diversity of the gut microbiota, which could, in turn, promote either beneficial or detrimental consequences on host health through alterations of the physiological functions of the gut microbiota [8–10]. Diets rich in animal-based foods such as the “Western-style diet” increase the levels of bile-tolerant bacteria including Bacteroidetes (e.g. Bacteroides and Alistipes), Proteobacteria (Bilophila) and decrease levels of fiber-degrading bacteria such as Firmicutes (e.g. Eubacterium and Ruminococccus) [11, 12]. In contrast, plant-

based diets such as the “Mediterranean diet” high in dietary fiber promote fiber-degrading bacteria, mainly including genera of the Firmicutes phylum, with increased overall diversity of the gut microbiota [13–15].

The majority of studies that have investigated associations between diet and the gut microbiota to date have primarily focused on “Western-style” or Mediterranean diets and have been conducted mainly in European and American populations [15]. In contrast, little is known about the associations between dietary habits and the gut microbiota in the Korean population [16]. Traditional Korean diets are characterized by high intakes of specific fermented vegetables, e.g. kimchi, and legumes, e.g. fermented soybean [17, 18]. Fermented foods are known to contain large amounts of microorganisms, and their strains are phylogenetically similar to probiotic strains, which could affect the composition and diversity of the gut microbiota, thus affecting human health [19, 20].

Recent human microbiome data from the International Human Microbiome Consortium (IHMC) and the European Metagenomics of the Human Intestinal Tract (MetaHIT) consortium have indicated that the human gut microbiota could be classified in distinct “enterotypes” [8]. Each of the three identified enterotype was distinguished by different microbial composition at the genus level, with prominent variation in Bacteroides, Prevotella, and Ruminococcus. So far, only two studies examined associations between these enterotypes and habitual diets in American [21] and in Korean [16] adults.

In a collaborative study between the National Institute of Agricultural Sciences of Korea and the International Agency for Research on Cancer (NAS – IARC), we investigated associations of long-term intake of both food/food groups and nutrients with the taxonomic composition and diversity of the gut microbiota in Korean adults aged 18–60 years. We also aimed to identify dietary patterns associated with gut microbiota within-sample (α-) diversity and to explore whether different enterotypes based on gut microbiota between-sample (β-) diversity were associated with the long-term intake of both food/food groups and nutrients.

Results

A total of 222 Korean adults (49% males) aged 18 ~ 58 years were included in this study. The main characteristics of the study population are shown in Table 1. The mean BMI of the study population was 22.9 kg/m^2 (5–95 percentiles: 19.1–28.5 kg/m^2) and was slightly higher in males than females.

Table 1

General characteristic and lifestyle factors of the healthy Korean adults (n = 222 participants)

	Total (n = 222)		Males (n = 108)		Females (n = 114)	
Age [a] (years)	29.6	20–51	26.9	21–48	32.2	20–51
BMI [a] (kg/m^2)	22.9	20.2–28.8	23.6	20.2–28.8	22.3	18.8–27.0
Alcohol intake [a] (g/day)	9.7	0.0-39.7	14.1	0.0-52.9	5.5	0-27.1
Dietary supplement intake (n,%)						
Yes	76	34.2%	26	24.1%	50	43.9%
No	143	64.4%	80	74.1%	63	55.3%
Don't know	3	1.4%	2	1.9%	1	0.9%
Regular physical activity (n,%)						
Yes	92	41.4%	49	45.4%	43	37.7%
No	130	58.6%	59	54.6%	71	62.3%
Smoking status (n, %)						
Ever	54	24.3%	47	43.5%	7	3.5%
Never	168	75.7%	61	56.5%	107	93.9%
Education (n, %)						
<University graduation	110	49.5%	60	55.6%	50	43.9%
≥University graduation	112	50.5%	48	44.4%	64	56.1%
Household Income (n, %)						
<4,000 USD/month	79	35.6%	40	37.0%	39	34.2%
≥4,000 USD/month	93	41.9%	41	38.0%	52	45.6%
Don't know	50	22.5%	27	25.0%	23	20.2%

[a] Mean and range (5–95 percentiles)

The dominant phyla in the study population were Bacteriodetes, Firmicutes, Proteobacteria, and Actinobacteria, of which medians of the relative abundance (% OTUs) were 54.3%, 37.7%, 3.8%, and 0.4%, respectively and encompassed a total of 96.2% of the overall microbiota (Table S1 in Additional file 1). The differences in the relative abundance of these four major phyla and Firmicutes-to-Bacteriodetes (F/B) ratio by general characteristics and lifestyle factors of study subjects are shown in Table S1 (in Additional file 1). The F/B ratio was significantly higher in females with higher abundance of Firmicutes

(39.6% in females vs. 33.0% in males, P-value < 0.01) and in never-smokers with lower abundance of Bacteroidetes (53.1% in never-smokers vs. 57.4% in ever-smokers, P-value = 0.02). Age and dietary supplement intake were both significantly associated with the abundance of Proteobacteria. Sex, age, BMI, and smoking status were significantly associated with within-sample (α-) and/or between-sample (β-) diversity of the gut microbiota (Table S2 in Additional file 1).

Association of dietary intake with gut microbial composition at phylum and genus levels

Partial Spearman correlations between the intakes of 22 food groups and relative abundance (%OTUs) of the gut microbiota at phylum and genus levels are shown in Figs. 1A and 1B. At the phylum level (Fig. 1A), intakes of vegetables (r = 0.20), fermented legumes (r = 0.15), and potatoes (r = 0.14) were statistically significantly positively correlated with the F/B ratio, while intakes of other cereal products such as noodles (r=-0.15) and non-alcoholic beverage including mainly carbonated and sweet beverages (r=-0.15) were significantly inversely correlated with the F/B ratio. The intake of dairy products (r = 0.14) was positively correlated with the relative abundance of Actinobacteria. In contrast, Proteobacteria was not associated with the intake of any food group.

Of a total of 261 genera within the four major phyla, partial Spearman correlations of 82 genera which had higher coefficients (|r| > 0.15) with at least one food group are shown in Fig. 1B. The intakes of potatoes, fermented legume, vegetable, nuts/seeds, and multi/whole grains showed higher positive correlations with similar genera such as Firmicutes(F)_ Eubacterium, F_Lactobacillus, F_Ruminococcus, and/or F_Gracilibacter. In contrast, the consumption of non-alcoholic beverages and other cereal products showed negative correlations with F_Ruminoccocus.

The correlations of gut microbial composition with intakes of macronutrients are shown in Figure S1 (in Additional file 1). At the phylum level (Figure S1A in Additional file 1), intakes of dietary fiber (r = 0.18), plant protein (r = 0.17), and PUFA (r = 0.17) showed significant positive correlations with the F/B ratio. At the genus level (Figure S1B in Additional file 1), the intake of dietary fiber showed highest positive correlations with F_Eubacterium, F_unclassified_Buminococcaceae, and F_ Lactobacillus ($r \geq 0.19$). In addition, the intakes of animal protein, animal fat and SFA showed similar patterns - positive correlations with F_Streptococcus and F_Granulicatella (r > 0.15) and negative correlations with Bacterodites(B)_Sphingobacterium (r<-0.16).

Association of dietary intake with the within-sample diversity of gut microbiota

The intakes of vegetables (r = 0.21), fermented legumes (r = 0.20), potatoes (r = 0.20), nuts/seeds (r = 0.14), and fruit/fruit juice (r = 0.13) among food groups and dietary fiber (n = 0.20) among nutrients were positively correlated with the within-sample (α-) diversity of the gut microbiota (Table S3 in Additional file 1). We also identified a dietary pattern that explained the within-sample (α-) diversity of the gut microbiota using RRR analysis (Fig. 2). This high α-diversity dietary pattern (HiαDP) was characterized by

greater intakes of fermented legumes, vegetables, potatoes, tea, and fruit/fruit juice and low intakes of non-alcoholic beverages (e.g. carbonated and sweet beverages). The score was significantly positively correlated with the F/B ratio (r = 0.24) (Table 2). At the genus level, the HiαDP score showed significant positive correlations with a number of genera within Firmicutes, such as Lactobacillus (r = 0.29), Ruminococcus (r = 0.23), Eubacterium (r = 0.22), Lachrnospira (r = 0.21), and Caloramator (r = 0.20).

Table 2

Spearman correlations [a] between the score of high α-diversity dietary pattern (HiαDP) [b] and relative abundance (% OUT) of gut microbiota at the phylum and genus levels in the NAS-IARC cross-sectional study (n = 222 participants).

	HiαDP score
Phylum level	
Actinobacteria	0.107
Bacteroidetes (B)	-0.232
Firmicutes (F)	0.233
Proteobacteria	-0.055
F/B ratio	0.238
Genus level	
Bacteroidetes(B_)	
B_Odoribacter	0.134
B_Parabacteroides	0.179
B_Prevotella	-0.179
B_Alistipes	0.158
Firmicutes(F_)	
F_Catenibacterium	0.135
F_Holdemania	0.156
F_Lactobacillus	0.285
F_Gracilibacter	0.196
F_Caloramator	0.203
F_Eubacterium	0.218
F_Coprococcus	0.195

F/B ratio, Firmicutes-to-Bacteriodetes ratio

[a]. Partial Spearman correlation analysis adjusted for sex, age, BMI, dietary supplement intake, smoking status and sample batch.

[b]. The high α-diversity dietary pattern (HiαDP) identified by Reduced Rank Regression (RRR) analysis was characterized by high intakes of fermented legume, vegetable, potato, tea, and fruit/fruit juice and the low intake of non-alcoholic beverages (e.g. carbonated and sweet beverages).

	HiαDP score
F_Lachnospira	0.206
F_Oscillibacter	0.157
F_Desulfonispora	0.134
F_Acetanaerobacterium	0.134
F_Ethanoligenens	0.165
F_Ruminococcus	0.233
F_Sporobacter	0.151
F/B ratio, Firmicutes-to-Bacteriodetes ratio	
[a]. Partial Spearman correlation analysis adjusted for sex, age, BMI, dietary supplement intake, smoking status and sample batch.	
[b]. The high α-diversity dietary pattern (HiαDP) identified by Reduced Rank Regression (RRR) analysis was characterized by high intakes of fermented legume, vegetable, potato, tea, and fruit/fruit juice and the low intake of non-alcoholic beverages (e.g. carbonated and sweet beverages).	

Enterotypes of gut microbiota and their association with dietary intake

Enterotypes of gut microbiota among the Korean healthy adults are shown in Fig. 3. Among the study population, three different enterotypes were identified and labelled enterotype 1 (37%), enterotype 2 (29%), and enterotype 3 (34%) (Fig. 3B). Each of these enterotypes was identifiable by one of these dominant genera; Bacteroides in enterotype 1, Prevotella in enterotype 2, and Ruminococcus in enterotype 3 (Fig. 3C). Ruminococcus enterotype showed significantly higher scores of the HiαDP and was more strongly associated with intakes of vegetables and nuts/seeds at the food group level and dietary fiber at the nutrient level compared to the two other enterotypes (Table 3). The Bacteroides enterotype showed higher intakes of legumes compared to the two other enterotypes. There were no significant differences with respect to sex, age, BMI, or smoking status (Table S4 in Additional file 1).

Table 3

Difference in dietary intake – the high α-diversity dietary pattern (HiαDP) and intakes of food groups and nutrients among three enterotypes of the healthy Korean adults (total subjects, n = 222)

	Enterotype 1		Enterotype 2		Enterotype 3		P-value [a]
	(n = 82, 36.9%)		(n = 64, 28.8%)		(n = 76, 34.2%)		
	Mean	SE	Mean	SE	Mean	SE	
Dietary pattern							
HiαDP score	0.02	0.11	-0.30	0.14	0.23	-0.12	0.006
Food groups (g/day) [b]							
Potatoes	31.4	3.5	28.9	3.8	34.1	4.0	0.336
Vegetables	160.4	13.9	126.4	11.2	180.0	22.7	0.042
Fermented vegetables	80.9	8.3	86.6	8.0	100.7	9.9	0.169
Legumes	51.1	6.9	33.6	4.6	47.0	5.9	0.029
Fermented legumes	3.8	0.4	3.2	0.5	4.2	0.6	0.161
Fruit/Fruit Juice	194.6	22.6	214.5	41.0	217.8	42.6	0.560
Nuts/Seeds	1.7	0.5	0.8	0.2	2.8	0.6	0.009
Dairy	158.2	19.0	121.6	14.2	139.6	15.2	0.732
Refined grains	447.3	21.9	437.8	24.3	443.3	20.7	0.965
Multi/whole grains	3.7	0.54	3.6	0.6	5.2	0.7	0.176
Other cereal products	78.5	6.2	82.3	7.5	72.7	5.1	0.677
Meats	117.4	10.9	113.6	18.3	111.4	12.3	0.759
Fish/Seashells	37.7	3.5	31.1	3.2	36.7	3.8	0.435
Eggs	25.2	2.4	22.5	2.2	24.6	2.3	0.958
Vegetable oils	2.0	0.2	1.8	0.2	2.0	0.2	0.114

SF, saturated fat; MUSF, monounsaturated fat; PUSF, polyunsaturated fat

[a] Differences of dietary pattern scores and intakes of food groups and nutrients among three enterotypes were examined by general linear models (GLMs) with sex, age, BMI, dietary supplement intake, smoking status, and sample batch as covariates

[b] The intakes of food groups and nutrients were log-transformed and adjusted for total energy intake using the residual method

	Enterotype 1		Enterotype 2		Enterotype 3		P-value [a]
	(n = 82, 36.9%)		(n = 64, 28.8%)		(n = 76, 34.2%)		
	Mean	SE	Mean	SE	Mean	SE	
Other fats	0.7	0.1	0.5	0.1	0.5	0.1	0.415
Sugar/Confectionary	3.5	0.4	4.0	0.8	3.3	0.4	0.812
Cakes/Sweets	23.0	2.3	20.3	5.0	20.6	2.1	0.307
Coffee	2.5	0.3	4.0	0.7	2.6	0.3	0.479
Tea	32.4	9.2	22.7	5.9	18.1	4.1	0.503
Non-alcoholic Beverages	96.4	13.9	120.5	34.2	56.5	6.7	0.343
Salty snacks	21.7	2.17	19.5	2.5	20.1	1.9	0.505
Nutrient intake (g/day) [b]							
Plant Protein	35.9	1.6	34.7	1.9	36.5	1.6	0.069
Animal Protein	38.4	2.9	34.6	4.0	36.5	3.1	0.284
Plant Fat	17.8	1.2	17.4	1.6	17.4	1.0	0.326
Animal Fat	30.3	2.4	30.4	4.6	29.0	2.5	0.644
SF	11.3	0.9	12.1	2.0	11.7	1.0	0.301
MUSF	11.5	0.9	12.5	2.2	12.1	1.0	0.240
PUSF	5.5	0.4	5.4	0.6	5.7	0.4	0.053
Carbohydrate	310.8	13.8	313.0	21.5	303.3	13.1	0.940
Dietary Fiber	17.5	1.1	16.4	1.2	18.6	1.4	0.037
Alcohol	9.4	2.4	11.1	2.7	8.7	1.4	0.546

SF, saturated fat; MUSF, monounsaturated fat; PUSF, polyunsaturated fat

[a] Differences of dietary pattern scores and intakes of food groups and nutrients among three enterotypes were examined by general linear models (GLMs) with sex, age, BMI, dietary supplement intake, smoking status, and sample batch as covariates

[b] The intakes of food groups and nutrients were log-transformed and adjusted for total energy intake using the residual method

Discussion

In the NAS-IARC cross-sectional study among Korean adults, we found that a traditional Korean dietary pattern characterized by higher intakes of plant-based and fermented foods and lower intakes of noodle products and carbonated and sugar-sweetened beverages were favorably associated with gut microbial composition and diversity. Specifically, this HiαDP was positively associated with the F/B ratio and some genera within the Firmicutes phylum, such as Lactobacillus, Ruminococcus, Eubacterium, and Lachrnospira. Further, we identified three distinct enterotypes, which differed in their HiαDP and habitual of specific foods and nutrients.

Our findings on the HiαDP characterized by higher intake of plant-based foods are in line with previous studies reporting that the gut microbial diversity of populations consuming plant-based diets in rural areas in Africa and South Ameria, was greater compared to western populations [12, 22, 23]. Plant-based diets are rich in dietary fiber, which is the main source of microbiota-accessible carbohydrates (MACs), which are the major source of energy for gut microbiota [24]. An animal model study showed that low MACs diets led to an irreversible depletion of gut microbiota diversity [25]. MACs can be metabolized by “fiber-degrading bacteria” such as Roseburia, Lactobacillus, Eubacterium, Ruminococcus, and Bifidobacterium, mostly belonging to the Firmicutes and Actinobacteria phyla [11, 26]. Human intervention studies also found that intakes of MACs-rich foods such as whole-grain foods enhanced the presence of some bacteria of the Firmicutes phylum [27, 28]. Consistent with these previous studies, when examining associations between long-term intakes of individual food groups and the gut microbiota in our study, we found that the plant-based foods including vegetables, fermented legumes, potatoes, multi/whole grains, and nuts/seeds were positively associated with the F/B ratio, especially having positive associations with some genera within the Firmicutes phylum including Eubacterium, Lactobacillus, and Ruminococcus, known as “fiber-degrading bacteria” [11, 26]. These fiber-degrading bacteria produce short-chain fatty acids including butyrate in the human intestine [29], which have been shown to be inversely associated with obesity, diabetes and colorectal cancer [30, 31].

Another interesting finding in the current study was the association of fermented foods such as fermented legumes with taxonomic composition and diversity of the gut microbiota, which has been rarely investigated in previous studies among western populations. Fermented foods such as fermented vegetables (e.g. Kimchi), and fermented legume products, mainly based on soybean, (e.g. Cheonggukjang, Doenjang, and Ganjang) are typical dishes of traditional Korean diets [17]. These fermented foods contain living microorganisms including probiotic bacteria (e.g. Lactobacillus, Bifidobacterium or Streptococcus) and bioactive compounds generated during the fermentation process which could affect the microbial composition and diversity in the human gut [19]. However, to date, there is very limited evidence on the impact of habitual intake of fermented foods on overall gut microbiota. To the best of our knowledge, this is the first observational study investigating associations between fermented legumes as part of habitual diets and gut microbial composition and diversity. In our study population, the group of fermented legumes, mainly fermented soybean pastes, was one of the major components of the HiαDP and was positively associated with the Firmicutes phylum and its genus Eubacterium, Lactobacillus, and Ruminococcus, while this association was not found in the non-fermented legume intake. Therefore, this difference may come from the fermentation process. In previous

studies on microbial communities in Korean fermented soybean pastes [32, 33], the dominant microbes were found to be Bacillus and other lactic acid bacteria, which mostly belong to the Firmicutes phylum. An intervention study [34] which examined the effect of a typical Korean diet including fermented foods like Kimchi and American-style diets on the gut microbiota in 61 overweight/obese Korean adults, also showed that the F/B ratio and some genera within the Firmicutes phylum including Weissella increased and after the consumption of the typical Korean diet with explaining as the effect of the fermented food intake. However, in our study, no significant association was found between higher intakes of fermented vegetables including Kimchi and the gut microbiota. It indicated that fermented vegetables like Kimchi, which is one of the most frequently (almost every day and with every meal) and widely consumed foods in the population, could not be a discriminant dietary factor in gut microbial diversity and composition among the Korean population.

In addition, we explored gut microbial enterotypes in the Korean adults applying a modified multivariate cluster analysis [8]. The retained enterotypes and their dominant genera - Bacteroides, Prevotella and Ruminococcus respectively, were similar as in the previous study [8]. Another study in American adults [21] suggested that gut microbial enterotypes were strongly associated with long-term diets compared to short-term diets, showing that the Bacteroides enterotype was strongly associated with high intake of protein and animal fat and the Prevotella enterotype with high intake of carbohydrate. Our study also showed that the enterotypes were strongly associated with habitual diet, but there was no significant difference in the intakes of protein, animal fat, and carbohydrate across enterotypes. In our study, there were significant differences in the intake of plant-based foods high in dietary fiber like vegetables and nuts/seeds across enterotypes. In particular, subjects of the Ruminococcus enterotype were more adherent to the HiαDP, which was characterized by higher intakes of vegetables, nut/seeds and dietary fiber compared to the two other enterotypes.

A major strength of our study was that it provides a first comprehensive overview of the Korean diet associated with the composition and diversity of human gut microbiota, considering not only individual intakes of food groups and nutrients but also specific dietary patterns. The dietary pattern analysis accounted for synergistic and correlated effects of food groups [35] on the host diet-gut microbiota. The relatively large sample size is another strength compared to previous microbiome studies (mostly, less than 100). The finding of this study should be interpreted in light of the following limitations. The 16 s rRNA sequencing may be associated with measurement error including limited resolution and lower sensitivity compared to metagenomic sequencing data, even though it enables to capture of broad snapshots to understand the gut microbial community in human gut [36]. Also, while we adjusted for confounding by factors known to affect the gut microbiota such as sex, age, BMI, intake of dietary supplements including probiotics, and smoking status, we were unable to account for the mode of birth delivery due to lack of data [37]. Last, since this was a cross-sectional study with convenience sampling from a southern part of Korean, we cannot determine a causal relationship and generalization of the study findings should be made cautiously.

Conclusions

We conclude that the habitual diet characterized by higher intakes of plant-based and fermented foods and lower intakes of noodle products and carbonated and sugar-sweetened beverages in Korean adults was associated with gut microbial taxonomic composition and diversity. A higher intake of plant-based and fermented foods rich in dietary fiber, was associated with distinct gut microbial enterotypes. Further studies are needed to investigate potential health effects of belonging to a specific enterotype.

Methods

Study design and subjects. Within the NAS-IARC cross-sectional study, participants were residents aged 18-60 years in the local vicinity (within 20 km) of the NAS, the Republic of Korea, between March and October 2018. We excluded volunteers who, prior to recruitment 1) were underweight (body mass index (BMI) < 18.5 kg/m^2) or obese (BMI ≥ 30 kg/m^2), 2) reported any chronic disease such as inflammatory bowel disease, hypertension, diabetes, hyperlipidemia, or cancer, 3) had taken medication including antibiotics within the past 2 weeks, 4) had taken hormone replacement therapy or used oral contraceptives within the past 2 weeks, or 5) were pregnant or breastfed within the past 6 months. Volunteers who had taken any dietary supplements within the past 3 months were not excluded, but this information was collected using lifestyle questionnaires. The study participants were initially invited to an information meeting, approximately one week prior to the start of the study, where anthropometric data including height and weight were measured by trained research assistants, and exclusion criteria were ascertained. Those eligible for the study were provided with a lifestyle questionnaire (physical activity, alcohol intake, smoking, and socioeconomic status) and a food frequency questionnaire (FFQ) with instructions, and were asked to fill in and return on the study day. During the study day, on-site fecal samples were collected and FFQ and lifestyle data of participants were reviewed by trained research assistants following standardized protocols. Of a total of 229 eligible participants, seven participants failed to collect fecal samples, leading to a sample size of 222 healthy Korean adults (49% males) for this study.

All procedures and protocols of the study were approved by the Public Institutional Review Boards Institutional Review Board of the Ministry of Health and Welfare, Korea (Approval no: P01-201801-11-003), and were registered at the Clinical Research Information Service (CRIS) of the Centers for Disease Control and Prevention of Korea (KCT0002831). All study participants provided written informed consent.

Dietary data collection. Long-term dietary intake data from participants were collected with a semi-quantitative FFQ, which was developed and validated for the Korean diet by the Korea National Institute of Health (KNIH) [38]. The FFQ included 106 food/dish items, including 9 Korean staple dishes (rice and noodles), 25 soups and stews, 54 side dishes, 9 non-alcoholic beverages, and 9 fruits. Subjects were asked to report the consumption frequency and average portion size of each item during the previous year. During the visit of the participants, trained research assistants reviewed the questionnaires with participants together for completeness. The 106 food/dish items were classified into 22 food groups –

potatoes, vegetables, fermented vegetables, legumes, fermented legumes, fruit/fruit juice, nuts/seeds, dairy, refined grains, multi/whole grains, other cereal products, meats, fish/seashells, eggs, vegetable oils, other fats, sugar/confectionery, cakes/sweets, coffee, tea, non-alcoholic beverages, salty snacks based on their recipe. In particular, vegetable and legume groups were divided into two sub-groups such as non-fermented and fermented to take into account fermentation, which could affect gut microbial composition and diversity. Intakes of macronutrients including protein, fat, carbohydrates (CHO), and dietary fiber were also estimated based on the FFQ data. Protein and fat intake were classified as either plant-based or animal-based separately. Additionally, saturated fatty acids (SFA), monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA) were estimated separately. The intakes of food groups and macronutrients were calculated as gram per day (g/day) based on the consumption frequency and average portion size based on a food composition database established for the FFQ [38]. Alcohol intake of the previous year was collected with a lifestyle questionnaire and converted into g/day.

Fecal sample collection. The fecal specimens were collected on-site on the study day at the NAS. We provided a collection tube (SARSTEDT AG & Co., Germany) for the fecal sample to each participant. Following the collection, the samples were immediately delivered to the laboratory for processing. Each fecal specimen was mixed manually using a spatula, and approximately 1-2 g of feces for each participant was aliquoted, representing a full scoop of feces, into stool nucleic acid collection tubes (Norgen Biotek Co., Canada). Samples were then frozen and stored at 4 until further processing (average time between sample collection and storage: approx. 12 mins).

16s rRNA gene sequencing and taxonomic assignment. All procedures from extracting bacterial DNA from the collected fecal samples to generating the gut microbial composition and diversity data have been performed by a biotechnology company (Macrogen Inc.) in Seoul, the Republic of Korea. On a weekly basis, the fecal samples collected for one week period were transferred to Macrogen Inc., and bacterial DNA from each sample was extracted using PowerSoil® DNA Isolation Kit (Cat. No. 12888, MO BIO) according to the manufacturers' protocol and stored at -80 until all samples were collected for further analysis. DNA quantity and quality were measured by PicoGreen and Nanodrop (ThermoFisher Sci. Inc. Waltham, MA, USA). The 16S rRNA amplicons covering variable regions V3-V4 were generated using the primers (forward: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCCTACGGGNGGCWGCAG-3' and reverse: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GACTACHVGGGTATCTAATCC-3') incorporating multiplexing indices and Illumina sequencing adapters. The final products were normalized and pooled using PicoGreen, the size of libraries were verified using TapeStation DNA screentape D1000 (Agilent Tech., Santa Clara, USA), and the amplicons were sequenced using the MiSeq™ platform (Illumina, San Diego, USA). In order to achieve the high quality of data on Illumina sequencing platforms, optimal cluster densities were created across every lane of every flow cell. The Rapid library standard Quantification solution and calculator (Roche, Basel, Switzerland) were used to generate a standard curve of fluorescence readings and calculate the library sample concentration. Using the QIIME 1.8.0 pipeline, the sequences were binned into operational taxonomic units (OTUs) from phylum to species levels with 97% identity [39].

In total, 5.7 million sequence reads from 222 subjects were obtained with an average of 25,852 (5-95 percentiles: 11,129-47,363) reads per subject, which were clustered into OTUs, and subsequently assigned taxonomy at different levels from phylum to genus levels. The gut microbial taxonomic composition and diversity data generated by this procedure included individual-level information on 1) relative abundance (proportion (%) of OTU) at different bacterial taxonomic levels, 2) within-sample (α-) diversity to understand the number (richness) and distribution (evenness) of species within a single subject by estimating a widely used α-diversity index [40] – Shannon index [41], and 3) between-sample (β-) diversity to understand differences of gut microbial composition in one subject compared to another [40] by measuring the phylogenetic distance between microbial communities of two subjects with weighting the relative abundance of species [42] – weighted phylogenetic UniFrac distance matrix.

Statistical analysis. Dietary intake data were log-transformed to render the distributions symmetrical and to approximate normality and were adjusted for total energy intake using the residual method. The Shannon α-diversity index was also log-transformed. The differences of relative abundance (% OTU) of the four major phyla and of the *Firmicutes-to-Bacteroidetes* (F/B) ratio, which are the two major phyla in human gut microbiota and are known to be modulated by diet [9, 11], by basic characteristic and lifestyle factors (sex; age group: <40 years vs. ≥ 40 years; BMI group: <23 kg/m^2 vs. ≥ 23 kg/m^2; dietary supplement intake within 3 month prior to the enrolment: yes vs. no; regular physical activity: yes vs. no; smoking status: ever vs. never, education: < university graduation vs. ≥ university graduation; household income: <4,000 USD/month vs. ≥ 4,000 USD/month) were examined by Wilcoxon-Mann-Whitney tests. Associations of within-sample (α-) diversity and between-sample (β-) diversity of gut microbiota with basic and lifestyle factors of study populations were examined by general linear models (GLMs) and permutational multivariate analysis of variance (PERMANOVA), respectively.

In order to examine the gut microbial composition in relation to dietary intake, partial Spearman's correlation coefficients of relative abundance (% OTU) of the four major phyla, the F/B ratio, and genera within the major phyla of human gut microbiota with the intakes of food groups and macronutrients were estimated. Adjustment for sex, age, BMI, dietary supplement intake, smoking status, and sample batch was performed. Correlation values were displayed using heatmaps after false discovery rate corrections.

Partial Spearman's correlation coefficients of the Shannon index with the intakes of food groups and macronutrients were estimated after adjustment for sex, age, BMI, dietary supplement intake, smoking status, and sample batch. To identify dietary patterns associated with high within-sample (α-) diversity, reduced rank regression (RRR) was used to derive patterns of 22 food groups (predictor variables) maximizing the explained variability of gut microbiota diversity (Shannon index as response variables). We then examined partial Spearman's correlation coefficients between the score of the high α-diversity dietary pattern (HiαDP score) and relative abundance (% OTU) of major phyla including F/B ratio and genera within the major phyla of human gut microbiota with sex, age, BMI, dietary supplement intake, smoking status, and sample batch as covariates.

Enterotypes of gut microbiota in healthy Korean adults were explored by a modified method to determine enterotype discovery (9) with a combination of principal coordinate analysis (PCoA) based on the weighted UniFrac distance matrix as a between-sample (β-) diversity index, and then k-means cluster analysis based on the PCoA scores of the first two principal coordinates (PCos). The optimal number of clusters was determined by visual inspection of clusters derived by three different methods – Elbow [43], Silhouette [44] and Gap statistic [45] methods (**Figure S2 in Additional file 1**) and by a priori knowledge [8]. The differences of general characteristics and lifestyle factors by enterotypes were examined by GLMs for continuous variables and chi-square test for categorical variables, and the differences in dietary intake – the HiαDP score and intakes of food groups and macro-nutrients by enterotypes were examined by GLMs with sex, age, BMI, dietary supplement intake, smoking status, and sample batch as covariates.

All analyses were performed using the R statistical software (version 3.6.1, R Development Core Team, 2019) for PCoA and k-means cluster analyses (using cmdscale, kmeans, and fviz_nbcluster functions) and generating heatmaps and boxplots, and SAS (version. 9.4, The SAS Institute, Cary, NC) for the rest of analyses.

Abbreviations

BMI, body mass index; *F/B* ratio, *Firmicutes-to-Bacteroidetes* ratio; HiαDP, Higher α-diversity dietary pattern; NAS-IARC, National Institute of Agricultural Sciences of Korea and the International Agency for Research on Cancer; FFQ, food frequency questionnaire; CHO, carbohydrates; SFA, saturated fatty acid; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acid; OTU, operational taxonomic units; GLM, general linear models; PERMANOVA, permutational multivariate analysis of variance; RRR, reduced rank regression; PCoA, principal coordinate analysis; MACs, microbiota-accessible carbohydrates;

Declaration

- **Ethics approval and consent to partipate**
  All procedures and protocols of the study were approved by the Public Institutional Review Boards Institutional Review Board of the Ministry of Health and Welfare, Korea (Approval no: P01-201801-11-003), and were registered at the Clinical Research Information Service (CRIS) of the Centers for Disease Control and Prevention of Korea (KCT0002831). All study participants provided written informed consent.
- **Consent for publication**
  Not applicable
- **Availability of data and materials**
  The datasets used and/or statistical anlaysis code for the current study are available from the corresponding author (Dr. H. Freisling, FreislingH@iarc.fr) on reasonable request
- **Competing interests**
- All authors declare that they have no conflict of interest

- **Funding**
  This work was supported by a research program for Agricultural Science & Technology Development (Project No. PJ01308801) from NAS-RDA
- **Authors'contributions**
  HN, HHJ, MJG, AS, and HF designed research; HHJ, SYC, and HJK conducted research; JK provided essential materials including dietary assessment tools and database; HN, HHJ, GK, and SZ prepared preprocessed data including microbiome data for statistical analysis; HN performed statistical analysis and wrote paper; HN, MJG, PF, AS and HF interpreted study results; HN, HHJ, and HF had primary responsibility for final content; All authors read and approved the final manuscript
- **Acknowledgements**
  We thank all study participants in the NAS-IARC cross-sectional study.
- **Disclaimer**: *Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization*

References

1. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: Metagenomic analysis of the human distal gut microbiome. Science 2006, 312:1355-1359.
2. Thursby E, Juge N: Introduction to the human gut microbiota. Biochem J 2017, 474:1823-1836.
3. Natividad JM, Verdu EF: Modulation of intestinal barrier by intestinal microbiota: pathological and therapeutic implications. Pharmacol Res 2013, 69:42-51.
4. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM: The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of Lipid Research 2013, 54:2325-2340.
5. Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I: Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. Nutrients 2015, 7:2839-2849.
6. Baumler AJ, Sperandio V: Interactions between the microbiota and pathogenic bacteria in the gut. Nature 2016, 535:85-93.
7. Belkaid Y, Hand TW: Role of the microbiota in immunity and inflammation. Cell 2014, 157:121-141.
8. 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, 473:174-180.
9. Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B, Nakamura M, Zhu TH, et al: Influence of diet on the gut microbiome and implications for human health. J Transl Med 2017, 15:73.

10. Zmora N, Suez J, Elinav E: You are what you eat: diet, health and the gut microbiota. Nat Rev Gastroenterol Hepatol 2019, 16:35-56.
11. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al: Diet rapidly and reproducibly alters the human gut microbiome. Nature 2014, 505:559-563.
12. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P: Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A 2010, 107:14691-14696.
13. De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, Serrazanetti DI, Di Cagno R, Ferrocino I, Lazzi C, et al: High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut 2016, 65:1812-1821.
14. Swiatecka D, Narbad A, Ridgway KP, Kostyra H: The study on the impact of glycated pea proteins on human intestinal bacteria. Int J Food Microbiol 2011, 145:267-272.
15. Telle-Hansen VH, Holven KB, Ulven SM: Impact of a Healthy Dietary Pattern on Gut Microbiota and Systemic Inflammation in Humans. Nutrients 2018, 10.
16. Lim MY, Rho M, Song YM, Lee K, Sung J, Ko G: Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and diet. Sci Rep 2014, 4:7348.
17. Patra JK, Das G, Paramithiotis S, Shin HS: Kimchi and Other Widely Consumed Traditional Fermented Foods of Korea: A Review. Front Microbiol 2016, 7:1493.
18. Woo HD, Shin A, Kim J: Dietary patterns of Korean adults and the prevalence of metabolic syndrome: a cross-sectional study. PLoS One 2014, 9:e111593.
19. Bell V, Ferrao J, Pimentel L, Pintado M, Fernandes T: One Health, Fermented Foods, and Gut Microbiota. Foods 2018, 7.
20. Dimidi E, Cox SR, Rossi M, Whelan K: Fermented Foods: Definitions and Characteristics, Impact on the Gut Microbiota and Effects on Gastrointestinal Health and Disease. Nutrients 2019, 11.
21. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al: Linking long-term dietary patterns with gut microbial enterotypes. Science 2011, 334:105-108.
22. Martinez I, Stegen JC, Maldonado-Gomez MX, Eren AM, Siba PM, Greenhill AR, Walter J: The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell Rep 2015, 11:527-538.
23. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al: Human gut microbiome viewed across age and geography. Nature 2012, 486:222-227.
24. Sonnenburg ED, Sonnenburg JL: Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab 2014, 20:779-786.

25. Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL: Diet-induced extinctions in the gut microbiota compound over generations. Nature 2016, 529:212-215.
26. Ganesan K, Chung SK, Vanamala J, Xu B: Causal Relationship between Diet-Induced Gut Microbiota Changes and Diabetes: A Novel Strategy to Transplant Faecalibacterium prausnitzii in Preventing Diabetes. Int J Mol Sci 2018, 19.
27. Carvalho-Wells AL, Helmolz K, Nodet C, Molzer C, Leonard C, McKevith B, Thielecke F, Jackson KG, Tuohy KM: Determination of the in vivo prebiotic potential of a maize-based whole grain breakfast cereal: a human feeding study. Br J Nutr 2010, 104:1353-1356.
28. Walter J, Martinez I, Rose DJ: Holobiont nutrition: considering the role of the gastrointestinal microbiota in the health benefits of whole grains. Gut Microbes 2013, 4:340-346.
29. Schwiertz A, Taras D, Schafer K, Beijer S, Bos NA, Donus C, Hardt PD: Microbiota and SCFA in lean and overweight healthy subjects. Obesity (Silver Spring) 2010, 18:190-195.
30. Karlsson F, Tremaroli V, Nielsen J, Backhed F: Assessing the human gut microbiota in metabolic diseases. Diabetes 2013, 62:3341-3349.
31. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al: A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 2012, 490:55-60.
32. Kim TW, Lee JH, Kim SE, Park MH, Chang HC, Kim HY: Analysis of microbial communities in doenjang, a Korean fermented soybean paste, using nested PCR-denaturing gradient gel electrophoresis. Int J Food Microbiol 2009, 131:265-271.
33. Nam YD, Lee SY, Lim SI: Microbial community analysis of Korean soybean pastes by next-generation sequencing. Int J Food Microbiol 2012, 155:36-42.
34. Shin JH, Jung S, Kim SA, Kang MS, Kim MS, Joung H, Hwang GS, Shin DM: Differential Effects of Typical Korean Versus American-Style Diets on Gut Microbial Composition and Metabolic Profile in Healthy Overweight Koreans: A Randomized Crossover Trial. Nutrients 2019, 11.
35. Hu FB: Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002, 13:3-9.
36. Poretsky R, Rodriguez RL, Luo C, Tsementzi D, Konstantinidis KT: Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 2014, 9:e93827.
37. Scepanovic P, Hodel F, Mondot S, Partula V, Byrd A, Hammer C, Alanio C, Bergstedt J, Patin E, Touvier M, et al: A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. Microbiome 2019, 7:130.
38. Ahn Y, Kwon E, Shim JE, Park MK, Joo Y, Kimm K, Park C, Kim DH: Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. Eur J Clin Nutr 2007, 61:1435-1441.
39. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al: QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010, 7:335-336.

40. Morgan XC, Huttenhower C: Chapter 12: Human Microbiome Analysis. Plos Computational Biology 2012, 8.
41. Shannon CE: The mathematical theory of communication. 1963. MD Comput 1997, 14:306-317.
42. Lozupone C, Knight R: UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005, 71:8228-8235.
43. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD: Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. 2nd International Conference on Vocational Education and Electrical Engineering (Icvee) 2018, 336.
44. Rousseeuw PJ: Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis. Journal of Computational and Applied Mathematics 1987, 20:53-65.
45. Tibshirani R, Walther G, Hastie T: Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society Series B-Statistical Methodology 2001, 63:411-423.

Figures

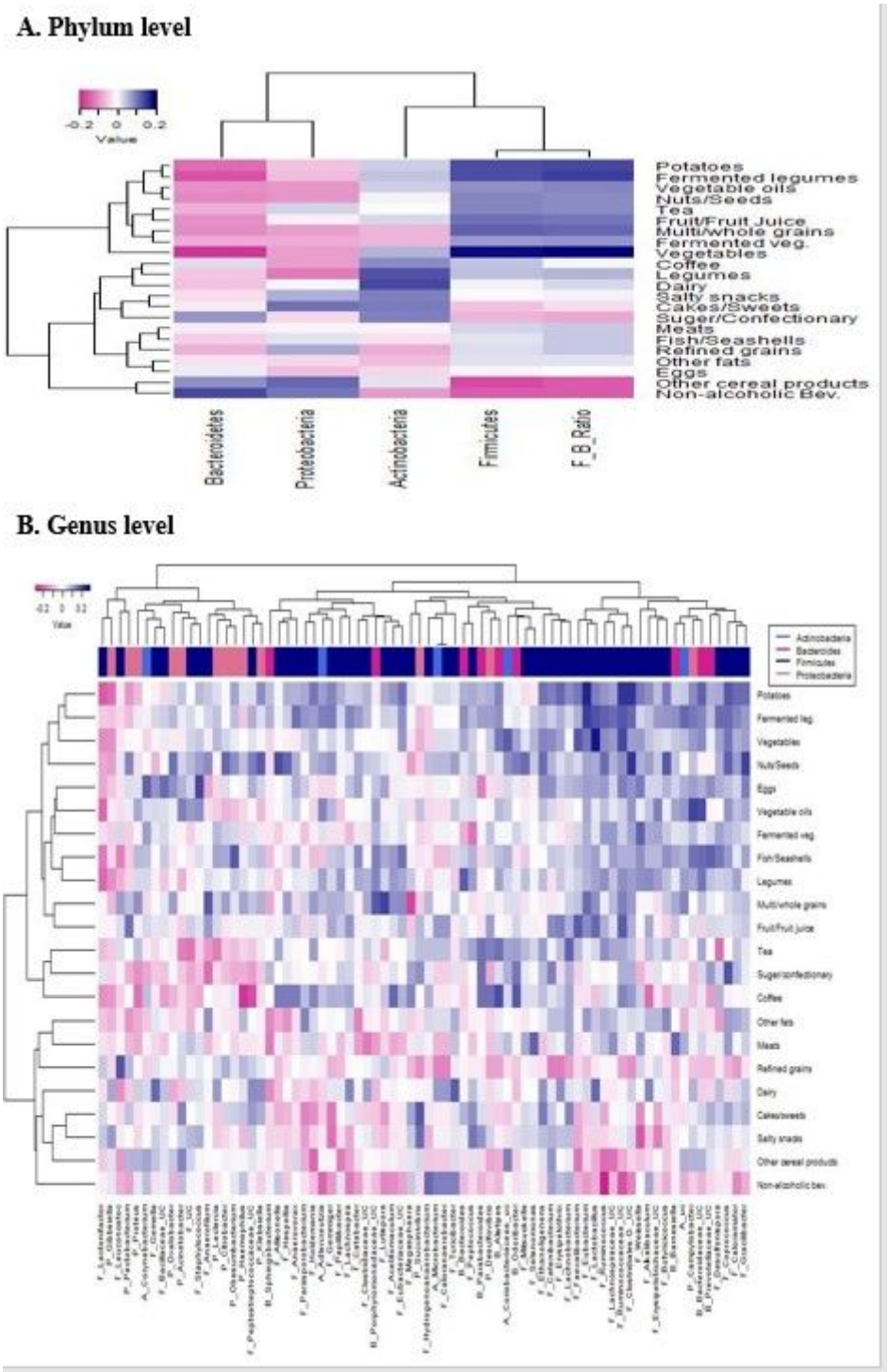

Figure 1

Heatmaps of partial Spearman correlations between intakes of food groups and relative abundance of gut microbiota at phylum (A) and genus (B) levels in the NAS-IARC cross-sectional study (n=222 participants). Partial Spearman correlation analysis adjusted for sex, age, BMI, dietary supplement intake, smoking status and sample batch; The intakes of food groups were log-transformed and adjusted for total energy intake using the residual method; Firmicutes-to-Bacteriodetes ratio was labeled as F_B_Ratio in the heatmap (A); The name of each genus was labeled as phylum initial (B, P, A or F)_genus in the heatmap (B)

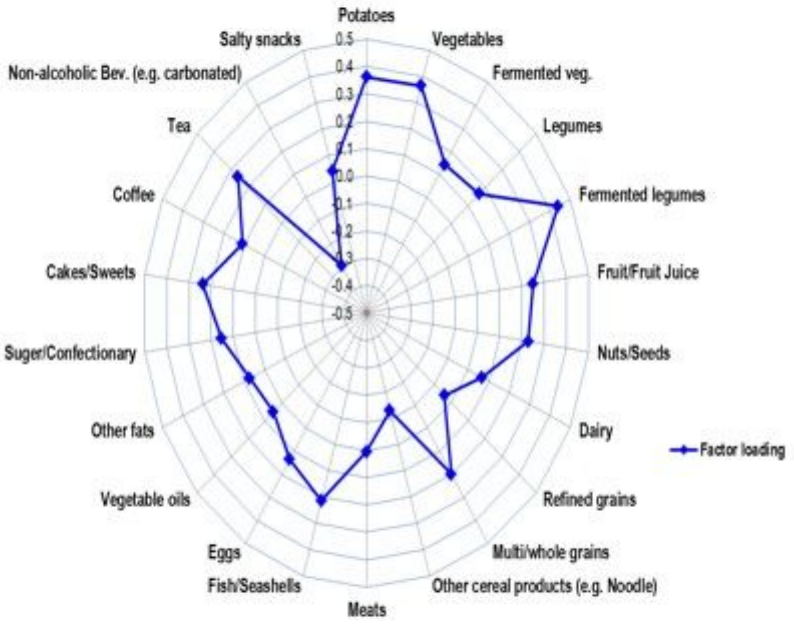

Figure 2

The high α-diversity dietary pattern (HiαDP) in the NAS-IARC cross-sectional study (n=222 participants). The factor loading of each food group of the HiαDP in the Korean adults was estimated in a Reduced Rank Regression (RRR) model with the intake of 22 food groups as predictor variables and the Shannon index (α-diversity index) as response variable; The Shannon index was log-transformed and the intakes of food groups were log-transformed and adjusted for total energy intake using the residual method; Of the variance of predictor variables, 11% was explained by the dietary pattern

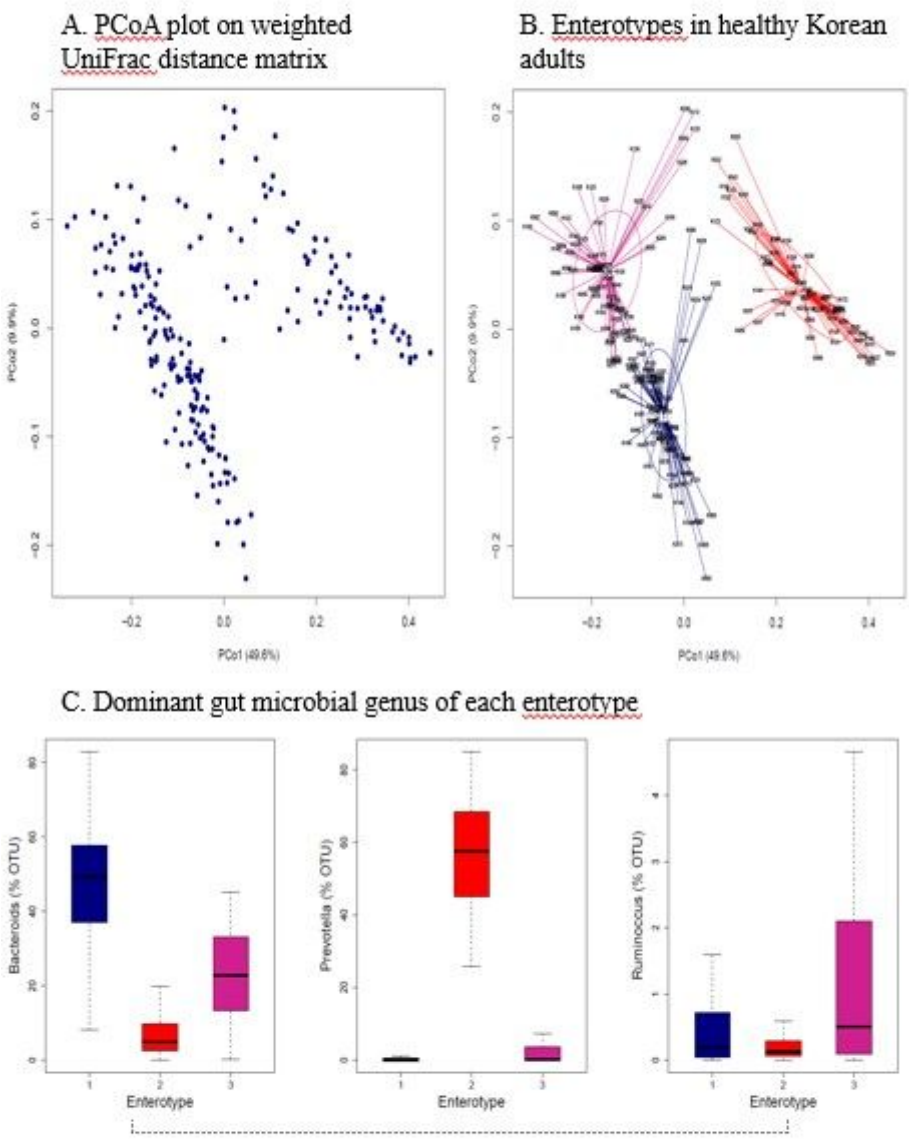

Figure 3

Enterotypes of gut microbiota in the NAS-IARC cross sectional study. A. The plot of two principal coordinates (PCos) were derived by Principal Coordinates Analysis (PCoA) based on weighted UniFrac distance matrix (β-diversity) of gut microbiota in the Korean healthy adults (n=222 participants) . B. Three enterotypes were identified by k-means clustering based on the scores of two first PCos – 82 subjects (37%), 64 subjects (29%), and 76 subjects (34%) in enterotype 1, 2, and 3, respectively. C. Each dominant bacteria genus of each enterotype – Bacteorides, Prevotella, and Ruminococcus in enterotype 1, 2, and 3, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx