npj | biofilms and microbiomes

Article

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-025-00769-9

A contributory citizen science project reveals the impact of dietary keys to microbiome health in Spain

Check for updates

Zaida Soler[1,2], Gerard Serrano-Gómez[1,2,4], Marc Pons-Tarín[1,2,4], Sara Vega-Abellaneda[1,2], Zixuan Xie[1], Isaac Manjón[1], Chloe Cognard[1], Encarna Varela[1], Francisca Yañez[1], Arnau Noguera-Segura[1], Melina Roca-Bosch[1] & Chaysavanh Manichanh[1,2,3]

Low consumption of whole grains, fruits, and vegetables has been identified as dietary risks for non-communicable diseases such as inflammatory bowel diseases (IBDs). We explore how individual and lifestyle factors influence these risks by shaping gut microbiome composition. 1001 healthy participants from all Spanish regions provided personal and dietary data at baseline, six, and twelve months, yielding 2475 responses. Gut microbiome data were analyzed for 500 healthy participants and 321 IBD patients. Our findings reveal that adherence to national dietary guidelines—characterized by diets rich in nuts, seeds, fruits, and vegetables—was associated with greater microbial diversity and reduced IBD-related dysbiosis. Finally, we observed variations in dietary patterns and microbiome diversity and composition across age groups, genders, regions, seasons, and transit time. This study is among the first to uncover dietary intake associated with IBD-related dysbiosis and to propose an interactive website for participants (https://manichanh.vhir.org/POP/en).

Habitual diet and geography have been suggested as among the strongest explanatory factors for human gut microbiota variation. A specific habitual diet may contribute to health or non-communicable diseases (NCDs), such as obesity, metabolic syndrome, and inflammatory bowel disorders (IBD). These conditions and associated mortality/morbidity have risen dramatically over the past decades, with the gut microbiome implicated as one of the potentially causal human-environment interactions[1].

In 2019, the Global Burden of Disease (GBD) Study assessed the impact of dietary habits on NCDs globally[2]. Using a comparative risk assessment approach, the researchers analyzed the consumption of major foods and nutrients across 195 countries. The findings revealed that in 2017, approximately 11 million deaths and 255 million disability-adjusted life-years (DALYs) were attributable to suboptimal dietary habits. Low intake of whole grains and low intake of fruits were identified as the leading dietary risk factors for both deaths and DALYs worldwide. Overall, the research emphasizes the urgent need for improving dietary patterns globally to mitigate the burden of NCDs.

Previous studies have identified significant variations in the gut microbial community among individuals, which has hindered the discovery of microbial species as reliable disease biomarkers. Various factors, including age, medication use, bowel habits, health status, anthropometric characteristics, habitual diet, and lifestyle, have been identified as potential contributors to this high microbiome variability[3]. Consequently, these variations necessitate a large cohort size to effectively discover and validate biomarkers.

Over the last decade, population studies have emerged to understand the role of habitual diets on health and disease through the modulation of the gut microbial community. These large-scale studies, involving hundreds to thousands of participants, included both non-European countries such as the USA[4,5], Canada[6], and China[7], and European countries such as Belgium[3], and the UK[8]. These studies exemplify large-scale projects that facilitate human microbiome hypothesis generation and testing on an unprecedented scale. They have uncovered associations between microbiome signatures and specific genetic variants, geographic variation, medication, and dietary habits.

Although the Spanish diet has been investigated in large-scale studies as part of the Mediterranean diet in relation to cardiovascular disease risk[9,10], no studies have yet comprehensively explored the association between the Spanish diet and both the gut microbiome using shotgun metagenomics at the population level.

[1]Microbiome Lab, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. [2]Departament de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain. [3]CIBER of Hepatic and Digestive Diseases (CIBERehd), Instituto de Salud Carlos III (ISCIII), Madrid, Spain; Departament de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain. [4]These authors contributed equally: Gerard Serrano-Gómez, Marc Pons-Tarín.
e-mail: cmanicha@gmail.com

This study investigates the relationship between diet and the microbiome, with the goal of understanding how national nutritional recommendations can influence the microbial ecosystem and, consequently, human health. We analyzed dietary and personal data from a large cohort of healthy individuals, calculated eating quality indexes based on national guidelines, and examined microbiome data for 500 participants. To further contextualize these findings, we developed a disease similarity index based on microbiome profiles from an IBD cohort of 321 patients. Our results reveal that lifestyle and demographic factors play a significant role in shaping dietary habits, which, in turn, influence microbiome profiles, potentially increasing their resemblance to those associated with IBD.

Results

Cohort characteristics and collected metadata, and samples

Between 2020 and 2024, we enrolled 1001 participants from four regions in Spain, covering all 17 autonomous communities (Fig. 1A, B). The cohort consisted of 458 men and 542 women, all over 18 years old. None of the participants had taken antibiotics for at least three months before the study began, and none had any diagnosed chronic intestinal disorders. Further details regarding the cohort's characteristics can be found in Supplementary Table S1. We employed an in-house[11] online short Food Frequency Questionnaire (sFFQ) to gather demographics, biometrics, lifestyle, and dietary data. Participants filled out the sFFQ at baseline ($n = 1001$), month six ($n = 822$), and month 12 ($n = 652$), resulting in a total of 2475 completed sFFQs. Additionally, stool samples were collected concurrently with the sFFQ for comprehensive analysis. Due to budget constraints, a random subset of 500 samples was selected from the total 1001 baseline samples for microbiome analysis. These samples underwent microbiome compositional and functional profiling through shotgun sequencing (Fig. 1C, D). An additional cohort of 321 IBD patients was included, with fecal microbiome composition data used exclusively to calculate the disease similarity index, as described below (also see the Methods section).

Personal traits, lifestyle decisions, and geography influence the quality of dietary intake ($n = 1001$)

The collected 58 food items from 2475 sFFQs were categorized into 24 food groups and 32 macro- and micronutrient contents (refer to the Methods section). We then investigated the relationship between covariates such as lifestyle, biometrics, and demographic factors on dietary intake using Permutational Multivariate Analysis of Variance (PERMANOVA). These self-reported covariates included age, geography, workplace (hospital or non-hospital), gender, body mass index (BMI), season, dietary types, smoking status, sweetener consumption, menstruation or menopause status (if applicable), and bowel habits. All covariates, except for workplace, were significantly associated with the composition of food items and food groups (Fig. 2A). Furthermore, seven covariates—region, gender, season, dietary types, smoking status, sweetener consumption, and bowel habits—were linked to variations in macro- and micronutrient intake (PERMANOVA, $P < 0.05$, Fig. 2A). These findings highlight the impact of personal traits and lifestyle choices on dietary patterns.

Taking advantage of the longitudinal setting of the study, we analyzed the intra- and inter-variability of food intake using the Bray-Curtis similarity index for food items, food groups, and nutrient data. As expected, we found that intra-individual variability (with sFFQs analyzed 6 months apart) was lower than inter-individual variability across all three dietary classifications ($P < 2{\cdot}2 \times 10^{-16}$, Supplementary Fig. S1). This suggests a relatively stable intra-individual dietary pattern across different seasons at all dietary levels.

Next, we examined how differences in population characteristics may explain variances in several eating quality indexes (EQIs), which were developed based on well-established national guidelines to evaluate the nutritional quality of individuals' diets and their adherence to recommended dietary patterns (refer to the "Methods" section for comprehensive explanations and abbreviations). To achieve this, we initially utilized the collected food items, food groups, and nutrients to calculate various EQIs (HEI-2015, IASE, HFD, hPDI, uPDI, and the aMED). Subsequently, we employed linear regression models, implemented in MaAsLin2, to assess the impact of different population characteristics on these EQIs while controlling for potential covariates mentioned above. Increasing age was found positively associated with several food groups, such as whole bread, nuts and seeds, fruits, and fruit products, which could explain its positive association with a healthier diet as indicated by two EQIs ($q(\mathrm{IASE}) = 0.03; q(\mathrm{hPDI}) = 7{\cdot}1 \times 10^{-07}$) (Fig. 2B). However, it was also found to be linked to a high intake of alcoholic beverages (Supplementary Table S2).

Men exhibited lower values of IASE, hPDI, aMED, and HFD, and higher values of uPDI compared to women, indicating poorer dietary habits compared to women (Fig. 2B). Men's dietary patterns were more associated with the consumption of ready-to-eat meals ($q = 0.038$) and alcoholic beverages ($q = 0.00014$), whereas women showed higher consumption of whole bread ($q = 0.013$), vegetables ($q = 5 \times 10^{-09}$), nonalcoholic drinks ($q = 0.002$), fruits and fruit products ($q = 0.002$), fish and shellfish ($q = 0.00024$), but also higher intake of fats and oils ($q = 2{\cdot}7 \times 10^{-07}$) (Supplementary Table S2).

Geographically, we divided Spain into four regional areas: the Mediterranean, the Interior, the North, and the Islands (Fig. 1B). This classification considers traditional Mediterranean diet patterns and geographical distribution all of which can influence dietary habits and patterns[12]. Compared to the Mediterranean region, the Interior exhibited a healthier dietary pattern based on the three eating quality indices (aMED, uPDI, HEI_2015) (Fig. 2B), characterized by a higher consumption of legumes ($q = 0.013$, Supplementary Table S2).

Interesting positive associations were identified between population behaviors and specific food groups (Supplementary Table S2). For instance, the use of sweeteners was correlated with the consumption of sugar ($q = 0.044$), ready-to-eat meals ($q = 0.00002$), sauces and condiments ($q = 0.0004$), and sausages and other meat products ($q = 0.018$). Additionally, smoking ($q = 0.001$) or past smoking ($q = 0.003$) habits were associated with alcohol consumption.

Partial alignment with recommendations from the GBD-2017

To evaluate whether the dietary intake of our population aligned with the recommendations of the Global Burden of Disease (GBD) Study 2017[2], we categorized our 58 sFFQ items ($n = 1001$; 2475 sFFQs) into 12 of the 15 GBD dietary risk factors (refer to the Methods section, Supplementary Table S3). Our cohort's intake of fruits, vegetables, and fiber met the recommended ranges set by the GBD study (Supplementary Table S4). However, we observed suboptimal intake levels for legumes, polyunsaturated fatty acids (PUFA), whole grains, nuts, milk, and calcium compared to GBD recommendations. Additionally, the intake of red meat, processed meat, and sugar-sweetened beverages exceeded the levels recommended by the GBD guidelines.

Demographic, anthropometric, and dietary data correlate with bacterial microbiome data

Next, to assess the effect size of population characteristics on the microbiome, we used Bray-Curtis distances with the adonis2 function from the R vegan package. Specifically, gender, age, and BMI demonstrated significant impacts on microbiome composition at the global level (Supplementary Fig. S2). These covariates were subsequently used as possible confounders in downstream analysis. A global microbiome profile of Spain at different taxonomic levels (phylum, genus, and species) can be found in Supplementary Fig. S3, as well as the same profile at the genus level across the four geographic areas.

Although there is no definitive evidence in the literature establishing a direct link between high gut microbial diversity and healthy status, several disorders, including inflammatory bowel diseases[13,14], obesity[15], and diabetes[16], have consistently been associated with low

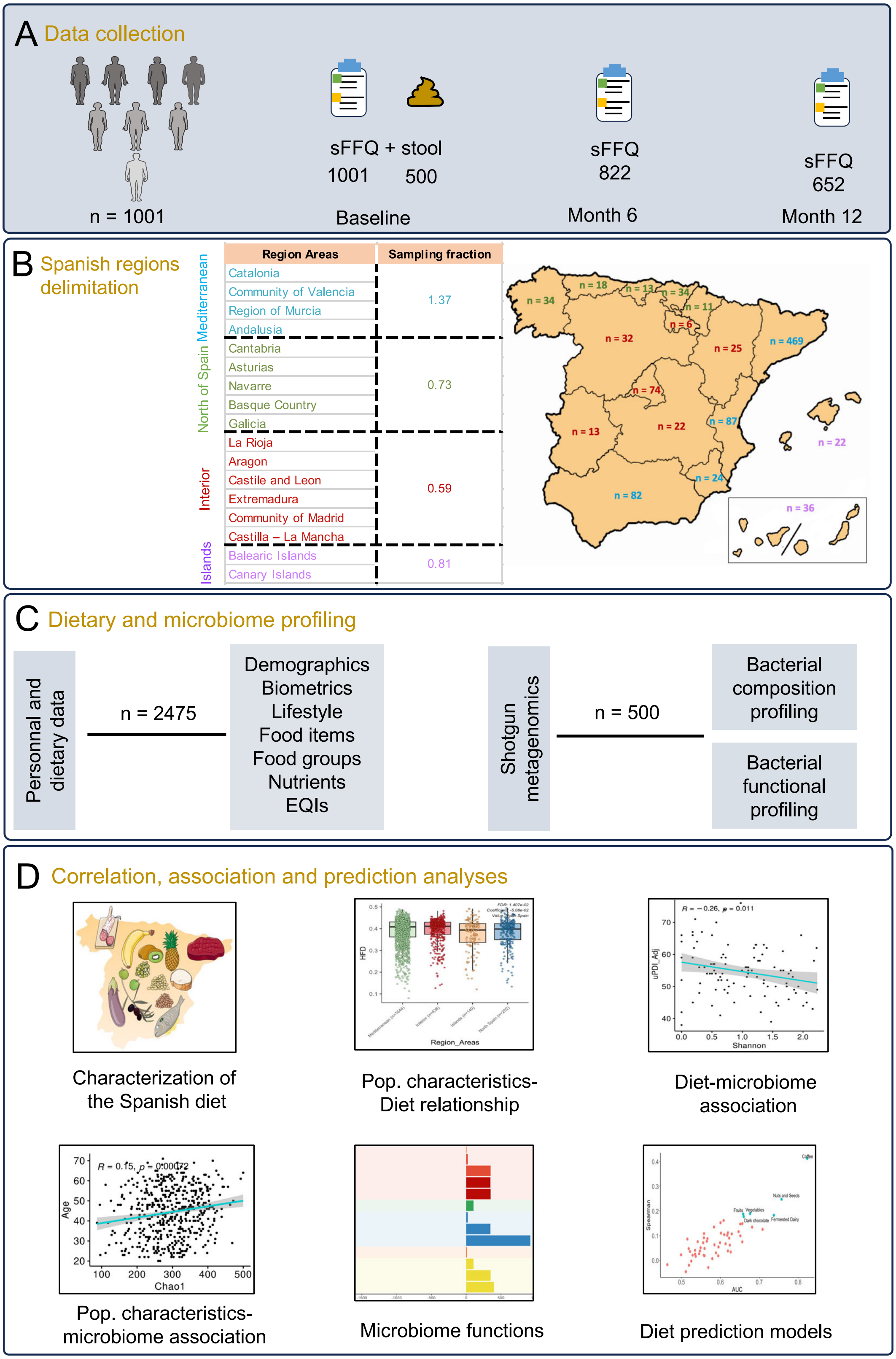

Fig. 1 | Study design. A 1001 participants reported their dietary intake and personal data through an in-house online short Food Frequency Questionnaire (sFFQ) at baseline, month six, and month 12 ($n = 2475$). Stool samples ($n = 500$) were processed at baseline for microbiome analysis. **B** Recruitment of participants from different autonomous regions of Spain and sampling fractions. The distribution of participants recruited from the 17 autonomous regions of Spain and the four regional areas is presented. The sampling fraction for each regional area was calculated based on the proportion of the population in each region, as reported by the Spanish government. **C** Information from the sFFQs was used to collect personal data and to calculate different Eating Quality indexes (EQIs). Extracted genomic DNA from stools was sequenced through a shotgun metagenomic approach, and sequences were processed to analyze microbiome composition and function. **D** Association analysis between microbiome and dietary data and diet prediction models. The association was performed using either the Spearman correlation test or the linear models implemented in the MaAsLin2 tool, and the predictions were performed using the random forest classification and regression algorithms.

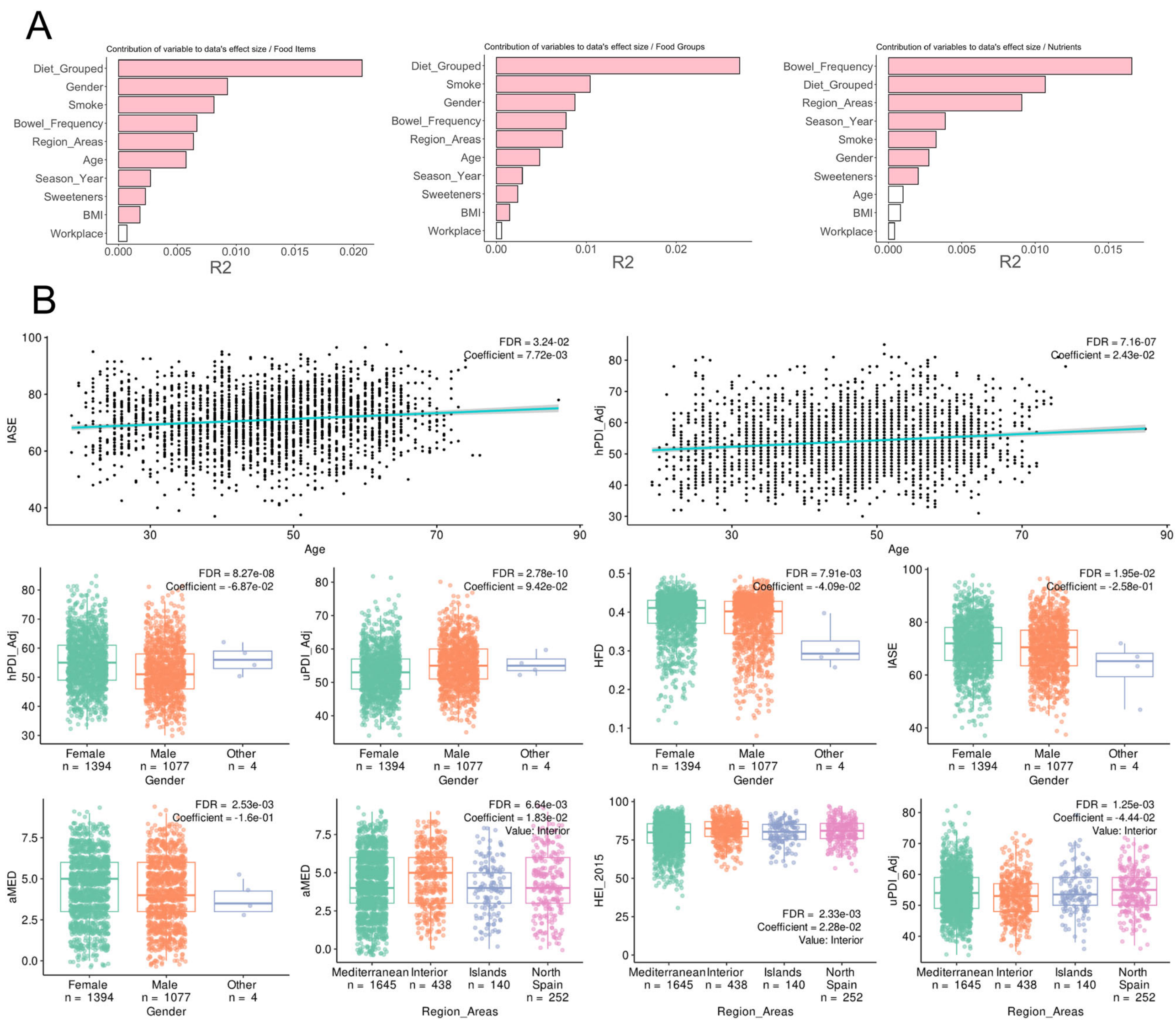

Fig. 2 | Relationship between population characteristics and dietary data. A Effect size of the population characteristics on dietary intake. The magnitude of the influence of specific characteristics on dietary intake was calculated using permutational analysis of variance (PERMANOVA), as implemented in the adonis2 function of the vegan R using the Bray-Curtis method. **B** Relationship between Eating Quality Indexes (EQIs) and population characteristics (age, gender, and region areas) was calculated using the MaAsLin2 tool. Data were plotted only when comparisons were significant. Correlation plots are shown for continuous data variables such as age and dietary indices, while boxplots are shown for categorical data variables such as gender.

microbial diversity. These associations suggest that a diverse gut microbiome plays a role in maintaining health. Using the Spearman test, we assessed the correlation between population characteristics, dietary data, and microbiome diversity. The results showed that diversity (based on Chao1 and Shannon indexes) was positively associated with vegetable intake (rho = 0.118, $P = 0.009$), fruits (rho = 0.160, $P = 0.0003$), and nuts and seeds (rho = 0.122, $P = 0.007$), while white bread and white grains were negatively linked to microbial diversity (rho = −0.152, $P = 0.0007$ and rho = −0.169, $P = 0.0002$, respectively) (Fig. 3A). This is further supported by the positive correlations between the Shannon index and dietary indexes such as the HEI-2015 (rho = 0.119, $P = 0.007$), the hPDI (rho = 0.138, $P = 0.002$), the aMED index (rho = 0.130, $P = 0.004$), which emphasize fruit and vegetable consumption (Fig. 3A). These results suggest that adherence to national dietary guidelines and recommendations was associated with increased microbial diversity.

Additionally, diversity (Shannon index, rho = 0.128, $P = 0.004$) and richness (Chao1 index, rho = 0.162, $P = 0.0003$) positively correlated with age, reinforcing the connection between older age and healthier eating habits (Fig. 3A). On the contrary, BMI (rho = −0.117, $P = 0.009$), and uPDI (rho = −0.142, $P = 0.001$) index were found negatively correlated with both richness and diversity. Given that age was also associated with BMI (rho = 0.31, $P < 0.05$), these findings suggest that higher diversity is linked to older age and lower BMI. We did not observe a seasonal effect on dietary intake and microbiome diversity (Fig. 3B).

Association analysis between metabolic pathways and dietary data revealed significant correlations between the L-arginine biosynthesis II and sucrose biosynthesis II pathways and the consumption of fruits, nuts, and seeds. At the nutrient level, significant associations were also found with fiber intake (Supplementary Table S5). These findings suggest that diet can influence not only the composition of the gut microbiome but also its functional capabilities.

The extent to which transit time (bowel movement) influences the microbiome is still not well understood. To address the question related to the impact of transit time on the microbiome community, we examined the association between defecation frequencies obtained from the sFFQs (categorized as 1.5 times/week, >3 times/week, 1 time/day, 2 times/day, and >2 times/day) on microbiome

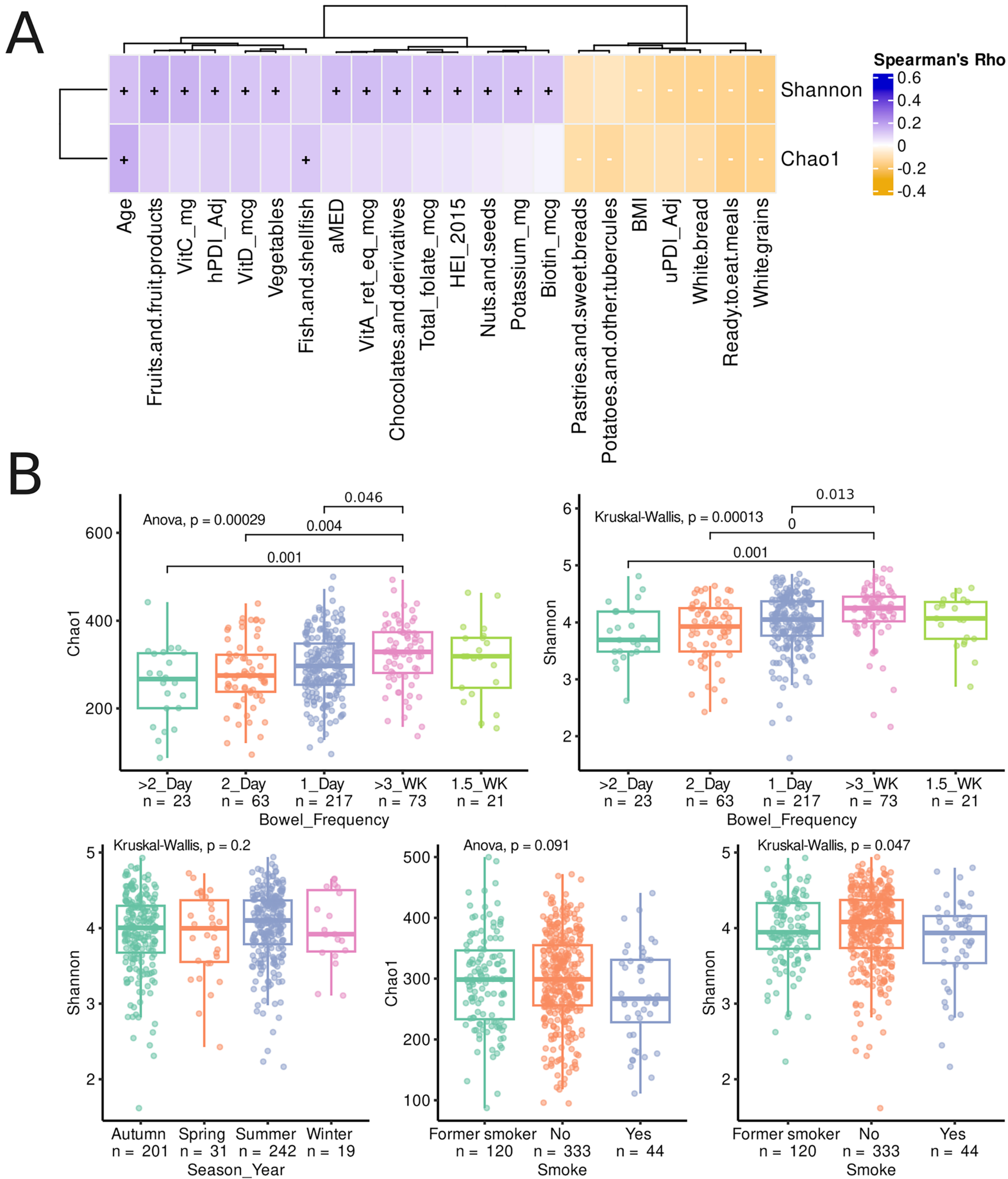

Fig. 3 | Population characteristics-microbiome alpha diversity association analysis. **A** Correlation between Eating Quality Indexes (EQIs), food groups, and personal data with alpha diversity (Chao1 and Shannon) using the Spearman correlation test ($n = 500$). Symbols + and − indicate significant correlations ($P < 0.05$). Only correlations with $P < 0.05$ and absolute rho > 0.11 are shown. **B** Differences in categorical population characteristics in relation to bacterial alpha diversity (Chao and Shannon indices), analyzed using the ANOVA test for normal data and Kruskal-Wallis test for non-parametric data, with the corresponding post-hoc tests ($n = 500$).

diversity and composition using the ANOVA or Kruskal-Wallis test and general linear models (MaAsLin2), respectively. Our results indicated that longer transit times were associated with higher diversity ($P < 0.05$, Fig. 3B). Additionally, we observed that microbiome diversity appeared to stabilize at a defecation frequency of more than 3 times per week, as indicated by non-significant differences in the Chao1 and Shannon indexes between defecating more than 3 times per week and 1.5 times per week. At the compositional level, using one defecation per day as a reference, 20 bacterial species (including *Akkermansia muciniphila*) were positively associated, while three species (including *Lacrimispora amygdalina* and *Blautia wexlerae*) were negatively associated with longer transit times (more than three times and only 1·5 times per week). Conversely, three species (*Ruthenibacterium lactatiformans*, *Eubacterium siraeum*, and *Alistipes putredinis*) were negatively associated with short transit times (>2 times per day) (Supplementary Table S6).

At the functional level, longer transit times were associated with more pathways than shorter transit times. These pathways include fermentation, glycan, amine degradation, amino acids degradation and biosynthesis, and lipid biosynthesis, while shorter transit times were more linked to carbohydrate degradation (Fig. 4, Supplementary Table S7, S8). Other correlations were found between microbiome diversity and demographic and biometric data including age, BMI, gender, season, and smoking (Fig. 3A, Supplementary Table S9). Interestingly, BMI, which correlated with three bacterial species, also correlated with 39 pathways (26 positive and 13 negative correlations).

Relationship between diet and IBD-related dysbiosis

To explore the link between diet and dysbiosis, we analyzed the microbiomes of 321 patients with IBD, comprising 208 with Crohn's disease and 113 with ulcerative colitis—two extensively studied non-communicable diseases associated with microbiome alterations. This shotgun metage-

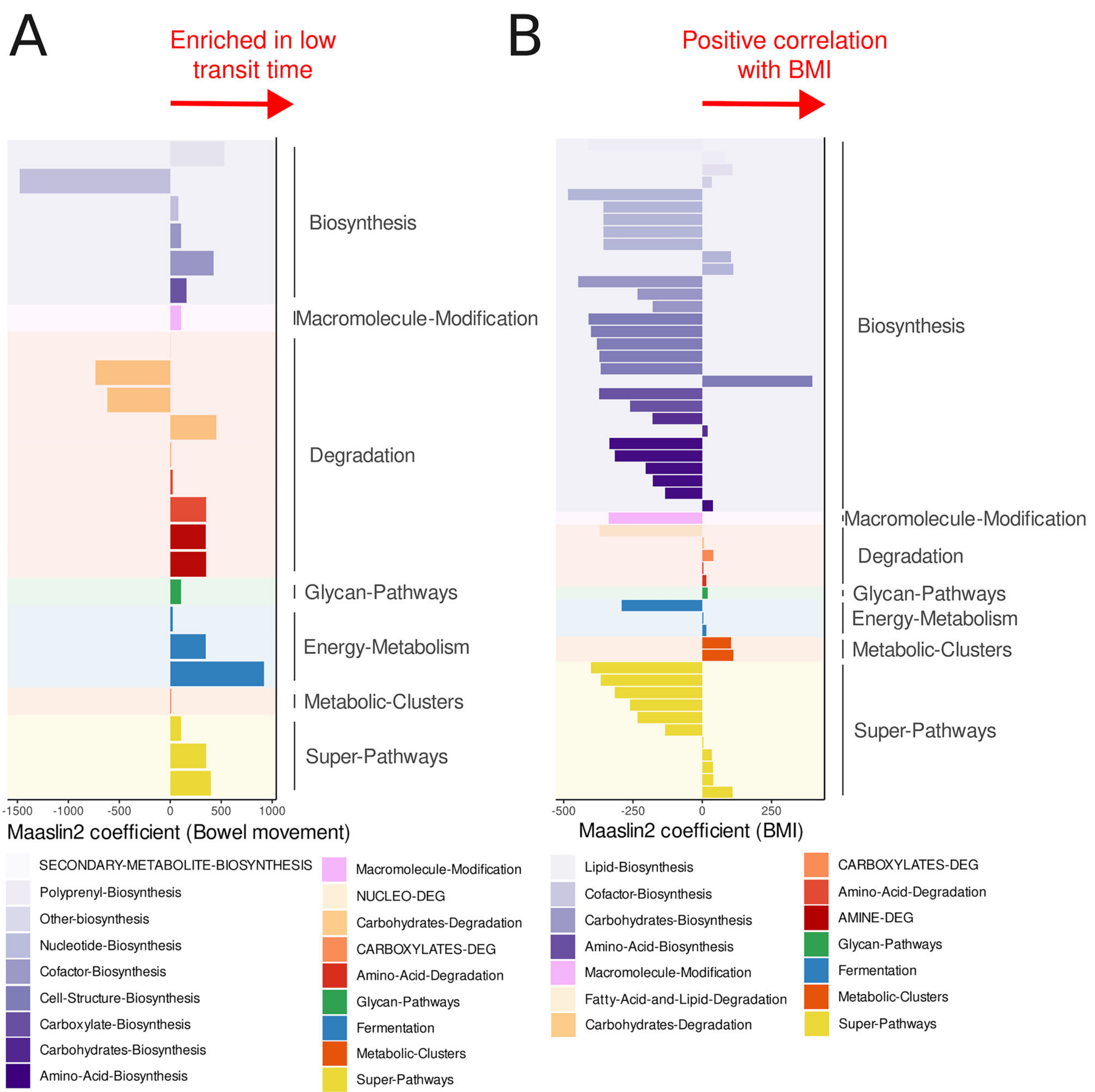

Fig. 4 | Differentially abundant metabolic pathways. Differentially abundant pathways in the microbiome of healthy individuals using to Maaslin2 depending on several conditions. Pathways were grouped according to their level 1 MetaCyc classes (broader functionality) and colored by their level 2 MetaCyc classes (more specific). Level 2 pathways assigned to more than one broader level parent were assigned to each of their level 1 functionality for plotting and interpretation purposes. Differentially abundant pathways were compared between low transit time (>3 times per week, 1 or 2 times per week) and the reference (once a day). **A** Positive coefficients reflected pathways enriched in low transit time, whereas negative coefficients represented their depletion. **B** Differentially abundant pathways were also associated with BMI, with positive coefficients indicating a positive correlation between pathway abundance and BMI.

nomic dataset was available from our previous projects[17]. To quantify microbial community disruption, we developed the IBD-similarity index, a metric measuring divergence from the microbiomes of 500 healthy individuals (see Methods section for detailed explanations). Higher index scores indicate a greater resemblance to IBD-associated microbial profiles. This approach effectively stratified the cohort by the degree of dysbiosis (Fig. 5A), explaining 36.3% and 15.3% of the variance in microbial composition along the first axis using weighted and unweighted UniFrac distances, respectively.

Spearman's correlation analysis revealed that higher alpha diversity in our cohort was associated with lower similarity to IBD microbiome profiles (Fig. 5B). Moreover, reduced consumption of vegetables, nuts, seeds, and fruits, combined with a higher intake of soft drinks, was linked to greater microbiome disruption (Fig. 5C). Personal traits such as age and BMI exhibited contrasting associations, with higher BMI correlating with increased dysbiosis. Correlation analysis between specific bacterial species, alpha diversity, and the disease-similarity index revealed that *Flavonifractor plautii* and *Ruminococcus gnavus* exhibited the strongest positive correlations with microbiome alterations. In contrast, the strongest negative correlations were associated with unidentified Clostridia and Bacilli species, as well as *Methanobrevibacter smithii*. Notably, all species that positively correlated with the disease-similarity index were inversely associated with alpha diversity metrics and vice versa.

Prediction of dietary intake by the gut microbiome

The "GBD 2017 Diet Collaborators" reported in 2019 that high intake of sodium and low intake of whole grains and fruits were the leading dietary risk factors for deaths and years of life adjusted for disability[2]. In our study, sodium was not properly evaluated in the questionnaire, as we did not add any specific question related to the added sodium

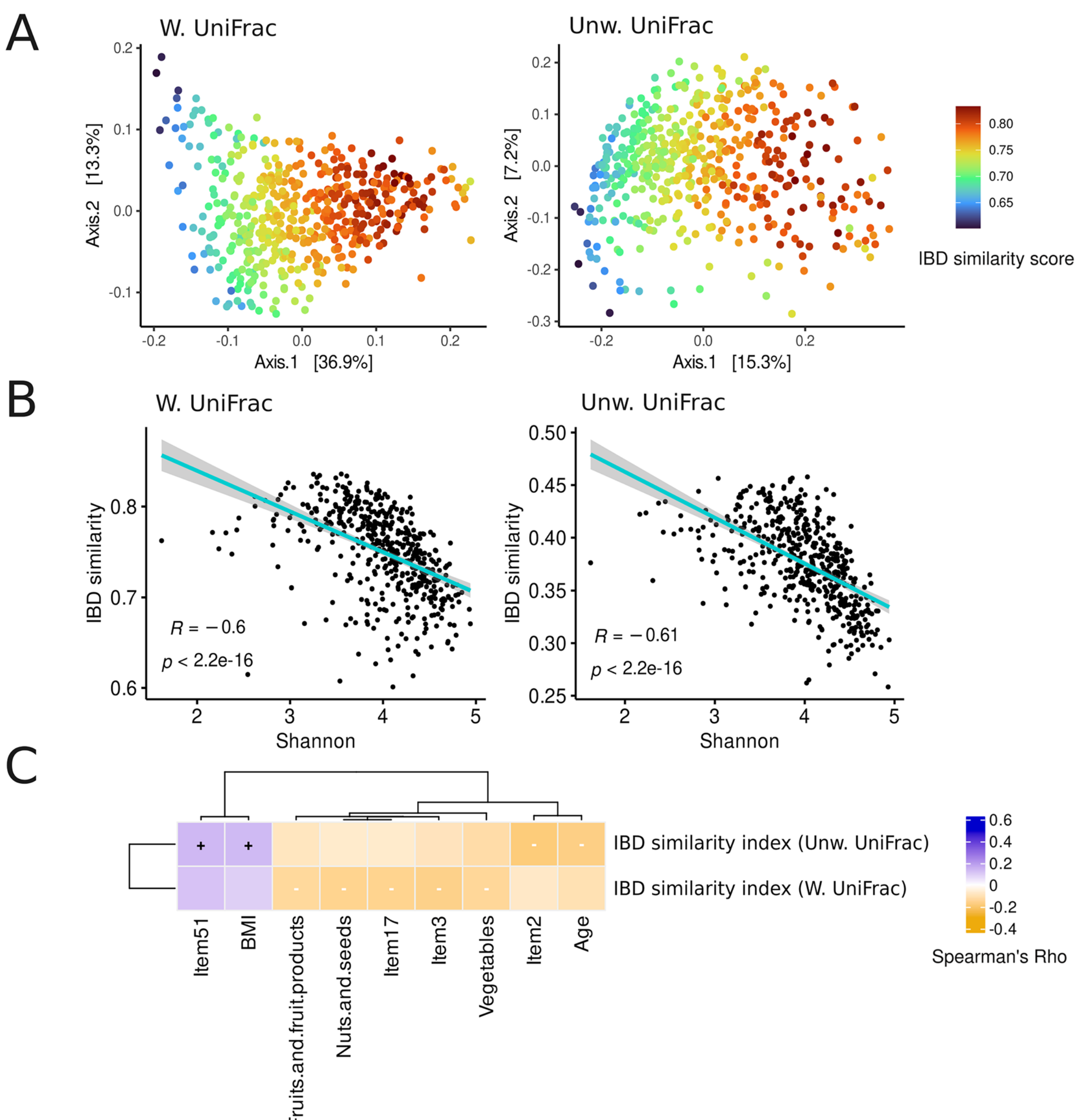

Fig. 5 | Disease similarity index and population characteristics. A Weighted (W.) and unweighted (Unw.) UniFrac distances of our cohort of healthy individuals ($n = 500$) colored by IBD-similarity score. IBD-similarity score was calculated as 1-median of a healthy sample to all samples in IBD plane ($n = 208$ CD and 113 UC) and can be a measure of how microbiome from a healthy individual resembles to the dysbiotic microbiome of IBD patients, which is widely accepted as an example of non-communicable disease. **B** Spearman correlation considering IBD similarity index and two different measures of alpha diversity. **C** Integrated heatmap representing food groups, items, EQIs, and personal traits that significantly impact the IBD similarity index. The more positive the IBD similarity value, the greater the resemblance to the IBD microbiome. FDR < 0.05. Item 2: Cooked leafy vegetables; Item 3: Tomato; Item 17: Nuts and seeds; Item 51: Soft drinks (see Supplementary Table S10 for more detailed diet information).

during the cooking process therefore, we cannot assess the impact of salt on the microbiome. Using a machine learning approach on microbiome features and the reported dietary data, as proposed by Manghi et al.[18], we showed that the consumption of several food items can be robustly predicted by the microbiome composition. These food items included coffee with and without caffeine (rho=0.41, AUC = 0.82), nuts and seeds (rho = 0.25, AUC = 0.76), vegetables (rho = 0.19, AUC = 0.67), fruits (rho = 0.19, AUC = 0.66), fermented dairy (rho = 0.18, AUC = 0.74), and dark chocolate (rho = 0.18, AUC = 0.66) (Fig. 6A). The analysis using food groups validated the findings with nuts and seeds (rho = 0.24, AUC = 0.75), fruits (rho = 0.20, AUC = 0.68), milk and dairy (rho = 0.20, AUC = 0.65), vegetables (rho = 0.19, AUC = 0.67), yogurt (rho = 0.17, AUC = 0.73), and chocolates (rho = 0.16, AUC = 0.66) (Fig. 6B). Furthermore, the composition of the microbiome was found to predict adherence to the Mediterranean diet (aMED score) as well as the intake of healthy vegetable proteins (hPDI index) (Fig. 6C).

Using Spearman correlation test, aMED and hPDI were the EQIs that correlated with the highest number of bacterial species, including *Ruminococcus torques* (aMED: rho = −0.22, q = 0.005; hPDI: rho = −0.20, q = 0.007), *Blautia massiliensis* (aMED: rho = −0.18, q = 0.030; hPDI: rho = −0.20, q = 0.010), and *Flavonifractor plautii* (aMED: rho = −0.19, q = 0.020).

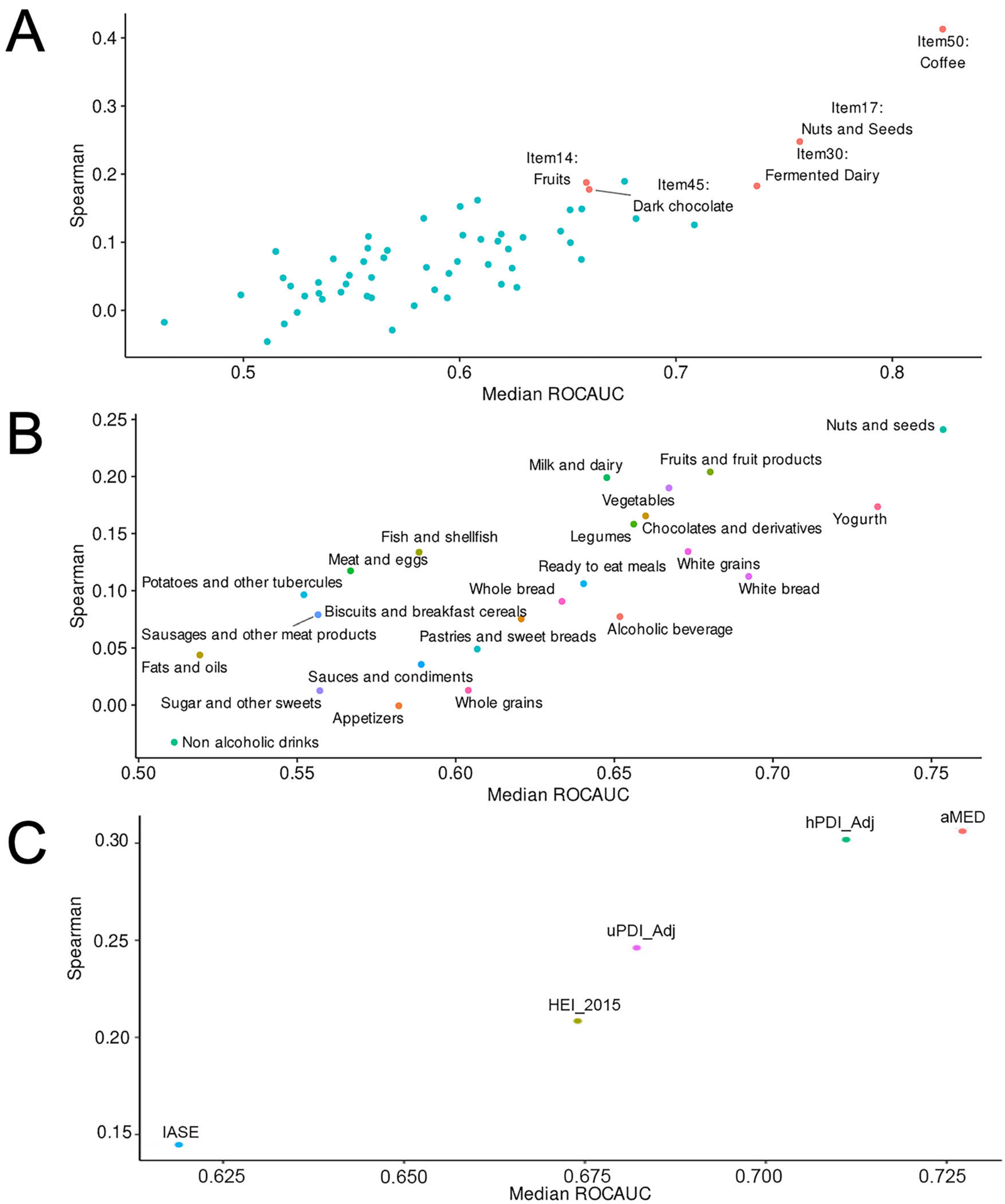

Fig. 6 | Prediction using machine learning technique. Prediction of different food items (**A**), food groups (**B**), and EQIs (**C**) using species-level genome bin (SGB)-level features information estimated by MetaPhlAn4. Y axis represents, for each variable, the median Spearman's correlation between observed values and predicted values from the random forest regression model. X axis represents, for each variable, the median receiver operating characteristic area under the curve (ROCAUC) from the random forest classifier. Further details on both random forest models can be found in the Methods section.

Website for the Citizen science project

This project was also designed to engage the public in data collection and raise awareness about scientific research. Participants contributed by providing their dietary data and stool samples. Through the website created for this project (https://manichanh.vhir.org/POP/en), participants were able to collect their dietary information using the sFFQ and ship their stool samples to the microbiome lab. Participants were provided with an overview of the study findings and received access to their personalized dietary and microbiome profiles at no cost. The website also offers resources to help participants understand the significance of their contributions and the impact of the research. The webpage is organized into two sections: "Study Results" and "Your Personal Traits," available in Catalan, Spanish, and English. Accessible to all, the "Study Results" section provides general information about the microbiome, diet, and their relationship. It outlines the study's objectives and methods, emphasizing the significant impact of participant involvement. In the "Your Personal Traits" section, personalized dietary information from the sFFQ is shared, helping participants track dietary changes over time and evaluate adherence to Spanish dietary recommendations. When shotgun sequencing results were available, participants could access them under the "Microbiome" section, which included: (1) Bacterial composition of prevalent species in the stool sample, from kingdom to species level, and (2) Population medians and α-diversity metrics (Chao1 and Shannon indices). Examples of how this data is presented in the webpage can be found in the Supplementary Fig. S4.

Discussion

This study provided new insights into the intricate relationships between Eating Quality Indexes, personal traits, geography, and diet, and their collective impact on the gut microbial community. It also highlighted how national dietary recommendations can shape this community. Notably, the study introduced a comprehensive web tool designed to help participants understand the influence of diet on their gut microbiome.

EQIs have been developed to serve as comprehensive tools for evaluating diet quality and guiding dietary recommendations. Researchers use EQIs to facilitate research on how diet affects the risk of chronic diseases, such as obesity, diabetes, cardiovascular diseases, and certain cancers[19]. In the present study, the assessment of the impact of the population characteristics on the nutritional quality revealed crucial insights into how age, gender, geographical location, and lifestyle shape eating habits. Our findings, reporting healthier dietary habits as we age, are validating previous works that showed that older adults have a more "prudent" dietary pattern characterized by higher intakes of vegetables, fruits, whole grains, nuts, and seeds[20,21]. In our study, we excluded individuals older than 75 years to avoid potential confounding factors, such as age-related undiagnosed diseases like frailty or early-stage neurodisorders.

Using a machine learning approach, the study identified key food items and food groups strongly associated with microbiome composition. Coffee, nuts and seeds, vegetables, fruits, fermented dairy, and dark chocolate emerge as significant predictors of microbial composition. As a Mediterranean country, Spain's traditional diet is rich in fruits, legumes, whole grain cereals, vegetables, nuts, and healthy unsaturated fats primarily from olive oil. It also includes frequent fish intake, moderate consumption of dairy products and fermented beverages, and a low intake of meat and meat-derived products[22]. Despite its benefits, adherence to the Mediterranean diet (MD) in Spain has decreased over time, shifting towards a more Western dietary pattern[23–25].

The influence of regional dietary habits, particularly within Mediterranean countries, is well-known. Our study's division of Spain into the Mediterranean, Interior, North, and Islands, and its identification of healthier dietary patterns in the Interior region, aligns partially with prior research showing geographical variability in adherence to the Mediterranean diet and other dietary patterns[26]. Moreover, our study showed that individuals from the Interior region were characterized by higher consumption of legumes, which offer a range of health benefits due to their rich nutrient content and bioactive compounds, including protein, fiber, vitamins, and minerals.

Among the dietary variables proposed by the Global Burden of Disease study, our Spanish cohort complied with only 3 out of the 12 food groups analyzed: vegetables (321·48 g/day), fruits (225·6 g/day), and fiber (27·32 g/day). These groups were related to higher alpha diversity and lower IBD related dysbiosis, and correlated with bacterial species with potential health implications. For instance, vegetables were negatively correlated with *Flavonifractor plautii*, a flavonoid-degrading bacterium associated with less healthy diets, lower scores in EQIs, and related to disease outcomes such as IBD.

The association analysis of food group consumption reveals gender-specific dietary behaviors. It is recognized that women generally exhibit healthier dietary patterns than men, consuming more fruits, vegetables, and whole grains, while men consume more meat and alcohol[27,28]. These findings are validated by our study, which shows that men have a higher consumption of ready-to-eat meals and alcoholic beverages.

Low microbial diversity and alterations in the microbiome composition have been linked to various disorders, suggesting a connection between health status and high microbial diversity[13,14]. This study highlights that adherence to national dietary guidelines—especially increased consumption of fruits, vegetables, fiber, nuts, and seeds—is positively associated with greater microbial diversity and lower levels of dysbiosis. In contrast, following an unhealthy diet, characterized by high intake of white bread, negatively impacts microbial richness and diversity, while excessive consumption of soft drinks adversely affects microbial composition. These findings align with previous reports indicating that a high-fiber diet enhances alpha diversity, while a low-fiber diet, such as one high in white bread, reduces it[29]. Additionally, they suggest that a high-fiber diet may help prevent IBD-related dysbiosis.

A key component of this project was the development of a website, which allowed participants to efficiently collect and submit their dietary information using a structured Food Frequency Questionnaire (sFFQ). As part of the growing trend in citizen science projects, the website also provides participants with private access to both the overall study findings and their personalized dietary and microbiome profiles, enhancing their understanding of their contributions. Additionally, we ensured that the website offers comprehensive resources to help participants appreciate the significance of their involvement in a citizen science project and the broader impact of the research. This integrated approach not only facilitated data collection but also strengthened the connection between the participants and the scientific community.

While our findings provide valuable insights, certain limitations should be acknowledged to contextualize the results. First, sFFQ relies primarily on self-reported and subjective data based on participants' memory, which can lead to over- or underestimation of dietary intake. Consequently, some of the results, particularly those related to nutrient intake, should be interpreted with caution, as they may reflect misreporting. This limitation is consistent with findings from our previous study[11]. Second, the nutritional tables used for dietary assessment may be considered limited, as they do not include information on certain components such as additives, cooking methods, and preservatives, which may influence microbiome composition. Third, using an IBD cohort to calculate a dysbiosis score for each participant may represent another limitation of this study, as it does not account for microbial alterations associated with other chronic disorders. While our method offers valuable insights into the links between diet and IBD-related microbiome alterations, further investigation is necessary to determine whether similar associations exist in other chronic conditions and to clarify how these microbial patterns may differ from or overlap with those observed in IBD. Fourth, although several of our findings revealed noteworthy correlations between dietary patterns and microbiome, some of which align with results from previous observational studies, it is essential to stress that correlation does not imply causation. Establishing causality requires experimental validation. Finally, due to budgetary constraints, we were unable to sequence all collected samples. However, we intend to complete the sequencing of the remaining samples as soon as additional funding becomes available. This will enable us to explore more specific research questions, including those related to particular dietary patterns (e.g., vegetarian and vegan diets) or ethnic backgrounds.

Despite these limitations, our study represents one of the most comprehensive investigations of gut microbiome composition and function in relation to dietary patterns and lifestyle factors in the Spanish population. By integrating metagenomic sequencing with dietary, clinical, and sociodemographic data in a large citizen science framework, we provide a robust foundation for future studies aiming to unravel diet–microbiome–health relationships. The depth and diversity of the dataset offer valuable opportunities for further hypothesis-driven and translational research, particularly as we continue to expand the cohort and validate key findings experimentally.

Methods

Participant's recruitment

We conducted a prospective longitudinal study in accordance with the Declaration of Helsinki, approved by the local Ethics Committee of Vall d'Hebron University Hospital, Barcelona (PR(AG)84/2020). Participants were enrolled in the study between December 2020 and March 2024 through announcements on social platforms such as Facebook, LinkedIn, and Instagram, as well as the Hospital Vall d'Hebron website. We recruited 1001 participants from different regions of Spain, aged 18–75, who had not taken antibiotics for at least three months and had no diagnosed chronic intestinal disorders, including inflammatory bowel diseases, type 2 diabetes, and autoimmune diseases, before entering the study. All participants signed a consent form.

To calculate the sampling fraction for each region area, we first downloaded the data from the Instituto Nacional de Estadística (INE) (https://www.ine.es/jaxiT3/Tabla.htm?t=2853&L=0) regarding the number of males and females between 18 and 75 years old in each autonomous community. We then calculated the population size for the selected region areas (Interior, North of Spain, Mediterranean, and Islands) by summing up the individuals from the corresponding autonomous communities. Using these values, we estimated the theoretical percentage for a sample size of 1000 individuals as follows: Theoretical percentage = (1000 x Population in each region area)/Total population in Spain. To evaluate how accurately we achieved our recruitment goal, we divided the actual number of individuals recruited in each region area by the theoretical values. This resulted in a ratio ranging from 0 to 1, where a ratio closer to 1 indicates more accurate recruitment.

Metadata and sample collection

Participants filled out an in-house validated short food frequency questionnaire (sFFQ)[11], which provided demographic, lifestyle, clinical, and dietary data, and shipped their stool samples to the microbiome laboratory at baseline, month six, and month 12. The questionnaire was administered online (https://manichanh.vhir.org/sFFQ/login.php). It included 58 food items divided into 13 sections (Supplementary Table S10): vegetables, legumes, and potatoes; fruits and dried fruits; cereals and derivatives; milk and derivatives; eggs, fish, and meat; selfish; oils and fats; bakery and pastry; sauces; non-alcoholic drinks; alcoholic drinks; processed food and others. Frequency of consumption was categorized into six possible options: "Never", "1 or 3 times per month", "1 or 2 times per week", "3 or more times per week", "once per day", and "2 or more times per day". Serving size consisted of a "standard portion" estimated using the ENALIA2 Survey[30] as well as our own expertise, "half of the standard", and "double of the standard". To facilitate the estimation of the amount of food consumed by the participants, we added colored photographs. Additional information such as age, sex, weight, height, birth type, smoking, blood type, specific diet, consumption of ready-to-eat food or sweeteners, liquids and supplements, or medication was also recorded. The participants also self-collected their stool samples. The samples were preserved in 97% ethanol and stored in a domestic freezer until they were shipped by the participants to the microbiome lab, where they were maintained at -80°C until DNA extraction.

Dietary data processing

The first step in converting the dietary information collected from the sFFQ was to transform monthly consumption into daily consumption: for instance, a consumption response of 1-2 times per week was interpreted as an average consumption of 1·5 times per week, which, when divided by the seven days of the week, yielded an average daily consumption of 0.21. Subsequently, this consumption value was multiplied by the weight associated with the selected serving size. For instance, for the legume item with a serving size of 150 g and the aforementioned consumption frequency, the final value of grams per day would be 0.21 × 150 g = 31·5 g/day. The values for the other consumption frequencies were as follows: 1–3 times per month = 0.066; +3 times per week = 0.64; once per day = 1; +2 times per day = 3. Using this gram-per-day information, the energy and nutritional value of each item in the sFFQ were then calculated utilizing a custom-developed food composition database[11].

We calculated the magnitude of the influence of specific participant's characteristics on dietary intake using permutational analysis of variance (PERMANOVA), as implemented in the adonis2 function of the vegan R package (https://cran.r-project.org/web/packages/vegan/index.html) with the Bray-Curtis method. The correlation between eating quality indexes and continuous population characteristics was calculated using the Spearman correlation test. For categorical data, the Mann-Whitney U test was used.

Dietary indexes

We utilized various eating quality indexes to assess the nutritional quality of diets. These indexes encompass the Healthy Eating Index-2015 (HEI-2015), the IASE (derived from its Spanish acronym 'Índice de Alimentación Saludable para la Población Española´), the plant-based dietary indexes PDI, uPDI (u=unhealthy), hPDI (h=healthy), the Healthy Food Diversity Index (HFD-index), and the Alternative Mediterranean Diet (aMED) score.

The HEI-2015, developed by the United States Department of Agriculture (USDA), is a scoring system designed to provide recommended nutritional guidelines to promote health and prevent chronic diseases[31]. It assesses the intake of different food groups and nutrients, assigning scores to components such as fruits, vegetables, whole grains, dairy, protein foods, fatty acids, refined grains, sodium, added sugars, and saturated fats. Higher scores indicate better adherence to dietary guidelines, with the maximum score for each component representing optimal intake according to the guidelines.

The IASE is a modified version of the HEI-2005, specifically tailored to assess the dietary quality of the Spanish population in 2011[32]. Similar to the HEI-2005, the IASE evaluates dietary patterns and adherence to dietary guidelines, but with considerations for the specific food choices and dietary habits commonly found in Spain. The IASE takes into account various components of the diet, including the consumption of fruits and vegetables, cereals and grains, proteins, dairy products, fats and oils, sweets, pastries, and alcoholic beverages. It assesses the quality of these food groups based on recommended intake levels and patterns that are more relevant to the Spanish diet and nutritional guidelines.

Introduced by Satija et al. in 2017[33], the PDI, uPDI, and hPDI evaluate the quality of a person's diet based on various aspects of dietary intake in the US. The PDI assesses the proportion of plant-based foods consumed relative to animal-based foods. A higher PDI score indicates a diet richer in plant-based foods like fruits, vegetables, whole grains, nuts, and seeds, with lower consumption of animal-based foods such as meat and dairy. The uPDI focuses on less healthy plant-based items like refined grains, potatoes, and sweets, with a higher score suggesting an increased intake of these less nutritious plant-based foods. In contrast, the hPDI emphasizes the consumption of healthier plant-based foods within a plant-based diet, such as fruits, vegetables, whole grains, nuts, and legumes, with a higher hPDI score reflecting a diet rich in these nutrient-dense plant-based food groups.

The HFD, developed by Dresher et al. in 2007[34], measures food intake diversity by evaluating the intake of various food groups including fruits, vegetables, whole grains, lean proteins, and healthy fats. A higher HFD-index score generally indicates a more diverse and nutritious diet.

The aMED score corresponds to a scoring system developed by Fung et al.[35], which is based on the original Mediterranean diet scale proposed by Trichopoulou et al.[36]. The aMED score ranges from 0 (indicating minimal adherence) to 9 (representing perfect adherence) points and evaluates adherence to nine food groups: 1) All kinds of vegetables excluding potatoes; 2) Legumes including tofu, beans, and peas; 3) Fruits and fruit juices; 4) Nuts including peanut butter; 5) Whole grains; 6) Red and processed meat; 7) Fish and shellfish; 8) Ratio of monounsaturated to saturated fat; 9) Alcoholic drinks. For each category, including the fatty acid ratio, the median intake was calculated in grams per day. Healthy food groups (vegetables, legumes, fruits, nuts, whole grains, fish, and the fatty acid ratio) were scored with 1 if the participant's intake was above the median and 0 if it was below. Conversely, for red and processed meats, 1 point was assigned if participants reported lower intake compared to the median, while 0 points were given for higher intake. Alcoholic drinks were scored differently. For men, consumption between 10 and 50 g per day or 5–25 g per day for females received 1 point, while intake outside these ranges received a score of 0.

Microbiome analysis

Genomic DNA was extracted following the recommendations of the International Human Microbiome Standards (IHMS; http://www.microbiome-standards.org). Briefly, a frozen aliquot (200 mg) of each sample was suspended in 250 µL of guanidine thiocyanate, 40 µL of 10% N-lauryl sarcosine, and 500 µL of 5% N-lauryl sarcosine. Mechanical disruption of the microbial cells with beads was applied, and nucleic acids were recovered from the lysates through ethanol precipitation[17].

The DNA shotgun library was prepared and sequenced using the Illumina NovaSeq6000 platform. The sequencing process provided an average of 5 Gb of sequence data per sample. The KneadData v0.7.4 pipeline was used to pre-process and decontaminate the sequence reads (https://huttenhower.sph.harvard.edu/kneaddata). KneadData performed a quality filtering of the reads using trimmomatic and then mapped them against a human reference genome database using Bowtie 2. Reads with lengths below 50% of the total input length and also those that mapped with the human genome were discarded from further analysis. Taxonomic profiles were provided by the MetaPhlan4's intermediary output file in the HumanN3 pipeline and functional profiles from the final output[37]. Taxonomic profiles, the outputs of MetaPhlan4, were generated in stratified relative abundance, from phylum to SGB level. For this reason, no normalization was applied, but the stratified relative abundances were extracted according to the taxonomic species level. Alpha and beta diversity analyses were computed using Chao1 and Shannon indexes[17] and the adonis2 function (Permutational Multivariate Analysis of Variance), respectively.

Functional profiles, the output of HumanN3, provided gene families and MetaCyc pathways. MetaCyc pathways were filtered to remove unmapped and unintegrated reads. All features that did not achieve 0.001 abundance and 0.1 prevalence (pathways that did not achieve 0.1% of the total sample abundance in at least 10% of the samples) were also discarded. Then, pathways were sum-normalized to counts per million (CPM) before further analysis.

Comparison of dietary intake with recommendations from the GBD-2017 consortium

To compare major food and nutrient consumption within the context of the Global Burden of Disease (GBD) study, we grouped our semi-quantitative sFFQ items into 12 out of the 15 proposed dietary risk factors defined by the GBD, aiming to align with their dietary profiles. We calculated the median intake in grams per day for fruits, vegetables, legumes, whole grains, nuts and seeds, milk, red meat, processed meat, sugar-sweetened beverages, fiber, and calcium, and compared these values with the optimal and optimal range of intake defined in the GBD study. For polyunsaturated fatty acids (PUFAs), we calculated their consumption percentage relative to the total energy intake and compared it with the GBD recommended values. Sodium was omitted from our analysis as our data only reflected sodium present in food and did not account for sodium added during cooking. Additionally, seafood omega-3 and trans fatty acids were not evaluated due to the absence of these variables in our sFFQ. Supplementary Table S2 listed the clustering of items into the dietary risk factors as suggested by the GBD consortium.

Development of a disease similarity index

We developed an estimator (disease similarity index) to quantify the similarity between the microbiome composition of healthy individuals and those of patients with non-communicable gastrointestinal diseases. To do so, we included a cohort of patients with IBD. Sequence data for this cohort were obtained from a previous study[38]. This index is defined as one minus the median weighted or unweighted UniFrac distance between a healthy sample and a reference set of 321 IBD samples (208 from Crohn's disease patients and 113 from ulcerative colitis patients). To compute this, we first calculated both weighted and unweighted UniFrac distances between a plane of IBD-affected individuals and each of the healthy participants of the study. To determine the distribution of these distances, we used the Shapiro-Wilk normality test. For the unweighted UniFrac metric, only 6 out of 491 participants exhibited normally distributed distances, while 491 did not. For the weighted UniFrac, 497 out of 497 distance distributions were non-normally distributed. Given the widespread non-normality, we selected the median rather than the mean as a more robust and representative measure of central tendency for defining the IBD-similarity index.

Statistical analyses

Microbiome sequence data were performed in R (v4.3). Covariates such as gender, age, body mass index (BMI), region areas, smoking habit, season, and workplace were tested for their impact on microbiota variation using the PERMANOVA test on weighted and unweighted UniFrac distance indexes.

We evaluated the gut microbiome's capacity to predict individual food items, food groups, and nutrient intakes using both Random Forest classifiers and regressors. For each task, we performed 100 bootstrap iterations with an 80/20 split between training and test sets to ensure robust performance estimates. Classification setup: Frequencies of food items, groups, and nutrients were divided into "low" (first quartile) and "high" (fourth quartile) consumption classes. We trained Random Forest classifiers on species-level genome bin (SGB) relative abundances generated by MetaPhlAn4. Model discrimination was assessed by the median area under the ROC curve (AUC) across the 100 test folds. Regression setup: Continuous intake values were predicted with Random Forest regressors, also trained on MetaPhlAn4 SGB relative abundances. Performance was quantified by the median Spearman correlation between observed and predicted values in the held-out data.

Given the compositional nature of the sequence data, differential abundance (DA) analysis of the microbial community was performed using MaAsLin2 (Multivariate Association with Linear Models)[39]. The analysis tested for differences in population microbiome while including bowel movement, gender, BMI, age, smoking habit, and season as fixed effects, as they showed a significant effect on the microbiome composition. To control the false discovery rate (FDR), the resulting p-values were adjusted using the Benjamini–Hochberg (BH) method and, when applicable, referred to as q-values. Associations identified by MaAsLin2 were considered significant if the coefficient, measuring the strength and direction of the association, was greater than 1 (in most cases) and the q-value was less than 0.05. Spearman tests were used to correlate dietary data with microbiome data.

For functional analysis, Spearman's correlation between alpha diversity indexes (Chao1 and Shannon) and pathway abundances was computed, and *p*-values were FDR (BH) corrected and referred to as q-values. Correlations with $-0.4 <=$ rho or $>= 0.4$ and FDR < 0.05 were considered significant and kept for further analyses. Association analysis was performed between these pathways and food items, food groups, and nutrients using the Spearman correlation test.

To assess changes in the potential pathways of the microbial community depending on personal information, we used linear models as implemented in MaAsLin2, adjusting for variables that showed significant effects on the microbiome composition, such as bowel movement (transit time), gender, BMI, age, smoking habit, region area, and season years as fixed-effects, using MetaCyc pathways information. To increase the interpretability of these results, pathways were grouped into their MetaCyc parent instances up to 7 levels, in which each level represents a broader biological function, with level 1 being the broadest and 7 the most specific. Pathways with more than one parent instance were duplicated and assigned to different parents for plotting and interpretation purposes.

Website construction: initiative for the general public

We built a website dedicated to this study (https://manichanh.vhir.org/POP/en/), where participants can access an overview of the results of this research, as well as their personal information on nutrient intake and dietary indices (based on the short food frequency questionnaire), and, if available, their microbiome sequencing results, including bacterial composition, and measures of alpha diversity. Nutrient intake data are compared to the guidelines established by the Scientific Committee of the Spanish Agency for Food Safety and Nutrition (AESAN), while dietary indices and alpha diversity scores are compared to the population median found in this study. Nutrient intake data and EQIs can be visualized across the different time points when each participant completed the sFFQ survey, allowing for the tracking of their progression over the 12-month period. Participant reports are produced dynamically in the form of a Shiny app (https://shiny.posit.co/), which is run in R language and hosted in our local Shiny server. All personal results are anonymized and password-protected, ensuring each participant may only access their own information.

Data availability

The statistical analyses and data visualization in this study were performed using R (version 4.3). Custom R scripts used for diversity metrics, random forest modeling, differential abundance testing, correlation analyses, and the development of the disease similarity index were home-made and tailored specifically for this project. These scripts are relatively short and not organized as reusable packages; therefore, they have not been deposited in any public repository. They are available from the corresponding author upon reasonable request. Parameters and software versions used for key analyses are detailed in the "Methods" section. The Shiny web application was built using the Shiny framework in R and is hosted locally for secure access by study participants. Data collected for the study include individual participant data and microbiome sequence data. Participants were codified. Shotgun metagenomic sequencing raw data (short-read archives, SRA) are available via NCBI Project Number PRJNA1146994.

Received: 28 January 2025; Accepted: 25 June 2025;
Published online: 11 July 2025

References

1. Collaborators, G. B. D. I. B. D. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* **5**, 17–30 (2020).
2. Collaborators, G. B. D. D. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **393**, 1958–1972 (2019).
3. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
4. McDonald, D. et al. American Gut: an open platform for citizen science microbiome research. *mSystems* **3**, https://doi.org/10.1128/mSystems.00031-18 (2018).
5. Walker, R. L. et al. Population study of the gut microbiome: associations with diet, lifestyle, and cardiometabolic disease. *Genome Med.* **13**, 188 (2021).
6. Turpin, W. et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
7. Sun, S. et al. Does geographical variation confound the relationship between host factors and the human gut microbiota: a population-based study in China. *BMJ Open* **10**, e038163 (2020).
8. Asnicar, F. et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat. Med.* **27**, 321–332 (2021).
9. Estruch, R. et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N. Engl. J. Med.* **368**, 1279–1290 (2013).
10. Muralidharan, J. et al. Effect on gut microbiota of a 1-y lifestyle intervention with Mediterranean diet compared with energy-reduced Mediterranean diet and physical activity promotion: PREDIMED-Plus Study. *Am. J. Clin. Nutr.* **114**, 1148–1158 (2021).
11. Yanez, F. et al. Integrating dietary data into microbiome studies: a step forward for nutri-metaomics. *Nutrients* **13**, https://doi.org/10.3390/nu13092978 (2021).
12. Herrera-Ramos, E. et al. Trends in adherence to the Mediterranean diet in Spanish children and adolescents across two decades. *Nutrients* **15**, https://doi.org/10.3390/nu15102348 (2023).
13. Halfvarson, J. et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2**, 17004 (2017).
14. Pascal, V. et al. A microbial signature for Crohn's disease. *Gut* **66**, 813–822 (2017).
15. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
16. Pedersen, H. K. et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376–381 (2016).
17. Serrano-Gomez, G. et al. Dysbiosis and relapse-related microbiome in inflammatory bowel disease: a shotgun metagenomic approach. *Comput. Struct. Biotechnol. J.* **19**, 6481–6489 (2021).
18. Manghi, P. et al. Coffee consumption is associated with intestinal Lawsonibacter asaccharolyticus abundance and prevalence across multiple cohorts. *Nat. Microbiol.* **9**, 3120–3134 (2024).
19. Shang, X. et al. Healthy dietary patterns and the risk of individual chronic diseases in community-dwelling adults. *Nat. Commun.* **14**, 6704 (2023).
20. Hiza, H. A., Casavale, K. O., Guenther, P. M. & Davis, C. A. Diet quality of Americans differs by age, sex, race/ethnicity, income, and education level. *J. Acad. Nutr. Diet.* **113**, 297–306 (2013).
21. Nicklett, E. J. & Kadell, A. R. Fruit and vegetable intake among older adults: a scoping review. *Maturitas* **75**, 305–312 (2013).
22. Davis, C., Bryan, J., Hodgson, J. & Murphy, K. Definition of the Mediterranean diet; a literature review. *Nutrients* **7**, 9139–9153 (2015).
23. Leon-Munoz, L. M. et al. Adherence to the Mediterranean diet pattern has declined in Spanish adults. *J. Nutr.* **142**, 1843–1850 (2012).
24. Moreira, A. C. et al. Nutritional status influences generic and disease-specific quality of life measures in haemodialysis patients. *Nutr. Hosp.* **28**, 951–957 (2013).
25. Varela-Moreiras, G. et al. Evaluation of food consumption and dietary patterns in Spain by the Food Consumption Survey: updated information. *Eur. J. Clin. Nutr.* **64**, S37–43 (2010).
26. Abellan Aleman, J. et al. Adherence to the "Mediterranean Diet" in Spain and Its Relationship with Cardiovascular Risk (DIMERICA Study). *Nutrients* **8**, https://doi.org/10.3390/nu8110680 (2016).
27. Fenton, S., Ashton, L. M., Lee, D. C. W. & Collins, C. E. Gender differences in diet quality and the association between diet quality and BMI: an analysis in young Australian adults who completed the Healthy Eating Quiz. *J. Hum. Nutr. Diet.* **37**, 943–951 (2024).
28. White, A. M. Gender differences in the epidemiology of alcohol use and related harms in the United States. *Alcohol Res.* **40**, 01 (2020).
29. Wang, Y., Wymond, B., Tandon, H. & Belobrajdic, D. P. Swapping white for high-fibre bread increases faecal abundance of short-chain fatty acid-producing bacteria and microbiome diversity: a randomized, controlled, decentralized trial. *Nutrients* **16**, https://doi.org/10.3390/nu16070989 (2024).
30. Marcos Suarez, V., Rubio Mañas, J., Sanchidrián Fernández, R. & Robledo De Dios, T. Spanish National dietary survey on children and

adolescents. *EFSA Supporting Publications* **12**, https://doi.org/10.2903/sp.efsa.2015.en-900 (2015).
31. Reedy, J. et al. Evaluation of the Healthy Eating Index-2015. *J. Acad. Nutr. Diet.* **118**, 1622–1633 (2018).
32. Norte Navarro, A. I. & Ortiz Moncada, R. Spanish diet quality according to the healthy eating index. *Nutr. Hosp.* **26**, 330–336 (2011).
33. Satija, A. et al. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S. adults. *J. Am. Coll. Cardiol.* **70**, 411–422 (2017).
34. Drescher, L. S., Thiele, S. & Mensink, G. B. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J. Nutr.* **137**, 647–651 (2007).
35. Fung, T. T. et al. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *Am. J. Clin. Nutr.* **82**, 163–173 (2005).
36. Trichopoulou, A. et al. Diet and overall survival in elderly people. *BMJ* **311**, 1457–1460 (1995).
37. Blanco-Miguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
38. Serrano-Gomez, G. Microbiome multi-omics analysis reveals novel biomarkers and mechanisms linked with CD etiopathology. *Biomark Res* **13**, 85 (2025).
39. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).

Acknowledgements
A preprint version of the current manuscript is available on Research Square at the following link: https://doi.org/10.21203/rs.3.rs-4990604/v1, under the title: 'Personal traits, lifestyle decisions, and geography shape our dietary intake and consequently our bacterial and fungal gut microbiome'. This work was financially supported by the Instituto de Salud Carlos III and co-funded by the European Union (ERDF/ESF, "A way to make Europe"/"Investing in your future" (PI20/00130; FI21/00262) and by the AGAUR (2021 SGR 00459). Francisca Yáñez was supported by a fellowship from ANID, BECAS Chile, No. 72190278. None of the funders were private companies.

Author contributions
Z.S. contributed to literature searches, data collection, data analysis, data interpretation, writing, review, and editing. M.P.-T., I.M., C.C., E.V. and F.Y. contributed to data curation, sample processing, review, and editing. G.S.-G., S.V.-A., M.R.-B., Z.X., A.N.-S. contributed to bioinformatics analysis and website building, review & editing. C.M. contributed to study design, fundraising, conceptualization, data analysis, data interpretation, writing, review, and editing. All authors are from the academic team. Z.S. and C.M. had accessed and verified the data reported in the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-025-00769-9.

Correspondence and requests for materials should be addressed to Chaysavanh Manichanh.