UAB
Universitat Autònoma de Barcelona

Dietary habits, microbiome profiling and bacterial isolation: A tri-focal approach in a Spanish cohort

Zaida Soler Luque

Dietary habits, microbiome profiling and bacterial isolation: A tri-focal approach in a Spanish cohort

Doctoral thesis 2025

Zaida Soler Luque

Dietary habits, microbiome profiling and bacterial isolation: A tri-focal approach in a Spanish cohort

A doctoral thesis presented by **Zaida Soler Luque** to aim the degree of Doctor

Director and tutor: Dra. Chaysavanh Manichanh

Doctoral Program in Medicine
Department of Medicine
Universitat Autònoma de Barcelona
Barcelona, 2025

AGRADECIMIENTOS

Después de 5 años de duro trabajo al final ha llegado el momento de cerrar un capítulo de mi vida. Todavía recuerdo cuando hace 6 años decidí venir a Barcelona de forma temporal con el fin de cursar el máster del VHIR y buscar luego trabajo en una farmacéutica. No había forma de saber que decidiría quedarme y que iba a conocer a gente maravillosa por el camino.

Primero de todo, me gustaría darle las gracias a mi tutora y directora, la Dra, Chaysavanh Manichanh, por proponerme, en primer lugar, la posibilidad de realizar las prácticas de máster con ella y ofrecerme la oportunidad de realizar una tesis después. Gracias por preocuparte siempre por mi y por aconsejarme cuando estaba perdida, pero también por darme la libertad suficiente de poder explorar y decidir qué investigar y cómo hacerlo. Gracias por obligarnos a hacer los lab meetings en inglés, los cuales me han ayudado a ganar fluidez en este idioma y en las presentaciones, pese a que era lo que más odiaba. Y gracias por alentarme a hacer una estancia en Francia, a la cual no quería ir y de la que luego me daba mucha pena volverme. Echaré de menos tus desayunos cuando no venía Gerard, tus recetas exprés que siempre hacías en 10 minutos, pero a nosotros nos costaba horas, y tus consejos en en ámbito profesional pero también personal.

Agradecer también de todo corazón a todas las personas que he tenido la oportunidad de conocer a lo largo de mi proyecto y que han hecho del trabajado un lugar agradable en el que me gustaba estar. Empezando por Fran y Encarna, por enseñarme los primeros pasos en el laboratorio e introducirme en el mundo de la microbiota y las cuales siempre serán mi modelo a seguir. A Gerard, por ser mi eterno compañero del café de las mañanas y con el que compartía la responsabilidad de ser “los mayores del lab”, aunque yo hubiese nacido en el 96. También por haberme ayudado tanto en la parte bioinformática del proyecto, el cuál acabó convirtiéndose en lo que juramos destruir: el nuevo METAHIT. A Sara y a Arnau, a los cuales bombardeaba con mil millones de preguntas y marrones, incluso estando en Francia. A Laura, porque pese a haber venido como postdoc más tarde, fue de las primeras personas con quien hablé cuando llegué al laboratorio y se quedó al pié del cañón con su “Isaac protocol!”.

Y no podía faltar mi minion favorito, compañero de gym y de piso, Marc, el cual me asignaron para que hiciese las prácticas del máster y el cuál ha decidido quedarse para continuar, en parte, mis pasos. Gracias por escuchar mis preocupaciones y hacerme siempre reír. Gracias por ayudarme en el proyecto y en las extracciones que nunca acababan…Sin ti no lo habría logrado ni me habría animado a conocer al resto de los chicos…

A Elba, por sus temas de conversación tan interesantes y proponerme siempre planes, aunque siento haber rechazado tus citas (excepto si eran para practicar catalán o inglés). A Isaac, por traer la mejor extraction playlist y la energía. Al grupo de Lupus (Mulli, Pere, Magí y Maria) por apuntarse a todos los planes que surgiesen y hacerme compañía cuando todo mi grupo teletrabajaba. Habrá que repetir lo de Valencia y las tardes de cine…

A Ana, Cat, Zixuan, Pau, Iñigo, Chloé, Luis, Blanca, Anna, Borja, Jordi, Carlos etc los cuales han pasado por el lab brevemente, pero han puesto su granito de arena. A Ray, por gestionar la enorme

cantidad de envios para los participantes, a Flor por empezar cada mañana con “¿Qué es lo que pasa aquí?” mientras reía. A coordinación de laboratorios, celadores, personal de limpieza etc por contribuir al buen ambiente de este lugar.

De mi estancia en Francia me quedo con el Dr. Lionel Rigottier-Gois, el cuál me acogió con las manos abiertas e hizo que me sintiese siempre como en casa. A Alan, con quien puse a punto un laboratorio en 3 meses. Fuiste la persona que me enseñó todo lo que sé sobre culturomics, pero también cambiaste mi perspectiva de los franceses e hiciste que me plantease volver. No olvidaré nuestros “coffe times” y “cigarette breaks” ni lo bien que lo pasábamos tomando cervezas por Paris. A Francis, por intentar hablar español conmigo y por los debates filosóficos a cualquier hora del día. Por siempre tomarte las cosas de buen humor y por tus consejos acerca de qué hacer cuando vives en otro país y estás perdido.

Agradecer también a mi grupo de amigos fuera del VHIR, a Bernat por ser uno de mis confidentes y darme ánimos siempre que lo necesitaba, sin important cuándo o dónde. A Sergi, por nuestros coffee breaks de 30 min cada tres meses para comentarnos qué tal nos iba la vida, pese a conocernos desde el instituto y de haber terminado en el mismo centro. A Laura por entenderme siempre tan bien y nunca juzgarme y por buscar siempre tiempo para vernos pese a vivir algo lejos. A Raquel y a Dani por siempre acogerme en su casa de Mollet y darme su apoyo en los momentos complicados.

A Maria y a Aina, por ser mis pilares en Barcelona y estar conmigo en las épocas buenas y en las muy malas. Por animarme a seguir adelante en todos los aspectos de mi vida y consolarme cuando no sabía qué hacer. A Morgane, que pese a estar muy lejos, seguíamos en contacto siempre que podíamos. A Raquel C., Andrea, Ana, Gerardo. Sin todos ellos, mi tiempo en Barcelona no habría sido el mismo. A Ángela, que empezó siendo mi profesora de idiomas y terminó siendo una amiga…

No podría terminar sin agradecer a mi familia, especialmente a mis padres, Ramón y Rosa y a mi hermana melliza, Mireya, los cuales se han preocupado siempre por mi bienestar y me han dado siempre todo su apoyo y con los que me he mantenido en contacto casi a diario pese a vivir en lugares distintos. Quiero que sepáis que significáis mucho para mi, aunque a veces no lo demuestre.

Me gustaria agradecer finalmente al Instituto Carlos III (ISCII) por la financiación recibida (Beca PFIS; PI20/00130) sin la cual no me habría sido posible realizar dicho proyecto ni la estancia posterior. Al instituto INRAE-MICALIS en Francia por permitirme aprender la técnica de Culturomics. Al VHIR por darme la oportunidad de desarrollar mi doctorado y finalmente a la UAB por incluirme en su programa y por el soporte recibido.

A todos vosotros, mi yo del pasado, del presente y del futuro os llevará siempre en un lugar de la memoria y del corazón.

Llegó al fin el fin de la Zaidocracia… veremos qué nos deparará el futuro ahora.

LIST OF ABBREVIATIONS

16S rRNA: 16S ribosomal RNA

24HR: 24-h dietary recall

AECOSAN: Spanish Agency for Consumer Affairs, Food Safety and Nutrition

AESAN: Agencia Española de Seguridad Alimentaria y Nutrición

AGP: American Gut Project

AI: Adequate Intakes

AIEC: Adherent-Invasive *Escherichia coli*

aMED: Alternative Mediteranean Score

ARGs: Antibiotic Resistance Genes

AUC: Area Under the Curve

BEDCA: Base de Datos Española de Composición de Alimentos

BHI: Brain Hearrat Infusion

BMI: Body Mass Index

CCAA: Autonomous Communities

CD: Crohn's Disease

CHO: Carbohydrates

CKD: Chronic Kidney Disease

CPM : Counts Per Million

CVD: Cardiovascular disease

DA: Differential Abundance

DQIs: Dietary Quality IndexesIndices

ENALIA2: Encuesta Nacional de Alimentación en la Población Adulta, Mayores y Embarazadas 2

EuroFIR: European Food Information Resource

FCDB: Food Composition Databases

FCT: Food Composition Tables

FDR: False Discovery Rate

FFQ: Food Frequency Questionnaire

FGFP: The Flemish Gut Flora Project

FISH: Fluorescence *in situ* hybridization

FMT: Fecal Microbiota Transplantation

FNDDS: USDA´s Food and Nutrient Database for Dietary Studies

GBD: Global Burden of Disease Study

GI: Gastrointestinal tract

HEI: Healthy Eating Index

HFD: High fat diet

HFD-index: Healthy Food Diversity Index

hPDI : Hhealthful Plant-based Diet Index

IASE: Healthy Eating Index for Spanish population

IBD: Inflammatory Bowel Disease

ICC: Intraclass Correlation Coefficient

IHMS: International Human Microbiome Standards

INR: Nutrition Reference Intake

IS: Immune System

ITS: Internal Transcribed Spacer

KNHANES: Korea National Health and Nutrition Examination Survey

LLDeep: Dutch LifeLies-DEEP Study

LPS: Lipopolysaccharides

MaAsLin: Multivariate Association with Linear Models

MALDI-ToF MS: Matrix-Aassisted Llaser Desorption mass spectrophotometry

MAR: Mean Adequacy Ratio

MDI: Meat Dietary Index

MedDiet: Mediterranean Diet

MS: Mass Spectrometry

NAR: Nutrient Adequacy Ratio

NASH: Non-Alcoholic Steatohepatitis

NGS: Next Generation Sequencing

NMR: Nuclear Magnetic Resonance

NRI: Nutritional Reference Intakes

OD: Optical density

PBS: Phosphate Buffered Saline

PDI: Plant-based Diet Index

POP Study: Population longitudinal study in Spain

PPI: Proton Pump Inhibitors

PREDIMED: Prevention with Mediterranean Diet

PUFAs: Polyunsaturated Fatty Acids

PVPP: Polyvinylpolypyrrolidone

qPCR: Quantitative Polymerase Chain Reaction

RF: Rumen fluid

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SCD: Single Cell Dispenser

SDGs: Sustainable Development Goals

SFCAs: Short Chain Fatty Acids

sFCDB: Spanish Food Composition Database

sFFQ: Short Food Frequency Questionnaire

SGB: Species-level Genome Bin

TAE: Tris Acetate EDTA

TMAO: Trimethylamine N-oxide

UC: Ulcerative Colitis

uPDI: Uunhealthful Plant-based Diet Index

USDA: United States Department of Agriculture

V4: Hypervariable region 4

VHIR: Vall d´Hebrón Research Institute

WD: Western-type diet

Index

SUMMARY

The human gut microbiome is a dynamic biomarker shaped by diet, lifestyle, and environmental factors. This doctoral research, based on the longitudinal recruitment of 1,017 healthy Spanish volunteers and the use of shotgun metagenomic sequencing, represents one of the largest studies in Spain exploring the diet-microbiome relationship at functional and species levels.

The study demonstrates that adhering to diverse, high-quality diets, rich in vegetables, fruits, legumes, whole grains, nuts, and seeds, correlated with increased bacterial diversity and a microbiome profile distinct from that associated with inflammatory bowel diseases. This dietary pattern correlates with increased beneficial bacterial species such as *Akkermansia muciniphila*, recognized for its protective metabolic properties, while simultaneously reducing the abundance of species like *Flavonifractor plautii*, a flavonoid-degrading bacterium linked to poor dietary quality, and *Ruminococcus torques*, a mucin-degrading bacterium that can compromise the intestinal barrier.

Multifactorial analysis revealed that beyond diet, factors such as smoking, higher body mass index, residence in the Mediterranean region, and infrequent bowel movements were associated with reduced bacterial diversity. Conversely, older age correlated with both better dietary habits and greater microbial diversity, and women exhibit healthier dietary patterns.

Despite Spain's Mediterranean tradition, adherence to the traditional Mediterranean diet, was low (median aMED score: 4.0/ 9.0), reflecting a gradual transition toward Western patterns. Only three of twelve Global Burden of Disease Study dietary targets were met: vegetables, fruits, and fiber intake. Mycobiome exploration through enrichment protocols showed limited associations with dietary and bacterial patterns, identifying *Saccharomyces cerevisiae* as the most prevalent fungal species.

Non-targeted culturomics on selected samples isolated 27 different bacterial species, including *Bifidobacterium animalis* and *Bacteroides uniformis*. While optimization is required for extremely oxygen-sensitive species such as *Faecalibacterium prausnitzii*, reformulating this method gives access to bacterial strains for future mechanistic studies. This work also integrated citizen science through personalized reports and open-access reporting, enhancing public engagement and democratization of scientific knowledge.

Overall, these findings underscore the need for holistic, personalized approaches that integrate diet, lifestyle, and an individual context in microbiome research.

RESUMEN

El microbioma intestinal humano es un biomarcador dinámico influenciado por la dieta, el estilo de vida y los factores ambientales. Esta tesis doctoral, basada en el reclutamiento longitudinal de 1.017 voluntarios sanos españoles y el empleo de secuenciación metagenómica shotgun, representa uno de los estudios más amplios en España que exploran la relación entre la dieta y el microbioma a nivel funcional y de especies.

El estudio demuestra que seguir dietas diversas y de alta calidad, ricas en verduras, frutas, legumbres, cereales integrales, frutos secos y semillas, se correlacionaron con una mayor diversidad bacteriana y un perfil microbiano distinto al asociado con enfermedades inflamatorias intestinales. Este patrón dietético se relaciona con un aumento de especies bacterianas beneficiosas como *Akkermansia muciniphila*, reconocida por sus propiedades metabólicas protectoras, al tiempo que reduce la abundancia de especies como *Flavonifractor plautii*, una bacteria degradadora de flavonoides vinculada a una dieta de baja calidad, y *Ruminococcus torques*, una bacteria degradadora de mucina que puede comprometer la barrera intestinal.

El análisis multifactorial reveló que, más allá de la dieta, factores como el tabaquismo, un mayor índice de masa corporal, residir en la región mediterránea y frecuencia de deposiciones extremas se asociaron con una reducción en la diversidad bacteriana. Por el contrario, una mayor edad se correlacionó tanto con mejores hábitos dietéticos como con una mayor diversidad microbiana, y las mujeres mostraron patrones dietéticos más saludables.

A pesar de la tradición mediterránea presente en España, la adherencia a la dieta mediterránea tradicional fue baja (puntuación mediana del aMED: 4.0/9.0), lo que refleja una transición progresiva hacia patrones occidentales. Solo se cumplieron tres de los doce objetivos dietéticos propuestos por la Global Burden of Disease Study: consumo de verduras, frutas y fibra. La exploración del micobioma mediante protocolos de enriquecimiento mostró asociaciones limitadas con patrones dietéticos y bacterianos, identificando *Saccharomyces cerevisiae* como la especie fúngica más prevalente.

La culturómica no dirigida en muestras seleccionadas permitió aislar 27 especies bacterianas diferentes, incluidas *Bifidobacterium animalis y Bacteroides uniformis*. Aunque se requiere una mayor optimización para especies extremadamente sensibles al oxígeno como *Faecalibacterium prausnitzii*, la reformulación de este método nos da acceso a especies bacterianas para futuros estudios mecanísticos. Este proyecto también integró la ciencia ciudadana mediante informes personalizados y difusión en acceso abierto, fomentando el compromiso de la ciudadanía y la democratización del conocimiento científico.

En conjunto, estos hallazgos ponen de manifiesto la necesidad de enfoques holísticos y personalizados que integren la dieta, el estilo de vida y el contexto individual en la investigación del microbioma.

1 INTRODUCTION

1.1 The Human gut microbiome

1.1.1 *The gut microbiome: not just bacteria*

The human microbiome is extensive and comprises genomes of a wide range of microorganisms such as protozoa, bacteria, fungi, viruses, and archaea. Traditionally, bacteria have been the most widely investigated microorganisms (1,2). However, significant efforts have recently been devoted to study other microbial components, including fungi and viruses (3,4).

Microbes are located everywhere along the body, however, the gut constitutes one of the largest interfaces (approximately 400 m^2) where the microbiome interacts with the host immune system (IS) and the environment (5). The collections of microorganisms in the gut, referred as the gut microbiota, is believed to play a critical role due to the wide range of essential functions that it performs.

1.1.1.1 Bacteria

In the gut of healthy human adults, approximately 10^{13} bacteria with a diversity of more than 4500 different species including an average of 300-400 species per individual can be found (6,7). Commensal bacteria are mainly composed by two dominant phyla (Bacillota and Bacteroidetes) and followed by less dominant ones such as Actinobacteria and Proteobacteria, which remain highly stable over time (8).

Among the wide range of functions, commensal bacteria can synthesize *de novo* essential vitamins such as vitamin K, B5, B9 and B12, some of which the human body cannot produce on its own. Folate or vitamin B9 are mainly produced by *Bifidobacteria* and are known to be involved in important metabolic pathways, including DNA synthesis and repair (5). Vitamin K is well known for being involved in the synthesis of blood clothing factors in the liver and may play a protective role in coronary heart disease (9,10). Additionally, B5 and B12 play a key task in the assembly of the neurotransmitter acetylcholine and hormone cortisol, required for the correct functioning of the nervous system (11).

Bacteria are also involved in the fermentation of indigestible carbohydrates (CHO) that come from diet, mostly fibers and resistant starch, and their transformation into short-chain fatty acids (SCFAs), mainly butyrate, propionate, and acetate. These SFCAs are famous for their anti-inflammatory properties and serve as the primary energy source for colonic cells, named as colonocytes (12,13). Besides, they play crucial roles in IS development, maintaining intestinal homeostasis, and strengthening the physical barrier, which helps to prevent the colonization of harmful bacteria through both direct and indirect competition (11,14).

1.1.1.2 Fungi

Beyond bacteria, fungi, collectively known as the gut mycobiome, represent a small but significant part of the human gut microbiome (approx. 0.1%). Ascomycota, Basidiomycota and Zygomycota appear to be the most prevalent phyla in the gut of healthy adults. However, unlike bacteria, their composition varies significantly between individuals and within the same individual over time (15). Functions performed by the mycobiome are still not fully understood, partially due to the challenges associated with its identification. Even so, it has been shown that it facilitates fiber and oligosaccharides digestion, similar to bacteria, while also modulating both the host and microbial metabolism (16,17). Moreover, another of its major roles seems to be the regulation of the IS. Fungal antigens, such as β-glucans, are recognized by specific immune cell receptors, triggering defense

mechanisms against harmful pathogens while promoting tolerance to commensals. They also might play a central role in protecting the mucosal layer through Th17 cells mediation and in immune homeostasis maintenance (18,19).

1.1.1.3 Viruses

Human gut is estimated to harbour around 10^9-10^{10} viral-like particles/g of feces, comprising a diverse range of virus types. These include DNA and RNA, both single and double stranded, ranging from plant viruses ingested through the diet to viruses capable of infecting other microorganisms (bacteriophages) or human cells (eukaryotic viruses).

More than 90% of the human virome is constituted by bacteriophages with the ability to infect bacteria and archaea. Shotgun sequencing revealed *Caudovirales* and *Microviridae* as the most abundant viral taxa (4). Most phages can undergo two distinct life cycles: lytic and lysogenic. Briefly, in the lytic cycle, the virus infects the cell, replicates, and triggers cell lysis to release new viral progeny. In contrast, in the lysogenic cycle, the virus enters the cell and integrates into the host's chromosome, allowing it to persist in a latent state, creating a mutualistic relationship between the phage and the bacterial cell. It is now well-established that phages play a crucial role in regulating bacterial populations via lytic cycles, thereby preventing bacterial adhesion to mucosal surfaces. Additionally, phages help bacteria to adapt to their environment and enhance survival by transferring DNA, such as antibiotic resistance genes (ARGs), between cells. Moreover, phages may also promote immune tolerance and facilitate commensal colonization (20,21).

1.1.2 *Technical approaches for the analysis of the microbiome*

Microbiome profile is unique to each individual and its composition has been assessed over time using different techniques, including fluorescence *in situ* hybridization (FISH), cultured-based methods, 16S ribosomal RNA (16S rRNA) sequencing for bacteria, internal transcribed spacer (ITS) sequencing for fungi, and RNA arrays. However, in recent years, omics approaches, such as metagenomics and culturomics, have emerged as the leading techniques for microbiome analysis (22–24).

1.1.2.1 Next Generation Sequencing and marker gene analysis

Targeted Next Generation Sequencing (NGS) is a culture-independent method that uses primers to target and amplify a specific region of a gene of interest. This gene commonly has a hypervariable region, which enables the distinction between species or genera, and a conserved region that serves as the primer binding site (25). For bacteria and archaea, hypervariable regions V1 to V6 of the 16S gene are commonly used (26–30), while for fungi, regions ITS1 to ITS3 regions and/or 18S gene are suitable options (17,31–33). However, primer selection can introduce amplification bias, as primers may not have equal affinity for all regions, and this method often fails to identify organisms beyond the genus level. Additionally, while some programs can infer metabolic functions from taxonomic data, the shotgun sequence strategy is preferred for achieving more comprehensive gene coverage and greater reliability.

In contrast, NGS is highly recommended for assessing low biomass samples (e.g., vaginal, saliva, tissue etc) as the microbial DNA content is often too low in these samples for shotgun sequencing.

Furthermore, large public 16S databases are available, and the cost of sample processing and analysis is relatively low compared to the shotgun sequencing approach (25).

1.1.2.2 Shotgun metagenome analysis

Shotgun metagenomics is an untargeted technique that involves sequencing the entire collection of DNA fragments extracted from a sample. For this purpose, DNA obtained from samples of interest is sequenced, without prior amplification, thereby avoiding possible PCR biases.

Compared to targeted NGS, this procedure enables a more comprehensive view of the microbial community, capturing both the taxonomic and functional diversity, as well resolving species and, in some cases, strain-level disctinctions. However, after DNA extraction, downstream data analysis is time- and resource intensive, making this approach more costly than targeted NGS methods. Moreover, contamination by host DNA can occur and less databases are currently available (3,34).

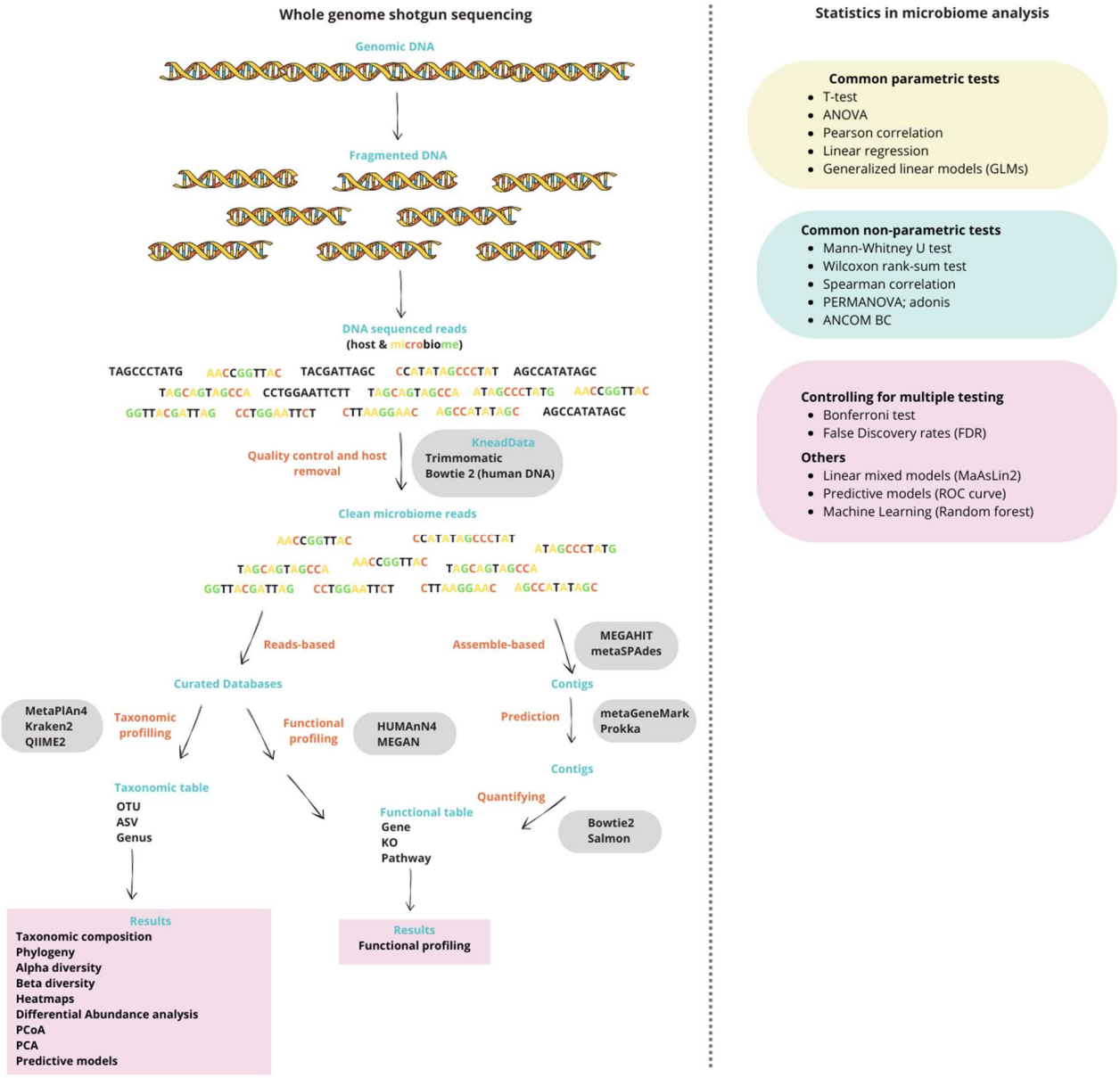


Figure 1. Summary of whole shotgun metagenome analysis workflow and statistical analysis commonly performed in microbiome studies. Adapted from (34–36).

First, raw DNA sequenced reads obtained from Illumina platform undergo quality control and decontamination of the host DNA using pipelines such as KneadData (https://huttenhower.sph.harvard.edu/kneaddata) which contains Bowtie2 and Trimmomatic tools (37,38). Clean microbial reads are then used to generate functional and taxonomic tables, which can be achieved using either read-based or assemble-based approaches. The read-based method determines taxonomy by aligning clean reads to curated databases, commonly utilizing tools such as MetaphlAn4 (39) or Kraken2 (40). For functional analysis , the most common softwares freely available include HUMAnN4 (41) and MEGAN (42). The assembly-based methods assemble the clean reads obtained into longer units called contigs based on overlapping regions, inferring in the end the original genome. To achieve this, tools such as MEGAHIT (43) and metaSPAdes (44) are employed.

Assembled contigs are then annotated using tools like metaGenMark (45) and Prokka (46), while gene abundance is quantified using additional alignment tools such as Salmon and Bowtie2 (47). From this computational step, functional and taxonomic tables are recovered and used in further analyses including α- and β-diversity, correlation and prediction analyses. A brief summary is presented in Figure 1.

1.1.2.3 Culturomics

Metagenomics has revealed the vast biodiversity present within the human microbiota, with numerous metagenomics species surpassing the species and strains cataloged in current culture collections. However, accessing the uncultivated fraction, also known as “microbial dark matter”, which is estimated to represent about 35 to 65% of species, poses a significant challenge. Overcoming this obstacle is crucial for moving beyond mere correlations and towards establishing causation with cultivable microorganisms (48–50). To achieve this, the systemic cultivation of new strain collections is necessary. However, isolating and cultivating microorganisms from human gut requires high-throughput technique like culturomics.

1.1.2.3.1 History

Early culturomics efforts relied on traditional microbiology techniques involving selection and/or enrichment, processes known for being both time-consuming and labor-intensive due to the necessity of multiple media and culture conditions. These methods were complemented by molecular identification (51). Nowadays, culturomics has evolved into a high-throughput approach that enables the rapid and efficient cultivation and isolation of microorganisms. It is often combined with matrix-assisted laser desorption mass spectrophotometry (MALDI-ToF MS) for cost-effective and rapid identification. When strain-level resolution or additional information is required, 16S sequencing is performed for bacteria (52), while the ITS region is analyzed for fungi (53).

In 2012, *Lagier et al.,* successfully implemented the first culturomics workflow in stool samples from three volunteers, testing 212 different culture conditions. As a results, 174 previously undescribed species from the human gut were successfully isolated and identified, demonstrating that new species can be cultivated under proper culture condition (51). The same group later expanded the human gut repertoire by adding 531 novel species (54). Furthermore, this technique also enables the isolation of other microorganisms, including fungi. In 2017, *Hamad et al.,* isolated 10 novel fungi species from stool samples of both healthy and patients, demonstrating that culturomics is a suitable technique for the discovery and isolation of a wide range of human gut commensals (53).

Since then, more research groups have utilized this technique for cultivation and/or discovery of species, not detected by metagenomics or not isolated before, potentially having a significant impact on human health. Automation of the current procedure was also attempted to expand existing microorganism collections (52,53,55–60). The successful isolation of *Facalibacterium prausnitzii* and *Akkermansia miciniphila*, two important gut commensals, serves as a good example for this approach (61,62).

1.1.2.3.2 Targeted vs non-targeted culturomics

When applying the culturomics technique, two main procedures can be used, with their key characteristics summarized in the following table:

Table 1. Differences between targeted and non-targeted culturomics (56,59,63)**.**

	Targeted culturomics	***Non-targeted culturomics***
Definition	Deliberate isolation of specific groups of microorganisms based on prior knowledge or hypothesis	An exploratory approach that enables the cultivation of a wide range of microorganisms without any specific objective
Method	Utilization of specific culture media and conditions that selectively promote the growth of the targeted microorganisms	Use of a wide range of culture media and conditions to try to capture the maximum diversity present in a sample
Advantages	Enables the acquisition of specific species for further study of their morphology, interactions with the environment, and metabolism.	An efficient approach for discovering new species, while also providing insights into the broader ecosystem.
Limitations	Relies primarily on previous knowledge and may lead to an underestimation of the total community if selection conditions are too limited. Screening devices are required.	Due to the need to test different conditions and screen a large number of colonies, it can be time-consuming

1.1.2.3.3 Performing culturomics: step by step

Nowadays, culturomics englobes several key steps: sample collection and processing, media selection and incubation, microbial isolation, cultivation, identification, and preservation. Briefly, samples are collected, homogenized, and subjected to serial dilutions. The diluted samples are then plated on different culture media and incubated under a wide range of oxygen levels, temperature, and time conditions. For targeted culturomics, selective media and conditions specific to bacteria of interest are directly applied, based on prior litterature or experience. Following incubation, microbial growth is assessed and distinct phenotypic colonies are isolated and grown individually. Once colonies are expanded, isolated species are identified. In routine, MALDI-ToF MS is commonly used for identification. However, if this method fails or new species are suspected, an additional step involving 16S or ITS sequencing is performed. Finally, once pure isolates of interest are obtained, viable cells are stored at -80°C in cryogenic tubes containing glycerol (63–65) (Figure 2).

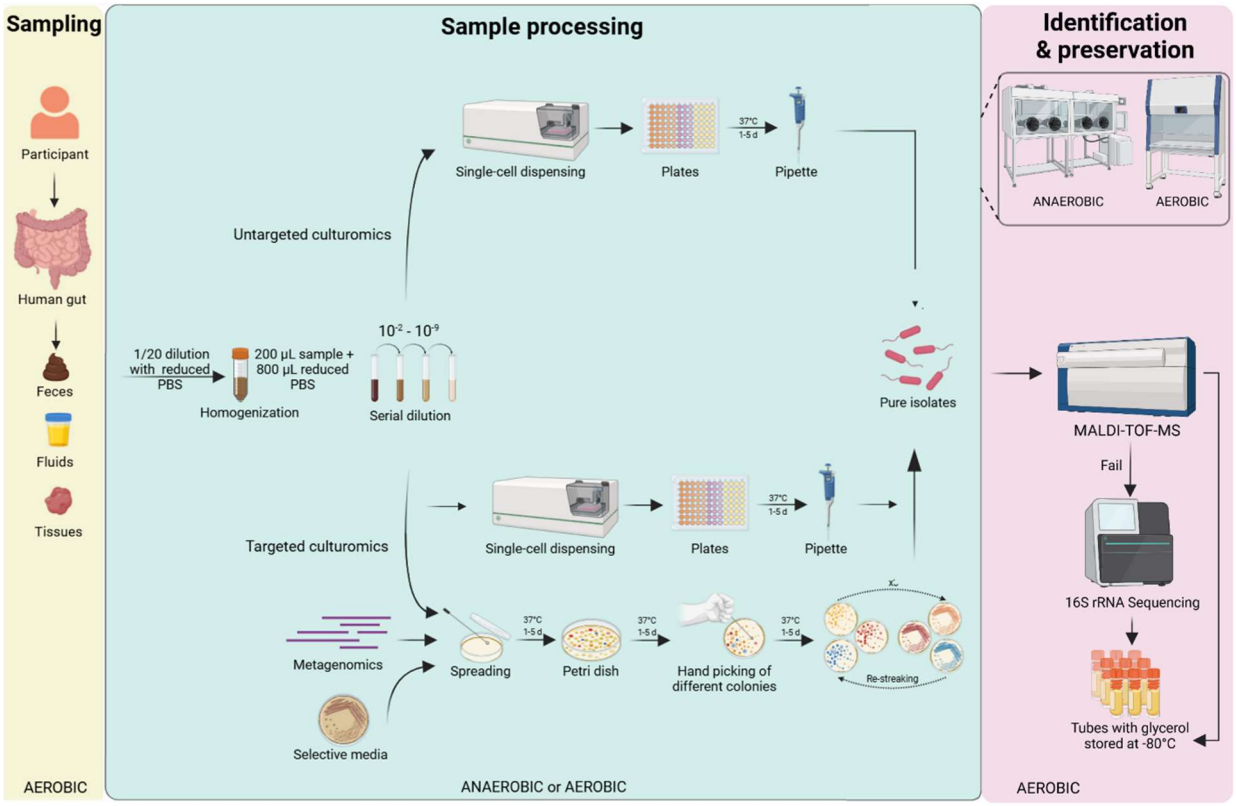


Figure 2. Workflow of targeted and non-targeted culturomics. Created by Biorender

1.1.2.3.4 Advantages of culturomics over metagenomics

Metagenomics has several limitations that can be addressed through culturomics. One of the primary challenges is the precise taxonomic assignment of the sequences at the strain level, which is hindered by short read lengths and heavily dependent on the quality of reference databases. Additionally, metagenomics analysis is often descriptive, generating hypotheses that require further validation through *in vitro* or *in vivo* studies to establish causality, which necessitates the isolation of microorganisms. Many of the dominant or relevant species uncovered through correlation analysis still lack a cultured representative strain. Moreover, low-abundant species are often difficult to detect using sequencing methods, even when a prior enrichment step is applied (48,52,65).

1.1.3 *Role of the gut microbiome in health and disease*

The equilibrium and healthy diversity state of intestinal microbiome is known as eubiosis and its regulation is highly complex (66–68). Disruptions in this regulation can result in dysbiosis, a state characterized by alterations in microbiome composition and metabolic capacity. This imbalance leads to a decrease in beneficial bacterial products, while promoting the growth of opportunistic and pathogenic bacteria (69). Several external factors, such as drug intake, stress and diet, can contribute to this homeostasis disruption (70,71).

It is now well accepted that microbiome plays a key role in non-communicable diseases such as Crohn´s disease (CD) and Ulcerative Colitis (UC), two main forms of Inflammatory Bowel Disease (IBD). Compared with healthy controls and UC patients, CD patients presented lower bacterial diversity and

a decrease in butyrate-producing bacteria such as *Faecalibacterium prausnitzii* in combination with an increase in *Escherichia coli* (72,73). Eukaryotic viruses from *Pneumoviridae* family in combination with *Caudovirales* bacteriophages were also found to be altered in intestinal mucosa of IBD patients with an increase in UC patients compared with healthy controls (74). At fungal level, a common fungi, *Candida albicans*, was enriched in IBD patients (75). Microbiota may be involved in other diseases including *C. difficile* infection (76), Celiac Disease (77,78), type 1 diabetes (79), obesity (80,81), chronic kidney disease (CKD) (82), etc.

1.1.4 *Factors modulating the gut microbiome*

Various factors have been suggested to influence the composition, structure and function of the gut microbiome, with the most relevant ones illustrated in Figure 3.

1.1.4.1 Genetics

Some studies suggest that host genetics account for approximately 1.9 - 8.1% of the variation in the human microbiome (83,84). Among the numerous genetic associations with microbial composition, the ABO and LCT genes appear to be the most consistently linked. The ABO gene encodes a glycosyltransferase involved in ABO blood group determination and it is expressed in various cell types. Several loci of the gene have been repeatedly correlated with *Collinsella*, *Bifidobacterium* and *Faecalibacterium* species (85,86). Similarly, the LCT gene, which plays a key role in lactase synthesis has been associated with *Bifidobacterium* species in multiple studies (86–88).

1.1.4.2 Age

The relationship between the gut microbiome and age is complex and dynamic. In early life (up to 1 year), the bacterial microbiome is dominated by two main phyla: Actinobacteria and Proteobacteria, with low bacterial diversity. Over time diversity gradually increases, leading to a more diverse and stable adult-like microbiome by the age of 2-3 years old (5). During adulthood, the gut microbiome remains relatively composed of Bacillota and Bacteroidetes, followed by less dominant phyla such as Actinobacteria and Proteobacteria (89). While the adult microbiome is generally stable, age-related changes have been reported in individuals over 65 years old. For instance, at 80 years old, a decrease in α-diversity is often noted, likely due to poorer diet and increased frailty (90,91). Unlike bacteria, fungal stability varies significantly among individuals and across life stages. In early life, the mycobiome is dominated by *Candida*, *Malassezia* and *Mycospharella* at three months, but at 12 months, it shifts towards *Saccharomyces* dominance, accompanied by a decrease in overall fungal α-diversity (92). As individuals progress through adolescence and adulthood, mycobiome composition is largely influenced by diet and Body Mass Index (BMI) (93). Under favorable conditions, a healthy adult mycobiome is dominated by *Saccharomyces cerevisiae*, *Dacryopinax primogenitus*, *Yarrowia lipolytica*, *C. parapsilosis* and *C.albicans* (94).

1.1.4.3 Sex

"Sex" refers to the biological categorization of a species based on reproductive systems and functions influenced by chromosomal types or hormones. The terms 'male' and 'female' are employed when discussing the sex of human participants or other sex-related variables (95). In contrast, "gender" encompasses socially constructed characteristics that define what its means to be a woman and man,

including norms, roles and their behaviors. Since gender is shaped by cultural and historical contexts, it can vary significantly across different societies and time periods (102).

In general, the impact of sex on bacterial α-diversity remains controversial. While some studies suggest a higher α-diversity in females (96,97), other found no significant differences (98–100). What seems to be more consistent is the fact that sex has been correlated with differential abundant species. Specifically, females tend to have a greater abundance of genera such as *Akkermansia*, *Bifidobacterium* and *Bilophila*, whereas males are more associated with an increased presence of genera like *Prevotella* (Table 2) (96–101). These differences may be partially driven by sex hormones such as estrogens or androgens (102–105), demonstrating the importance of considering sex as biological variable in experimental design as well as a possible confounder when performing microbiome studies.

Table 2. Differential abundant bacterial genera enriched based on sex.

Males ↑	***Females ↑***
Faecalibacterium (99)	*Oscillibacter* (99)
Ruminococcus_gauvreauii (99)	*Anaerostipes* (99)
Mitsuokella (99)	*Bilophila* (98,100)
Veillonella (100)	*Akkermansia* (97,101)
Methanobibracter (100)	*Bifidobacterium* (96,101)
Prevotella (101)	*Ruminococcus* (101)
Megamonas (101)	*Flavonifactor* (99)
Fusobacterium (96,99,101)	
Megasphaera (101)	

1.1.4.4 Medication

Antibiotics are known for being one of the main gut microbiota disruptors as they reduce commensal diversity and alter microbial composition, while promoting the overgrowth of opportunistic bacteria (106). Several studies have investigated the short and long-term impact of different antibiotics (107,108). For instance, ciprofloxacin treatment has been linked to a depletion of *Faecalibacterium* and *Alistipes,* accompanied by an increase in *Bacteroides* between 11 and 30 days post-treatment (109). Moreover, azithromycin has been associated with the lowest recovery of pre-treatment microbial levels compared to other antibiotics (110).

A further consequence of antimicrobial use resides on the increase of ARGs, collectively known as the gut resistome. Healthy individuals may contain approximately 100 unique ARGs, with the most common conferring resistance to β-lactams and tetracycline. However, antibiotic treatment increases ARG prevalence by selecting resistant bacterial strains (110–112).

Beyond antibiotics, non-antibiotic drugs can also significantly influence the gut microbiome. *Maier et al.* showed that nearly 24% of human-targeted drugs can inhibit the growth of at least one bacterial species *in vitro* (113). Among these, proton pump inhibitors (PPI), metformin, and laxatives appear to exert the strongest effects on the microbiome composition, although further research is required to elucidate their impact (114,115).

1.1.4.5 Environmental factors

1.1.4.5.1 Horizontal transmission

Horizontal transmission refers to the acquisition of microorganisms from the external environment (116). This process occurs through various pathways including person-to-person interactions and cohabitation (115,117–119), the ingestion of microbes from food sources (120), and also microbial exchange with our pets (121,122). To explore microbiome transmission patterns, *Vallés-Colomer et al.,* analyzed 9700 human metagenomes, identifying key contributors to microbial sharing. Among the various transmission routes, cohabitation emerged as the dominant factor, with strains- sharing rates ranging from 11 up to 71% among individuals living together compared to non-cohabiting individuals from the same population. Interestingly, the number of shared microbial species declines with age, with younger individuals exhibiting the highest levels of microbial exchange (119). These findings align with research by *Gracesa et al.,* who reported that 48.6% of microbial taxa were influenced by cohabitation, whereas 6.6% were considered heritable (115).

1.1.4.5.2 Geography

Each geographical region is shaped by unique dietary habits, cultural factors, and topographical features that may affect gut microbiome composition and function. These effects are particularly pronounced when comparing geographical distant regions or contrasting under-developed and developed countries. One of the most consistent findings is a reduction in α-diversity, which appears to correlate with the degree of industrialization. Hunter-gatherer communities exhibit the highest microbial diversity, followed by traditional farming/fishing populations, with the lowest diversity observed in highly urbanized, westernized societes (123–126). At compositional level, distinct microbial profiles emerge. Hunter-gatherer populations are characterized by a high abundance of *Treponema*, an opportunistic pathogen, along with higher *Prevotella, Clostridium*, *Oscillibacter*, *Lachnospira,* and others (124,127). In contrast, Westernized populations show an enrichment in *Bacteroides*, *Escherichia*, *Proteobacteria*, *Clostridium, Faecalibacterium*, *Ruminococcus*, *Dorea*, *Blautia*, *Roserburia,* and *Oscillospira* (123,125,127). Interestingly, traditional farming or fishing populations exhibit a microbiome profile that integrates characteristics of both extremes, with *Prevotella* and *Eubacterium* who are present in less urbanized regions and *Ruminococcus*, *Blautia*, *Dorea* and *Clostriudium* species, more common in industrialized areas. However, the relative abundance of these taxa varies (123,127)(Table 3). However, divergences in microbial composition could be in part attributed to their dietary specialization making geography still a variable that needs to be explored in further research.

Table 3. Differences in α-diversity and bacterial composition and diet based on geography.

Non-industrialised traditional populations	*Rural agrarian populations from low-to-middle income countries"*	*Westernised urban-industrialised populations*
Very high α-diversity	*High α-diversity*	*Low α-diversity*
↑ *Prevotella, Treponema, Clostridium, Catenibacterium, Eubacterium, Lachnospira* (127)	↑*Bacteroidetes, Ruminococcus, Blautia, Dorea, Treponema* (127)	↑ *Bacteroides* (127)
↑ *Prevotella, Eubaacterium, Oscillibacter, Butyricoccus, Sporobacter, Succinivibrio and Treponema* (124)	↑ *Bacteroidetes (Prevotella, Xylanibacter), Actinobacteria* (123)	↑ *Bacillota, Proteobacteria, Shigella, Escherichia, Bacteroides, Alistipes* (123)
	↑ *Prevotella, Lactobacillus, Ruminococcus, Oscillospira, Eubacterium, Dialister, Clostridium* (125)	↑ *Bifidobacterium, Bacteroides, Blautia, Dorea, Roserburia, Faecalibacterium, Ruminococcus* (124)
		↑ *Bacteroides, Alistipes, Clostridium, Oscillospira* (125)
Diet		
↓ CHO	↓ Animal protein and fat (123)	White CHO, sugar, meat and fish (124)
Wild food such as meat, honey, berries and tubers (124)	↑ Vegetables, legumes, cereals (123)	CHO, vegetables and animal protein (125)
	Rice, bread and lentils (125)	

In green the common species between hunter-gatherer and traditional farming or fishing populations are highlighted. Orange represents common bacteria among traditional farming or fishing population and western urban populations.

1.1.4.6 Lifestyle

Recent evidence suggests that physical activity may influence the human gut microbiome, however the lack of proper controls had led to inconsistent findings (128). A few interventional studies claimed that exercise was associated with an increase in the abundance of butyrate producers (129–131), though its effects on α- and β-diversity remain contradictory (130,132,133).

Smoking habits have been previously associated with distinct β-diversity profiles, but studies report only modest decreases or non-significant changes in α-diversity indices when comparing smokers to non-smokers. In terms of relative abundance, smokers generally exhibit higher abundance of Bacteroidetes and a reduction in Bacillota and Proteobacteria (134–137). Additionally, smoking has been associated with an increased susceptibility to *Clostridium difficile* infection (138).

1.1.4.7 Transit time

Recently, gut transit time has drawn interest due to its potential impact on the human microbiome composition (139–141). Several studies have associated longer transit time with higher α-diversity (140,142,143), possibly because prolonged transit allows for greater substrates availability and

fermentation, such as CHO and proteins, creating a favorable environment for slower-growing species (140,144).

Moreover, although further research is needed to determine the underlying mechanisms, very low stool frequency has been linked to an increased risk of cardiovascular and all-cause death (145–147).

At the microbiome level, certain bacteria appear to be linked to transite time. *A. muciniphila* or *F. prausnitzii* have been associated with longer transit times and better stool consistency (140,148,149). *A. muciniphila* is a mucin degrading bacteria that plays a role in maintaining gut mucosal integrity (62,150). Consequently, its reduction has been proposed as a potential disease biomarker, as lower levels have been reported in conditions such as IBD and obesity compared to healthy controls or patients in remission (151–154).

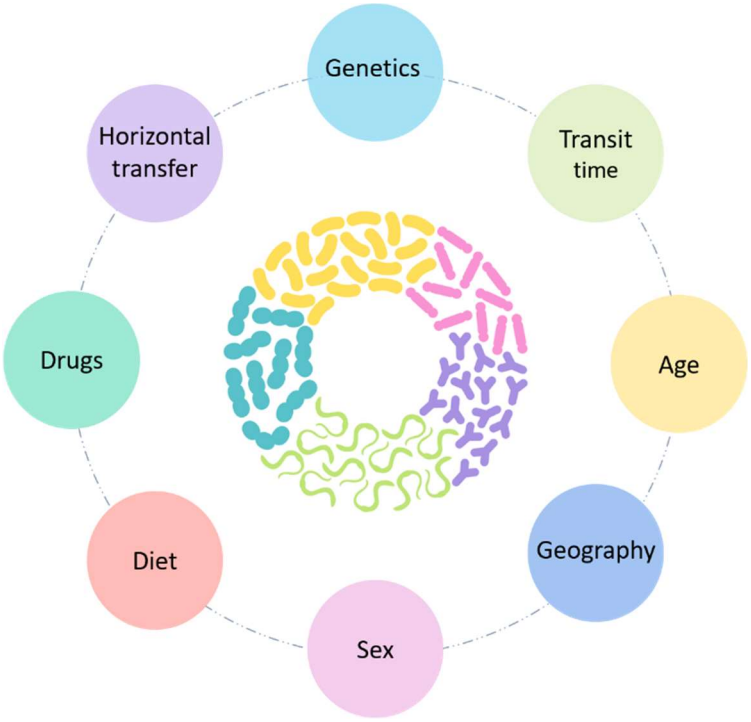


Figure 3. Summary of factors capable of influencing gut microbiota composition

1.2 Diet

1.2.1 *Dietary assessment methods*

When assessing the food intake of micro- and macronutrients in a population, biochemical markers are considered the gold standard option. Unlike dietary recall methods, they do not rely on an individual´s memory or reporting accuracy. Additionally, they can be acquired through non-invasive samples such as urine, saliva, blood and stool, making them a practical and objective tool for nutritional assessment (155). In this line of research, metabolomics, which comprise mainly two primary techniques (Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS)), can be promising for objective identification of dietary biomarkers. Both techniques can follow targeted or non-targeted approaches. Non-targeted metabolomics captures the full spectrum of metabolites present in a sample, including unidentified compounds. However, due to its high cost, complex data analysis and statistical challenges, targeted strategies are often preferred (156). To date, numerous nutritional studies have focus on identifying metabolomic signatures associated with dietary patterns, specific foods, or nutrients (157–161). Nevertheless, identifying universal metabolic markers remains challenging as they can be influenced by factors such as disease, stress, or age. Therefore, metabolomic approaches are often combined with traditional dietary assessment tools for comprehensive analysis (155).

Traditional methods include a huge variety of approaches (see Table 4). One reliable method for assessing dietary intake is the use of dietary records or food diaries. In this approach, participants or trained staff record all foods and beverages consumed over a specified period of time (between 2-7 days), including brand names, weighted portion sizes and cooking methods. However, accurate recording requires prior training and familiarity with the process, necessitating a high level of motivation from volunteers. Additionally, a single record reflects only short-term intake rather than habitual dietary patterns, requiring multiple recordings to estimate usual consumption, an aspect that poses challenges for large-scale population studies (162).

The 24-h dietary recall (24HR) method records the individuals´ food intake over the previous 24h. It is a versatile approach, as it can be performed online or offline, either self-administered or with the assistance of trained staff, thereby reducing costs. Similar to other dietary assessment methods, details such as food preparation methods, brand names, and portion sizes must be recorded, often with the aid of visual materials to enhance accuracy. Still, this method has limitations, as it relies, for example, on participants' memory and may be prone to recall bias. To improve the accuracy of habitual dietary intake assessment, at least three 24h recalls are recommended (two on random weekdays and one on a weekend). Additionally, trained staff professionals are needed for macro- and micronutrients quantification, making this method costly for population studies (163).

For large cohort studies, the Food Frequency Questionnaire (FFQ) is often prefered due to its cost-effectiveness and ability to be self-administered to capture habitual dietary intake. Moreover, it is a tool available in various formats, including quantitative, semi-quantitative, and qualitative versions, sometimes incorporating images to aid facilitating food portion size selection. However, like the 24HR, the FFQ relies on participants' memory and requires prior validation for the target population, typically through comparison with 24HR or dietary records (164).

Table 4. Most commonly used dietary assessment methods.

	Recalls	*Records or food diaries*	*FFQ*[1]
Options	1 or 7 day	1, 3, 4 or 7-day	Previous month or year, 4-day
Data Collection	Subjective current intake over a defined timeframe	Subjective current intake over a defined timeframe	Typical intake estimation over a span (typically 6 months or 1 year). Subjective
Advantages	Validity, extensive participant cooperation, and accuracy.	Reliable reference technique for validation studies due to its validity.	Economical method suitable for large-scale epidemiological studies. Minimal expertise required for quantification or data collection.
Limitations	An accurate assessment typically requires between 3 to 4 24HR. Trained staff is essential. Relies on memory that can introduce bias	Subjects must be highly motivated, and volunteers may inaccurately report proportions they perceive as correct rather than actual proportions.	Tailored to the study population and research objectives; utilizes a closed-ended questionnaire; prone to low accuracy (due to recall bias); necessitates precise evaluation of questionnaire development.

[1] Food Frequency Questionnaire

1.2.2 *Food composition databases (FCDB) and food composition tables (FCT)*

FCDB and FCT are fundamental in nutrition research, providing detailed information on the macronutrient and micronutrient content of foods and beverages. Food composition data are typically derived from either quantitative chemical food analysis of commonly consumed foods within a country or extracted from scientific literature (165).

FCDB are widely used by researchers, the food industry, governmental institutions, and consumers as they enable the calculation of energy and nutrient intake based on dietary assessed data. This, in turn, provides valuable insights into the dietary quality of individuals and populations. The primary sources of FCDBs include datasets released by national governmental agencies, as well as contributions from research institutions and private companies (166).

In Spain, 18 FCTs/FCDB have been developed to date. Early FCTs were available in book format and lacked of updated versions. It was not until 2010 that the Spanish Agency for Consumer Affairs, Food Safety and Nutrition (AECOSAN) launched the first official, free access and unified FCDB, known as "Base de Datos Española de Composición de Alimentos" (BEDCA; https://www.bedca.net/), designed in accordance with European Recommendations However, BEDCA present certain limitations, such as a relatively limited number of food items compared to international FCDBs such as the United States Department of Agriculture (USDA) database. Additionally, it does not include information on commonly consumed dietary supplements, vitamins or recipes (167).

In a effort to harmonize European FCDBs, the European Food Information Resource (EuroFIR) developed FoodExplorer, a search interface that enables users to access nutritional information from 33 international FCDB covering data from 31 different countries (168).

1.2.3 *Dietary Quality metrics or Dietary Quality indices (DQIs)*

Diet corresponds to a highly complex variable that is often simplified through the use of DQIs. The indices are nutrition-based metrics designed to assess diet quality by considering the consumption of specific nutrients or food groups, either individually or collectively, in terms of healthiness or unhealthiness. By summarizing overall dietary patterns into a single measure, DQIs offer an efficient way to account for dietary factors without introducing excessive complexity into analytical models. DQIs typically evaluate diet quality across three main categories: adherence to dietary guidelines, recommended food intake, and dietary variety (169). Various DQIs have been applied in research studies focusing on diet, each with distinct characteristics, summarized in Table 5.

Table 5. Summary of some of the most common indices used to measure diet quality at population level.

DQIs[1]	*Full name*	*Developed by*	*Key Components*	*What does it measure*	*Score*
HEI-2015	Healthy Eating Index 2015 (170–172)	USDA[2]	13 calorie-adjusted components divided into "adequacy" components which are food groups whose intake is recommended and "moderation" components, whose intake must be controlled. The latest version corresponds to HEI-2015	Degree of alignment with Dietary Guidelines for Americans to assess overall diet quality	Score A: 100-90 points Score B: 89-80 points Score C: 79-70 points Score D: 69-60 points Score F: 59-0 points
IASE	Healthy Eating index for Spanish population (98,173)	University of Alicante, Spain	Frequency of consumption by means of 10 variables that divide diet into "dairy", "weekly" and "occasional" consumption plus diet variety points that refers to diet diversity.	Adaptation of HEI index to Spanish population recommendations	Healthy: > 80 points Need changes: 80-50 points Not healthy: < 50 points
HFD	Healthy Food Diversity Index (174)	Christian-Albrecht's University of Kiel	Considers consumption of nutrient-dense foods such as fruits, vegetables, whole grains, lean proteins, and healthy fats to assess overall diet quality.	Measures diversity and nutritional quality of foods consumed based on German guidelines	Higher score indicates a more diverse and nutritious diet, associated with better health outcomes.
MAR	Mean Adequacy Ratio (151,175)	University of Oslo	Assessment of 16 essential micro/micronutrient defined as NAR[3] which includes proteins, fiber, vit. A, thiamine, riboflavin, niacin, vit. B6, 12, vit.C, D, E, folate, calcium, potassium, iron and magnesium. Average value of the 16 NARs is defined as MAR.	Individual adequacy ratios for essential nutrients by comparing actual intake to recommended intake of the country of assessment.	Higher MAR indicates greater proportion of recommended nutrient intake, suggesting better nutritional adequacy
PDI	Plant-Based Dietary Index (176–179)	Harvard T.H Chan School of Public Health, US	Intake of healthy and less healthy plant food groups included in the uPDI and hPDI	Evaluates diet quality based on amount of plant-based foods relative to animal-based foods	Higher score indicates richer intake of plant-based foods like fruits, vegetables, whole grains, nuts, and seeds, and lower intake of animal-based foods like meat and dairy
aMED	Alternative Mediterranean score (180,181)	Department of Nutrition, Simmons College, Boston, MA	Vegetables, legumes, fruits, nuts, whole grains, red and processed meat, fish and selfish, MUFA/PUFA[4], alcoholic drinks	Measure the adherence to a typical Mediterranean Diet (MedDiet)	Higher score indicates a more Mediterranean type diet. Scores from 0 to 9

DQIs[1]	Full name	Developed by	Key Components	What does it measure	Score
uPDI	Unhealthful Plant-Based Diet Index (177–179,182)	Harvard T.H Chan School of Public Health, US	Evaluation of intake of refined grains, potatoes, and sweets within a plant-based diet.	Version of PDI that focuses on intake of less healthy plant-based items	Higher score suggests higher intake of these less nutritious plant-based foods within an otherwise plant-based diet.
hPDI	Healthful Plant-Based Diet Index (177–179,182)	Harvard T.H Chan School of Public Health, US	Intake of healthy foods like fruits, vegetables, whole grains, nuts, and legumes within a plant-based diet	Version of PDI that emphasizes consumption of healthier plant-based foods	Higher score reflects diet rich in these nutrient-dense plant-based food groups.
MDI	Meat index (177–179,182)	Harvard T.H Chan School of Public Health, US	Intake of animal fat, dairy, eggs, fish and shellfish and animal-based foods	Version of PDI that is based on the consumption of meat and/or meat derived products	Higher score reflects higher consumption of meat.
MEDAS	14-item Mediterranean Diet Adherence Screener (183,184)	PREDIMED[5] Study	12 questions related to food consumption (olive oil, vegetables, fruit, meat, butter, sweetened or carbonated beverages, wine, pulsed, fish and shellfish, commercial pastry etc) and two questions of food intake habits	Focuses on Spanish Mediterranean diet patterns	Range: 0 to 14 points Good adherence: ≥9 points Poor adherence: ≤8 points

[1]Dietary Quality Indices
[2]United States Department of Agriculture
[3]Nutrient Adequacy Ratio
[4]Monounsaturated/polyunsaturated ratio
[5]Prevention with Mediterranean Diet

1.3 Diet and gut microbiota

As mentioned earlier, multiple factors shape the composition of the human gut microbiota, but diet has been suggested as a key determinant of interindividual differences (185,186). In this line of research, Westernized diets, characterized by high sugar and fat intake, have been linked to gut dysbiosis, whereas diets rich in vegetables and fruits are suggested to have anti-inflammatory properties (187). Additionally, previous findings indicated that higher DQI scores, which denote better diet quality, have been linked to an increase in bacterial α-diversity, with some exceptions like Unhealthful Plant-Based Diet Index (uPDI), that shows a negative association due to its focus on less healthy plant-based sources (182,183).

The importance of maintaining a healthy diet is further highlighted by a 2017 meta-analysis conducted within the Global Burden of Disease Study (GBD), which identified 15 dietary risk factors associated with morbidities and mortality worldwide (190).

In order to elucidate the relationship between diet and microbiome, two main types of studies are commonly conducted: observational studies monitor participants without altering their diet, allowing for the identification of natural associations, whereas interventional studies actively modify dietary patterns through supplementation or control diet modifications to assess causal effects (191). In this thesis, we primarily focus on observational population studies, as they provide a cost-effective approach to gathering preliminary data. These findings can act as a foundation for designing future interventional studies aimed at establishing causal relationships between diet and the microbiome (Table 6).

1.3.1 *The size matters: observation studies at population level*

1.3.1.1 Habitual diet

1.3.1.1.1 Europe

Early large-scale observational studies investigating the relation between diet and microbiota were conducted by *Zhernakova et al.* and *Falony et al.* (185,186). The first study, led by *Zhernakova et al.,* analyzed habitual dietary patterns in 1,135 volunteers from the Dutch LifeLies-DEEP Study (LLDeep) cohort. Significant associations were observed between microbiota and 60 dietary variables (False discovery rate (FDR) <0.1). Notably, high bacterial diversity was linked to the consumption of low-fat milk, coffee, tea and red wine, likely due to the polyphenolic content of the latter three. Additionally, red wine was associated with *F. prausnitzii*, a known butyrate producer. Conversely, whole milk, sweetened beverages, snacks, and diets high in CHO were correlated with lower bacterial diversity (186). Similarly, *Falony et al.* explored the association between diet and gut microbiome composition using The Flemish Gut Flora Project (FGFP) cohort (n= 1,106). They identified significant correlations between 10 dietary factors and overall microbiome community variation (β-diversity). Using an independent cohort (the previously named LLDeep), they partially validated five of them, more specifically the effect of coffee, beer, alcohol, fruits and soda (185).

Partula et al. confirmed similar dietary-microbiota association in a French cohort (n= 862), reporting that sugary drinks and high-fat foods such as fried products were related with lower α-diversity, while fish and raw fruits consumption were associated with greater microbial diversity. Expanding on previous

studies, they also analyzed β-diversity, identifying cheese, ready-to-eat meals, cooked fruits, raw fruits and fried products as responsible from up to 0.77% of the microbiome variation. At the species level, their findings diverged from prior research after multiple testing corrections. Notably, dairy products and raw fruits were positively associated with *Streptococcus salivarius* and *Lachnospira eligens,* respectively. In contrast, cheese consumption showed a negative correlation with *A. muciniphila*, a bacterium linked to gut barrier integrity. Additionally, meat, previously suggested to promote inflammation, was negatively correlated with *Blautia* and positively correlated with *Clostridium* (26).

More recently, the gut microbiota of the Spanish population was characterized for the first time by *Latorre-Pérez et al.* in a cohort of 530 volunteers from across the country. While no significant associations were found at the species level, several noteworthy relationships were identified at the genus level. Among different food groups, nuts presented the highest number of microbial associations (23 genera), likely due to their high polyphenol and fiber content, which may contribute to gut microbiome modulation. The genus *Flavonifactor* was positively correlated with sweetened drinks and inversely with nuts, fruits and some vegetables, showing an opposite direction when compared with *Akkermansia*. Similarly, *Ruminococcus* was identified as an indicator of meat consumption, with its abundance decreasing in response to "healthy choices" such as nuts and vegetables. In contrast, *L. eligens* displayed an opposite pattern (99). More recently, *Qin et al.* (n= 8,798) expanded on this knowledge, identifying a strong positive correlation between *Bifidobacterium* and dairy consumption, further supporting the role of diet in shaping the gut microbiota composition (87).

1.3.1.1.2 USA and UK

The American Gut Project (AGP) was launched in 2012 in the US and extended to other countries, enabling one of the largest population-based microbiome studies to date (n= 10,699). Findings suggested that the number of different plants consumed, specifically more than 30 types per week, was more strongly correlated with *F. prausnitzii* and *Oscillospira* (both SCFAs producers) than the overall dietary classification. Additionally, higher plant diversity in the diet was associated with a reduction in some ARGs, highlighting another potential benefit of plant-rich diets (27).

Asnicar et al. (n= 1,098) further explored diet-microbiome interactions, finding a positive correlation between α-diversity and the consumption of shellfish and white fish (188). Also, a total of 42 species were associated with at least five dietary exposures, with the strongest association observed between *Lawsonibacter asaccharolyticus* and coffee/tea consumption. At a broader level, two major clusters emerged: one primarily composed of butyrate-producing bacteria associated with healthy plant-based food choices, and another dominated by *Clostridium* species (among other), which were linked to animal products and less healthy plant-based choices. As a novel approach, they proposed to use microbiome composition to predict health and dietary patterns (188). A similar trend was partially validated by *Walker et al.,* (n= 1,423) who found that species richness was most strongly associated with fish, vegetables, fruit, tea, and coffee (28). A step further, *Berry et al.,* used gut microbiome composition as a part of a model for postprandial biochemical prediction with acceptable results. Findings were independenly validated using an independent cohort (192).

1.3.1.1.3 China

China has also contributed to microbiome research and long-term diet studies with a cohort of 1,920 individuals. Within food groups, positive correlations were found between dairy intake and *Bifidobacterium*, fish and seafood food consumption and *Coprococcus,* and processed meat and *Acinetobacter*. Conversely, *Roseburia* showed an inverse association with processed meat (193). *Zhang et al.* included a cohort of 702 participants from six different cities in China. While no significant differences in α-diversity across food groups except for eggs was obtained, β-diversity analysis revealed that whole grains and vegetables accounted for up to 1.46% of total inter-individual variation. At the taxonomic level, healthy food choices such as whole grains positively correlated with *Megasphaera*, while vegetables were negatively associated with *Eubacterium coprostanoligenes* and *Leuconostoc*. Additional correlations included a negative association between red meat and *Weissella*, a positive link between red meat and *Coprobacter*, and positive associations between dairy and *Anaerostipes*, as well as refined grains and *Lactobacillus.* (30).

Taking the research a step further, *Sun et al.* (n= 942) also explored the relationship between diet and fungi. Blueberries and buttermilk tea exhibited the highest number of fungal associations. Specifically, blueberries positively correlated with *Tetrapisispora blattae*, *Sugiyamaella lignohabitans*, *Kazachstania africana,* and *Kazachstania naganishii*, while buttermilk tea was linked to *Naumovozyma castellii*, *Botrytis cinerea,* and *Penicillium chrysogenum*. Within the fruit group, grapefruit correlated positively with *Zygosaccharomyces parabailii* and *Candida glabrata*, Papaya with *Saccharomycopsis fibuligera*, plum with *Multicellular* and *Sclerotinia sclerotiorum*; watermelon with *Scheffersomyces stipitis*. Coffee consumption was associated with *Saccharomyces paradoxus* (3).

1.3.1.2 Dietary patterns

1.3.1.2.1 Western diet

With globalization, the Western-type diet (WD) has spread worldwide. This diet is characterized by high consumption of saturated fats, ultra-processed foods, sugars, and salt often accompanied by a reduced intake of fiber, fruits, and vegetables (194). *Shikany et al*., investigated microbiome differences between the WD and a "Prudent diet" (rich in vegetables and fruits) in a study involving 517 men from the USA. Participants who reported higher adherence to WD presented higher BMI although no effects in bacteria richness were reported. At the genus level, *Alistipes*, *Anaerotruncus*, *Collinsella*, *Coprobacillus*, *Desulfovibrio*, *Dorea*, *Eubacterium,* and *Ruminococcus* were positively associated with the WD. In contrast, *Coprococcus*, *Faecalibacterium*, *Haemophilus*, *Lachnospira*, *Paraprevotella,* and *Prevotella* showed an inverse correlation (195).

Similar research has been conducted in Korea to compare dietary patterns. *Lim et al.* (n= 890) found that a traditional Korean diet (characterized by high intake of vegetables, seaweed and soybean) was associated with higher relative abundance of *Sutterella*, *Coprococcus,* and *Paraprevotella.* In contrast, a WD with consumption of instant noodles, meat, and snacks induced an increase in *Lachnospiraceae* and *Dorea* and a decrease in *Streptococcus* and *Haemophilus parainfluenzae* (196). Another Korean cohort (n= 1,199), was investigated the same year, comparing not only WD and traditional diet but also a rice-based diet. At the genus level, no clear separation between dietary patterns was observed, but significant correlations emerged between specific taxa and food groups. For example, *Prevotella* was

positively associated with bread, noodles, fish, nuts and refined grains, while *Ruminococcus* correlated with egg, fruits, milk, noodles and refines grains as well. Finally, *Bacteroides* was related with bread, legumes, seasonings, fast food and noodles (197).

1.3.1.2.2 Mediterranean diet

Few studies have assessed the effect of different dietary patterns at the population level. Among those available, the Mediterranean Diet (MedDiet) has been consistently associated with positive health outcomes and a reduced risk of cardiovascular (CVD) and metabolic diseases (198–201). At the population level, adherence to the MedDiet is commonly measured using DQIs. *Latorre-Pérez et al.* (n= 530) found that adherence to MedDiet was associated with three bacterial taxa proposed as biomarkers. An inversed correlation with *Flavonifractor plautii* (99), a species previously shown to decrease with the consumption of fruits, nuts, whole grains but to increase with sugary drinks (99,188). *F. plautii* plays a role in flavonoid degradation, potentially reducing their bioavailability (202). A similar inverse association was seen with *Ruminococcus torques*.

Conversely, *L. eligens* was positively linked to higher MedDiet adherence and to vegetable and fruits consumption, while negatively associated with meat intake (26,99,188). *Asnicar et al.* observed a similar relationship between *F. plautti* and *L. eligens* with MedDiet. Other bacterial taxa showing strongest associations with foods or food groups included: *Roserburia hominis* (positively with whole grains), *Agathobaculum butyriciproducens* (positively with coffee, tea, vegetables and nuts; negatively with in desserts), *Ruminococcus lactaris* (positively with vegetables and nuts), *H. parainfluenzae* (negatively with coffee, tea, meat, sugary drinks and alcohol; positively with fruits, whole grains and legumes), *F. prausnitzii* and *Bifidobacterium animalis* (both positively correlated with adherence to MedDiet)(188). *F. prausnitzii* is well known for producing SCFAs and has been linked to plant and vegetables consumption, as well as red wine, a polyphenol-rich beverage (27,186). In contrast, taxa negatively associated with the MedDiet included *E. coli*, *Ruminococcus gnavus*, *Ruthenibacterium lactatiformans*, *Pseudoflavonifractor* and several *Clostridium* species (*C. spiriforme, C. symbosyum, C. leptum, C. innoculum*), most of which were linked to less healthy dietary choices such as sugary drinks and desserts.

Despite these findings, population-level studies focusing on specific dietary patterns remain limited. Further research is needed to clarify how diets can be used for gut microbiota modulation.

1.3.1.3 Specific components or food groups

In addition to examining overall dietary patterns, studies have also explored the impact of individual foods and dietary components. For instance, moderate intake of red wine is often recommended as part of the MedDiet for promoting healthy aging (203). *Le Roy et al.* (n= 916) in the context of the Twins UK cohort, studied the effect of wine and other alcohol beverages on gut microbiota composition and diversity. Their findings showed that wine consumption was positively correlated with higher α-diversity levels and an increased proportion of three bacterial genera (*Phascolarctobacterium*, *Barnesiella* and *Prevotellaceace*). Most importantly, the association with α-diversity was replicated in two independent cohorts (FGPF and the AGP), supporting the robustness of these findings (204).

Another example of reproducible results relates to the intake of coffee and its link to *L. asaccharolyticus*, a butyrate-producing bacterium. This association was observed in both caffeinated and decaffeinated coffee drinkers (205). However, the implication of *L.asaccarolyticus* for health and disease remain to be fully understood (206–208)

Table 6. Summary of observational population studies carried out in the context of diet and gut microbiota in humans.

Author	*Nº participants*	*Dietary assessment method*	*Population*	*Sequencing method*	*Type of diet*
			Habitual diet		
Partula et al., 2019 (26)	862	FFQ[1]	Adults from France	16S rRNA	Habitual
Zhernakova et al., 2016 (186)	1,135	FFQ[1]	Dutch cohort	Shotgun	Habitual
Falony et al., 2016 (185)	3,948	FFQ[1]	Belgium FGFP[2] and Dutch LLDeep[3] cohorts	16S rRNA	Habitual
McDonald et al., 2018 (27)	>10,699	FFQ[1] and primary diet survey	Adults from USA, UK and Australia	16S rRNA	Habitual
Yu et al., 2021 (193)	1,920	FFQ[1]	Adults from two China cohorts	16S rRNA	Habitual
Walker et al., 2021 (28)	1,423	FFQ[1]	American FHS[4] cohort	16S rRNA	Habitual
Latorre-Pérez et al., 2021 (99)	530	FFQ[1]	Spanish population (99)	16S rRNA	Habitual
Asnicar et al., 2021 (188)	1,098	FFQ[1]	PREDICT 1[5] study in UK and 100 Americans for validation	Shotgun	Habitual
Qin et al., 2022 (87)	8,798	FFQ[1]	FINRISK study in Finland	Shotgun	Habitual
Koponen et al. 2021 (209)	4,930	FFQ[1]	FINRISK 2002 study in Finland	Shotgun	Habitual
Zhang et al., 2022 (30)	702	FFQ[1]	TARGET-C8[6] study in China	16S rRNA	Habitual
Sun et al., 2021 (3)	942	FFQ[1]	Chinese volunteers	Shotgun	Habitual
Delavy et al., 2023 (31)	821	FFQ[1] and electronic case report	French healthy volunteers	Shotgun and ITS3-4	Habitual
Shuai et al., 2022 (210)	1,244	FFQ[1]	Middle age and elderly from Chinese GNHS[7] cohort	ITS2, 16S rRNA and Shotgun	Habitual
Gacesa et al., 2022 (115)	8,208	FFQ[1]	DMP[8]	Shotgun	Habitual
			Dietary patterns		
Asnicar et al., 2021 (188)	1,098	FFQ[1]	PREDICT 1[5] study in UK and 100 Americans for validation	Shotgun	MedDiet[9]
Berry et al., 2021 (192)	1,002	FFQ[1]	PREDICT 1[5] study in UK and 100 Americans for validation	16S rRNA	Habitual
Latorre-Pérez et al., 2021 (99)	530	FFQ[1]	Spanish population	16S rRNA	MedDiet[9]
Wu et al., 2021 (197)	1,199	FFQ[1]	Adults from KNHANES[10]	16S rRNA	Korean traditional diet vs rice-based vs WD[11]

Author	Nº participants	Dietary assessment method	Population	Sequencing method	Type of diet
Lim et al., 2021 (196)	890	FFQ[1]	Korean volunteers	16S rRNA	WD[11] traditional diet
Shikany et al., 2019 (195)	517	FFQ[1]	Men from MrOs[12] study US	16S rRNA	WD[11] vs prudent diet
Shen et al., 2024 (189)	705	FFQ[1]	BLSA[13] Cohort	Shotgun	Healthy Plant-based diet vs unhealthy plant-based diet
Specific components or foods					
Manghi et al., 2024 (205)	22,347	FFQ[1]	Coffee consumption	Shotgun	Coffee
Le Roy et al., 2020 (204)	916	FFQ[1]	Alcohol consumption	16S rRNA	Beer, cider, red and white wine and spirits

[1]Food frequency questionnaire
[2]Flemish Gut Flora Project
[3]Dutch LifeLies- DEEP Study
[4]Framingham Heart Study
[5] Personalized Responses to Dietary Composition Trial-1
[6] Comparative evaluation of novel screening strategies for colorectal cancer screening in China
[7] Guangzhou Nutrition and Health Study
[8] Lifelines Dutch Microbiome Project
[9] Mediterranean diet
[10]Korea National Health and Nutrition Examination Survey
[11]Western-type diet
[12]The Osteoporotic Fractures in Men
[13] The Baltimore Longitudinal Study of Aging

1.3.2 *What´s next? Interventional studies*

Once correlations are identified at the population level, the next step is to demonstrate causality. This is typically achieved through gold-standard approaches such as human intervention or animal models' studies.

Human intervention studies are more limited due to ethical constraints, as harmful components can be tested in human trials. Additionally, these type of studies present challenges in controlling variables such as inter-individual variability, diet and other potential confounders including lifestyle, medication and supplements, demographics (211). Since interventional studies are over the scope of the present thesis, just some of them have been introduced in detail. However, a more extense summary, containing some of the most recent human intervention studies, can be found in Table 7.

1.3.2.1 Specific dietary patterns

Among the limited human intervention studies available, dietary fiber is the one of most extensively studied components for gut microbiota modulation. Defined as non-digestible CHO that escape absorption in the small intestine and reach the colon, dietary fibers serve as substrate for the gut microbiota (212).

Holscher et al., trial showed that 21 g of soluble fiber shifted Bacillota/Bacteroidetes (F/B) ratio while a 12-week prebiotic study increased both the F/B ratio and beneficial *Bifidobacterium* levels (213,214). *Bifidobacterium* produces SCFAs that lower GI tract pH, inhibit pathogenic bacteria, and enhance calcium and magnesium bioavailability (215).

β2-fructan supplementation increased Bifidobacterium, enhanced SCFA production, and reduced inflammation (216). Whole grain intervention decreased Enterobacteriaceae while increasing Lachnospira, improving stool metrics and SCFA production (217). Higher doses of fiber (50 g/day) consumption significantly enriched Bifidobacterium (218). Polyphenol-rich diets increased *Clostridium leptum* while reducing other bacterial species, though with limited analysis (219).

1.3.2.2 Nuts

Nuts, containing digestion-resistant compounds rich in fiber, polyphenols, and fatty acids, contain compounds that are resistant to digestion and reach the colon, where they can be metabolized by gut microbiota (220). Walnut consumption altered gut microbiota composition by increasing certain bacteria (*Faecalibacterium*, *Clostridium*, *Roserburia*, *Dialister*) while decreasing others (*Ruminococcus*, *Dorea*, *Oscillospira*), and improved cardiovascular markers (221). A larger study confirmed diversity findings but showed different microbial shifts, including increased *Bifidobacteria* (222). Beyond maintaining health, walnuts intake may also serve as complementary strategy to treat CVD. A recent article by the American Heart Association highlighted the potential cardiovascular benefits of walnut consumption (223).

Almond studies yielded mixed results with *Liu et al.* showing increased *Bifidobacterium* and *Lactobacillus* (224), while *Holscher et al*. found decreased *Bifidobacterium* but increased *Lachnospira*, *Dialister*, *Clostridium* and *Roserburia* (225). A more recent human intervention trial conducted in California expanded on these findings, observing increased *Lachnospira*, decreased pathogens, and

notably increased α-diversity (226). Mixed nut supplementation studies showed minimal or inconsistent microbiome changes despite cognitive improvements (227,228).

1.3.2.3 Cocoa and tea

Cocoa (*Theobroma cacao*) is rich in polyphenols with global consumption averaging 0.9 kg/year (highest in Switzerland at 11.6 kg/year) (229). Its flavonoids may reduce the risk of CVD (230–235) and hypercholesterolemia (236). *Tzounis et al*. found chocolate drinks increased beneficial bacteria proportional to flavonoid content (237). While 85% dark chocolate increased microbial diversity and improved mood (238), certain chocolate varieties reduced *Faecalibacterium* and microbial diversity in postmenopausal women (239).

Tea (*Camellia sinensis*) contains significant polyphenols (100-200 mg flavonoids per 250 mL) and, as the world's second most consumed beverage, has attracted attention for health benefits (240,241). While unabsorbed green tea polyphenols are converted to beneficial phenolic metabolites by gut bacteria (241–245), tea's effects on microbiota composition show inconsistent results compared to chocolate. *Li et al.,* study with Oolong tea demonstrated an increase in α-diversity and impact on bacterial populations (246). Other studies have reported an increase in *Dorea, Faecalibacterium, Roserburia, Bifidobacterium spp.* and *Eubacterium* following green tea consumption (245), although results remain inconsistent across different studies (244,247).

1.3.2.4 Coffee

Coffee arabica and *Coffee canephora* are the two most widely consumed coffee species globally, valued for their sensory properties and physiological effects. Notably, coffee is rich in antioxidants, fiber, caffeine, nicotinic, and chlorogenic acids (248,249). In a human trial involving 16 healthy individuals, the consumption of three cups of coffee per day led to a significant increase in *Bifidobacterium spp*. following the intervention period (250). Similar increase in *Bifidobacterium* was reported in an intervention trial involving non-alcoholic steatohepatitis (NASH) and diabetic patients after the administration of chlorogenic acid and caffeine, two of the key components of coffee (251). In contrast, in a study involving 30 volunteers the intake of a single dose of coffee did not achieve a significant impact on gut microbiota, suggesting that the microbiome-modulating effects of coffee may require sustained consumption rather than a one-time intake (252).

Table 7. Summary of intervention studies in humans in which microbiome has been analyzed.

Author	*Nº participants*	*Component*	*Dose (g/day)*	*Study duration (weeks)*	*Microbiome analysis method*	*Comparison*
Oliver et al., 2021 (218)	20	Diet rich in fiber	40-50	3	Shotgun	Individuals before vs after the intervention
Vanegas et al., 2017 (217)	81	Fiber from whole grains	35	8	NA	Whole grains vs refined grains
Vetrani et al., 2020 (219)	78	Diet rich in polyphenols	2.9	8	DGGE[1], qPCR[2]	Diet rich in polyphenols vs rich in PUFA[3] vs low polyphenols & PUFA vs high in polyphenols and PUFA[3]
Specific components						
Clarke et al., 2016 (216)	30	Fiber	15	10	qPCR[2]	Placebo vs β2-1 fructan group
Alfa et al., 2018 (253)	84	Fiber	30	14	16S rRNA	Resistant starch vs placebo in elderly and mid group
Holscher et al., 2015 (213)	21	Fiber	21	9	Shotgun	Placebo vs polydextrose vs soluble fiber
Holscher et al., 2018 (221)	18	Nuts (walnut)	42	7	16S rRNA and 18S sequencing	Placebo vs walnut supplementation
Bamberger et al., 2018 (222)	194	Nuts (walnut)	43	24	16S rRNA	Nut-free control diet vs walnut-enriched diet
Tindall et al., 2020 (223)	42	Nuts (walnut)	57-99	8	16S rRNA	Walnut diet to replace SFA[4] vs two vegetable oils
Liu et al., 2014 (224)	48	Nuts (almonds)	56	10	Culture	Almond group, control group and almond skin group vs baseline

Author	Nº participants	Component	Dose (g/day)	Study duration (weeks)	Microbiome analysis method	Comparison
Holscher et al., 2018 (225)	18	Nuts (almonds)	42	20	16S rRNA, Archaea F/R and 18S	Natural/whole roasted/chopped roasted/almonds and almond batter vs baseline
Dhillon et al., 2019 (226)	73	Nuts (almonds)	56.7	8	16S rRNA	Almond group vs control cracker group
Haskell-Ramsay et al., 2023 (228)	79	Nuts (mixed)	30	12	16S rRNA	Mixed nut group vs placebo
Rosas et al., 2020 (227)	20	Nuts (mixed)	42	3	16S rRNA	Mixed nut group vs placebo
Shin et al., 2022 (238)	48	Cocoa	30	3	16S rRNA	Control vs 85% dark chocolate vs 70% dark chocolate
Tzounis et al., 2011 (237)	22	Cocoa	150 mL	4	FISH[5]	High cocoa flavonol drink (494 mg) vs low-cocoa flavonol drink (29 mg)
Wiese et al., 2019 (254)	30	Cocoa	10	4	16S rRNA	70% Dark chocolate vs 70% Dark chocolate + 7 mg GAL-MFSA[6] vs 30 mg GAL-MFSA[6] vs 30 mg GAL-PUFA[7]
Hernández-González et al., 2024 (239)	19	Cocoa	100	2	16S rRNA	100 mg of milk chocolate in the morning vs 100 mg of milk chocolate at night vs placebo
Li et al., 2023 (246)	28	Tea (Oolong)	2.5	3	16S rRNA	2.5 g of Oolong tea vs placebo group
Yuan et al., 2018 (245)	12	Tea (Green)	400 mL	3	16S rRNA	400 mL 2-weeks intervention + 1-week washout period

Author	Nº participants	Component	Dose (g/day)	Study duration (weeks)	Microbiome analysis method	Comparison
Huang et al., 2023 (244)	13	Tea (Pu-erh)	300 mL	4	16S rRNA	Tea intervention group
Jin et al., 2024 (247)	10	Tea (Green)	100 mL	10	qPCR[2], TRFLP[8]	Tea intervention group
Jaquet et al., 2009 (250)	16	Coffee	10.2	6	qPCR[2], DGGE[1], FISH[5]	Coffee intervention group
Chong et al., 2020 (252)	30	Coffee	8	3 days	16S rRNA	Single dose coffee consumption
Mansour et al., 2020 (251)	26	Caffeine/chlorogenic acid	0.4 g	12	qPCR[2]	Patients with Diabetes/NASH[9]: 1. chlorogenic + caffeine; 2. chlorogenic + placebo; 3. caffeine + placebo. 4. Placebo

[1]Denaturing gradient gel electrophoresis

[2]Quantitative polymerase chain reaction

[3]Polyunsaturated fatty acids

[4]Short Chain Fatty Acids

[5]Fluorescence in situ hybridization

[6]GA lycopene formulated with medium saturated fatty acids

[7] GA lycopene formulated with polyunsaturated fatty acids

[8] Terminal restriction fragment length polymorphisms

[9] Non-alcoholic steatohepatitis

1.4 Contributory Citizen Science

Citizen science is a research approach that actively involves the general population in the collection, analysis, and dissemination of scientific data, usually as a part of a collaborative project led by professional scientists. Notably, data collection by participants often aligns with Sustainable Development European Goals (SDGs) such as quality education, good health and well-being, parterships for the goals and reduced inequalities, among others (255). Among the different citizens models available, we will focus on contributory citizens science, which specifically involve public as data gatherers only, while health departments are in charge of study design, data collection and results analysis (256).

This democratization of science, promotes public understanding, encourages engagement, and expands research capacity by enabling non-experts to contribute to gathering valuable data across diverse disciplines. Traditional research teams often face challenges in collecting large datasets across broad spatial and temporal scales, making citizen science a highly effective solution (257–259). Additionally, volunteers gain firsthand experience with the scientific process, improving their critical thinking skills and supporting a greater appreciation for evidence-based decision-making. Citizen science initiatives can also influence health policies by providing localized data that supports informed decision-making and helps communities to identify and address risk factors, ultimately strengthening societal resilience (255,258).

Notable examples of citizens science initiatives in microbiome research include the AGP (260) or the British Gut (now ZOE programme) (261) and more recently, The Microsetta Initiative (https://microsetta.ucsd.edu/). These large-scale projects have been made possible through the participation of volunteers who contribute stool samples and dietary information. The collected data has facilitated mapping gut microbial variation across diverse populations and has helped explore the intricate relationship between human gut microbiome, diet, and health outcomes.

2 HYPOTHESIS

HYPOTHESIS

Over the last years, large-scale studies have been conducted in several countries with the aim of understanding the effect of habitual diet on health and disease state through the modulation of the gut microbiome community. However, to the best of our knowledge, no population studies have been yet performed characterizing Spanish diet, demographic data and microbiome using shotgun sequencing data. More importantly, correlations derived from these studies doesn´t demonstrate causation, manifesting the necessity of isolation of viable species to further support a direct effect and understand the mechanisms behind. Additionally, results obtained are commonly shared through scientific publication that are accessible just for a minority, without taking in consideration the importance of general population, who is a key variable when carrying out this type of studies.

First, we believe that performing shotgun sequencing in a large cohort of individuals coming from different Spanish regions can add valuable knowledge to previous carried studies in Spain. Together with dietary data collected at several timepoints, this study will allow us to better understand how national dietary recommendations are able to influence the microbiome ecosystem and in turn, health.

Second, we theorize that isolation of viable species from human gut could be crucial when studying causal-effect relationships, providing deeper knowledge about the implication of the microbiome in the health-disease context.

Third, we hypothesize that enhancing open-science through the development of a webpage for sharing results could increase awareness of general public about the importance of science and their participation. Making science understandable and accessible to everyone might motivate volunteers for future participation in population studies.

3 OBJECTIVES

OBJECTIVES

The main objetive for the present thesis would be to explore the interplay between host-related factors, dietary patterns and gut microbial communities, with a particular focus on how national and international dietary guidelines shape the intestinal microbiome and their subsequent impact on human health outcomes.

As secondary objetives, we plan to:

First of all, identify how personal traits, geography, diet and quality of diet influence gut microbial composition and diversity using a new Spanish cohort and try to use the microbiome data generated to predict intake of certain food groups or overall diet quality.

Second, to identify potential strategies for stratifying healthy individuals according to their gut microbiome profiles.

Third, to isolate viable bacterial species from healthy human gut for future study of their effect on health outcomes using *in vitro* models.

Fourth, to promote public engagement in scientific research by creating a digital platform for communicating study findings to volunteers.

4 MATERIALS AND METHODS

4.1 Study design

Population longitudinal study in Spain (POP Sudy) was conducted between December 2020 and August 2024 at the Vall d´Hebrón Research Institute (VHIR), Barcelona, Spain. This cohort aimed to define what constitutes a "healthy microbiome" at national level and its relationship with diet.

We used an updated version of our previously developed and validated semi-quantitative short FFQ (sFFQ) (98) to assess participants´ dietary intake. Healthy volunteers completed three self-administered sFFQs over the course of a year, alongside providing three fecal samples, from which a subset of samples collected at baseline was analyzed to determine bacterial and fungal load and composition.

4.2 Study population

A total number of 1017 of healthy volunteers coming from different Spanish Autonomous Communities (CCAA) were recruited between December 2020 and August 2024. Enrollment was facilitated through announcements on social media and on the official Vall d´Hebron Hospital webpage, managed by the Communication Department. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local Ethics Committee of Vall d´Hebrón University Hospital, Barcelona (reference number: PR(AG)84/2020). All participants conducted written informed consent form prior to study entry. Exclusion criteria included individuals under 18 and over 75 years of age, as well as those with chronic diseases associated with gut microbiota dysbiosis (e.g., IBD, diabetes mellitus, autoimmune diseases). Additionally, participants were required to be antibiotic-free for at least three months prior to enrollment to ensure proper microbiota recovery.

To ensure that the recruited cohort was representative of the general population, we calculated the sampling fraction for the four region areas considered (Interior, North of Spain, Mediterranean and Islands). 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 CCAA. We then calculated the population size for the selected region areas by summing up the individuals from the corresponding autonomous communities. Using these values, we estimated the theoretical percentage for a sample size of 1017 individuals as follows: Theoretical percentage = (1017 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 indicated more accurate recruitment.

4.3 Dietary assessment

Habitual diet was assessed three times over a one-year period using an updated version of a web-based semi-quantitative sFFQ previously developed and validated by the same research group (98). This questionnaire focuses on measuring food consumption over the past month. Briefly, the questionnaire included 58 food items (see ANNEX 1) divided into 13 sections: vegetables, legumes and potatoes, fruits and dried fruits, cereals and derivatives, milk and dairy products, eggs, fish and meat, selfish, oils and fats, bakery and pastry, sauces, non-alcoholic drinks, alcoholic drinks, processed food, and others. Consumption frequency was classified into six categories: "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 defined as a "standard portion" (based on estimates from "Encuesta Nacional de Alimentación en la Población Adulta, Mayores y Embarazadas 2", ENALIA2 Survey, (262) as well as our own expertise), "half of the standard", and "double of the standard". To help participants accurately estimate their food intake, we provided standardized color photographs (see ANNEX 2). Additional information was collected on factors that could potentially impact gut microbiota such as age, gender, weight, birth type, smoking, blood type, self-reported specific diet, consumption of ready-to-eat food or sweeteners, liquid intake and supplements or medication use.

This updated version allowed for automated quantification of micro- and macronutrient intake and includes extra questions on chronic illness, stool frequency, pregnancy status, average steps per day (for volunteers with a smartwatch), and CCAA.

The questionnaire is available in Spanish, Catalan, English, and French to facilitate its use in population-based studies and enhance international collaboration (see ANNEX 3). It can be accessed at https://manichanh.vhir.org/sFFQ/login.php

4.4 Analysis of dietary information

Monthly consumption data was transformed into daily consumption frequencies. To do so, we calculated the gram per day (g/day) as follows: a reported consumption of 1 or 2 times per week was averaged to 1.5 times per week, which, when divided by the seven days of the week, resulted in a daily consumption frequency of 0.21. This value was then multiplied by the weight corresponding to the standarized serving size. For example, if the serving size for legumes was 150 g, the final intake would be 0.21 x 150 g = 31.5 g/day. The conversion factors for other frequencies were: 1 or 3 times per month = 0.066; 3 or more times per week = 0.64; once per day = 1; 2 or more times per day = 3.

Using this g/day information, we then calculated the energy and nutritional value of each item in the sFFQ based on our own developed food composition database. All calculations were automated to minimize human error during the process.

The impact of specific participant characteristics on dietary intake (food groups, food items and food nutrients) was calculated through permutational analysis of variance (PERMANOVA), using the adonis2 function from the vegan R package (https://cran.r-project.org/web/packages/vegan/index.html) and employing the Bray-Curtis method. Concretely, the variables analyzed corresponded to age, geographic region, workplace (hospital vs. non-hospital), gender, BMI, season, self-reported diet type, smoking status, sweetener consumption, menstruation or menopause status (if applicable), and bowel habits.

The relationship between DQIs and population characteristics was assessed using linear regression models implemented in MaAsLin2 while adjusting for potential covariates resulting from PERMANOVA analysis (smoke, gender, bowel frequency, region areas, age, season, sweeteners, BMI, workplace, diet grouped).

4.5 Development of Spanish food composition database (sFCDB)

We developed our own version of the Spanish food composition database (sFCDB) in 2020 based on food codes to facilitate quantification and diminish bias. In order to do so, we integrated data from the BEDCA Spanish database (263), Moreira´s table (264), and selected entries from FoodData Central, corresponding to the USDA nutritional database (265). When foods or menus items were missing from these dabases, we manually added them using specific food tables or homemade recipes provided by participants. The initial sFCDB contained 1104 foods and mixed dishes, grouped into 13 food group nutrients plus energy per 100 g of food. However, in 2022, we updated the database to include total sugar content, which was previously unavailable. Total sugars were estimated using updated versions of BEDCA, FoodData Central and Moreiras´ table (263,265). When direct information was unavailable, mean sugar value was calculated based on at least five nutritional labels from different supermarket products.

Additionally, we manually added 777 dietary supplements based on participants' responses to question 59 of the sFFQ regarding food supplements and drug intake. Nutritional information was sourced from product labels and company-provided data. When possible, values were quantified per 100 g of product; otherwise, information per recommended daily intake was used.

4.6 Interindividual variability

Leveraging the longitudinal design of the study, we assessed both intra- and inter-individual variability in diet at three different timepoints (baseline, six and twelve months) by calculating the Bray-Curtis similarity index for food items, food groups, and nutrient data. Intra-individual variability refered to the distance between baseline and six months (different season) for each participant. Oppositelly, inter-individual variability was measured using two different approaches: 1. Global approach that consisted on, for a given sample, computing the median of the distances between this sample and all

the other samples; 2. Seasonal approach where for a concrete sample, the median of the distances between this sample and all other samples was computed taken within the same season.

4.7 Calculation of DQIs

Diet is highly complex, with substantial variability between and within countries and individuals. Furthermore, dietary assessment methods differ widely. To simplify the analysis and enhance comparability with other studies, we calculated several DQIs. While numerous dietary indices exist in the literature, we selected the most appropriate ones based on previous studies that examined the relationship between diet and gut microbiome (98,151,169,188).

4.7.1 *Healthy Food Diversity Index (HFD-index)*

Some evidence suggests that a more diverse diet is associated with better health and a more diverse microbiota (27). The HFD index is a dietary measure designed to capture diet diversity, providing an effective way to assess population diet quality. It ranges from 0 to 1-1/n, with 1-1/n (0.99 in our study) refers to a very diverse diet, while 0 indicates no diversity. It can be defined as:

$$HFD - index = BI\ x\ \frac{HV}{0.26}$$

Where:

$$Berry - Index\ or\ BI = (1 - \sum s_i^2)$$

$$Health\ value\ or\ HV = \sum (hf_i\ x\ s_i)$$

$$s_i = \frac{grams\ consumed\ of\ each\ food\ group\ i}{grams\ of\ total\ consumption}$$

$$hf_i = health\ factor\ of\ each\ food\ group\ i$$

If we observe the hf_i we can see that the maximum value that can be achieved is 0.26 (see Table 8), thus, the division of HV by its maximum ensures that the score obtained is comprised between 0 and 0.99 (174). As food groups considered by HFD-index differ from our sFFQ, first we reagrouped our sFFQ food items into HFD-index food groups and assigned the corresponding hf_i based on German Guidelines (Table 8).

Table 8. HFD-index food groups, the corresponding food items in our in house sFFQ and hf value for each food group considered.

HFD-index food groups	*sFFQ food items*	hf_i
Vegetables, fruits, leaf salads, juices	1-11, 14-17	0.2628
Wholemeal products	19, 23	0.2044
Potatoes	12	0.146
White-meal products/peeled rice	18, 20, 22, 57	0.0876
Snacks and sweets	21, 44- 46, 51- 52, 56	0.0292
Fish/low-fat meat/low-fat meat products	34, 36-39	0.09
Low-fat milk/low-fat dairy products	25-27, 29-30	0.07
Milk/dairy products	24, 28, 31	0.05
Meat products, sausages, eggs	32-33, 35, 58	0.03
Bacon	N/A	N/A
Oilseed rape/walnut oil	N/A	N/A
Wheat germ oil/soybean oil	42	0.0056
Corn oil/sunflower oil	41	0.004
Margarines/butter	43	0.0024
Lard/vegetable fat	40,48	0.0008

Next, we calculated the HDF-Index using the provided formula in R studio.

4.7.2 *Healthy Eating Index 2015 (HEI-2015)*

HEI is a measure used to assess how well a diet aligns with the Dietary Guidelines for Americans. Over the years, several updates of HEI have been developed, including HEI-1995, HEI-2005, HEI-2010 and HEI-2015 (170)(Table 9).

As with the previous case, components assessed in the HEI do not directly align with our sFFQ food items requiring a reclassification for index calculation (170,171). Additionally, the HEI-2015 originally comprised nine adequacy components that were reduced into eight in our index by combining “Total fruit” and “Whole fruit” components. Since our sFFQ cannot distinguish between them, we assigned a maximum score of 10 (5 points for each original component).

As an extra step, we converted gram into cup or ounce equivalents following Food and Nutrient Database for Dietary Studies (FNDDS) 2017-2018, provided by USDA Agricultural Service (https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds-download-databases/). We finally calculated the HEI-2015 index using R Studio.

Table 9. HEI-2015 components together with maximum scores per component, values considered for assigning maximum or minimum scores and formula used.

HEI-2015 Components	*Maximum points*	*Standard for maximum score (maximum points)*	*Standard for minimum score of zero*	*Score calculation in between values*
		Adequacy		
Total fruit	10	≥0.8 cup equiv/1,000 kcal	No fruit	10× (cup equiv total fruit per 1,000 kcal/0.8)
Total Vegetables	5	≥1.1 cup equiv/1,000 kcal	No vegetables	5× (cup equiv total vegetables per 1,000 kcal /1.1)
Greens and Beans	5	≥0.2 cup equiv/1,000 kcal	No dark green vegetables or beans and peas	5× (cup equiv greens and beans per 1,000 kcal /0.2)
Whole Grains	10	≥1.5 oz equiv/1,000 kcal	No whole grains	10× (oz. whole grains per 1,000 kcal /1.5)
Dairy	10	≥1.3 cup equiv/1,000 kcal	No dairy	10× (cup equiv total fruit per 1,000 kcal /1.3)
Total Protein food	5	≥2.5 oz equiv/1,000 kcal	No protein foods	5× (oz. total protein food per 1,000 kcal /2.5)
Seafood and Plant Proteins	5	≥0.8 cup equiv/1,000 kcal	No seafood or plant proteins	5× (cup equiv total fruit per 1,000 kcal /0.8)
Fatty acids	10	(PUFA[1]+MUFA[2])/SFAs[3] ≥ 2.5	(PUFA[1]+MUFA[2])/SFAs[3] ≤ 1.2	10× ((ratio result- 1.2)/(2.5-1.2))
		Moderation		
Refined Grains	10	≤1.8 oz equiv/1,000 kcal	≥4.3 oz equiv/1,000 kcal	10 – (10× (oz. refined grains per 1,000 kcal – 1.8 / 2.5 (4.3-1.8))
Sodium	10	≤1.1 g/1,000 kcal	≥2.0 g/1,000 kcal	10 – (10× (g per 1,000 kcal – 1.1) / (2.0-1.1))
Added Sugars	10	≤6.5% of energy	≥26% of energy	10 – (10× ((((g added sugar × 4)/total kcal) ×100)- 6.5)/ (26-6.5))
Saturated fats	10	≤8% of energy	≥16% of energy	10 – (10× ((((g saturated fat ×9)/total kcal) ×100) - 8)/ (16-8))

[1]Polyunsaturated fatty acids

[2] Monounsaturated fatty acids

[3]Saturated fatty acids

4.7.3 *Mean Adequacy Ratio (MAR)*

To assess the global nutrient adequacy of our population´s diet, we calculated the MAR, which is based on the Nutrient Adequacy Ratio (NAR), itself determined by dividing the mean intake of a nutrient by the corresponding Spanish intake recommendation for that nutrient. The result is then multiplied by 100 to express it as a percentage. The MAR was obtained by averaging all individual NAR values (151,175):

$$MAR = \frac{\sum NAR_n}{16}$$

$$NAR_n = \frac{mean\ intake\ of\ nutrient\ n}{Spanish\ recommendation\ nutrient\ n} x\ 100$$

. Table 10 presents the NAR variables and the corresponding columns in the sFFQ as well as the Nutrition Reference intakes (NRI) for the Spanish population. If NRI values were unavailable, Adequate Intakes (AI) were used as substitutes (266).

Table 10. MAR components along with their corresponding sFFQ columns and NRI values for the Spanish population.

MAR				
	Micronutrient	***NAR_n***	***sFFQ***	***Nutrition reference intake Spain (NRI[1]; (266))***
1.	*Proteins*	NAR_1	Total_protein_g	0.83 g/kg
2.	*Fiber*	NAR_2	Total_dietetic_fiber_g	25 g/day
3.	*Retinol equiv (Vitamin A)*	NAR_3	Vitamin_A_µg_retinol_equiv	M[2]: 750 µg RE/day F[3]: 650 µg RE/day
4.	*Thiamine*	NAR_4	Tiamin_mg	M[2]: 1.2 mg/day F[3]: 1.1 mg/day
5.	*Riboflavin*	NAR_5	Riboflavin_mg	M[2]: 1.5 mg/day F[3]: 1.2 mg/day
6.	*Niacin*	NAR_6	Total_niacin_equiv (mg)	M[2]: 17 mg/day F[3]: 14 mg/day
7.	*Vitamin B6*	NAR_7	Vitamin_B6_mg	M[2]: 1.7 mg/day F[3]: 1.2 mg/day
8.	*Folates*	NAR_8	Total_folate_µg	330 µg/day
9.	*Vitamin B12*	NAR_9	Vitamin_B12 µg	2.4 µg/day
10.	*Ascorbic acid (Vitamin C)*	NAR_{10}	Vitamin_C_mg	75 mg/day
11.	*Vitamin D*	NAR_{11}	Vitamin_D_µg	12.5 µg/day

	Micronutrient	*NAR*$_n$	*sFFQ*	***Nutrition reference intake Spain (NRI[1]; (266))***
12.	*Vitamin E*	NAR_{12}	Vitamin_E_mg_α-tocoferol	M[2]: 13 mg/day F[3]: 11 mg/day
13.	*Calcium*	NAR_{13}	Calcium_mg	950 mg/day
14.	*Potassium*	NAR_{14}	Potassium_mg	3500 mg/day
15.	*Iron*	NAR_{15}	Total_iron_mg	AI[4]: M[2]: 9.1 mg/day F[3]: 18 mg/day
16.	*Magnesium*	NAR_{16}	Magnesium_mg	M[2]: 350 mg/day F[3]: 300 mg/day

[1]Reference nutritional intake, understandable by the general population that covers 97-98% of the population

[2]Male

[3]Female

[4]Adequate intake

To prevent higher-than-recommended intakes from compensating for lower intakes, each NAR was limited to a maximum of 100. A higher MAR score indicates better overall nutrient adequacy for individual's diet. The MAR calculation was performed using R Studio.

4.7.4 *Healthy Eating index for Spanish population (IASE)*

The IASE is an index developed by *Navarro et al.* based on the HEI index and supplemented with data from Spanish nutrition surveys. An adaptation of these variables to our sFFQ is presented in Table 11 (98,173).

Next, each variable, was assigned a score based on the frequency of consumption recorded in the sFFQ (Table 12). The total score was then calculated by summing all variables, with a maximum score of 100 points. Based on the final score, individuals were classified into three cateogories: healthy (score ≥ 80), need some changes (score 80-50) and not healthy (score < 50). Analysis was done using R Studio.

Table 11. Adaptation of IASE variables to our sFFQ items divided by consumption frequency.

IASE Variables	*sFFQ food items*
Daily consumption	
Grains and derivatives	18-23
Vegetables	1 - 12
Fruits	14-16
Dairy products	24-26, 28-31
Weekly consumption	
Meats	32-34, 36-39
Legumes	13
Occasional consumption	
Cold-processed meats	35
Sweets	44-46
Beverage	51-52
Diet variety	2 points if you meet daily recommendations and 1 point if you meet weekly consumption.

Table 12. Scores of IASE index based on frequency of consumption

		Points assigned		
FFQ frequency of consumption	*Metadata equivalent*	*Daily consumption*	*Weekly consumption*	*Occasional consumption*
No consumption	0	0	0	10
1-3 times/month	1	2.5	5	7.5
1-2 times/week	2	5	10	5
+3 times/week	3	7.5	7.5	2.5
1 time/day	4	10	2.5	0
+2 times/day	5	10	2.5	0

4.7.5 *Plant-Based Dietary Index (PDI), unhealthful Plant-Based Diet Index (uPDI) and healthful Plant-Based Diet Index (hPDI)*

Plant based diets have been associated with a lower risk of several diseases (177–179,182). To study the effect of such diets in the microbiome, three different plant-based indices were calculated based on *Satija et al*. using our dietary data which corresponds to PDI, uPDI and hPDI. Additionally, we calculated the MDI to assess the consumption of meat and meat derived products relative to plant-based foods (178).

To compute these indices, sFFQ responses were first grouped into the 18 food groups proposed by the original method. Alcohol as well as vegetable and fats such as margarine were not included in the indices but adjusted for during the analysis. For each index, the classified food groups were divided into quintiles of consumption (g/day) and assigned positive or reverse scores (Table 13). For positive scores (marked in the table as +), volunteers in the highest quintile received a score of 5, while those in the lowest quintile received a score of 1. For negative scores (-), the scoring pattern was reversed, with the highest quintile receiving 1 point and the lowest quintile receiving 5 points (Table 14). Once plant-based

and animal-based food groups were scored, all points were summed up to obtain the final indices. The food group classification and index calculation were performed using R Studio, applying an energy-adjusted method using residual method. A higher PDI score indicates greater plant consumption relative animal-based food, with a range of score that varies between 18 and 90 points.

Table 13. Summary of food groups and corresponding sFFQ items for PDI and its variants (hPDI, uPDI and MDI)

Food groups	sFFQ food items	NOTES	PDI	hPDI	uPDI	MDI
Plant food groups - Healthy						
Whole Grains	19, 23, 27×0.56	Remove soymilk	+	+	-	-
Fruits	14-16		+	+	-	-
Vegetables	1-9, 10-11, 47		+	+	-	-
Nuts	17		+	+	-	-
Legumes	13, 27×0.28	Consider soymilk	+	+	-	-
Vegetable oils	40-42		+	+	-	-
Tea and coffee	49-50		+	+	-	-
Plant food groups - Less healthy						
Fruit juices	52		+	-	+	-
Refined grains	18, 20-22, 57		+	-	+	-
Potatoes	12		+	-	+	-
Sugar sweetened beverages	51		+	-	+	-
Sweets and desserts	44-46, 56		+	-	+	-
Animal food groups						
Animal fat	43×0.70	Remove margarine	-	-	-	+
Dairy	24-26, 28-31		-	-	-	+
Egg	32		-	-	-	+
Fish or Seafood	36-39		-	-	-	+
Meat	33-35		-	-	-	+
Miscelanea o animal based foods	48, 58		-	-	-	+

Table 14. Summary of points assigned for each of the percentiles of the above classified food groups.

		Score			
	Quintile	PDI	hPDI	uPDI	MDI
Plant food groups - Healthy	> P20	1	1	5	5
	P20 - P40	2	2	4	4
	P40 - P60	3	3	3	3
	P60 - P80	4	4	2	2
	< P80	5	5	1	1
Plant food groups - Less healthy	> P20	1	5	1	5
	P20 - P40	2	4	2	4
	P40 - P60	3	3	3	3
	P60 - P80	4	2	4	2
	< P80	5	1	5	1

		Score			
	Quintile	PDI	hPDI	uPDI	MDI
Animal food groups	> P20	5	5	5	1
	P20 - P40	4	4	4	2
	P40 - P60	3	3	3	3
	P60 - P80	2	2	2	4
	< P80	1	1	1	5

4.7.6 *Alternative Mediterranean score (aMED)*

The aMED developed by *Fung et al.,* (180) is an adaptation of the original MedDiet scale proposed by *Trichopoulou et al.,* (181). aMED score ranges from 0 (minimum adherence) to 9 (perfect adherence) points by considering 9 food groups, which correspond to 1. For each category, including the fatty acid ratio, the median intake (g/day) was calculated. Healthy food groups (vegetables, legumes, fruits, nuts, whole grains, fish and fatty acids ratio) were scored with 1 if intake was above the median and 0 if consumption was below. For red and processed meats, participants received 1 point for intake lower intake below the median and 0 point if intake was higher. Alcoholic beverages followed a different scoring approach: For men, one point for consumption between 10-50g/day; For women, one point for consumption between 5-25 g/day (Table 15).

Table 15. aMED components, sFFQ items and criteria for maximum and minimum scores.

aMED Components	Items included	Standard for score 1 point	Standard for score 0 points
Vegetables	1-9,11	Greater than median intake (g/day)	Lower than median intake (g/day)
Legumes	10,13	Greater than median intake (g/day)	Lower than median intake (g/day)
Fruits	14-16, 52	Greater than median intake (g/day)	Lower than median intake (g/day)
Nuts	17	Greater than median intake (g/day)	Lower than median intake (g/day)
Whole Grains	19,23	Greater than median intake (g/day)	Lower than median intake (g/day)
Red and processed meats	33, 35	Lower than median intake (g/day)	Greater than median intake (g/day)
Fish and shellfish	36-39	Greater than median intake (g/day)	Lower than median intake (g/day)
MUFA[1]*/SFA*[2]	N/A	Greater than median intake (g/day)	Lower than median intake (g/day)
Alcoholic drinks	53-55	Men: 10-50g/day; Women: 5-25g/day	Values outside the corresponding ranges for men and women

[1]Monounsaturated fats

[1]Saturated fats

4.8 Analysis of individual responses

Finally, we consolidated the information from the three sFFQs. We classified foods and drinks into 13 groups based on the EUROCODE 2 classification. Nevertheless, we regrouped and subdivided them into 24 subgroups foods based on their potential and similar role on the gut microbiota modulation as shown in Table 16.

Table 16. Food groups and sFFQ items based on the EUROCODE and its potential effect on gut microbiota.

Food groups	*sFFQ food items*
Alcoholic beverage	53-55
Appetizers	57
Biscuits breakfast cereals and cereal bars	20-21
Chocolates and derivatives	45
Fats and oils	40-43
Fish and shellfish	36-39
Fruit and fruit products	14-16
Legumes	10, 13
Meats and eggs	32-34
Milk and dairy products except fermented milk	24-26, 28-29, 31
Non-Alcoholic beverage	27, 49-52
Nuts and seeds	17
Pastries and sweets breads	44
Potatoes and other tubers	12
Ready to eat meals	58
Sauces and condiments	47-48
Sausages and other meat products	35
Sugar and other sweets	46,56
Vegetables and vegetable products	1-9, 11
White bread	18
White grains and white pastas	22
Wholegrain or whole meal bread	19
Whole meal grains and whole meal pastas	23
Yogurt and fermented milk	30

4.9 Comparison with Global Burden of Disease (GBD) 2017

To compare major food and nutrient consumption assessed within the framework of the GBD study, we grouped our sFFQ items into 12 out of the 15 dietary risk factors defined by the GBD. For fruits, vegetables, legumes, whole grains, nuts and seeds, milk, red meat, processed meat, sugar-sweetened beverages, fiber and calcium, the median intake (g/day) was calculated and compared with the optimal and optimal range of intake defined by GBD study. For PUFA, we assessed their percentage contribution to total energy intake and compared it with recommended values. Sodium intake was not considered,

as our sFFQ data only captured sodium naturally present in foods, excluding sodium from added salt during cooking. Finally, seafood omega-3 and trans fatty acids were not evaluated due to the absence of these variables in our metadata. The classification of sFFQ items into the GBD dietary risk factors suggested is presented in Table 17.

Table 17. sFFQ items grouped by GBD dietary risk factors. Optimal level of intake as well as optimal range suggested by GBD is also provided.

GBD dietary risk factors	*sFFQ food items*	*GBD Optimal level of intake (optimal range of intake))*
Diet low in fruits	Items 14-16	250 g (200–300) per day
Diet low in vegetables	Items 1-8, 11	360 g (290–430) per day
Diet low in legumes	Items 10, 13	60 g (50–70) per day
Diet low in whole grains	Items 19,23	125 g (100–150) per day
Diet low in nuts and seeds	Item 17	21 g (16–25) per day
Diet low in milk	Items 24-26	435 g (350–520) per day
Diet high in red meat	Item 33	23 g (18–27) per day
Diet high in processed meat	Item 35	2 g (0–4) per day
Diet high in sugar sweetened beverages	Items 51-52	3 g (0–5) per day
Diet low in fiber	Fiber_g	24 g (19–28) per day
Diet low in calcium	Calcium_mg	1250 mg (1000–1500) per day
Diet low in polyunsaturated fatty acids	(((PUFA×9)/Energy_kcal) ×100)	11% (9–13) of total daily energy

4.10 Comparison of our population self-reported intake and the Nutritional Reference Intakes for Spanish Population

To assess whether volunteers met the Spanish dietary recommendations, we compared the macro- and micronutrient consumption data derived from our sFFQ with the NRI of the Spanish population, considering age and gender (266). The difference between reported intake and the required adequacy level was computed in comparison with 80% of the Spanish NRI (267), following the recommendation of the ANIBES study (268).

4.11 Sample collection and genomic DNA extraction

4.11.1 *Sample collection*

For Population study (n= 1017), collection kit was optimized (Figure 4) to facilitate stool sample collection across Spain. Briefly, the kit contained:

- One 15mL Falcon tube with 7mL of 95% ethanol, which both inactivates viruses and preserves the sample at room temperature during shipment.
- Instructions for sample collection
- An informed consent form to be signed (only in the kit at baseline)
- A spatula for transferring the sample into the tube
- A small garbage bag

- Gloves
- A biosecurity bag to ship the sample
- A pre-labeled mail envelope
- A prepaid postage stamp, allowing participants to send the sample at no cost

Participants were instructed to send the samples the same day after collection or to store them frozen in a household freezer until shipment was possible. Samples were mailed to our lab at room temperature using National post service (Correos, Spain) and, upon arrival, stored at -80ºC until further processing.

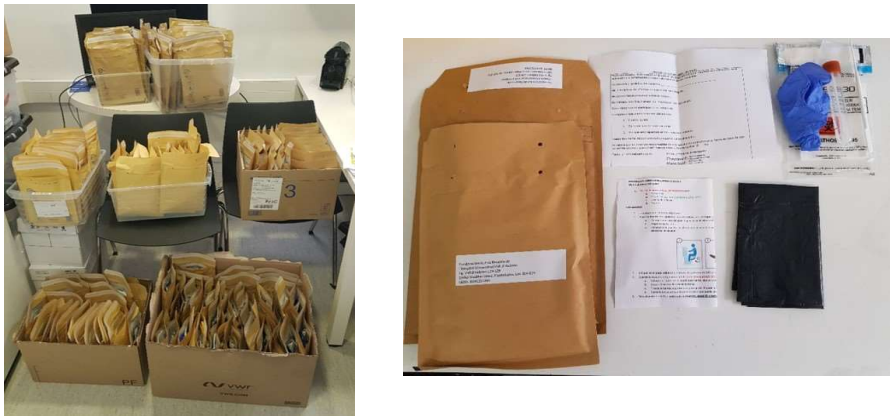


Figure 4. Collection kit used by participants in the population study.

4.11.2 *Sample processing and DNA extraction protocol*

In short, to prevent nucleic acid degradation, aliquots of feces (200 mg) were prepared on ice. Genomic DNA was then extracted from a randomized subset of 500 baseline samples, following the International Human Microbiome Standards (IHMS) guidelines (IHMS website; http://www.human-microbiome.org/; accessed on October 9, 2024) and as previously described (269). For extraction, each tube containing a fecal aliquot was supplemented with 800 mg of 0.1 mm sterilized zirconia beads and 250 µl of guanidine thiocyanate, followed by 40 µl of 10% N-lauroyl sarcosine and 500 µl of 5% N-lauroyl sarcosine to perform chemical lysis. Furthermore, guanidine thiocyanate also served as an effective SARS-CoV-2 inactivator. Samples were then incubated at 70ºC for one hour followed by mechanical lysis using a Beadbeater (Biospec Products ©) to disrupt gram-positive bacterial cell walls. Homogenized samples were further processed by washing with Polyvinylpolypyrrolidone (PVPP) and RNA was degraded using RNAse. Finally, DNA was precipitated with ethanol and the pellet was resuspended in 200µl of Tris-EDTA buffer.

DNA integrity was assessed by measuring absorbance ratios 260/280 and 260/230 using the NanoPhotometer® NP80 (IMPLEN). For DNA quality visualization, gel electrophoresis was performed using 1% agarose gel stained with RedSafe® and ran in 1X Tris Acetate EDTA (TAE) buffer for 1h at 70 V. 2 µl of DNA was mixed with 3 µl of loading dye and loaded into the solidified agarose wells, and InvitrogenTM TracklitTM 1Kb Plus DNA Ladder was used to evaluate DNA size (Thermo Fisher Scientific, Lithuania).

4.12 Enrichment protocol for fungi samples

To enrich fecal samples for fungi before DNA extraction a centrifugation step was performed on a randomly selected, gender-paired subset of samples (n= 100), based on the fact that bacterial and fungal cell differ in size. Estimation of centrifuge speed and time was calculated by Stoke´s law considering the following formula:

$$D = \left(\frac{18\eta \ln\left(\frac{R_f}{R_o}\right)}{(\rho_p - \rho_f)\omega^2 t} \right) 0.5$$

Where:

$D = particle\ diameter\ in\ cm; 4e^{-04}$ fungi and $4e^{-04}\ bacteria$
$\eta = fluid\ viscosity; 0.0089$
$R_f\ and\ R_o = final\ and\ initial\ radius\ of\ rotation\ respectively\ in\ cm;$
$\rho_p\ and\ \rho_f = density\ of\ the\ particle\ and\ fluid\ respectively\ in\ g/mL$
$\omega = rotational\ velocity\ in\ radians/second$
$t = time\ needed\ for\ sedimentation\ from\ R_o\ to\ R_f\ in\ seconds$

Briefly, 1.5 mL of sterile 1X phosphate buffered saline (PBS) (Sigma-Aldrich) and 10 glass beads of 2 mm (Merck) were added to 200 mg of fecal samples. Microbial cell walls were then disrupted for 60 s using a BeadBeater (Biospec Products®). The resulting homogenate was then filtered through a 40 µm cell strainer (Clearline®) to remove unwanted large particles. The filtered content was then centrifuged for 3 min at 201 g using the rotor Eppendorf A-4-62. The supernatant was discarded and the remaining pellet was resuspended in 15 mL of 1X PBS. To further remove bacterial cells and enrich the fungal fraction, the sample underwent a centrifugation step under the same conditions (201 g for 3 min). The supernatant was discarded, and the pellet resuspended in 1 mL of 1X PBS. A final centrifugation step (20 min at 10,000 g) was performed using an Eppendorf Centrifuge 5427R. The final pellet was frozen until DNA extraction was performed using standard in-house protocol (see Sample collection and genomic DNA extraction section).

4.13 Library preparation, sequencing and profiling

Between 10 and 20 ng/µl of DNA from each sample was sent to Novogene (UK) for library preparation and sequencing using the Illumina NovaSeq6000 platform (Novogene, UK).

Metagenomic sequencing process yielded an average of 5 Gb of data per sample. Pre-processing and decontamination of the sequence reads were conducted using the KneadData v0.7.4 pipeline (https://huttenhower.sph.harvard.edu/kneaddata). KneadData employed Trimmomatic for quality filtering and subsequently aligned the reads to a human reference genome using Bowtie 2. Reads shorter than 50% of the input length and those aligning with the human genome were excluded from further analyses. Taxonomic bacterial profiles were derived from the intermediate output of MetaPhlan4 within

the HuMANn3 pipeline, while functional profiles were generated from the final output (39). Taxonomic profiles, presented in stratified relative abundance from the phylum to Species-level Genome Bin (SGB) level, did not require normalization. Instead, species-level stratified abundances were extracted. α-diversity was assessed using the Chao1 and Shannon indices (269), and β-diversity was evaluated using PERMANOVA; adonis2 function, vegan R package.

Functional bacterial profiles generated by HuMANn3 provided gene families and MetaCyc pathways. To ensure data quality, MetaCyc pathways were filtered to exclude unmapped and unintegrated reads, and pathways with less than 0.001 abundance or 0.1 prevalence (those accounting for less than 0.1% of the total abundance in at least 10% of the samples) were also discarded. Pathways were then sum-normalized to counts per million (CPM) before further analysis.

Fungal taxonomic and functional profiles were obtained running FunOMIC2, an unpublished updated version of FunOMIC (94). Raw counts were normalized using the Counts Per Million (CPM) method, implemented in the "edgeR" package in R. Fungal α- diversity was measured by calculating the Chao1 index on raw fungal species counts and the Shannon index on CPM-normalized counts. β- diversity was evaluated by computing Bray-Curtis distances.

4.14 Bacterial and fungal load

Fungal and bacterial loads were estimated in fecal samples by targeting the V4 region of 16S rRNA for bacteria and ITS2 region for fungi. Amplification was performed using a 7500 Fast Real-time PCR system (Applied Biosystems, Foster City, CA, USA). The fungal ITS2 region was amplified using ITS2-fungi-sense (5'-GTG ART CAT CGA ATC TTT-3') and ITS2-fungi-antisense (5'-GAT ATG CTT AAG TTC AGC GGG T-3') primers. The V4 hypervariable region of the 16S rRNA gene (290 bp) was amplified using the following primers: V4F_517_17 (5'-GCC AGC AGC CGC GGT AA-3') and V4R_ 805_19 (5'-GAC TAC CAG GGT ATC TAA T-3').

The qPCR was performed in a 25 µl final volume, containing Power SYBR green PCR master mix (Fisher Scientific, Spain) and 100 nM of each primer. For the amplification of the hypervariable region V4, the reaction conditions were 50ºC for 2 min, 95ºC for 10 min followed by 38 cycles at 95ºC for 15 s and 60ºC for 1 min. For the amplification of the ITS2 region, the conditions were 50ºC for 2 min, 95ºC for 2 min followed by 40 cycles at 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 60 s, and a final extension cycle of 72ºC for 10 min. For bacterial load quantification, genomic DNA extracted from the fecal samples was diluted at 1/1000 and amplified in duplicate, while for fungi, samples were analyzed in triplicates at 1/100 to ensure accuracy. Additionally, mean values were calculated.

Melting curve analysis was performed to confirm amplicon specificity. To quantify microbial loads, standard curves were generated using calculated amounts of linearized plasmids containing the amplified region of the reference bacterium. Plasmid concentration was measured using a NanoPhotometer® NP80 (IMPLEN), and the number of gene copies was calculated based on the molecular weight of the plasmid. Serial dilutions of the template DNA were amplified to extrapolate

bacterial (from 10^{-2} to 10^{-7}) and fungal (from 10 to 10^{-6}) copy numbers. Results were expressed as copies per gram of stool.

4.15 Development of IBD-similarity index

To assess microbiome alterations associated with IBD, we developed a disease similarity index to quantify the grade of resemblance between the gut microbiome composition of healthy individuals (n= 500) and those diagnosed with non-communicable gastrointestinal diseases, such as IBD. The IBD-similarity index was defined as one minus the median weighted or unweighted UniFrac distance between any given healthy sample and a reference plane of 321 IBD samples from previous studies (208 from CD patients and 113 from UC patients)(269).

4.16 Culturomics Pilot Study

4.16.1 *Selection of suitable donors for bacterial culture and sampling*

All participants selected for culturomics belonged to POP study and should meet the following inclusion criteria:

- No antibiotic intake in the previous three months
- Residence or workplace located in Barcelona to ensure timely sample collection and transport

4.16.1.1 Experiment 1: Collection of healthy human gut species

This experiment aimed to isolate a collection of bacteria (non-targeted approach) from healthy individuals for subsequent testing on human explant tissue and animal models. To select the appropriate individuals, IBD-similarity index was calculated for the 500 healthy individuals with available microbiome data. Two volunteers with a very low dysbiosis score, high Chao1 and Shannon diversity scores, and aMED/hPDI scores above the median value were selected and asked to provide a new single, ethanol-free sample, stored in their domestic freezer (-20ºC) for up to 24h before being transported in dry ice by laboratory staff to maintain the cold chain. Upon arrival at the laboratory, five aliquots of 300 mg each were prepared on dry ice and were stored directly at -80ºC until shipment to the INRAE facility (France) on dry ice for further cultivation.

4.16.1.2 Experiment 2: Collection of IBD species

To investigate the role of microbiome in IBD, this experiment focused on previously reported bacterial species that are enriched or depleted in IBD patients, concretely *E. coli* and *F. prausnitzii.* One healthy volunteer and two CD patients with higher relative abundance of each of the targeted bacteria were selected to provide a new single, ethanol-free fecal sample. Again, the fecal samples were stored and prepared under the conditions as in Experiment 1 before shipment to INRAE, France.

4.16.2 *Update of MALDI-TOF database*

Identification of bacterial isolates are often carried out using MADI-TOF technique by comparison of the spectra generated by the interested pure culture against a pre-existing database. Our MALDI-TOF

default database (MALDI Biotyper Sirius, Bruker) contained spectra from 4320 species. Three additional free online spectra databases were added to improve spectra identification and bacterial diversity. ClostriTOF v2.0 (270) comprise 142 bacterial strains within Clostridia class, Zenodo v4.2 and EMBALIMB v3 enclosed 1,601 and 1,142 bacterial strains coming from Robert Koch Institute (271) and Japan collection of microorganisms in collaboration with the Medical Mycology Research Center, Center for Conservation of Microbial Genetic Resource (Gifu University) and Institute of Tropical Medicine, respectively.

4.16.3 *Culture media*

All media compositions were given per liter of osmotic water and prepared following the recommendations of the Leibniz Institute DSMZ and manufacturers' instructions:

Table 18. Recipe of all media used for culture of human derived gut samples.

BD Bacto™ Brain Heart Infusion supplemented (L-YHBHI.4) + 10% Rumen fluid	
Ingredient (1000 mL)	***Concentration***
Main solution (1000 mL)	*1x*
Brain Heart Infusion (ref 237500, BD Bacto™)	37 g
Bacto Yeast Extract	5 g
Hemine solution	10 mL
Rumen fluid	100 mL
Maltose	0.5 g
Cellobiose	0.5 g
Almidon soluble	0.5 g
Cysteine	0.5 mg
Osmotic water	890 mL
Agar	15 g for solid media
Resazurin	1 mL for liquid media

AccuDia™ Gifu Anaerobic Media (GAM) Broth	
Ingredient (1000 mL)	***Concentration***
Main solution (1000 mL)	*1x*
GAM Broth, Modified powder (ref 05433, Shimadzu AccuDia™)	41.7 g
Resazurin	1 mL for liquid media
Agar	15 g for solid media
Osmotic water	999 mL

BD Difco™ LB Broth, Miller (Luria-Bertani)	
Ingredient (1000 mL)	***Concentration***
Main solution (1000 mL)	*1x*
LB Broth, Miller (Luria-Bertani) (ref 244620, BD Difco™)	25 g
Agar	15 g for solid media
Osmotic water	1000 mL

BD BBL™ MacConkey Agar	
Ingredient (1000 mL)	***Concentration***
Main solution (1000 mL)	*1x*
MacConkey (ref 211387, BD BBL™)	50 g
Agar	15 g for solid media
Osmotic water	1000 mL

DSMZ Medium 1611- YCFA medium modified	
Ingredient (1000 mL)	***Concentration***
Main solution 1611 (1000 mL)	*1x*
Casitone	10 g
Yeast extract	2.5 g
Glucose	5 g
MgSO4 x 7 H2O	0.0448 g
CaCl2 x 2 H2O	0.0900 g
KH2PO4	0.4492 g
K2HPO4	0.4492 g
NaCl	0.9 g
Resazurin solution	1 mL for liquid media
Osmotic water	979 mL
Hemin solution	20 mL
NaHCO3	1 g
L-Cysteine HCl	0.5 g
Agar	15 g for solid media
Volatile fatty acids (2.7 ml)	*1x*
Acetic acid	1.9 mL
Propionic acid	0.7 mL
iso-Butyric acid	90 µl
n-Valeric acid	100 µl
iso-Valeric acid	100 µl
Vitamine solution (500 ml) we take 80 µL per Hungate tube or 10 mL for 1L	***20x***
Biotin	0.02 g
Folic acid	0.02 g
Pyroxidine hydrochloride	0.1 g
Thiamine-HCl x 2H2O	0.05 g
Riboflavin	0.05 g
Nicotinic acid	0.05 g
D-Calcium pantothenate	0.05 g
Vitamin B12 (10x)	0.01 g
p-Aminobenzoic acid	0.05 g
Lipolic acid	0.05 g
Distilled water	500 mL

Anaerobic media (DSMZ 1611, L-YHBHI.4 + 10% Rumen fluid (RF) and GAM) were prepared using two oxygen exclusion techniques; the Hungate technique for liquid cultures and the anaerobic chamber for plate cultures.

For the Hungate method, main solution for each of the medium (excluding cysteine) was prepared as described in Table 18. Resazurin was added as a redox indicator. The media were then deoxygenated by boiling in a vessel equipped with a chimney reflux to prevent overflow. Boiling continued until the resazurin changed color from blue to pink (oxidation) and to colorless (reduced), indicating oxygen removal. The medium was then cooled down with CO_2, introduced via a needle jet, to ensure no degradation of subsequent reagents will take place.

At this stage, cysteine and any required volatile fatty acids were added to the solution. The deoxygenated medium was then transferred into a 250 mL Schott bottle, influxed with N_2, and immediately sealed with butyl rubber bungs after the removal of the gassing needle. Bottles were then autoclaved at 121°C for 15 min. For liquid in 96-well plates, 200 µl of the prepared medium was distributed in each of the wells under anaerobic conditions. The remaining medium was aliquoted and stored inside the anaerobic chamber for future subculturing of isolated bacterial strains.

For agar plate culture, the anaerobic chamber was used. Main solution was prepared according to Table 18 and autoclaved at 121°C for 15 min. Notably, resazurin was excluded from solid media preparations. After cooling, 50 mL of medium was poured into each rectangular dish (Nunc OmniTray Single-Well Plate) under a Class II Biological Safety Cabinet. Plates were left at room temperature to solidify, then placed in the anaerobic chamber to reduce for approximately 48h prior to use.

4.16.4 *Single-Cell Dispenser (SCD)*

The SCD B. SIGHT (Cytena GmbH, Germany) employs a microfluidic system, that generates microdroplets from diluted samples and integrates optical components for real-time cell detection and image recording. Upon ejection from the cartridge's nozzle carrying the diluted fecal fluid, droplets containing single cells are dispensed into designated plates (agar or liquid), while droplets with no or multiple cells are discarded via vacuum. The SCD remained inside the anaerobic chamber throughout the entire experimental period and was used for all experiments except *E.coli*.

4.16.5 *Bacteria culture and isolation*

4.16.5.1 Escherichia coli

E. coli is a facultative anaerobe; therefore, its cultivation required slightly different conditions.

Briefly, 300 mg of fecal aliquots from CD patients with the most elevated levels of *E. coli* were homogenized in 5.7 mL of sterile PBS. Serial dilutions (10^{-1} to 10^{-8}) were prepared, and 100 µL of dilutions from 10^{-2} to 10^{-8} were plated onto LB (Luria-Bertani), Miller, and MacConkey agar plates. Media were prepared according to the manufacturer's instructions (see Table 18).

Plates were incubated at 37°C for 1-4 days under aerobic and anaerobic conditions by using two different incubators. Individual colonies were then hand-picked and subjected to multiple rounds of streaking on LB, Miller agar plates to obtain pure cultures.

4.16.5.2 *Faecalibacterium prausnitzii*

For isolation of *F. prausnitzii*, 300 mg of feces from a healthy volunteer known to hold high level of the species were processed inside the anaerobic chamber. The sample was homogenized in 5.7 mL of reduced PBS and subjected to serial dilutions (10^{-2} to 10^{-6}). A 100 µL aliquot of the 10^{-4} dilution was loaded into the B. SIGHT single-cell dispenser (SCD, Cytena GmbH, Germany) and distributed into twelve 96-well plates containing liquid DSMZ 1611 and L-YHBHI.4, supplemented with rumen fluid (RF).

Plates were then incubated at 37°C for 1-4 days in anaerobic conditions. Growth was monitored via optical density (OD) using the Infinite® 200 PRO plate reader (TECAN). Wells with OD > 0.2 were then plated onto agar containing the same media to ensure proper identification, followed by re-cultivation in 1 mL of liquid media. Plates were then incubated at 37°C for 1-4 days in anaerobic conditions and optical density (OD) measured (Infinite® 200 PRO, TECAN) for growing control.

4.16.5.3 *Collection of human gut species*

Fecal samples (300 mg) from two healthy volunteers were processed under anaerobic conditions. Each sample was homogenized with 5.7 mL of reduced PBS and serially diluted (10^{-2} to 10^{-6}). A 100 µL aliquot of the 10^{-4} dilution was dispensed using the SCD B. SIGHT into 96-well liquid and 384-well agar plates containing GAM and L-YBHI.4, supplemented with RF. For each medium type, one liquid and three agar plates were used.

Plates were then incubated anaerobically at 37°C for 1-4 days. Liquid cultures were monitored for growth by OD measurement. Wells with OD > 0.2 were then re-inoculated in 1 mL of the same media and spotted onto agar plates for identification. Agar plates were monitored for visible colony formation, and growing colonies were transferred to 1 mL of liquid media and re-spotted onto agar for identification.

4.16.6 *Bacterial strain identification*

Identification of pure isolates was carried out using the MALDI-Biotyper Sirius system (Bruker Daltonic). Briefly, cell biomass from single colonies was applied onto a MALDI target plate, treated with 1 µL of formic acid for cell wall disruption and protein extraction. Once the spot dried, 1 µL of MALDI matrix (Bruker IVD HCCA solution) was added. Spectra acquisition was performed according to the manufacturer's instructions.

Isolates with MALDI scores below 1.7, indicating low-confidence identification, were further analyzed via 16S rRNA sequencing. Bacterial DNA was extracted using FTA membrane and 700 bp including V3 region was amplified as described previously (272).

4.16.7 *Strain storage*

All identified isolates using the MALDI-Biotyper were cultured in appropriate liquid media to generate sufficient biomass. A 750 µL aliquot of each culture was transferred to a cryogenic tube and mixed with sterile glycerol at a 1:3 ratio to achieve a final glycerol concentration of 20%. The step was prepared inside the anaerobic chamber. All tubes were immediately stored at -80°C to ensure viability of the isolated species.

4.17 Website development

We built a website dedicated to this study (https://manichanh.vhir.org/POP/en/, username: reviewers, password: reviewers), 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 sFFQ), and, if available, their microbiome sequencing results, including bacterial composition, and measures of α-diversity. Nutrient intake data was compared to the guidelines established by the AESAN, while DQIs, food groups and α-diversity scores were compared to the population median found in this study. Nutrient intake data and DQIs could 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 were 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 were anonymized and password-protected, ensuring each participant may only access their own information.

4.18 Statistical Analysis

Analysis was conducted using RStudio v4.3. Covariates such as gender, age, BMI, geographic region, smoking status, season, and workplace were examined for their influence on microbiota variation using the PERMANOVA test implemented via the adonis2 function in the vegan package (https://cran.r-project.org/web/packages/vegan/index.html) on both weighted and unweighted UniFrac distance metrics.

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 5-fold cross-validation (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 sequencing data, differential abundance (DA) analysis of the microbial community was carried out using MaAsLin2 (Multivariate Association with Linear Models) (273). This analysis examined variations in categorical population characteristics, adjusting for confounding variables such as gender, BMI, and age. The resulting p-values were adjusted for the FDR. Associations were considered significant when the coefficient exceeded 1 (in most cases), and the q-value was less than 0.05. Spearman correlation and Mann Whitney U tests were employed to associate dietary data and numerical traits with microbiome profiles and diversity measures.

For functional analysis, Spearman's correlations between α- diversity indices (Chao1 and Shannon) and pathway abundances were calculated and adjusted for FDR. Only correlations with a rho value between -0.4 and 0.4 and an FDR < 0.05 were considered significant and retained for further examination. Association analyses were then conducted between these pathways and dietary variables (food items, food groups, and nutrients) using the Spearman correlation test.

To investigate changes in potential microbial community pathways based on personal data, linear models were used as implemented in MaAsLin2, adjusting for bowel movement (transit time), gender, BMI, age, smoking status, geographic region, and season as fixed effects, using MetaCyc pathway information. To enhance result interpretation, pathways were grouped into their MetaCyc parent categories up to seven levels, with level one representing the broadest biological function and level seven the most specific. Pathways with multiple parent categories were duplicated and assigned to each relevant parent for visualization and interpretation purposes.

5 RESULTS

5.1 POP Study: a comprehensive cohort for studying diet and microbiome

Out of 4,124 initially interested participants, only 1,241 met the inclusion criteria. A total of 216 individuals were further excluded for failing to provide either stool samples or complete the online sFFQ, while eight individuals were removed due to incomplete dietary information (more than one missing question regarding items 1-57). This resulted in a final baseline cohort (sFFQ_0) of 1,017 volunteers. At six months (sFFQ_1), 844 participants remained, and by twelve months (sFFQ_2), the cohort size was composed by 754 volunteers. The project was officially named the "POP" study (Figure 5).

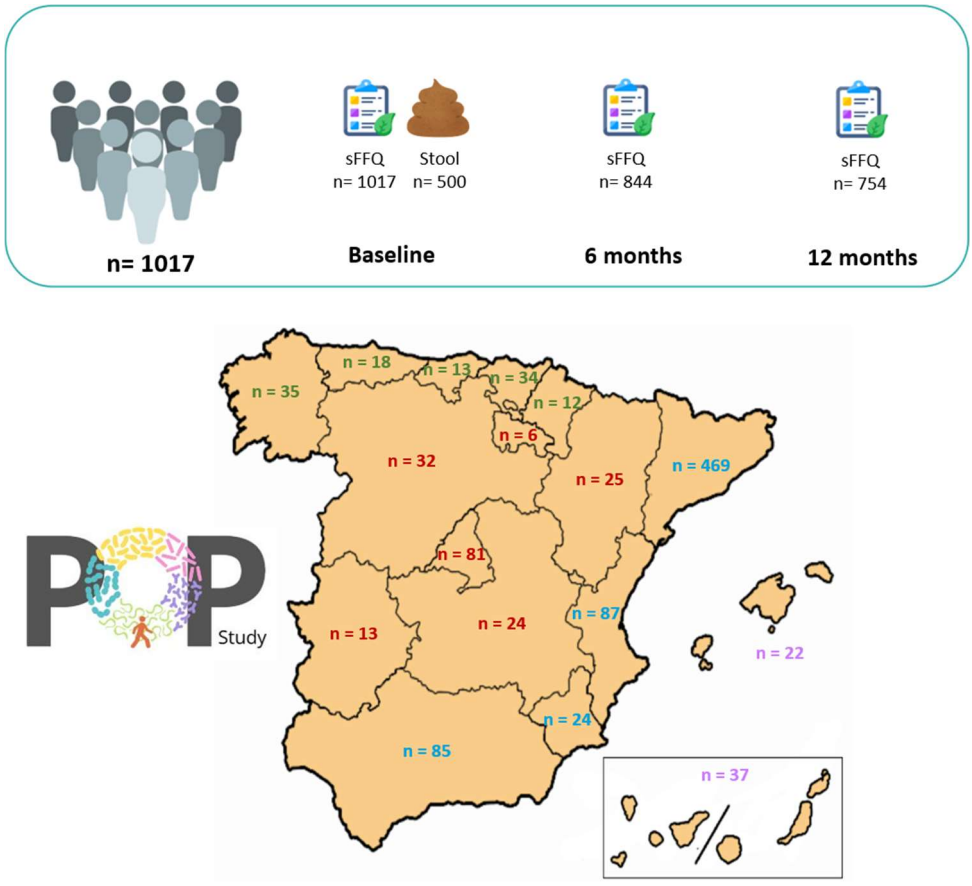


Figure 5. Spanish map with number of individuals coming from the 17 CCAA. Note that different colors represent the four region areas considered for further analysis that are Mediterranean, the Interior, the North and the Islands. Classification was done taking into consideration traditional dietary patterns in the Mediterranean country together with geographical distribution.

5.1.1 *Description of the POP cohort*

The median age of the POP cohort was 45.26 (range 19-75 years old) with a BMI of 24.46 ± 8.24 kg/m^2 (median ± SD). Additionally, 61.95% (n= 630) of the volunteers were classified as "normal weighted". Moreover, 54.18% of participants were females (n= 551), including 24.86% who were menopausal (n= 137) and 1.45% who were pregnant (n= 8).

Most of the participants were born via vaginal delivery (87.41%; n= 889) and had a blood type A (34.41 %; n= 350) or O (34.81%; n= 354). The majority were identified as non-smokers (66.27%; n= 674) and reported a stool frequency of once per day (49.95%; n= 508).

Participants belong to all 17 CCAA in Spain: Andalusia (8.36%; n= 8), Aragon (2.46%; n= 25), Asturias (1.77%; n= 18), Balearic Islands (2.16%; n= 22), Canary Islands (3.64%; n= 237), Cantabria (1.28%; n= 13), Castilla – La Mancha (2.36%; n= 24), Castile and León (3.15%; n= 32), Catalonia (46.21%; n= 470), Valencian Community (8.55%; n= 87), Extremadura (1.28%; n= 13), Galicia (3.44%; n= 35), La Rioja (0.59%; n= 6), Community of Madrid (7.96%; n= 81), Region of Murcia (2.36%; n= 24), Navarra (0.10%; n= 12) and Euskadi (3.34%; n= 34). No participants from Ceuta and Melilla cities were recruited. When considering region areas, the sampling fractions obtained were, 1.34 for Mediterranean region followed by Islands 0.80, North of Spain 0.74 and Interior having the lowest value, 0.61.

In terms of dietary habits, 86.23% followed a conventional diet (n= 877), which was considered as a diet that doesn´t adhere to specific dietary restrictions. 2.95% were vegetarian (n= 30) and 0.88% strict vegan (n= 9). Sweeteners consumption was common (28.81%; n= 293), and 46.31% regularly consumed ready-to-eat meals (n= 471) (Table 19).

Table 19. Description of POP cohort population characteristics. Note that for Intake of supplements or drugs, condition and diet type sections in the table, more than one category could be assigned for a single participant.

	Total		
	Baseline	***6 months***	***12 months***
n	***1017***	***844***	***754***
Age (years) mean ± SD	45.26 ± 11.81	46.65 ± 11.54	47.39 ± 11.47
18-29 years, n (%)	126 (12.39)	76 (9.00)	65 (8.62)
30-39 years, n (%)	189 (18.58)	145 (17.18)	116 (15.38)
40-49 years, n (%)	320 (31.47)	282 (33.41)	246 (32.63)
50-59 years, n (%)	255 (25.07)	221 (26.18)	210 (27.85)
≥ 60 years, n (%)	127 (12.49)	120 (14.22)	117 (15.52)
Gender, n (%)			
Male	465 (45.72)	379 (44.91)	332 (44.03)
Female	551 (54.18)	463 (54.86)	420 (55.70)
Other	2 (0.1)	2 (0.24)	2 (0.27)
Menstruation, n (%)	***551***	***463***	***420***
Currently menstruating	141 (25.59)	110 (23.76)	84 (20.00)
Non-menstruating	273 (49.55)	232 (50.11)	213 (50.71)
Menopause	137 (24.86)	121 (26.13)	123 (29.29)
Pregnant, n (%)	8 (1.45)	10 (2.16)	6 (1.43)
BMI (kg/m2) mean ± SD	24.46 ± 8.24	24.22 ± 4.36	24.55 ± 7.96

	Total		
	Baseline	***6 months***	***12 months***
n	***1017***	***844***	***754***
Weigth status, n (%)			
Underweight (< 18.5 kg/m2)	38 (3.74)	32 (3.79)	26 (3.45)
Normal (18.5-24.9 kg/m2)	630 (61.95)	522 (61.85)	470 (62.33)
Overweight (25-29.9 kg/m2)	257 (25.27)	217 (25.71)	188 (24.93)
Obese (≥ 30 kg/m2)	92 (9.05)	73 (8.65)	70 (9.15)
Birth type, n (%)			
Vaginal birth	889 (87.41)	727 (86.14)	650 (86.21)
C-section	98 (9.64)	81 (9.60)	75 (9.95)
Unknown	30 (2.95)	36 (4.27)	29 (3.85)
Blood type, n (%)			
A	350 (34.41)	282 (33.41)	257 (34.08)
B	87 (8.55)	70 (8.29)	66 (8.75)
AB	42 (4.13)	38 (4.50)	32 (4.24)
O	354 (34.81)	316 (37.44)	275 (36.47)
Unknown	184 (18.09)	138 (16.35)	124 (16.45)
Rhesus			
+	655 (64.41)	567 (67.18)	512 (67.90)
-	133 (13.08)	110 (13.03)	104 (13.79)
NA	229 (22.52)	167 (19.79)	138 (18.30)
Smoking status, n (%)			
Non-smoker	674 (66.27)	572 (67.77)	499 (66.18)
Smoker	88 (8.65)	67 (7.94)	59 (7.82)
Former smoker	255 (25.07)	205 (24.29)	196 (25.99)
Region, n (%)			
Andalusia	85 (8.36)	71 (8.41)	64 (8.49)
Aragon	25 (2.46)	22 (2.61)	21 (2.79)
Asturias	18 (1.77)	15 (1.78)	13 (1.72)
Balearic Islands	22 (2.16)	17 (2.01)	16 (2.12)
Canary Islands	37 (3.64)	32 (3.79)	27 (3.58)
Cantabria	13 (1.28)	12 (1.42)	11 (1.46)
Castilla – La Mancha	24 (2.36)	16 (1.90)	12 (1.59)
Castile and León	32 (3.15)	32 (3.79)	30 (3.98)

	Total		
	Baseline	***6 months***	***12 months***
n	***1017***	***844***	***754***
Catalonia	469 (46.21)	383 (45.38)	342 (45.36)
Valencian Community	87 (8.55)	79 (9.36)	68 (9.02)
Extremadura	13 (1.28)	12 (1.42)	9 (1.19)
Galicia	35 (3.44)	31 (3.69)	28 (3.71)
La Rioja	6 (0.59)	5 (0.59)	3 (0.40)
Community of Madrid	81 (7.96)	57 (6.75)	54 (7.16)
Region of Murcia	24 (2.36)	19 (2.25)	18 (2.39)
Navarre	12 (0.10)	9 (0.12)	10 (0.13)
Basque Country	34 (3.34)	33 (3.79)	28 (3.71)
Ceuta	0 (0.00)	0 (0.00)	0 (0.00)
Melilla	0 (0.00)	0 (0.00)	0 (0.00)
Working on health care system			
Yes	308 (30.29)	247 (29.27)	223 (29.58)
No	694 (68.24)	597 (70.73)	531 (70.42)
Unknown	15 (1.47)	0 (0.00)	0 (0.00)
Stool frequency			
1-2 times/week	44 (4.33)	35 (4.15)	35 (4.64)
more than 3 times/week	141 (13.86)	121 (4.34)	103 (13.66)
1 time/day	508 (49.95)	418 (49.53)	393 (52.12)
2 times/day	163 (16.03)	147 (17.42)	110 (14.59)
More than 2 times/day	59 (5.58)	52 (6.10)	47 (6.20)
Unknown	102 (10.03)	71 (8.41)	66 (8.75)
Liquid intake (l), mean ± SD	1.92 ± 0.71	1.92 ± 0.67	1.92 ± 0.68
	618	***541***	***499***
Mean steps/day mean ± SD	9127.09 ± 5540.31	9222.73 ± 5416.75	8839.99 ± 4321.77
Diet type, n (%)			
Conventional	877 (86.23)	721 (85.43)	638 (84.61)
Strict Vegetarian	30 (2.95)	28 (3.32)	28 (3.71)
Low in animal protein	26 (2.56)	20 (2.37)	17 (2.25)
Gluten free	21 (2.06)	21 (2.49)	22 (2.92)
Lactose free	13 (1.28)	12 (1.42)	12 (1.59)
Low gluten	12 (1.18)	12 (1.42)	6 (0.79)

	Total		
	Baseline	***6 months***	***12 months***
n	***1017***	***844***	***754***
Low carbohydrates	10 (0.98)	14 (1.66)	15 (1.99)
Vegan	9 (0.88)	7 (0.83)	8 (1.06)
Other[1]	45 (4.42)	38 (4.50)	33 (4.38)
Intake of ready-to-eat meals, n (%)			
Yes	471 (46.31)	388 (45.97)	364 (48.28)
No	546 (53.69)	456 (54.03)	390 (51.72)
Intake of sweeteners, n (%)			
Yes	293 (28.81)	231 (27.37)	204 (27.06)
No	724 (71.19)	613 (72.63)	550 (72.94)
Intake of supplements or drugs, n (%)			
None	430 (42.28)	317 (37.56)	248 (32.89)
Treatment			
Dietary supplements	366 (35.99)	347 (41.11)	335 (44.43)
Analgesic	48 (4.72)	29 (3.44)	16 (2.12)
Anti-inflammatory	38 (3.74)	47 (5.57)	38 (5.04)
Antihypertensive	33 (3.24)	29 (3.44)	39 (5.17)
Probiotic	30 (2.95)	32 (3.79)	31 (4.11)
Antianemic	23 (2.26)	23 (2.73)	21 (2.78)
Antidepressant	22 (2.16)	27 (3.20)	29 (3.85)
Decrease cholesterol levels or other lipids	22 (2.16)	24 (2.84)	28 (3.71)
Thyroid therapy	18 (1.77)	18 (2.13)	15 (1.99)
PPI	17 (1.67)	21 (2.49)	12 (1.59)
Benzodiazepine	14 (1.38)	13 (1.54)	17 (2.25)
Antihistaminic	13 (1.28)	16 (1.89)	12 (1.59)
Anticoagulant	12 (1.18)	8 (0.95)	6 (0.79)
Other[2]	94 (9.24)	106 (12.56)	124 (16.45)
Disease, n (%)			
Non-reported disease	813 (79.94)	661 (78.32)	563 (74.67)
Reported disease			
Hypertension	35 (3.44)	29 (3.44)	33 (4.38)
Thyroid disease	34 (3.34)	35 (4.15)	34 (4.51)
Asthma	18 (1.77)	12 (1.42)	7 (0.93)
Allergy	15 (1.47)	8 (0.95)	14 (1.86)

	Total		
	Baseline	***6 months***	***12 months***
n	***1017***	***844***	***754***
Hypercholesterolemia	13 (1.28)	13 (1.54)	11 (1.46)
Heart disease	10 (0.98)	11 (1.30)	12 (1.59)
Anxiety	9 (0.88)	10 (1.18)	7 (0.93)
Migraine	9 (0.88)	8 (0.95)	8 (1.06)
Hernia	8 (0.79)	7 (0.83)	9 (1.19)
IBS	8 (0.79)	8 (0.95)	5 (0.66)
Osteoarthritis	8 (0.79)	11 (1.30)	9 (1.19)
Other[3]	116 (11.41)	118 (13.98)	120 (15.92)

[1]Other diet type include: hypocaloric, restricted, eastern asian, anti-inflammatory diet, Herbalife diet, paleo diet, FODMAP, ayurvedic, low in sugar, low in processed food, dairy free, ketogenic, low in dairy, free of processed food, sugar free, salt free, low in salt, low in lactose, low in refined carbohydrates, free of refined carbohydrates and fructose free.

[2]Other medication include Acne treatment, adrenergic receptor agonist, aldosterone inhibitor, anthelmintic, antianginals, antiarrhythmic, antibiotics, antidiabetic, antiemetic, antiepileptic, antiflatulent drug, antifungal, anti-hyperuricemia, antimalarial, antimuscarinic, antiplatelets, antipsychotic, anti-rheumatics, antiviral, anxiolytic, aromatase inhibitor, bipolar disorder treatment, bronchodilator, cholinergic receptor antagonist, coagulant, corticoids, corticoids, diuretics, estrogen modulators, estrogen substitute therapy, eye pressure reduction treatment, gastric reflux treatment, GI stimulants, glucagon type 1 agonist, histamine antagonist, immunosupressant, kidney Stone treatment, laxative, monoclonal antibodies, mucolytics and antitussive, muscle relaxant, oral contraceptive, osteoporosis and Paget disease treatment, pancreatic enzymes, parkinson treatment, postbiotic, prebiotics, progesterone, SGLT2 inhibitor, syndrome of Menière treatment, testosterone inhibitors, treatment of excessive sleepiness, triptan, Typhus vaccine, uric acid treatment, vasodilator, venotonics, xanthine oxidase inhibitors.

[3]Other dsease include Abdominal distension, alopecia, anemia, apnea, Asperger, asthmatic bronchitis, atopic dermatitis, Autism, benign prostatic hyperplasia, Bipolar disorder, Birt Hogg Dube Syndrome, brain lesions, cancer, candidiasis, celiac disease, Chilaiditi's Syndrome, cholinergic urticaria, chondropathy, chronic fatigue, chronic urticaria, Clostridium difficile, constipation, depression, diabetes, diverticulitis, dysmetria, endometriosis, epilepsy, Epstein Barr, esophageal atresia, esophagitis, factor V Leiden, fibromyalgia, fructose malabsorption, gallstones, gastritis, GI discomfort, Gilbert syndrome, glaucoma, hearing loss, hemorrhoids, herpes, HIV, HPV, hyperuricemia, ictus, idiopathic angioedema, ischemia, kidney cyst, kidney stones, lactose intolerance, Lichen sclerosis, lipedema, Lyme disease, Meniere´s disease, menorrhea, Multiple sclerosis, muscular dystrophy, NASH, obesity, OCD, ocular hypertension, Parkinson, Parry Romberg Syndrome, persistent COVID, phimosis, pituitary adenoma, polycystic kidney disease, polycystic ovary syndrome, pre-diabetes, psoriasis, Raynaud's disease, reflux, refractory endometrium, renal insufficiency, retinal dystrophy, rhinitis, rosacea, scoliosis, SIBO, sinusitis, tensional cefalea, thrombophilia, tinnitus, tonsillar Ectopia, urinary incontinence, uterine polyp, vitiligo

5.2 Characterization of POP diet

5.2.1 *Individual diet is relatively stable over time*

Given the longitudinal nature of the study, we assessed intra- and inter-individual variability across three dietary categories: food groups, food items and macro- and micronutrients. This analysis was conducted using Bray-Curtis similarity index, where lower values indicate greater similarity between samples.

As expected, intra-individual variability (evaluated across the three timepoints: baseline, six and twelve months, as well as in pairwise comparisons) showed lower Bray-Curtis values when compared to inter-individual variability and seasonal effect. This pattern was consistent across the three categories, suggesting that participants maintained a relatively stable diet over the one-year study period (Figure 6).

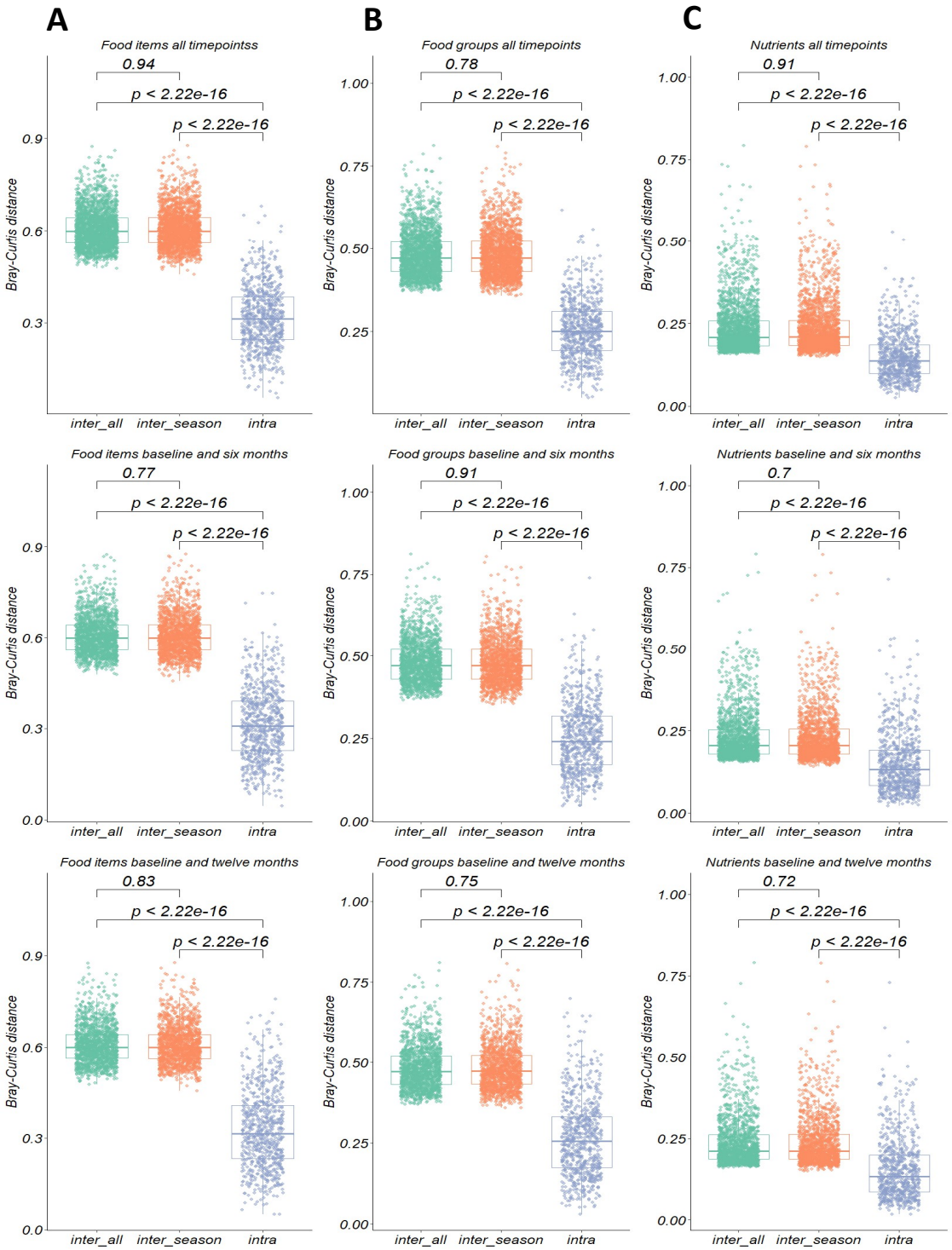


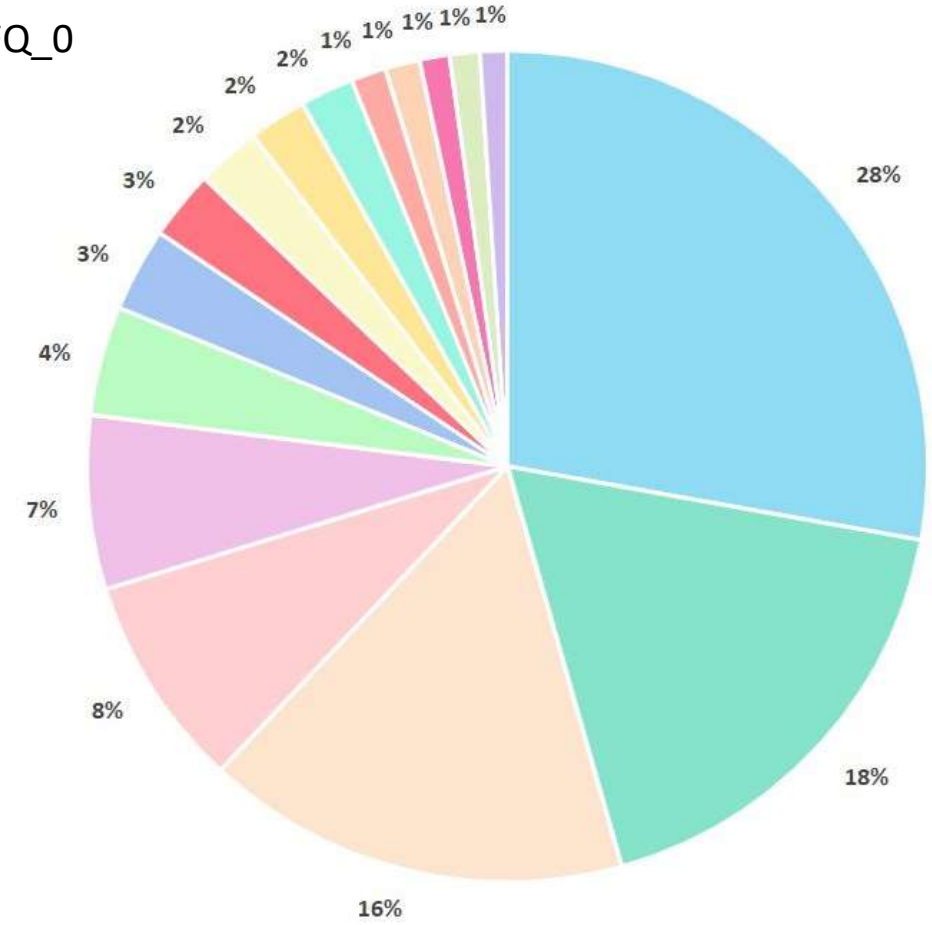
Figure 6. Intra- and Inter-individual variability of dietary intake across timepoints, baseline vs six months and baseline vs one year. We computed the Bray-Curtis dissimilarity index between samples, taking into account A) food items, B) food groups, and C) nutrients. Inter_all shows, for each sample, the median of distances between the samples and all the other samples. Inter_season shows, for each sample, the median of distances between the sample and other samples taken on the same season. Intra shows for each participant the distance between timepoints, baseline vs six months and baseline vs twelve months. Differences were tested for significance using Mann-Whitney test.

5.2.2 *Spanish food preferences*

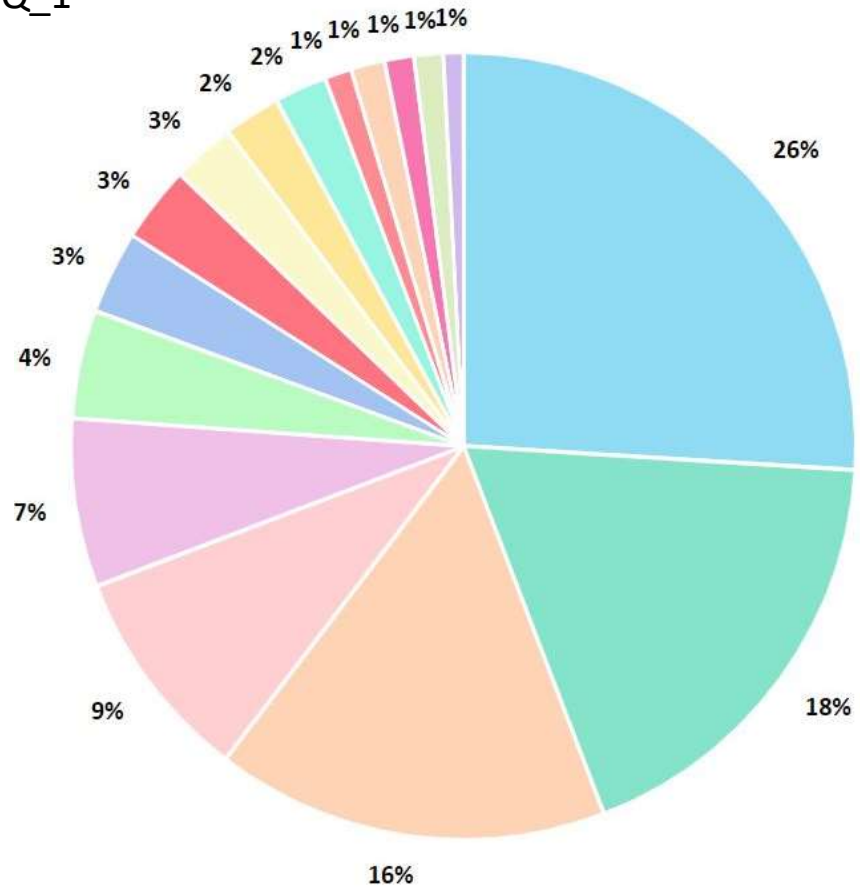
The median daily energy intake in our cohort was 1,787.03 kcal at baseline, 1,722.34 kcal at six months, and 1,703.74 kcal at twelve months. A detailed breakdown of median daily intake values (g/day), along with the 25th and 75th percentiles for food groups, energy, and nutrients across all three time points, is available in ANNEX 4.

Additionally, the most consumed food groups in the POP cohort were: vegetable and vegetable products (28%, 26% and 25% in sFFQ_0, sFFQ_1 and sFFQ_2 respectively), followed by fruit and fruit products (18%, 18%, 19%), non-alcoholic drinks (16%,16%, 17%), milk and dairy products with the exception of fermented products (8%, 9%, 9%), meat including eggs (7%), fish and shellfish (4%), and legumes (3%). Additional information on remaining food groups can be found in Figure 7.

sFFQ_0



sFFQ_1



sFFQ_2

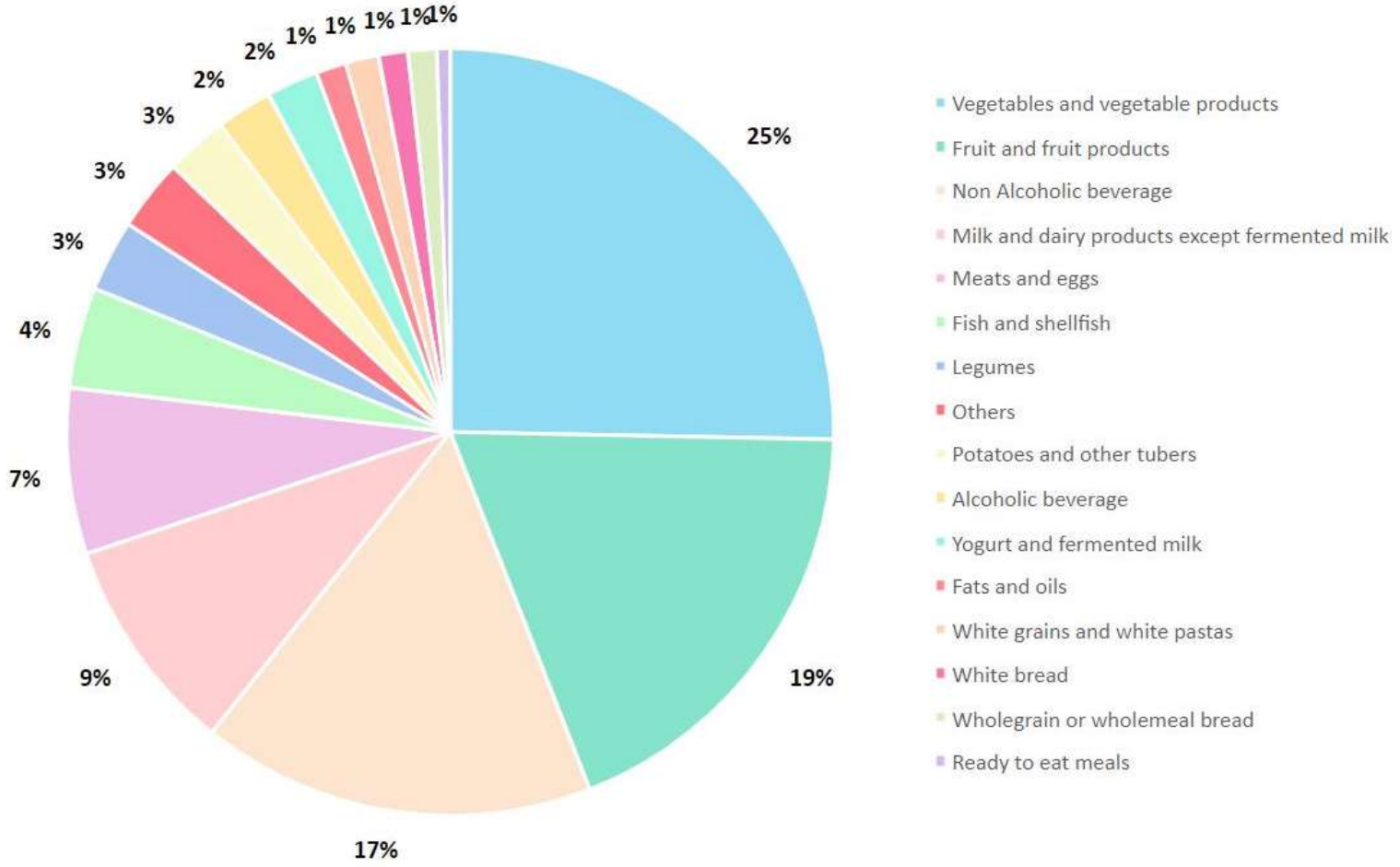


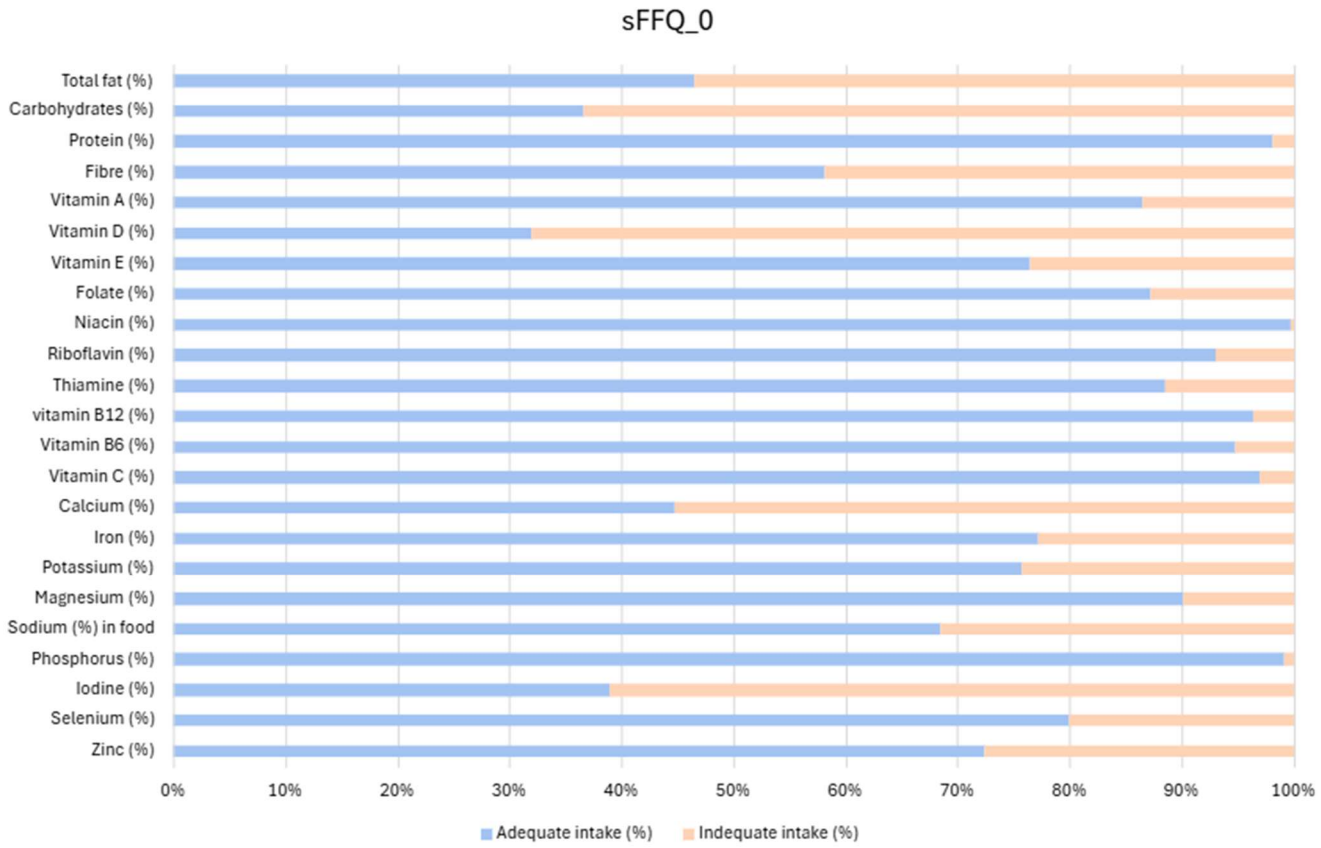
Figure 7. Percentage of food group most consumed by our population as assessed by sFFQ_0 (n= 1017), sFFQ_1 (n= 844) and sFFQ_2 (n= 754). Food groups whose % of consumption was lower than 1% were grouped as "Other" category.

5.2.3 *Do we meet the standard recommendations?*

5.2.3.1 Adequate and inadequate intake values in Spanish population

Based on the nutritional intake recommendations for the Spanish population proposed by AESAN (266), the proportion of volunteers meeting at least 80% of the recommended intake varied by nutrient: 98.01-98.70% met the recommendations for protein (0.83 g/kg per day), 53.05-58.21% for fiber (> 25 g/day), 42.57-46.51% for total fat (20-35% of total energy), and 36.58% for CHO (45-60% of total energy). Protein and total fat intake exceeded the recommended values, fiber was accomplished just for half of the volunteers while CHO fell below the recommended thresholds.

For micronutrients (Figure 8), over 70% of participants met standard recommendations, except for calcium (39.92-44.84%) and vitamin D (31.83-31.96%). Sodium and iodine intake were not reliably assessed, as added salt during cooking was not quantified, making it difficult to determinate adequate or inadequate intake for its calculation.



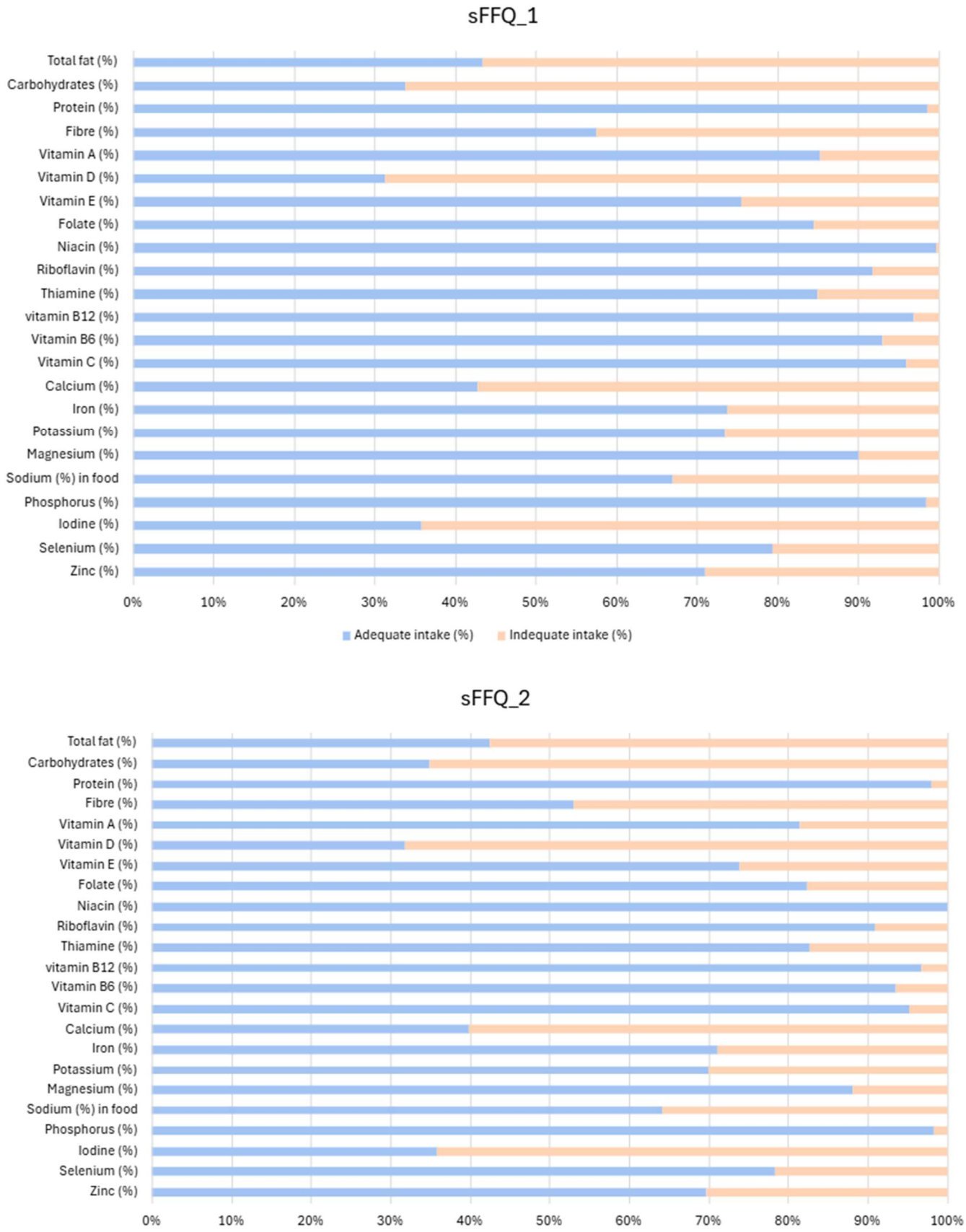


Figure 8. Percentage of volunteers in POP cohort who exhibit adequate and inadequate of daily macro- and micronutrient intake based on the 80% cut-off of the Spanish population recommendations. Values for sFFQ_0 (n= 1017), sFFQ_1 (n= 844) and sFFQ_2 (n= 754) are presented.

5.2.3.2 Comparison with GBD-2017 recommendations

In 2019 the GBD study highlighted sodium, whole grains and low fruit consumption as key risk factors for non-communicable diseases and mortality (190). To assess how well our cohort aligned with these recommendations, we mapped our initial 58 sFFQ items into 12 out of 15 GBD dietary risk factors (see Methods section, Table 17). Sodium was excluded from this analysis due to the lack of a specific question regarding salt added during cooking process.

Our cohort's intake of fruits (median intake of 225.6 g/day), vegetables (321.98 g/day) and fiber (27.32 g/day) fell within the recommended GBD ranges (see Table 20). However, legumes (41.4 g/day), wholegrains (22.65 g/day) nuts (9.6 g/day), milk (64 g/day), calcium (874.73 mg/day) and PUFA (6.4 g/day) presented a suboptimal intake compared to the GBD optimal values and ranges. Conversely, there was an excess intake of red meat (27.3 g/day), processed meat (4.72 g/day) and sugar-sweetened beverages (6.6 g/day).

Table 20. sFFQ items grouped by GBD dietary risk factors (n= 2615). Optimal level of intake as well as optimal range suggested by GBD was also provided.

Statistics	Fruits g/day	Vegetables g/day	Legumes g/day	Wholegrains g/day	Nuts g/day	Milk g/day
P25[1]	137.39	190.305	19.8	2.64	3.15	0
Median	225.6	321.98	41.4	22.65	9.6	64
P75[2]	607.35	522.3	57.15	70	30	200
GBD[3] *optimal value*	250	360	60	125	21	435
GBD[3] *optimal range*	200-300	290-430	50-70	100-150	16-25	350-520

Statistics	Red meat g/day	Processed meat g/day	Sugar-sweetened beverages g/day	Fiber g/day	Calcium mg/day	PUFA g/day
P25[1]	8.58	1.49	0	18.71	636.845	5.18
Median	27.3	4.72	6.6	27.32	874.73	6.4
P75[2]	27.3	9.45	26.4	38.44	1192.35	7.73
GBD[3] *optimal value*	23	2	3	24	1250	11
GBD[3] *optimal range*	18-27	0-4	0-5	19-28	1000-1500	9-13

[1]Percentile 25

[2]Percentile 75

[3]Global Burden of Disease

5.3 Personal traits, geography and lifestyle influence our food choices

To investigate the impact of lifestyle, biometric and demographic factors on dietary choices within the Spanish population, we used MaAsLin2 models. Our analysis revealed that food groups were influenced by all variables, except for workplace. Food items were affected by all covariates, while seven factors (Bowel Frequency, Region Areas, Diet Grouped, Season Year, Smoke, Gender and Sweeteners) showed association with micro- and macronutrient intake among volunteers (Figure 9).

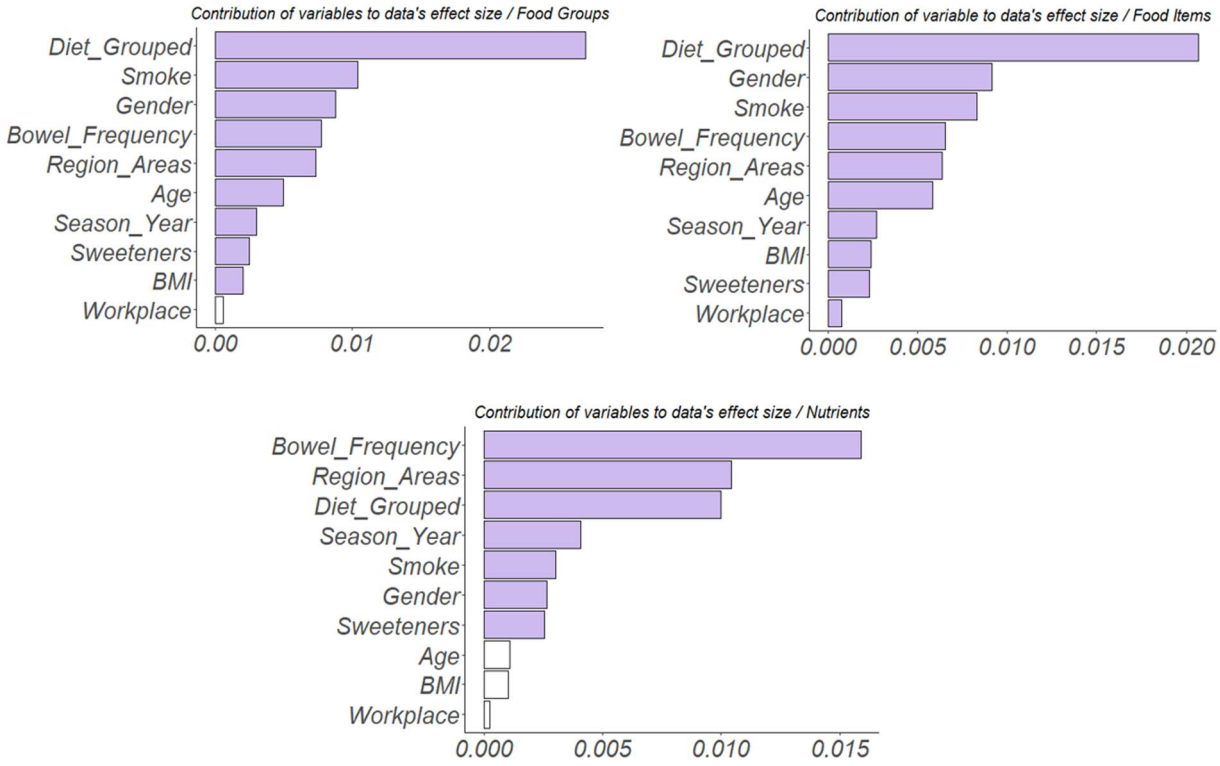


Figure 9. Effect size of the population characteristics on dietary intake. The magnitude of the influence of specific characteristics on dietary intake (n= 2615) was calculated using permutational analysis of variance (PERMANOVA), as implemented in the adonis2 function of the vegan R using the Bray-Curtis method. Y axis represents the variables while X axis refers to the contribution to effect size (R2). Significant results ($p < 0.05$) where purple colored.

The use of **DQIs** represent an alternative and simplified measure of diet quality in population studies. Thus, to elucidate the relationship between different DQIs and self-reported population characteristics, linear models implemented in MaAsLin2 were used. After adjusting for significant variables obtained by PERMANOVA several significant results emerged.

Age was positively correlated with DQIs indicative of healthier choices (q(IASE)= 0.00151; q(hPDI)= 1.22e-05; q(aMED)= 0.00679) and negatively correlated with meat and plant protein from unhealthier sources (q(MDI)= 0.0480; q(uPDI)= 0.0480). These results suggest that, with age individuals may become more careful about their habitual diet (Figure 10). Specifically, older individuals tended to consume more fruit and fruit products (q= 0.00411) and nuts (q= 0.00760) but also alcoholic beverages (q= 0.00266) and were characterized by a decrease of intake of poorer food groups such as ready to eat meals (q=

1.24e-07), white grains (q= 2.11e-06), pastries and sweet breads (q= 2.34e-06) or appetizers (q= 7.42e-06). Additional associations can be found in Table 21.

In contrast, **BMI** correlated negatively with aMED (q= 0.0023), PDI (q= 0.00505) and hPDI (q= 0.0479)

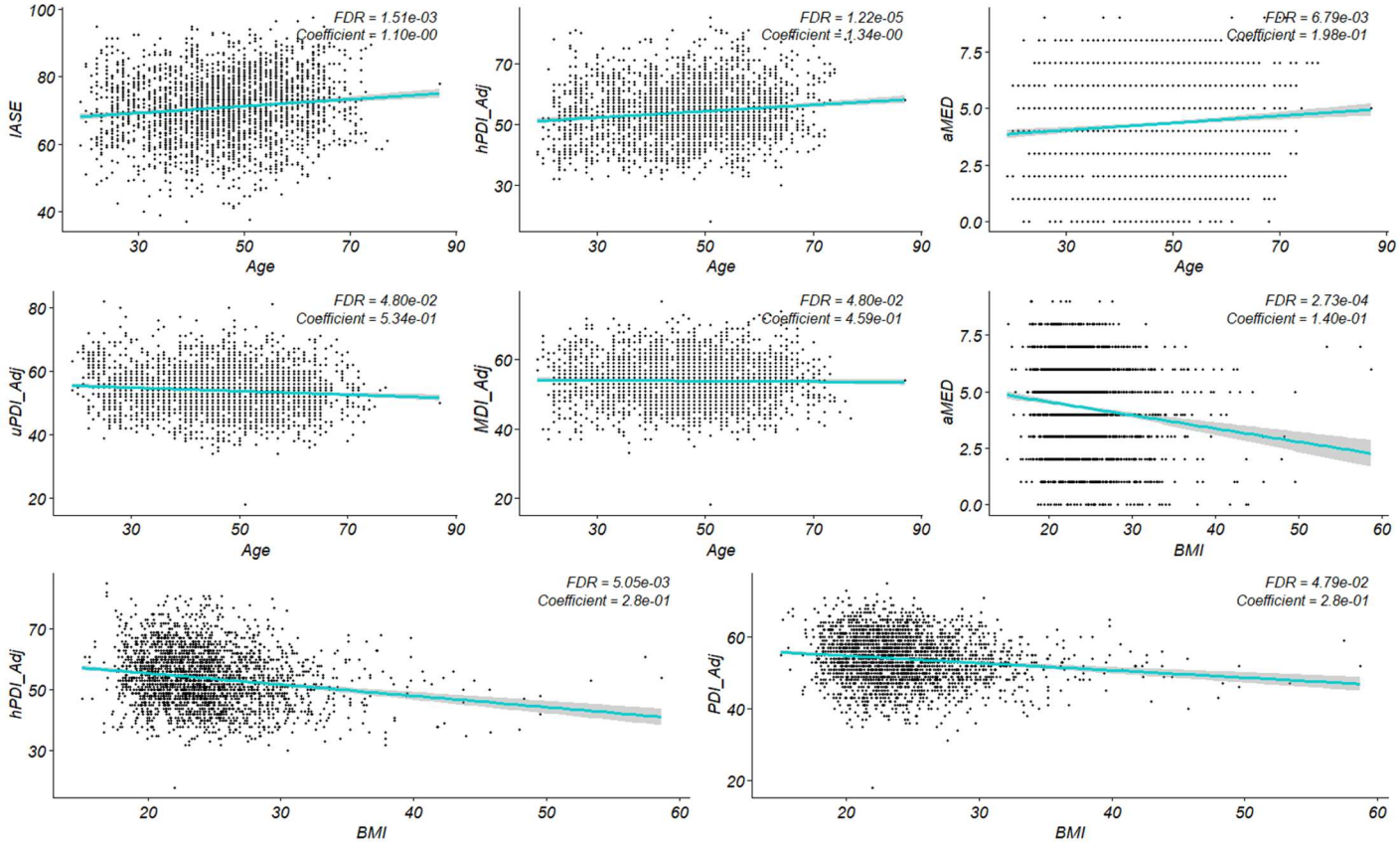


Figure 10. Relationship between DQIs and continuous population characteristics (age and BMI). Correlations were calculated using the MaAsLin2 tool (n= 2615).

scores, suggesting that a poorer diet is associated with an increase in BMI (Figure 10). One possible explanation could be that individuals with higher BMI presented an increased intake of less healthy foods, such as white bread (q= 0.00114) and ready to eat meals (q= 0.000411) (Table 21).

Current smokers reported a less healthy diet, as indicated by lower scores on DQIs compared with non-smokers and former smokers (q(HEI_2015) = 0.00884; q(IASE)= 0.00884) (Figure 11). Smokers also had a higher preference for alcohol (q= 0.000107) and lower intake of fruits and fruit derived products (q= 2.99e-05). In contrast, former smokers also reported higher alcohol consumption (q= 0.00427) but a decrease in biscuits and breakfast cereals (q= 0.023) in comparison with non-smokers (see ANNEX 5).

Table 21. Association between food groups and age/BMI. Output from MaAsLin2 (n= 2615).

Food group	*Variable*	*Coef*	*Stderr*	*N*	*p-value*	*q-value*
Ready to eat meals	Age	-0.40595107	0.06845741	2615	4.1793E-09	1.2422E-07
White grains	Age	-0.32352562	0.05991403	2615	8.3821E-08	2.1176E-06
Pastries and sweet breads	Age	-0.36038135	0.06708928	2615	9.7694E-08	2.3447E-06
Appetizers	Age	-0.34566815	0.06732395	2615	3.3996E-07	7.4174E-06
Meat and eggs	Age	-0.171732	0.03702107	2615	3.9973E-06	6.3957E-05
Potatoes and other tubercles	Age	-0.21784506	0.04833175	2615	7.3614E-06	0.00010096
Sauces and condiments	Age	-0.28975655	0.06552438	2615	1.094E-05	0.00014193
Alcoholic beverage	Age	0.35622271	0.09645022	2615	0.00023306	0.00266356
Fruits and fruit products	Age	0.17750776	0.05007341	2615	0.0004115	0.00411504
Nuts and seeds	Age	0.20383102	0.06122342	2615	0.00090347	0.00760814
Whole grains	Age	0.23300654	0.08116884	2615	0.00418543	0.02911602
Biscuits and breakfast cereals	Age	-0.20985783	0.07708515	2615	0.00660265	0.04225693
White bread	Age	-0.25045909	0.09379733	2615	0.00770598	0.04742144
Ready to eat meals	BMI	0.1396035	0.03944205	2615	0.00040977	0.00411504
White bread	BMI	0.16507582	0.05171791	2615	0.00143531	0.01167708

Females exhibited healthier dietary habits compared to men, reflected by higher values of HEI-2015 (q= 7.15e-03), hPDI (q= 2.28e-07), PDI (q= 2.41e-07) and aMED (q= 0.000173) and lower values of MDI (q= 1.31e-07) and uPDI (q= 2.41e-07). Thus, suggesting a better dietary habit of this group which might follow a more MedDiet type (more consumption of plant-based sources and lower meat intake). Indeed, at the food group level, this pattern is partially supported, with an increased consumption of vegetables (q= 4.95e-07), fruit and fruit products (q= 0.00304) and whole bread (q= 0.0237), fish and shellfish (q= 0.00271) but also higher intake of fats and oils (q= 4.13e-05) and non-alcoholic drinks (q= 0.0244). On the other hand, males preferred alcoholic beverages (q= 2.20e-05), white grains and bread (q= 0.0104; q= 0.0424) and ready to eat meals (q= 0.0162) (Figure 11, ANNEX 5).

Geographically, Spain was divided into four different **region areas** (Mediterranean, North of Spain, Interior, and Islands) based on geographical distribution, which were consider that could have an impact on dietary patters. The Mediterranean region included Catalonia, Community of Valencia, Region of Murcia and Andalusia; The North of Spain was formed by Cantabria, Asturias, Navarre, Basque Country and Galicia; the Interior region encompassed La Rioja, Aragon, Castille and Leon, Extremadura, Community of Madrid and Castilla-La Mancha. Finally, the Islands included only Balearic and Canary Islands (Figure 5). Compared to the Mediterranean region, the Interior of Spain exhibited higher aMED (q= 0.0111) scores and lower uPDI (q= 0.0479) values suggesting a healthier diet. This region exhibited higher intake of legumes (q= 0.0228) and milk and dairy (q= 0.0479), but lower intake of white grains (q= 0.0424) (Figure 11, ANNEX 5).

Interestingly, regular use of sweeteners was positively correlated with the intake of ready-to-eat meals (q= 1.80e-05), sauces and condiments (q= 0.00703), and sausages and other processed meat products (q= 0.0251) (ANNEX 5).

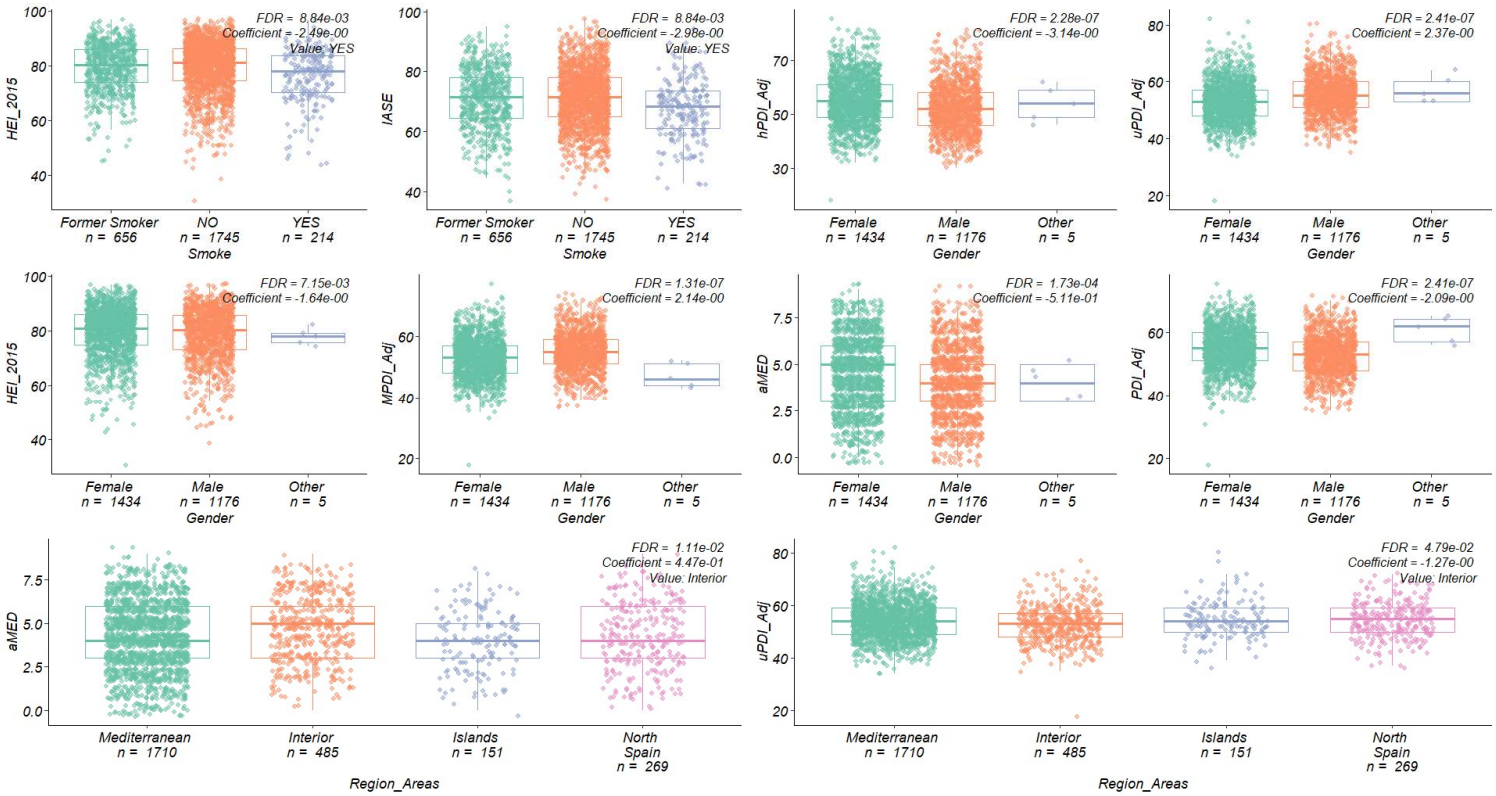


Figure 11. Relationship between Eating Quality Indices (DQIs) and population categorical values (smoke, gender, bowel frequency and region areas). Analysis performed using the MaAsLin2 tool (n= 2615).

5.4 Interplay between diet, lifestyle and microbiome

5.4.1 *Increased fungal recovery using an enrichment protocol*

To assess the impact of fungal protocol enrichment on recovery rate, we analyzed microbiome composition and diversity in a subset of 100 gender-paired individuals, comparing samples with and without enrichment (Figure 12). Two enriched samples failed quality control during library construction, leaving 98 samples for analysis. The median read count per sample was 25 in enriched samples, compared to just two in non-enriched samples. Enrichment significantly increased species detection, identifying 141 species versus 45 in non-enriched samples, also increasing diversity metrics such as Chao1 and Shannon indices, highlighting the effectiveness of the enrichment protocol.

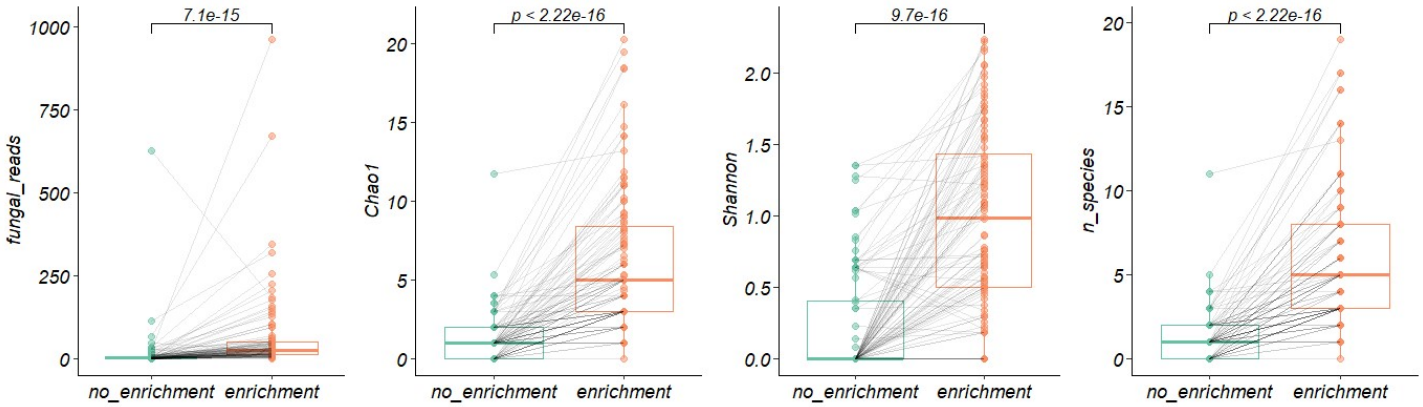


Figure 12. Fungal profiling comparison between enrichment and non-enrichment protocols (n= 100 baseline). The first panel represents the comparison of mapped fungal reads per sample between enrichment and non-enrichment protocols whereas the remaining panels indicate α-diversity metrics (Chao1, Shannon and number of species) between enrichment and non-enrichment protocols. Significance was tested using paired Wilcoxon test.

5.4.2 *Effect on bacterial and fungal α- diversity*

5.4.2.1 Bacterial α- diversity

Before assessing the effect of diet and lifestyle on bacterial microbiome (n= 500 baseline), we used adonis2 method to identify possible confounder variables. Our analysis revealed that gender, age, and BMI were significant factors ($p < 0.05$) influencing bacterial composition, and therefore, were considered in the subsequent analysis (Figure 13).

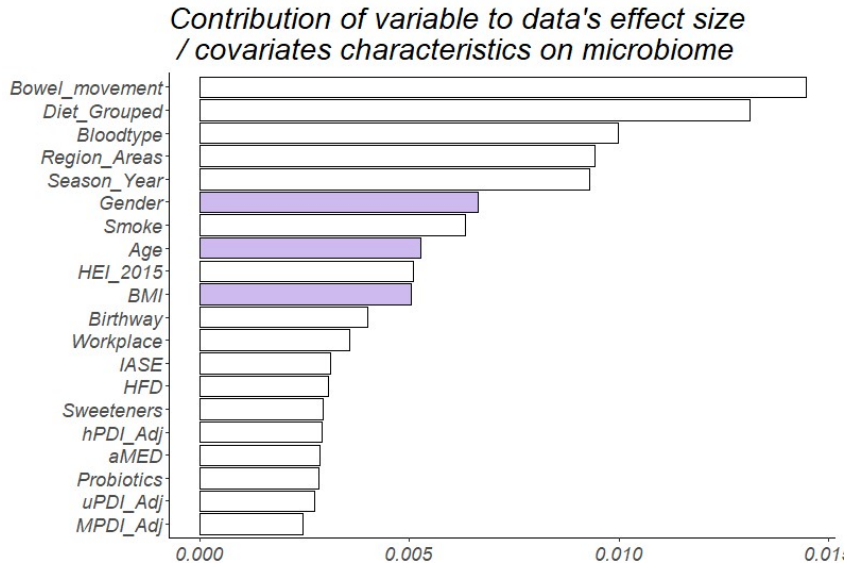


Figure 13. Effect size of the dietary data on the microbiome data. The magnitude of the influence of DQIs and personal traits on the microbiome was calculated using permutational analysis of variance (PERMANOVA), as implemented in the adonis2 function of the vegan R and using the Bray-Curtis method (n= 500 baseline). Y axis represents the variables while X axis refers to the contribution to effect size (R^2). Significant results ($p < 0.05$) where purple colored.

To explore the relationship between diet, POP characteristics and α-diversity, Spearman correlation test was applied. In general, **DQIs** representing healthier dietary patterns, such as HEI-2015 (Shannon,

q= 0.0183), hPDI adjusted (Shannon, q= 0.0183), and aMED (Shannon, q= 0.0210) exhibited positive correlations with richness and/or evenness. In contrast, uPDI adjusted showed a negative correlation with Shannon (q= 0.0184) and Chao1 (q= 0.0372). These results were further supported by examining correlations at the food groups, food items, and nutrients levels. For instance, nuts and seeds (Shannon, q= 0.0351), fruits and fruit products (Shannon, q= 0.0037) and vegetables (Shannon, q= 0.0351), exhibited the same direction as the healthy DQIs. Conversely, white grains (Shannon, q= 0.0079; Chao1, q= 0.0114) and white bread (Shannon, q= 0.0105; Chao1, q= 0.0214) showed correlations aligning with the uPDI direction (Figure 14, ANNEX 6). When looking at items, a similar pattern was obtained. While item 1 (raw leafy vegetables; Shannon, q= 0.002; Chao1, q= 0.016), item 2 (boiled leafy vegetables; Shannon, q= 0.016; Chao1, q= 0.016), item 3 (tomato; Shannon, q= 0.027), item 12 (potatoes; Shannon, q= 0.020), item 14 (fresh fruit; Shannon, q= 0.015), item 16 (dehydrated fruit; Shannon, q= 0.031), item 17 (nuts and seeds; Shannon, q= 0.040), item 36 (bluefish; Chao1, q= 0.016), item 37 (whitefish; Chao1, q= 0.032), item 39 (mollusk and crustacean; Chao1, q= 0.021) and item 45 (dark chocolate; Shannon, q= 0.03) showed positive correlations with diversity, item 18 (white bread; Shannon, q= 0.016; Chao1, q= 0.024), item 22 (refined cereals; Shannon, q= 0.015; Chao1, q= 0.016), item 41 (sunflower oil; Shannon, q= 0.020; Chao1, q= 0.040), item 44 (pastries; Chao1, q= 0.04), item 46 (confectionary; Shannon, q= 0.041), item 51 (soft drinks; Shannon, q= 0.021; Chao1, q= 0.020) and item 58 (processed food; Shannon, q= 0.021; Chao1, q= 0.015) presented negative tendencies (Figure 15, ANNEX 6).

Furthermore, cohort characteristics also exhibited correlations with Shannon and/or Chao1 indices. Notably, diversity was increased with **age** (Shannon, q= 0.0129; Chao1 q= 0.0028), which supports the previous result (Figure 10) that dietary habits tend to improve with age. In contrast, **BMI** negatively correlated with both α-diversity indices (Shannon, q= 0.0142; Chao1, q= 0.0197).

Interestingly, bacterial loads positively associated with both richness and diversity (Shannon, p= 9.5e-07; Chao1, p= 1.8e-07) (Figure 14).

In addition to continuous variables, the effect of three categorical factors (bowel frequency, smoke and seasonality) was assessed using the Mann-Whitney test.

In general, higher **bowel frequency** (classified as > 2 times per day, 2 times per day, 1 time per day, > 3 times per week, and 1.5 times per week) was associated with lower α-diversity values ($p < 0.05$). Specifically, the group with > 2 times per day defecation frequency showed the lowest diversity with diversity increasing gradually until the < 3 times per week category, where it seemed to stabilize, as indicated by the non-significant p-value for both Shannon and Chao1 indices (Figure 16). Additionally,

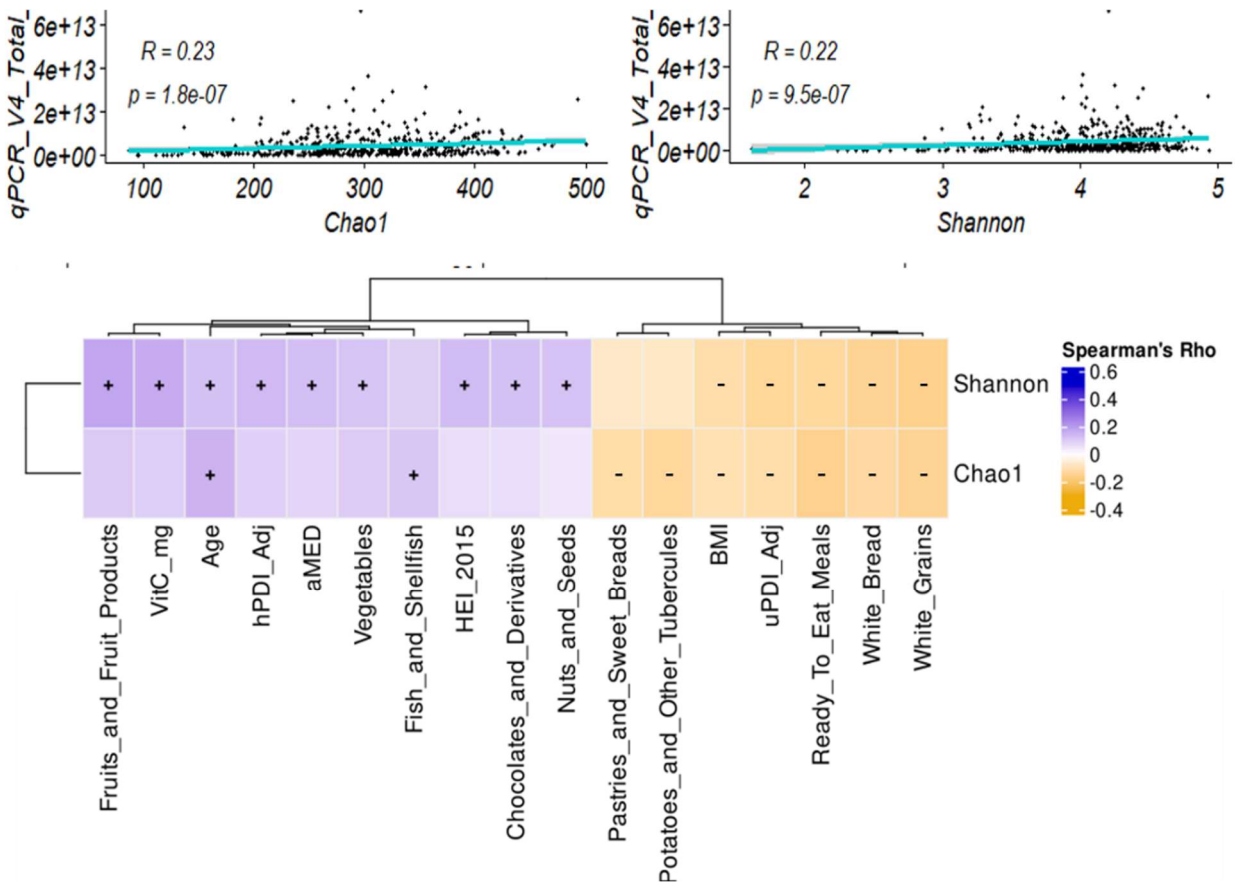


Figure 14. Upper part represents correlation between bacterial loads with bacterial α-diversity (Chao1 and Shannon) using the Spearman correlation test (n= 500 baseline). Lower part corresponds to an integred heatmap showing the significant factors afecting bacterial α- diversity, including food groups, DQIs, nutrients and personal traits (n= 500). Symbols inside the color squares denote significant associations (FDR < 0.05).

a small but significant effect of **season** on bacterial diversity was also observed, with diversity being higher in summer compared to winter ($p= 0.048$), which could be attributed to the increased consumption of fruits and vegetables during the warmer months. Finally, **current smokers** exhibited significantly lower diversity values in both Shannon ($p= 0.020$) and Chao1 ($p= 0.037$) indices compared to non-smokers (Figure 16). No significant effects were observed for other variables including gender, sweetener use, or regional differences.

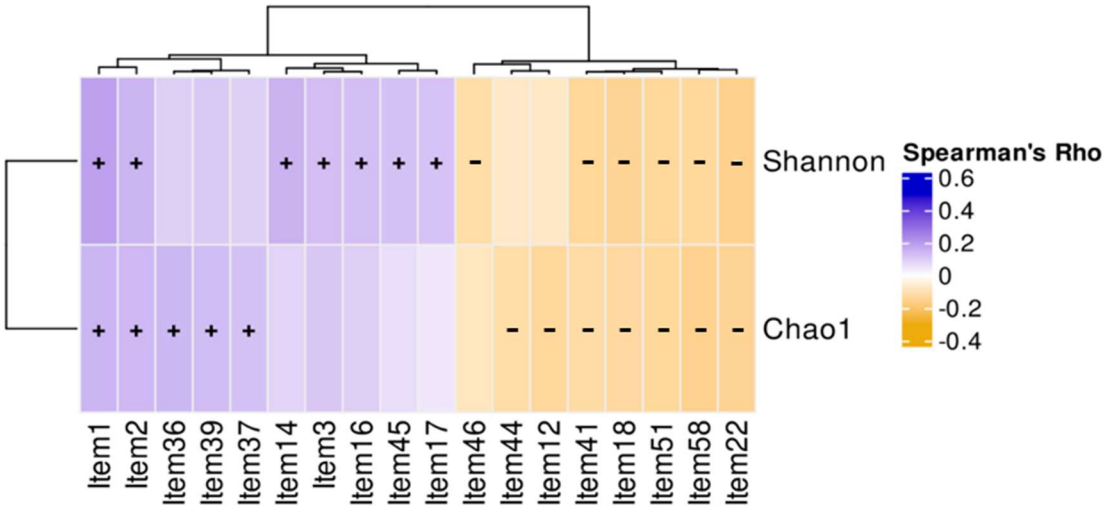


Figure 15. Correlation between sFFQ food items data with bacterial α- diversity (Chao1 and Shannon) using the Spearman correlation test (n= 500 baseline). Symbols inside the color squares denote significant associations (FDR < 0.05). **Item1**: Raw leafy vegetables; **Item2:** Boiled leafy vegetables; **Item3**: Tomato; **Item 12**: Potatoes; **Item 14**: Fresh fruit; **Item16**: Dehydrated fruit; Item**17**: Nuts and seeds; **Item18:** White bread; **Item22**: Refined cereals; **Item36:** Bluefish; **Item37**: Whitefish; **Item39**: Mollusk and crustacean; **Item41**: Sunflower oil; **Item 44:** Pastries; **Item 45**: Dark chocolate; **Item 46**: Confectionary; **Item51:** Soft drinks; **Item58:** Processed food.

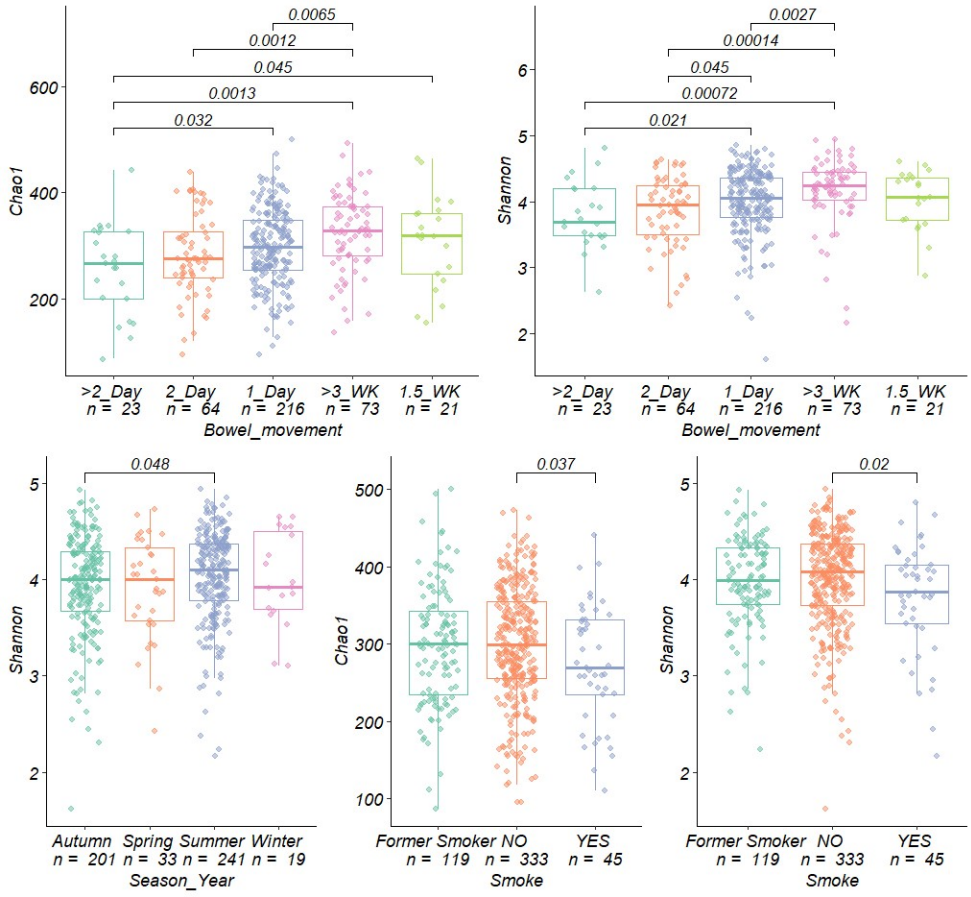


Figure 16. Population characteristics-microbiome α- diversity association analysis. Figures represent differences in categorical population characteristics in relation to bacterial α- diversity (Chao and Shannon indices), analyzed using the Mann-Whitney test (n= 500 baseline).

5.4.2.2 *Fungal α-diversity*

As with the bacterial microbiome, we applied the adonis2 method as the preliminary analysis to assess the potential impact of confounding variables on mycobiome composition (n= 100 baseline). No significant results were obtained for any of the variables considered as shown in Figure 17.

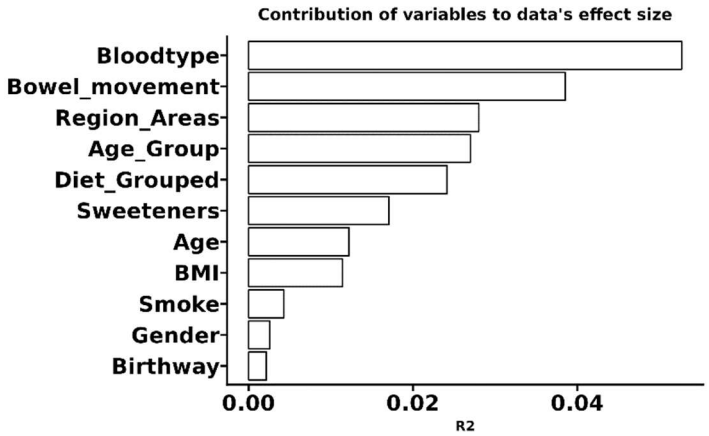


Figure 17. Effect size of the dietary data on the mycobiome. The magnitude of the influence of diet group and personal characteristics on the fungal microbiome was calculated using PERMANOVA, as implemented in the adonis2 function of the vegan R and using the Bray-Curtis method (n= 100 baseline). Y axis represents the variables while X axis refers to the contribution to effect size (R^2).

We next examined for potential associations between POP characteristics, dietary habits and fungal microbiome using PERMANOVA implemented in adonis2 function in R package. After adjusting for confounders, only a few significant associations were observed. Similar to bacteria, fungal diversity increased with **age** (p= 0.044) and was higher during summer compared to winter (Shannon, p= 0.017) and autumn (p= 0.011). Additionally, diversity decreased with uPDI (p= 0.011), an DQI that focus on unhealthy plant-based dietary sources. Food items had a minor impact on diversity, with Item 1, which corresponds to raw leafy vegetables, showing a positive correlation with both Shannon and Chao1 diversity (p= 0.017; p= 0.014 respectively). In contrast, two items (Item 44, pastry and 47, canned tomato sauce) were negatively associated with Shannon diversity (p= 0.014 and p= 0.01) (Figure 18). No significant correlations were observed at food group or nutrient level.

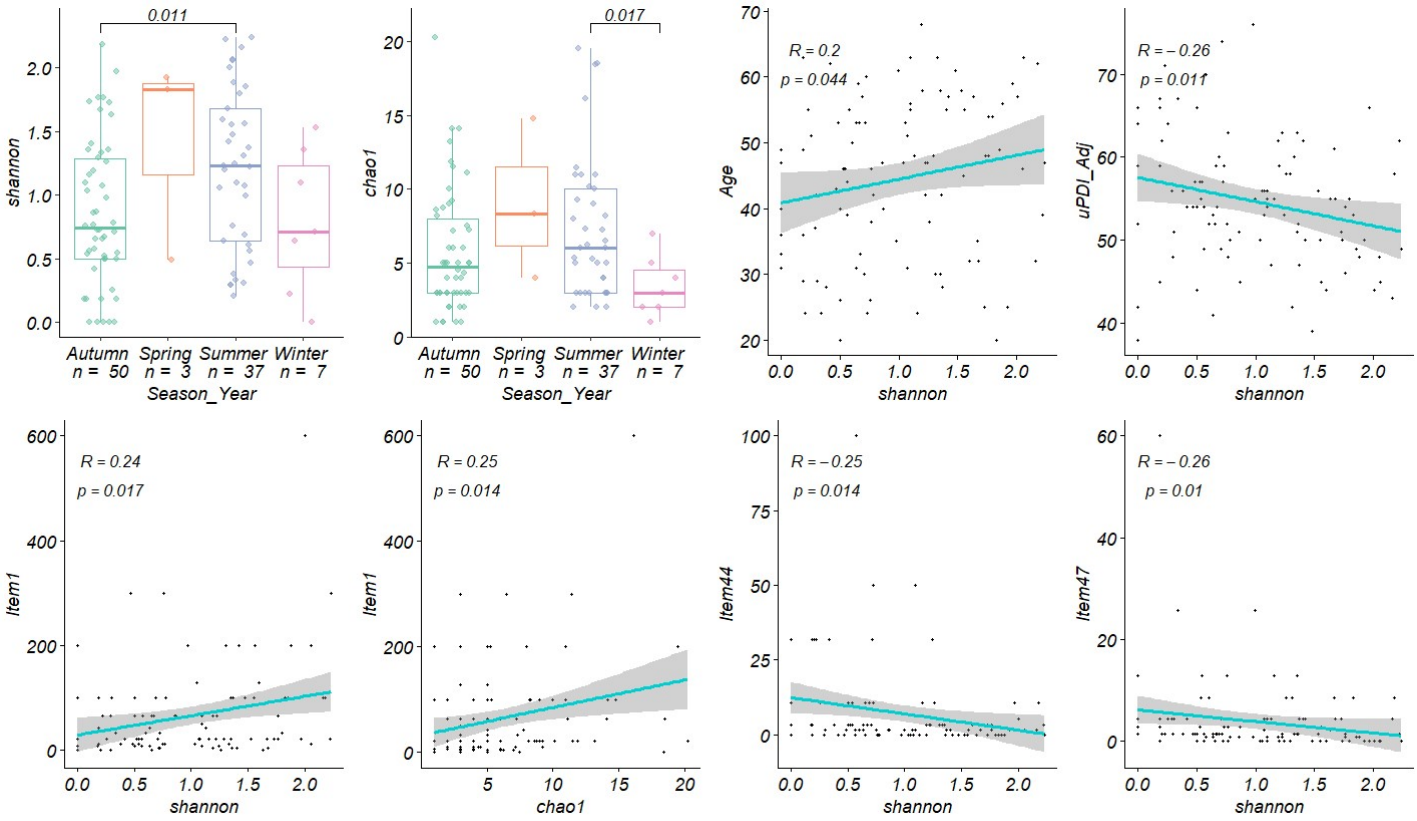


Figure 18. Population characteristics-mycobiome α-diversity association analysis. Figures represent fungal α-diversity in enriched samples (n= 100 baseline). Top left figures show Shannon and Chao diversity among different seasons (Mann-Whitney U test). Bottom and top right figures represent Spearman correlation between age and adjusted uPDI with Shannon. **Item1**: Raw leafy vegetables; **Item44:** Bakery and pastry; **Item47:** Canned and commercial tomato sauce.

5.4.3 *Effect on bacterial and fungal load*

Encouraged by the interesting results regarding the relationship between bacterial load and the increase in bacterial α-diversity, we further examined bacterial and fungal loads, assessed via qPCR in a subset of individuals (n= 500 baseline), using Spearman's correlation and the Mann-Whitney U test. However, no significant values were found between bacterial or fungal loads and population characteristics, food groups, food items, macro or micro-nutrients intake after FDR correction.

5.4.4 *Lifestyle and diet also affect our microbiome composition and function*

5.4.4.1 Bacterial composition

The most prevalent bacterial species in POP cohort (n= 500) included an unknown *Lachnospiraceace* (497 samples), *Blautia wexlerae* (496 samples), *unknown Clostidiaceae* (495 samples), *unknown Clostridia* (495 samples), *Anaerostipes hadrus* (495 samples), *Faecalibacterium prausnitzii* (493 samples), *Dorea formicigenerans* (491 samples), *Blautia faecis* (490 samples), *Anaerobutyricum hallii* (489 samples) and *Fusicatenibacter saccharivorans* (488 samples).

When examining relative abundance (Figure 19), *F. prausnitzii* accounted for 5.57% ± 4.09% of total relative abundance of the sample, followed by *Bacteroides uniformis* (4.28% ± 4.93%), *Prevotella copri clade A* (4.17% ± 9.48%), *Phocaeicola vulgatus* (3.72% ± 4.77%), *Clostridia bacterium* (2.55% ± 3.01%),

Phocaeicola dorei (2.17% ± 3.54%), *Eubacterium rectale* (1.72% ± 2.32%), *Pervotella marseillensis* (1.70% ± 5.49%), *Roserburia faecis* (1.55% ± 2.35%) and *Alistipes putredinis* (1.48% ± 1.57%).

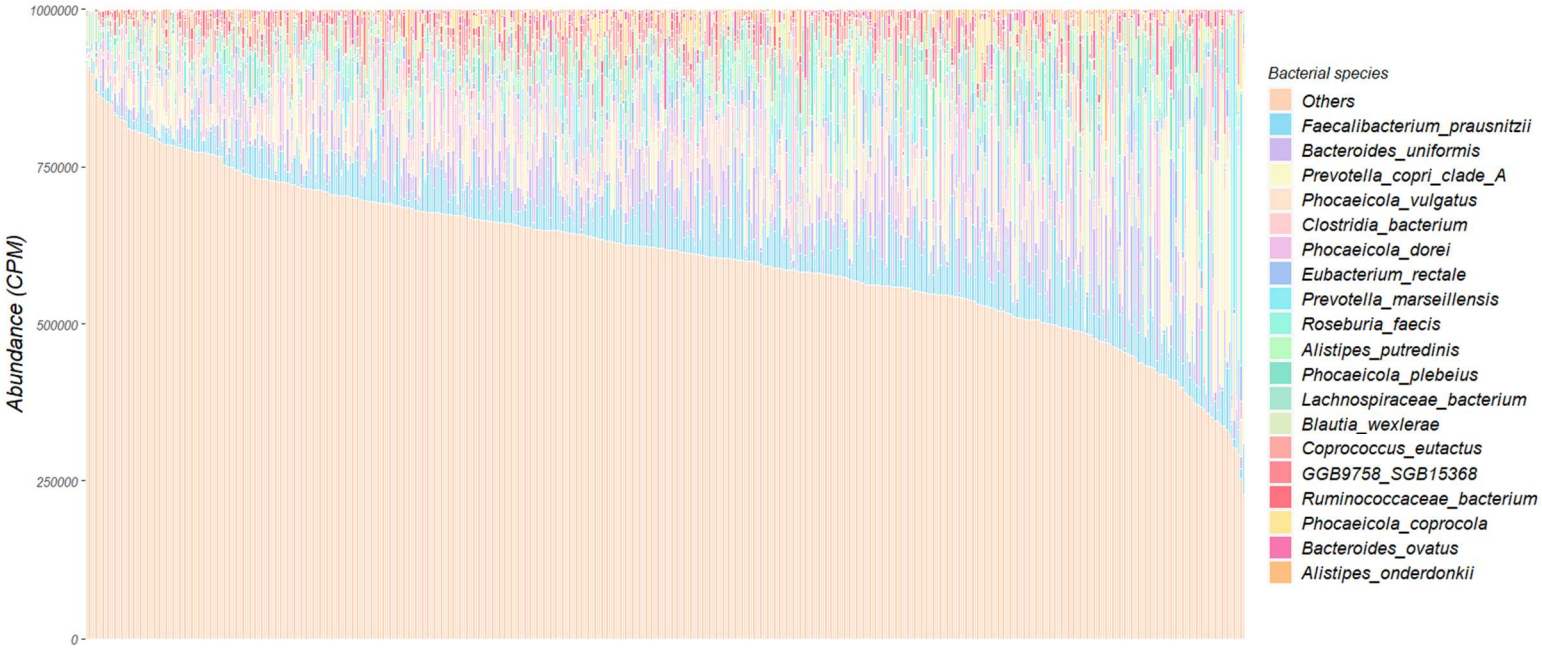


Figure 19. Bacterial profiling. Distribution of mapped bacterial reads per sample

5.4.4.1.1 Interplay between bacterial species, functions and personal traits

The next step was to explore correlation between distinct microbial profiles and functional pathways from MetaCyc with demographic data using linear models implemented in MaAsLin2. Fixed effects considered included "Bowel_movement", "Gender", "Smoke", "Region_Areas", "Season_Year". Additionally, biometric data ("Age" and "BMI") were correlated with microbial profiles using corrected Spearman correlations.

After correction, **age** was associated with 33 bacterial species. A positive correlation was observed with *A. muciniphila* (q= 0.031), while negative correlations were found with *Bifidobacterium bifidum* (q= 0.006) and *Flavonifractor plautii* (q= 0.046). Interestingly, some of the strongest correlations were found with unclassified species (Figure 20, ANNEX 7).

BMI showed linked with 26 bacterial species. Half of these exhibited negative associations, including *Intestinimonas gabonensis* (q= 0.017), *Intestinimonas massiliensis* (q= 0.026) and some unknown species (*Ruminococcaceae bacterium* D5 (q= 0.015); *Bacili bacterium* (q= 0.033)). The other half demonstrated positive associations, such as with *Coprococcus comes* (q= 7.69e-05) and *R. torques* (q= 0.019) (Figure 20, ANNEX 7).

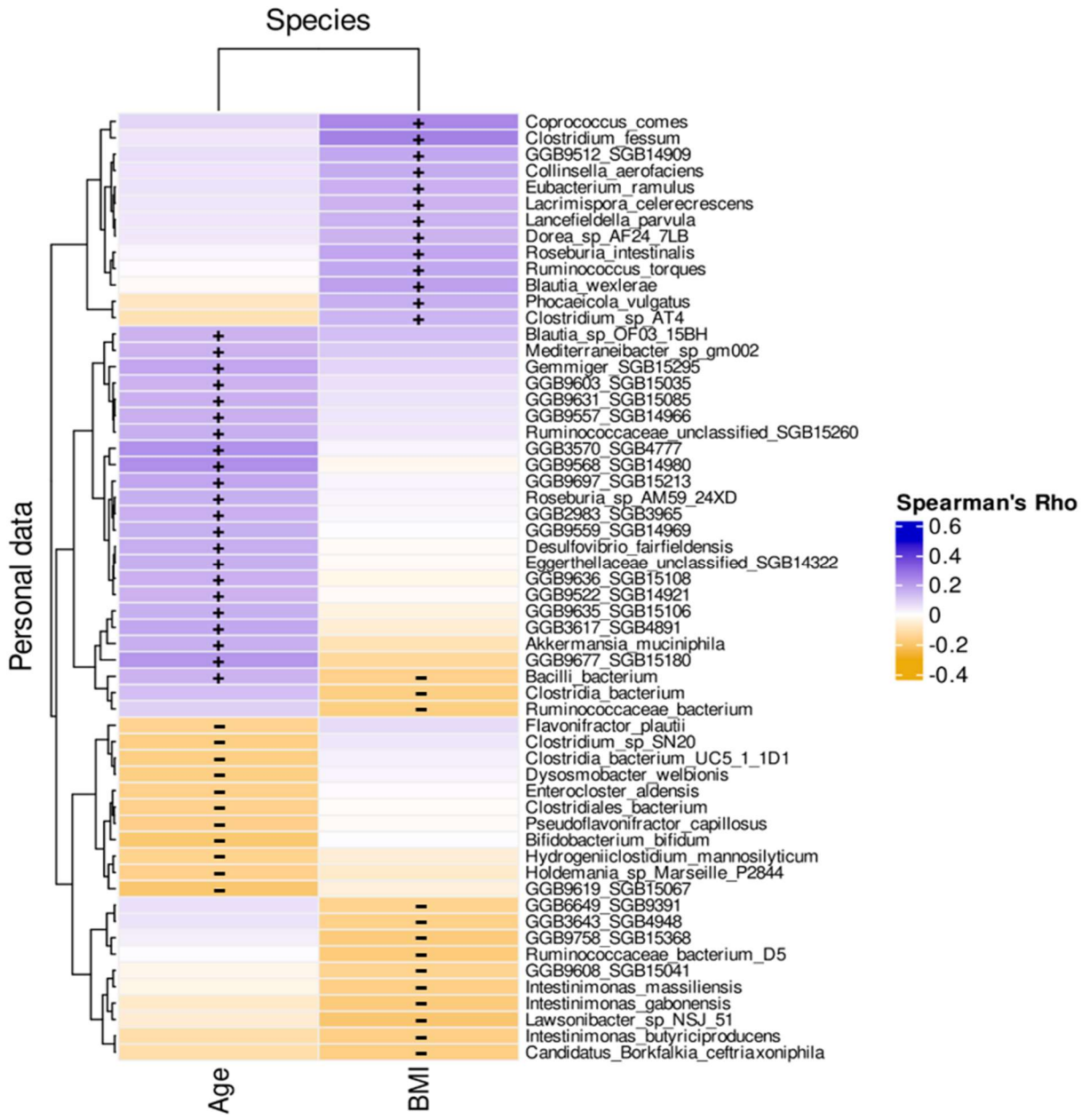


Figure 20. Correlation between age, BMI and bacterial species. Symbols inside the color squares denote significant associations (FDR < 0.05). Analysis were performed using the Spearman correlation test (n= 500 baseline).

At functional level, BMI was associated with 39 pathways (26 positive and 13 negative). Higher BMI values were linked to a general downregulation of biosynthetic pathways related to amino-acids, co-factors, CHO, and nucleotides, as well as a reduction in fermentation processes. In contrast, increasing functional pathways related to amino acid and carboxylic degradation, as well as peptidoglycan maturation involving meso-diaminopimelate were observed. No significant functional changes were observed with age (Figure 21, ANNEX 8).

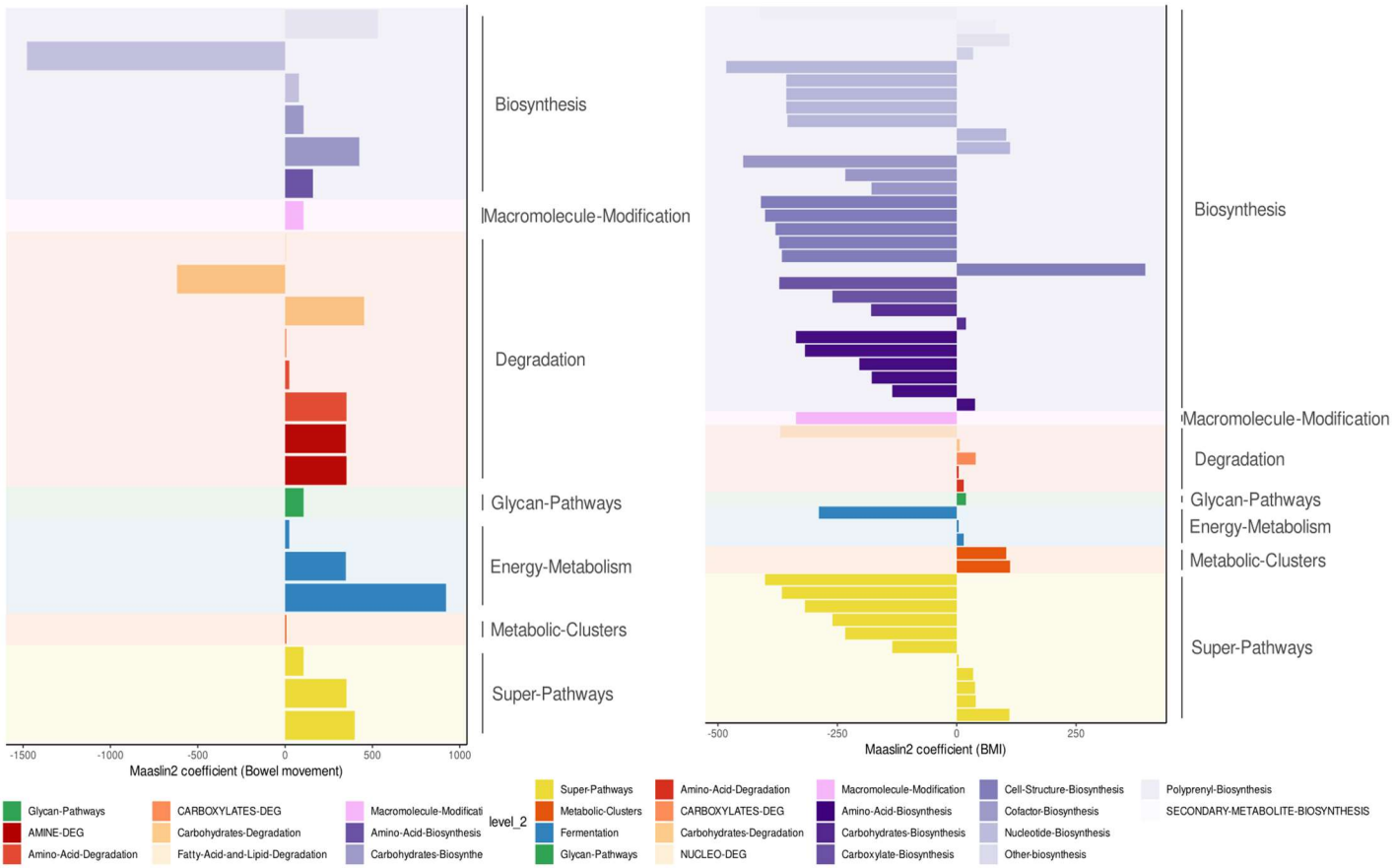


Figure 21. Differentially abundant pathways at the metagenomic level. Pathways are classified by their functionality according to the MetaCyc database and are influenced by several conditions. Differentially abundant pathways were compared between low transit time (>3 times per week, 1 or 2 times per week) and the reference (once a day). Positive coefficients reflected pathways enriched in low transit time (left side of the figure), while negative coefficients represented their depletion. Differentially abundant pathways also depend on BMI, with positive coefficients indicating a higher abundance of pathways in individuals with higher BMI (right side of the figure) (n= 500 baseline).

Analyzing categorical variables also revealed significant microbiome changes. The highest number of significant associations was found with **bowel frequency** (Table 22). At compositional level, using "1 day" as reference value, longer defecation intervals (> 3 times per week) were positively correlated with ten species, including *A. muciniphila*, q= 0.014 and *Intestinimonas massiliensis*, q= 0.011). In contrast, two species (*Blautia wexlerae*, q= 0.011 and unknown *GGB9614_ SGB15049*, q= 0.028) correlated negatively with longer transit times (1.5 times per week). On the other hand, one species (*Ruthenibacterium lactatiformans*, q= 0.018) was negatively associated with shorter transit times (> than 2 times per day). Longer transit times (1.5 times and > 3 times per week) were also linked to greater number of pathways, particularly those related to fermentation, lipid biosynthesis, glycan pathways, and the degradation and biosynthesis of amine and amino acids. Conversely, shorter transit times (> 2 times per day) were associated with increased number of CHO degradation pathways (Figure 21).

Table 22. Significant association (q < 0.05) between bacterial species and categorical variables (bowel frequency, gender, region areas and smoke). Output from MaAsLin2 (n= 500 baseline)

Feature	*Metadata*	*Value*	*Coef*	*Stderr*	*N*	*N.not.0*	*p-value*	*q-value*
Ruminococcaceae_bacterium_D5	Bowel_movement	1.5_WK	4.37507354	0.882868373	497	220	1.09061E-06	0.002155314
GGB9694_SGB15204	Bowel_movement	>3_WK	1.920275156	0.426136296	497	91	8.80967E-06	0.006964048
Clostridia_unclassified_SGB15402	Bowel_movement	>3_WK	1.501048276	0.348476429	497	130	2.10775E-05	0.01078727
Candidatus_Geddesella_stercoravicola	Bowel_movement	>3_WK	1.852320933	0.433039851	497	109	2.39735E-05	0.011147696
Intestinimonas_massiliensis	Bowel_movement	>3_WK	2.458752317	0.578402351	497	238	2.68522E-05	0.011171922
Blautia_wexlerae	Bowel_movement	1.5_WK	-1.693903398	0.397980557	497	496	2.62513E-05	0.011171922
GGB9620_SGB15068	Bowel_movement	1.5_WK	3.023575584	0.717451007	497	81	3.13674E-05	0.012397981
Akkermansia_muciniphila	Bowel_movement	>3_WK	2.950949946	0.713162815	497	311	4.32493E-05	0.014394309
GGB45491_SGB63163	Bowel_movement	>3_WK	1.363174811	0.329640632	497	74	4.37019E-05	0.014394309
Cloacibacillus_evryensis	Bowel_movement	>3_WK	1.558437572	0.382232082	497	50	5.55977E-05	0.017580006
Ruthenibacterium_lactatiformans	Bowel_movement	>2_Day	-2.439258171	0.601446464	497	468	6.07444E-05	0.018468621
GGB9694_SGB15203	Bowel_movement	1.5_WK	3.557723051	0.892079508	497	174	7.99665E-05	0.02082335
GGB9342_SGB14306	Bowel_movement	>3_WK	2.663542387	0.673283231	497	289	9.09998E-05	0.021157445
GGB9524_SGB14924	Bowel_movement	>3_WK	2.072198382	0.532808072	497	240	0.000118805	0.025382574
GGB9614_SGB15049	Bowel_movement	1.5_WK	-3.810633114	0.989177801	497	336	0.000137411	0.027852225
Dielma_fastidiosa	Bowel_movement	>3_WK	1.231343136	0.332201125	497	65	0.000241448	0.042414368
GGB3892_SGB5290	Bowel_movement	1.5_WK	1.806174996	0.494634234	497	60	0.000297405	0.047019757
Ruminococcaceae_unclassified_SGB15309	Gender	m	1.716202241	0.395537634	497	62	1.83998E-05	0.010389312
GGB9635_SGB15103	Gender	m	1.971739989	0.474725854	497	114	4.05239E-05	0.014394309
GGB3118_SGB4130	Gender	m	1.068404698	0.264371917	497	66	6.44217E-05	0.018861244
Allisonella_histaminiformans	Gender	m	1.78797674	0.446103879	497	124	7.37844E-05	0.020112606
Prevotella_copri_clade_C	Gender	m	2.107496662	0.531526648	497	69	8.77882E-05	0.021029258
Parabacteroides_merdae	Season_Year	Spring	-4.219562239	0.981460453	497	420	2.18338E-05	0.01078727
Faecalibacillus_intestinalis	Season_Year	Autumn	-1.603682447	0.43050418	497	392	0.000224988	0.041361184

The remaining significant correlations were observed with season of the year (two species and reduced isoprene biosynthesis) and gender (five species and seven pathways) (Table 22) (ANNEX 8).

5.4.4.1.2 Gut microbial associations and functions with diet related variables

Some of the strongest microbial associations with diet were obtained with uncultured or unidentified taxa (ANNEX 9). Among the DQIs, the aMED and hPDI adjusted significantly correlated with 14 out of 27 species, including negative correlations with *F. plautii* (hPDI, rho= -0.17, q= 0.042; aMED, rho= -0.19, q= 0.013) or *R. torques* (hPDI, rho= -0.18, q= 0.023; aMED, rho= -0.23, q= 0.003; PDI, rho= -0.21, q= 0.006) and positive with *A. muciniphila* (aMED, rho= 0.17, q= 0.041), *H. parainfluenzae* (aMED, rho= 0.17, q= 0.034) , *Intestinimonas gabonensis* (hPDI, rho= 0.19, q= 0.011) and *Clostridium saccharogumia* (aMED, rho= 0.17 , q= 0.042; hPDI, rho= 0.19 , q= 0.016). This highlights the impact of diet diversity and quality on gut microbiota responsiveness. Other indices, including HEI_2015, IASE, MAR, MDI, PDI and uPDI also influenced the human microbiome, although to a lesser extend. Interestigly, white bread food group presented the highest number of associations (6 out 34), all negative, including *Clostridium saccharogumia* (rho= -0.20, q= 0.02 and two unclassified *Clostridium* species (*sp AF20 17LB* and *sp AF36 4*). Fruits and fruit products showed the second highest number of correlations (5 out of 34), all positive like *Lachnospira eligens* (rho= 0.19, q= 0.025). The strongest food group-microbe association was found between yogurt and *Streptococcus thermophilus* (rho= 0.32, q= 0.00137) (Figure 22, ANNEX 9).

When analyzing food items at a broader level, two distinct clusters emerged, representing foods that correlated in opposite directions. Cluster one included items such as white bread, refined grains, pastries, soft drinks and processed food (Items 18, 22, 44, 51 and 58), while cluster two consisted of raw and cooked leafy vegetables, carrot, fresh fruit, dried fruit, nuts and seeds, blue fish and dark chocolate (Items 1, 2, 6, 14, 16, 17, 36 and 45). Less healthy food choices presented negative correlations with some *Clostridium sp*. (example *C. saccharogumia*, *Clostridium sp AF20 17LB*) as well as *Intestinimonas gabonensis*, *Bacteroides cellullosilyticus*, *Mediterranibacter butyricigenes* and unidentified GGBs. In contrast, healthier food choices correlated negatively with species such as *F. plautii*, *R. torques*, *R. gnavus* and positive associations with *Roserburia hominis* and *L. eligens* (Figure 23, ANNEX 9).

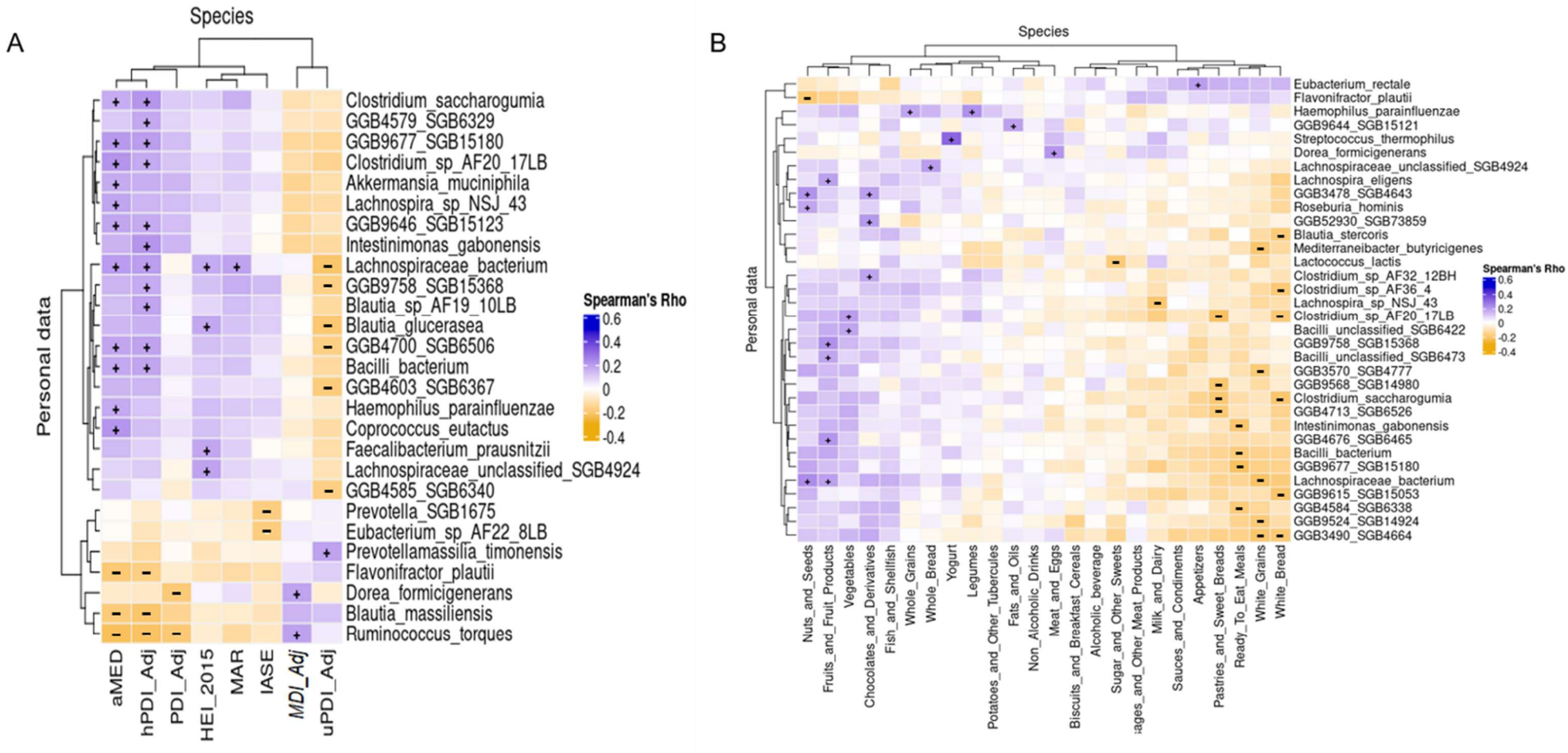


Figure 22. Single Spearman correlations adjusted by BH method between microbial species, DQIs (A) and Food groups (B). Symbols inside the color squares denote significant associations (FDR < 0.05) (n= 500 baseline).

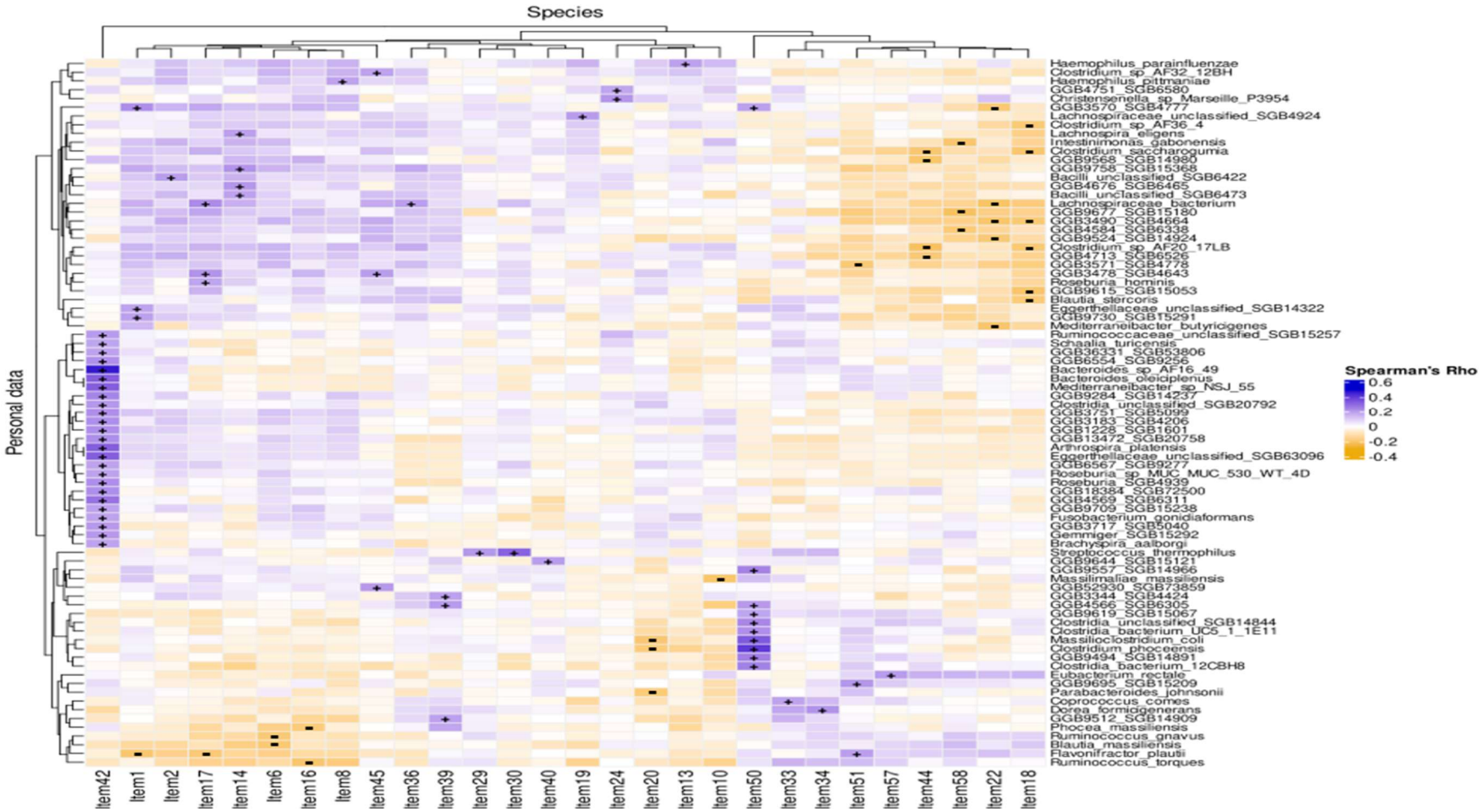


Figure 23. Single Spearman correlations adjusted by BH method between microbial species and food items (n= 500 baseline). Symbols inside the color squares denote significant associations (FDR < 0.05). **Item1**: Raw leafy vegetables; **Item2**: Boiled leafy vegetables; **Item6**: Carrot, pumpkin and beet ;**Item8**: Crucifers; **Item10**: Corn and fresh legumes; **Item13**: Cooked lentils, cooked kidney beans (pinto, white or black), and cooked chickpeas; **Item14:** Fresh fruit ; **Item16:** Dehydrated fruit; **Item17**: Nuts and seeds; **Item18**: White bread; **Item19**: Whole wheat bread ; I**tem20**: Breakfast cereal ; **Item22**: Refined cooked cereal and pasta; **Item24**: Whole-milk ; **Item29:** Low-fat cheese ; **Item30:** Fermented dairy; **Item33**: Red meat ; **Item34**: Lean meat ; **Item36:** Bluefish; Item39: Mollusk and crustacean; **Item40**: Olive oil ; **Item42**: Other oils such as those from corn, rapeseed, and grape seed; **Item44:** Pastries; **Item45:** Dark chocolate (> 50% cocoa) and cocoa powder; **Item50**: Coffee (with and without caffeine); **Item51**: Soft drinks; **Item57**: Fried potato, nacho, salted tortilla chips, snacks, salted pretzels, potato ring crisps, twiglets, and crackers; **Item58**: Processed food.

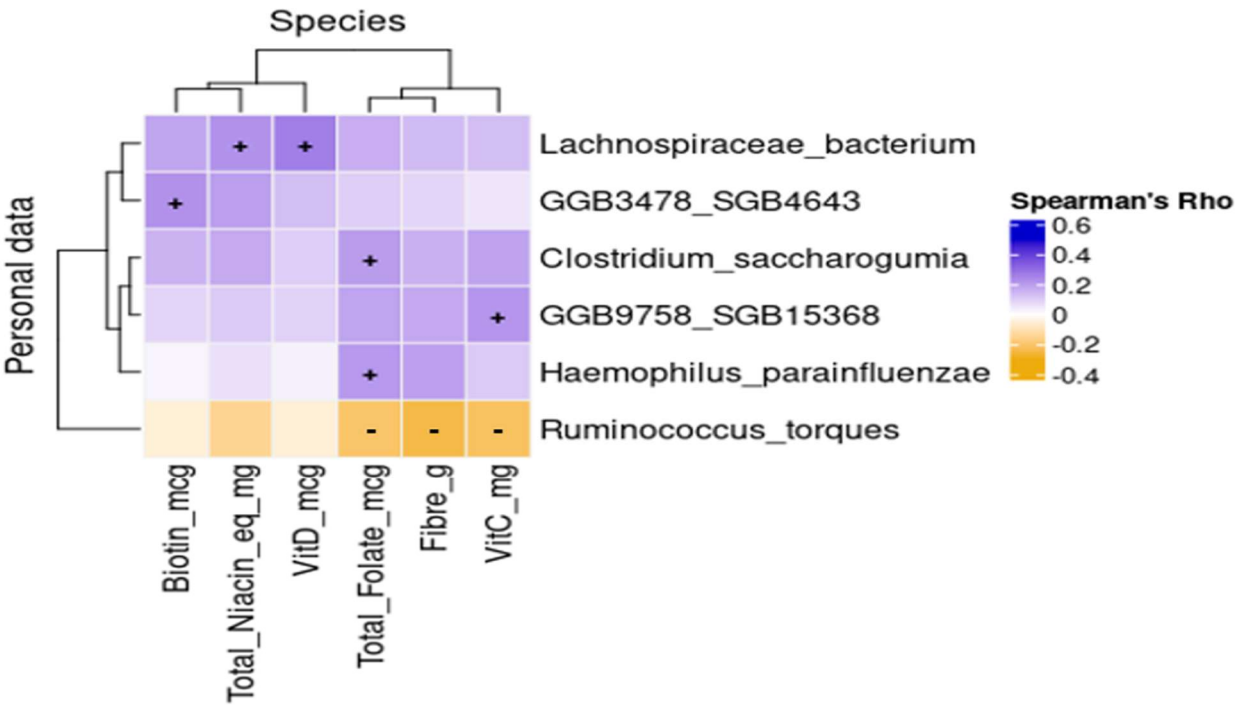


Figure 24. Single Spearman correlations adjusted by BH method between microbial species and different micro and macronutrients quantified using the in-house sFFQ (n= 500 baseline). Symbols inside the color squares denote significant associations (FDR < 0.05).

Surprisingly, Item 42, which includes "Other oils such as those from corn, rapeseed, and grape seed" exhibited the highest number of associations and some of the strongest, alongside with Item 50, which included coffee with and without caffeine. Specifically, oils correlated with 25 bacterial species while coffee was found associated with ten bacterial species including *Clostridium phoceensis* (rho= 0.42, q= 0) and *Massilioclostridium coli* (rho= 0.36, q= 0). More associations are available in Figure 23 and ANNEX 9.

At the nutrient level, two different clusters were identified. Negative correlations were found between diverse macro- and micronutrients and *R. torques*. Conversely, a positive correlation was observed with *H. parainfluenzae*, *Clostridium saccharogumia*, *Lachnospiraceae bacterium* and some GGBs (Figure 24, ANNEX 9).

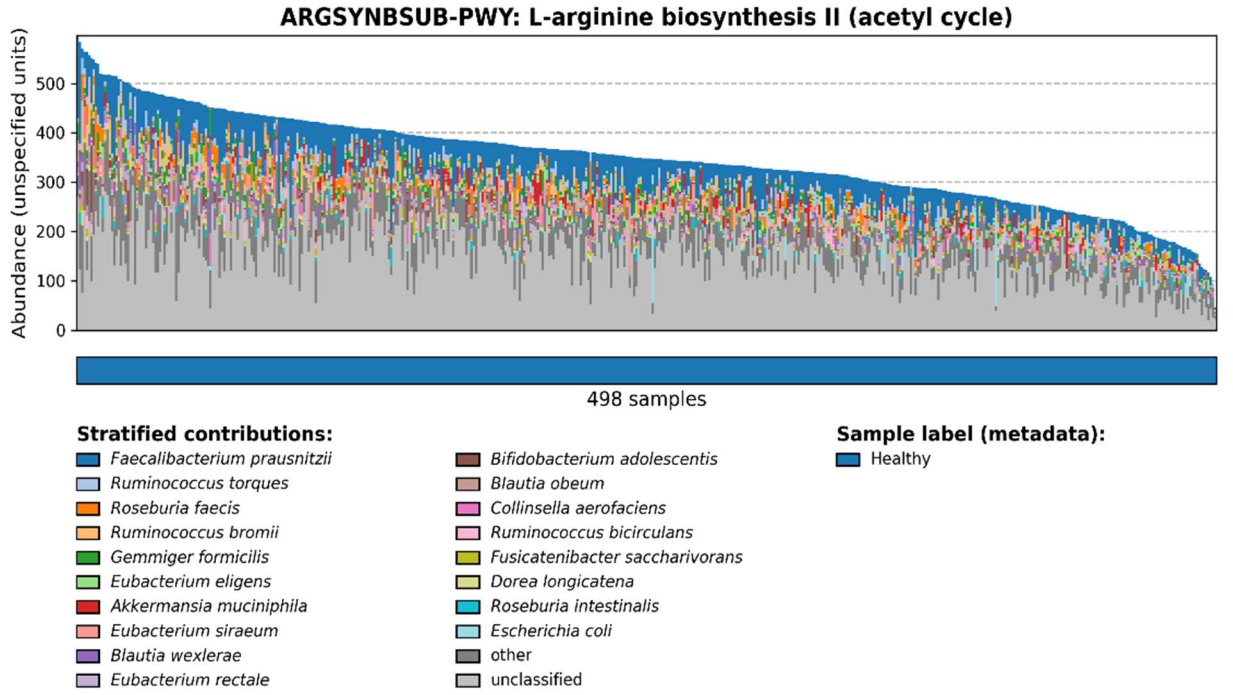
Finally, to assess whether dietary factors influence the functional properties of gut microbiome, we conducted linear association analysis between microbial pathways and dietary data. Significant relationships were identified between the intake of fruits, vegetables and nuts/seeds, and fiber with the L-arginine biosynthesis II and sucrose biosynthesis II pathways (Table 23).

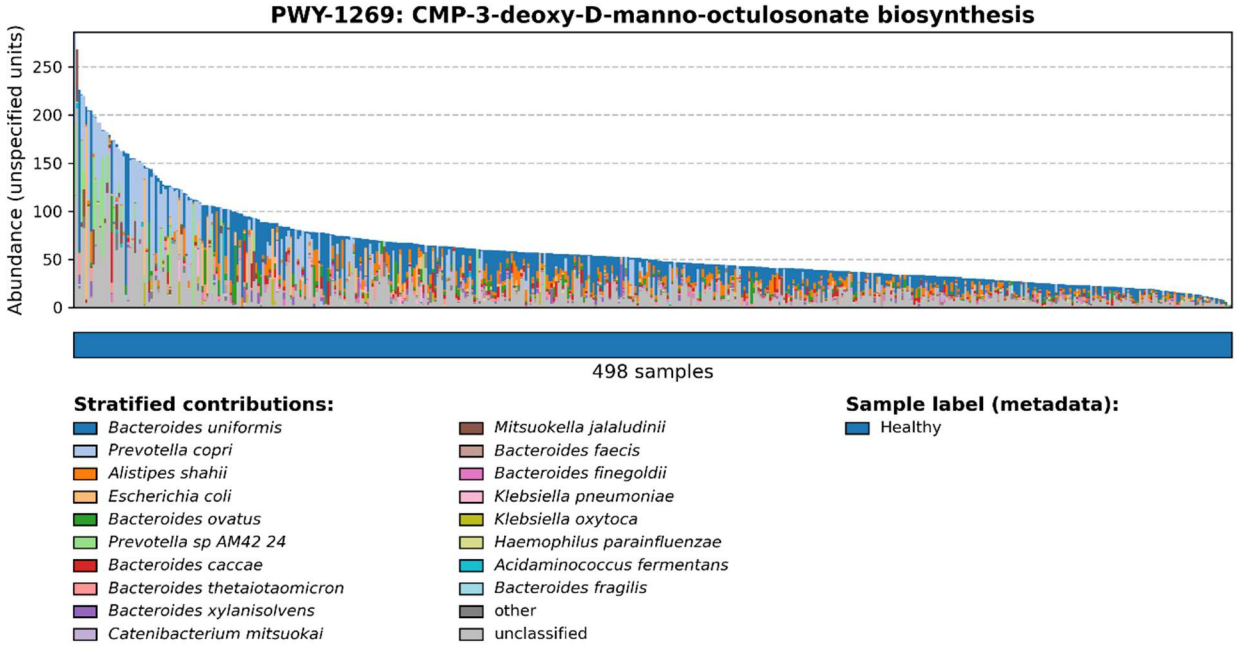
Upon further analysis of the bacterial species that contributed the most to L-arginine biosynthesis, we identified an unclassified taxon, followed by *F. prausnitzii*, a strictly anaerobic bacteria known for producing SCFAs, *R. torques*, *Roseburia faecis* and *Ruminococcus bromii* (amylolitic key-stone specie) as the top contributors.

Table 23. Significant Spearman correlations (FDR < 0.05) between dietary data and microbial composition.

Var1	*Var2*	*R*	*p-value*	*q-value*
Fibre_g	ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.19713608	9.5417E-06	0.00161665
Fibre_g	ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)	0.1854827	3.1713E-05	0.00161665
Fibre_g	PWY-7238: sucrose biosynthesis II	0.18499161	3.3307E-05	0.00161665
Fibre_g	ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	-0.15595741	0.00048401	0.00882174
Fruits_and_Fruit_Products	PWY66-399: gluconeogenesis III	0.1825066	4.2603E-05	0.00161665
Fruits_and_Fruit_Products	PWY-7238: sucrose biosynthesis II	0.18144171	4.7296E-05	0.00161665
Fruits_and_Fruit_Products	ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.17541306	8.4507E-05	0.0024442
Fruits_and_Fruit_Products	PWY0-1296: purine ribonucleosides degradation	0.16828378	0.00016384	0.00362369
Fruits_and_Fruit_Products	ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)	0.15514214	0.00051854	0.00883281
Fruits_and_Fruit_Products	PWY-7383: anaerobic energy metabolism (invertebrates, cytosol)	0.15187609	0.00068117	0.00987647
Fruits_and_Fruit_Products	PWY-7663: gondoate biosynthesis (anaerobic)	-0.1571815	0.00043614	0.00863107
Fruits_and_Fruit_Products	ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	-0.16966687	0.00014439	0.00361926
Nuts_and_Seeds	ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.18644805	2.8788E-05	0.00161665
Nuts_and_Seeds	PWY-7238: sucrose biosynthesis II	0.18197812	4.4874E-05	0.00161665
Nuts_and_Seeds	ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine)	0.17109894	0.00012654	0.0033986
Nuts_and_Seeds	GLYCOGENSYNTH-PWY: glycogen biosynthesis I (from ADP-D-Glucose)	0.15574712	0.0004927	0.00882174
Nuts_and_Seeds	PWY-7383: anaerobic energy metabolism (invertebrates, cytosol)	0.15465401	0.00054031	0.00883281
Nuts_and_Seeds	PWY66-399: gluconeogenesis III	0.15341569	0.00059937	0.00939019
Nuts_and_Seeds	PWY0-1586: peptidoglycan maturation (meso-diaminopimelate containing)	-0.18402045	3.6684E-05	0.00161665
Vegetables	PWY0-1296: purine ribonucleosides degradation	0.14806063	0.00093049	0.01239487
Vegetables	ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.14248501	0.00144877	0.01602168
Vegetables	PWY-7238: sucrose biosynthesis II	0.12396581	0.00565133	0.04086347
Item50	PWY-1269: CMP-3-deoxy-D-manno-octulosonate biosynthesis	-0.13003411	0.00368482	0.03078871

F. prauznitzii, R. faecis, B. uniformis, and *Blautia obeum* emerged as major contributors of the sucrose biosynthesis pathway. Another noteworthy significant correlation was found between item 50 in our sFFQ (coffee) and the CMP-3-deoxy-D-manno-octulosonate biosynthesis pathway. As with previous correlations, the most prevalent species was unclassified followed by *B. uniformis* and *Prevotella copri* (Figure 25).





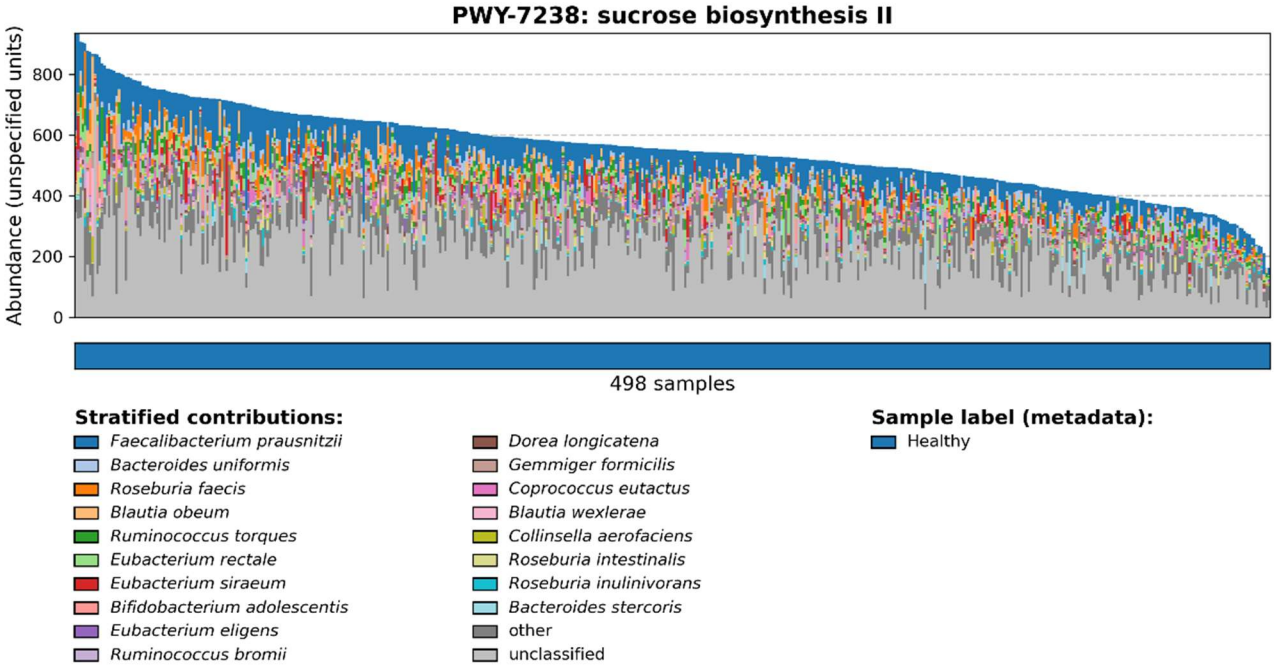


Figure 25.Top bacterial species with the highest level of contribution in terms of abundance to three functional pathways (L-arginine biosynthesis II, sucrose biosynthesis II and CMP-3-deoxy-D-manno-octulosonate biosynthesis)

5.4.4.2 Fungal composition

The five most prevalent fungal species in enriched samples (n= 100) were *Saccharomyces cerevisiae* (80 samples), *Malassezia restricta* (33 samples), *Debaryomyces hansenii* (25 samples), *Penicillium roqueforti* (21 samples) and *Meira nashicola* (21 samples).

When examining relative abundance of these species (Figure 26) *S. cerevisiae* remained the most abundant, accounting for a mean of 50.06%, followed by *Penicillium roqueforti* (5.77%), *Debaryomyces hanseii* (4.23%), *Geotrichum candidum* (3.41%) and *Rhamphospora nymphaeae* (3.34%).

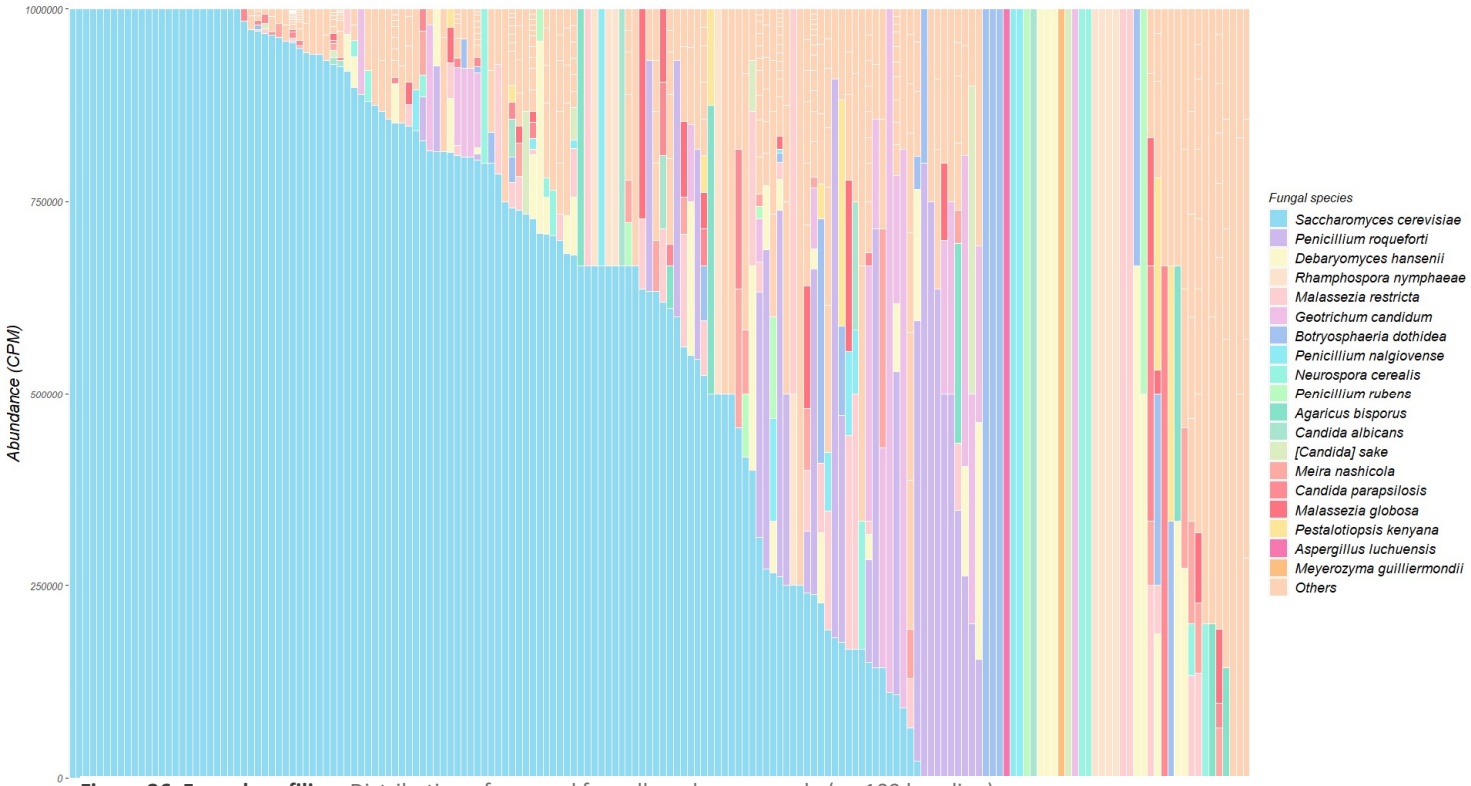


Figure 26. Fungal profiling. Distribution of mapped fungall reads per sample (n= 100 baseline)

5.4.4.2.1 Interplay between fungal species, functions and personal traits

Similar to our analysis of the bacterial microbiome, we assessed the effect of demographics and personal data on fungal microbiome using MaAsLin2 with "Bowel_movement", "Gender", "BMI", "Age", "Smoke", "Region_Areas", "Season_Year" as fixed effects. No significant associations were observed.

5.4.4.2.2 Associations of fungal species and functions with diet related variables

No significant associations were found between fungi and DQIs, food groups or food nutrients using the Spearman test. However, three food items significantly correlated with six fungal species. Skimmed milk (Item 26) showed positive correlation with four fungal species: *Talaromyces amestolkiae* (q= 0.0448), *Cyphellophora europea* (q= 0.0448), *Rhizopus delemar* (q= 0.0448) and *Brettanomyces sp.* (q= 0.0025). Confectionary (Item 46) was positively correlated with *Aspergillus penicilloides* (q= 0.0099), while red meat (Item32) exhibited the only negative correlation and the association with *Botryosphaeria dothidea* (q= 0.0007) (Figure 27).

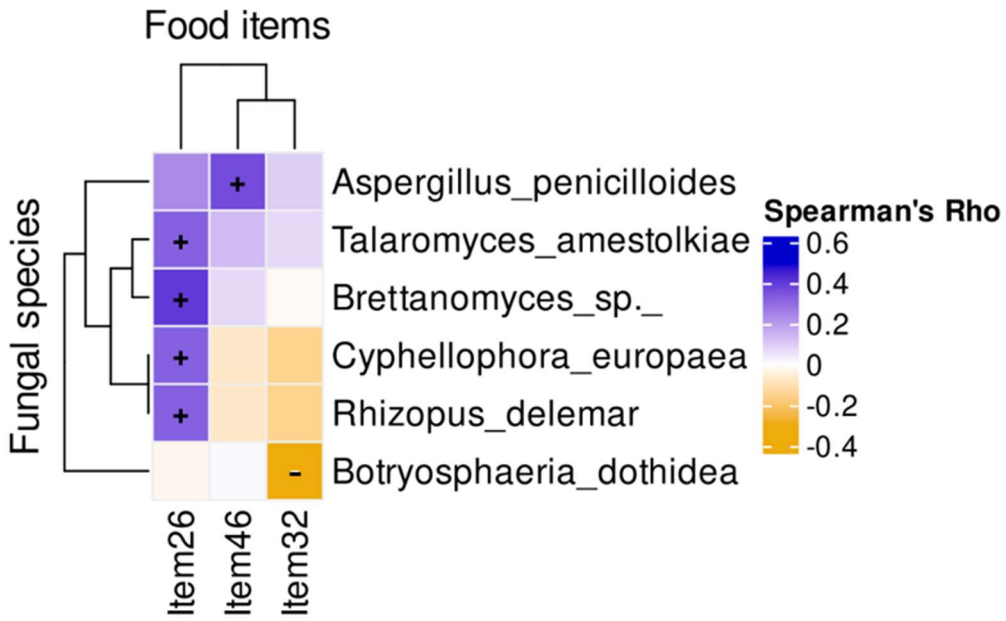


Figure 27. Correlations between fungal species and food items (n= 100 baseline). Analysis performed using the Spearman correlation test adjusted by BH method. Symbols inside the color squares denotes significant associations (FDR q< 0.05). **Item 26**: skimmed milk; **item 32**: red meat and **item 46**: confectionary.

5.4.5 *Fungal-bacterial interplay*

To investigate the interplay between bacteria and fungi, we used the Spearman correlation test. *Candidatus Metaruminococcus caecorum* presented the highest number of associations and the strongest positive correlation with *Penicillium spp* and *Penicillium camemberti* (q= 4.72e-07 and q= 7.46e-05, respectively). At fungal level, *Penicillium nalgiovense* and *Geotrichum candidum* showed the larger number of positive associations with 12 and 11 bacterial species, respectively, followed by *Penicillium nordicum* with ten correlations.

Additionally, few significant relatiosnhips involving pathobionts were observed, such as the association between *Candida albicans* with an unclassified *Clostidria* (q= 0.005). Extended data can be found in Figure 28 and ANNEX 10.

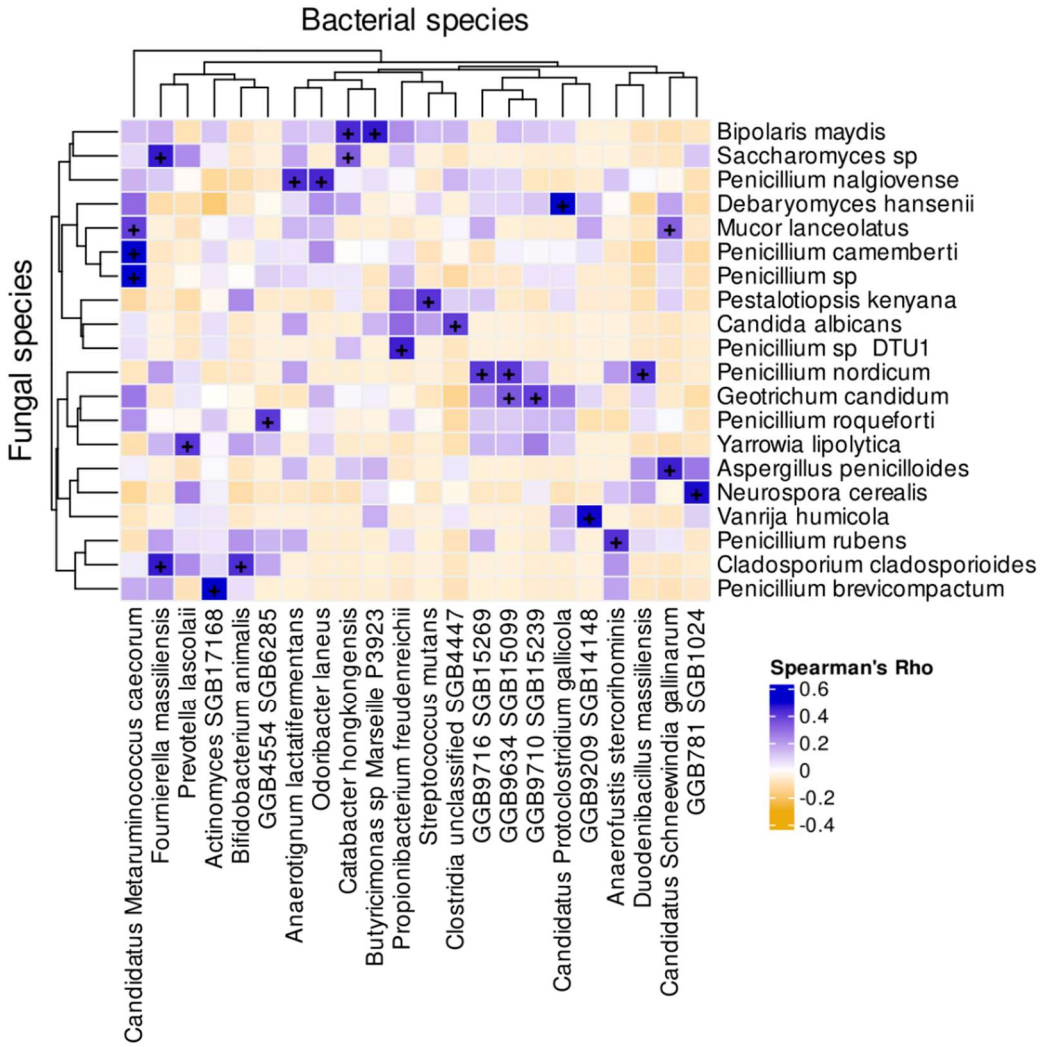


Figure 28. Correlations between bacterial and fungal species (n= 100 baseline). The analysis was performed using the Spearman correlation test, corrected by BH. Symbols inside the color squares denote significant associations (FDR, q< 0.05). Results were filtered by r > 0.4 and r < -0.4.

5.5 Relationship between diet and IBD-type microbiome

To investigate the relationship between diet and IBD-type dysbiosis, a well-known example of microbiome disruption in non-communicable diseases, we examined and compared our data with the microbiomes of 321 IBD patients, including 208 with CD and 113 with UC. Shotgun metagenomic dataset from previous projects was utilized for the analysis (269). We developed the IBD-similarity index, a metric that measures divergence from the microbiomes of our 500 healthy individuals from the microbiomes of IBD patients, to assess microbial community disturbance (see methods section for detailed details). Higher index values indicate greater similarity to microbial profiles associated with IBD.

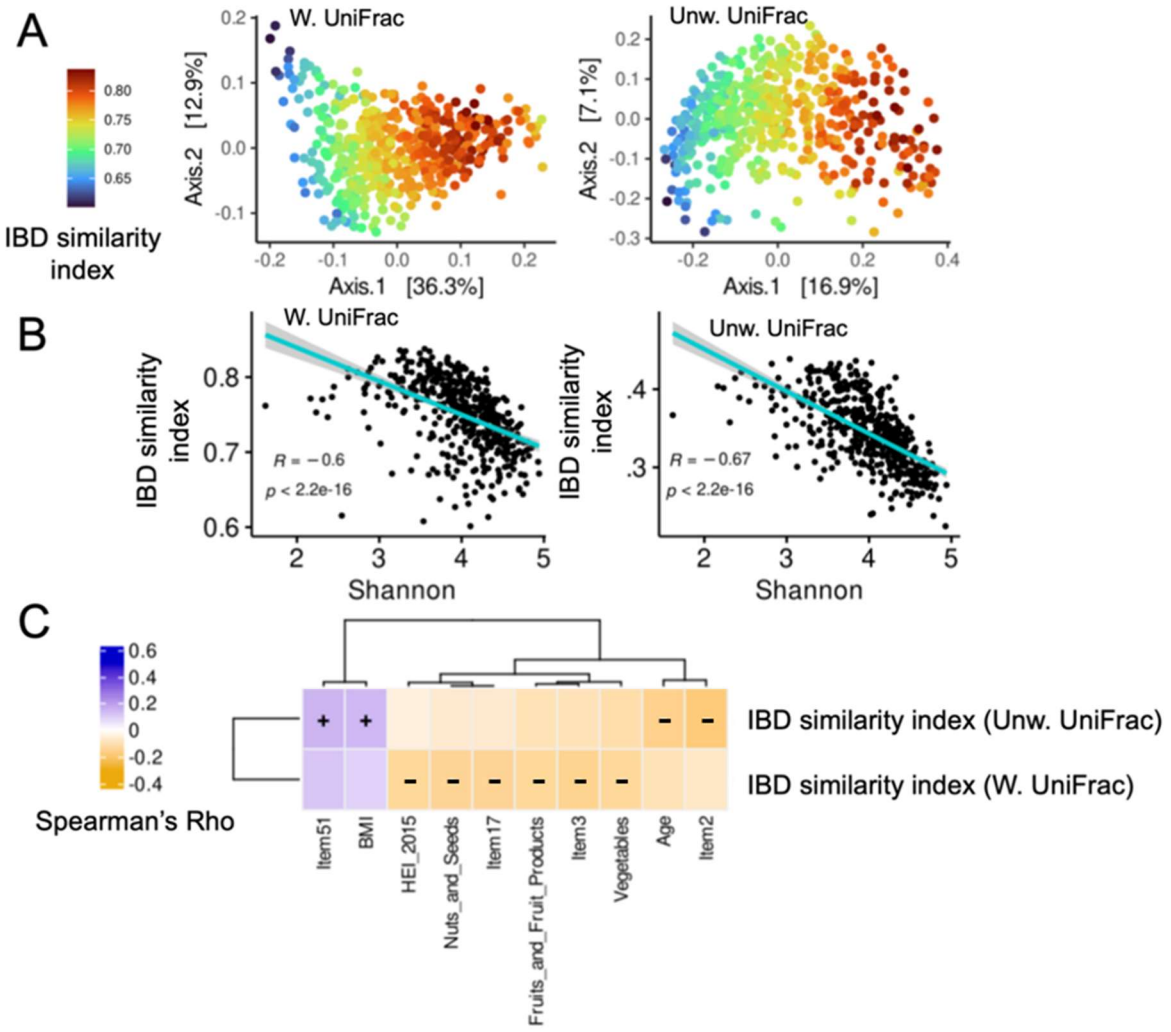


Figure 29. IBD similarity index and population characteristics. A). Weighted (W.) and unweighted (Unw.) UniFrac distances of our cohort of healthy individuals (n= 500 baseline) colored by IBD-similarity score. IBD-similarity score was calculated as 1- median of a healthy sample to all samples in IBD 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 α- diversity (Chao and Shannon). C) Integrated heatmap representing food groups, items, DQIs 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. Symbols inside the color squares denote significant associations (FDR < 0.05). **Item 2:** Cooked leafy vegetables; **Item 3:** Tomato; **Item 17**: Nuts and seeds; **Item 51:** Soft drinks).

Using this method, healthy individuals from POP cohort were successfully stratified based on their degree of similarity to the IBD microbiome, explaining up to 36.3 % and 16.9% of the variance in the first principal component when using weighted and unweighted UniFrac distances, respectively (Figure 29A).

Spearman correlation analysis revealed that high α-diversity was correlated with low similarity to IBD microbiome profiles (Figure 29B). Furthermore, lower disruption of the microbiome was correlated with higher intake of healty food choices such as vegetables, nuts and seeds or fruits. Conversely, consumption of soft drinks (Item 51) was linked to greater microbiome disruption (Figure 29C).

Among the dietary quality indices, only the HEI-2015 demonstrated a meaningful connection. Age and BMI showed conflicting relationships, with higher BMI associated with more IBD-type disturbance while age was negatively correlated.

Correlation analysis between specific bacterial species, α- diversity, and the IBD-similarity index revealed that *F. plautii* and *R. gnavus* exhibited the strongest positive correlations with microbiome alterations. In contrast, the strongest negative correlations were found with unidentified *Clostridia* and *Bacilli species*, as well as *Methanobrevibacter smithii*. Interestingly, all species positively correlated with the IBD-similarity index were inversely associated with α-diversity metrics, and vice versa (Figure 30).

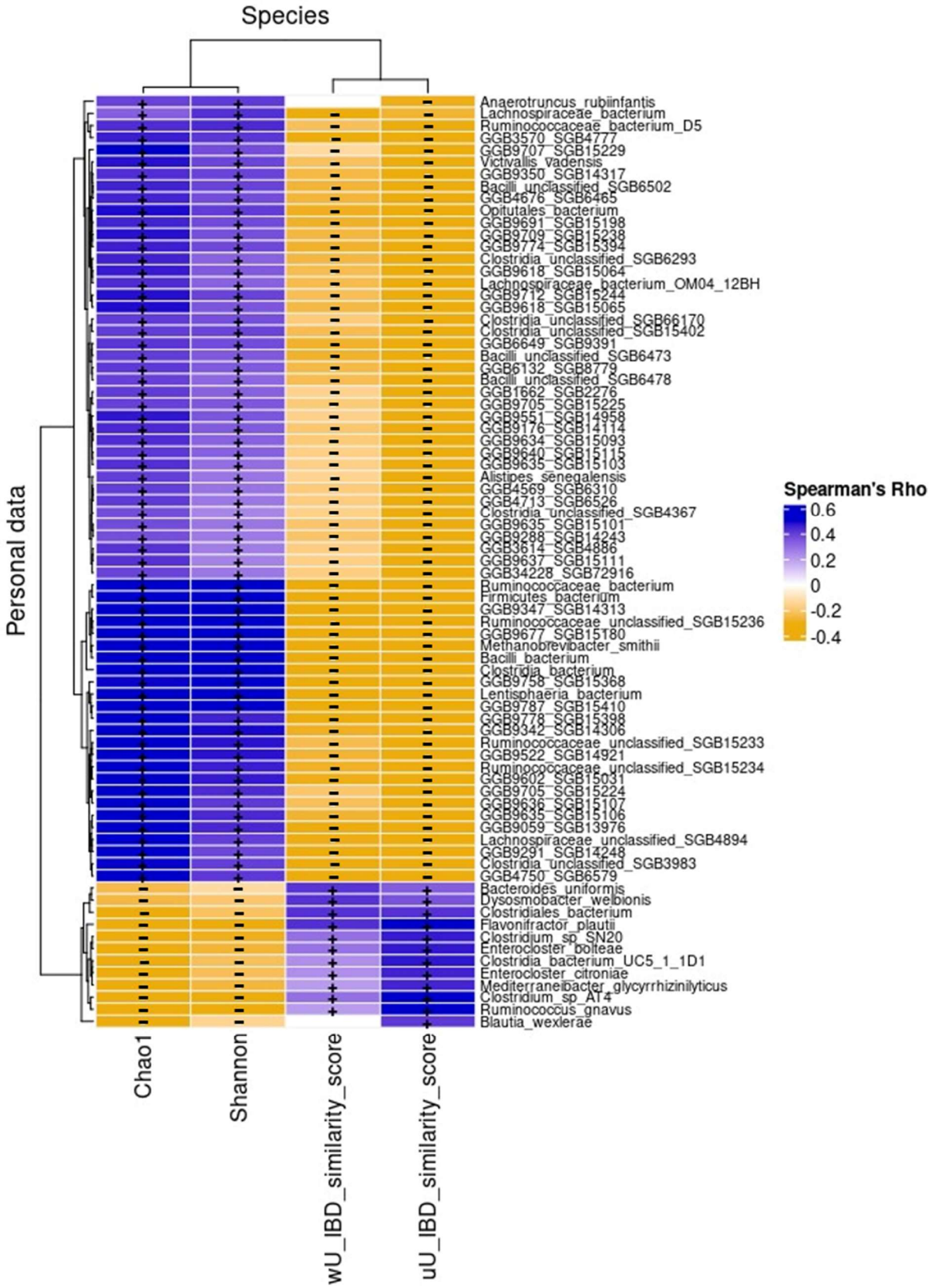
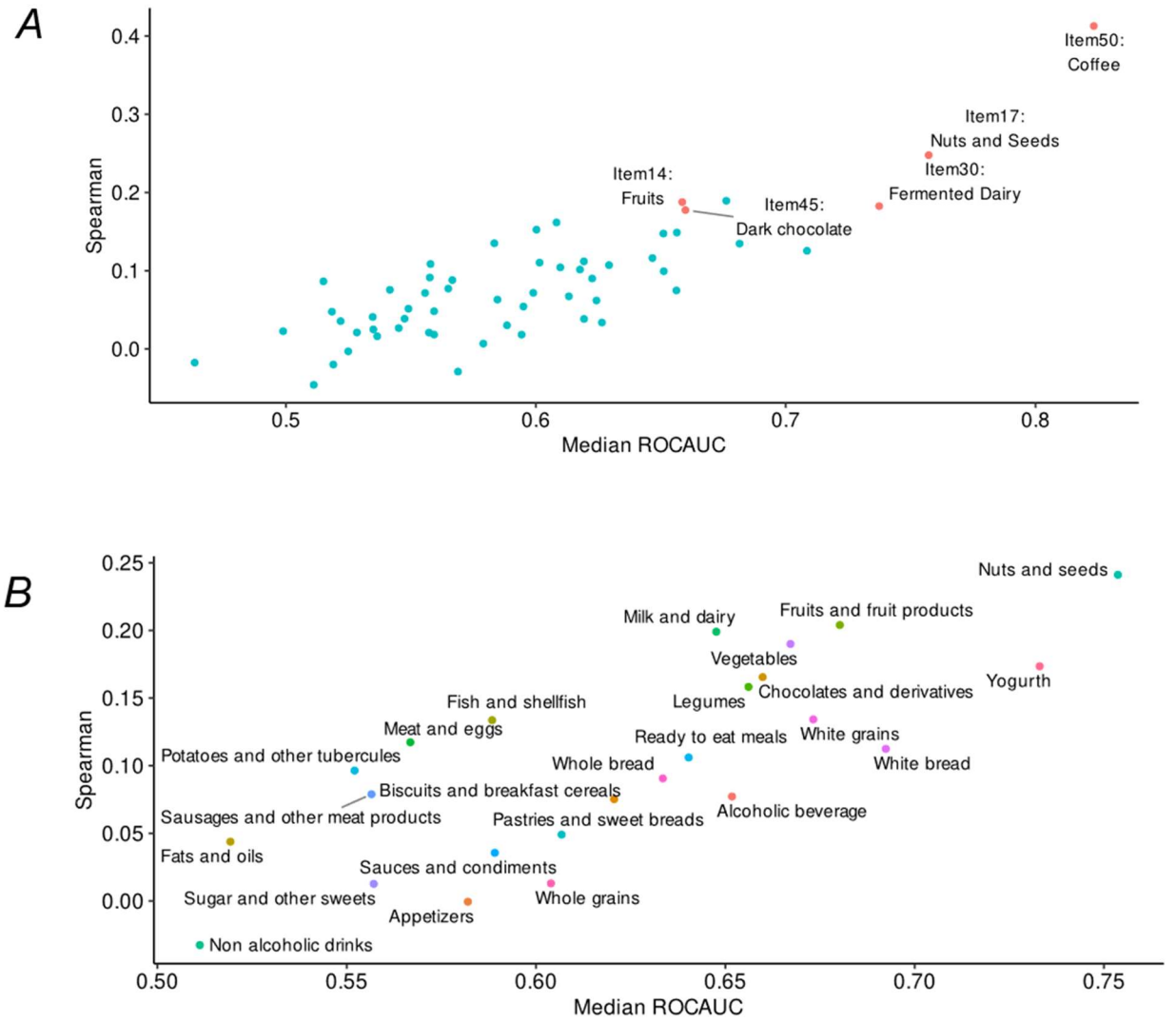


Figure 30. Significant Spearman correlation of bacterial species with IBD-similarity index (weighted and unweighted) and different α- diversity measures (Chao1 and Shannon). Symbols inside the color squares denote significant associations (FDR < 0.05) (n= 500 baseline).

5.6 Using microbiome as predictor of dietary intake

Using available dietary data and applying the random forest machine learning method to microbiome features, we demonstrated a strong connection between the composition of the bacterial microbiome and a variety of food items. These food items included **dark chocolate** (rho= 0.18, AUC= 0.66), **vegetables** (rho= 0.19, AUC= 0.67), **fruits** (rho= 0.19, AUC= 0.66), **coffee** (rho= 0.41, AUC= 0.82), **nuts and seeds** (rho= 0.25, AUC= 0.76), and **fermented dairy** (rho= 0.18, AUC= 0.74). The analysis utilizing food groups supported 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) (Figure 31). However, diet prediction using fungal data was not as accurate as bacteria. While some AUC values exceed 0.7, Spearman rho correlations were generally weak, with most values around 0.1 (ANNEX 11). Notably, the microbiome showed acceptable prediction accuracy for two DQIs, hPDI_Adj (rho= 0.30, AUC= 0.710) and aMED (rho= 0.32, AUC= 0.73) (Figure 31).



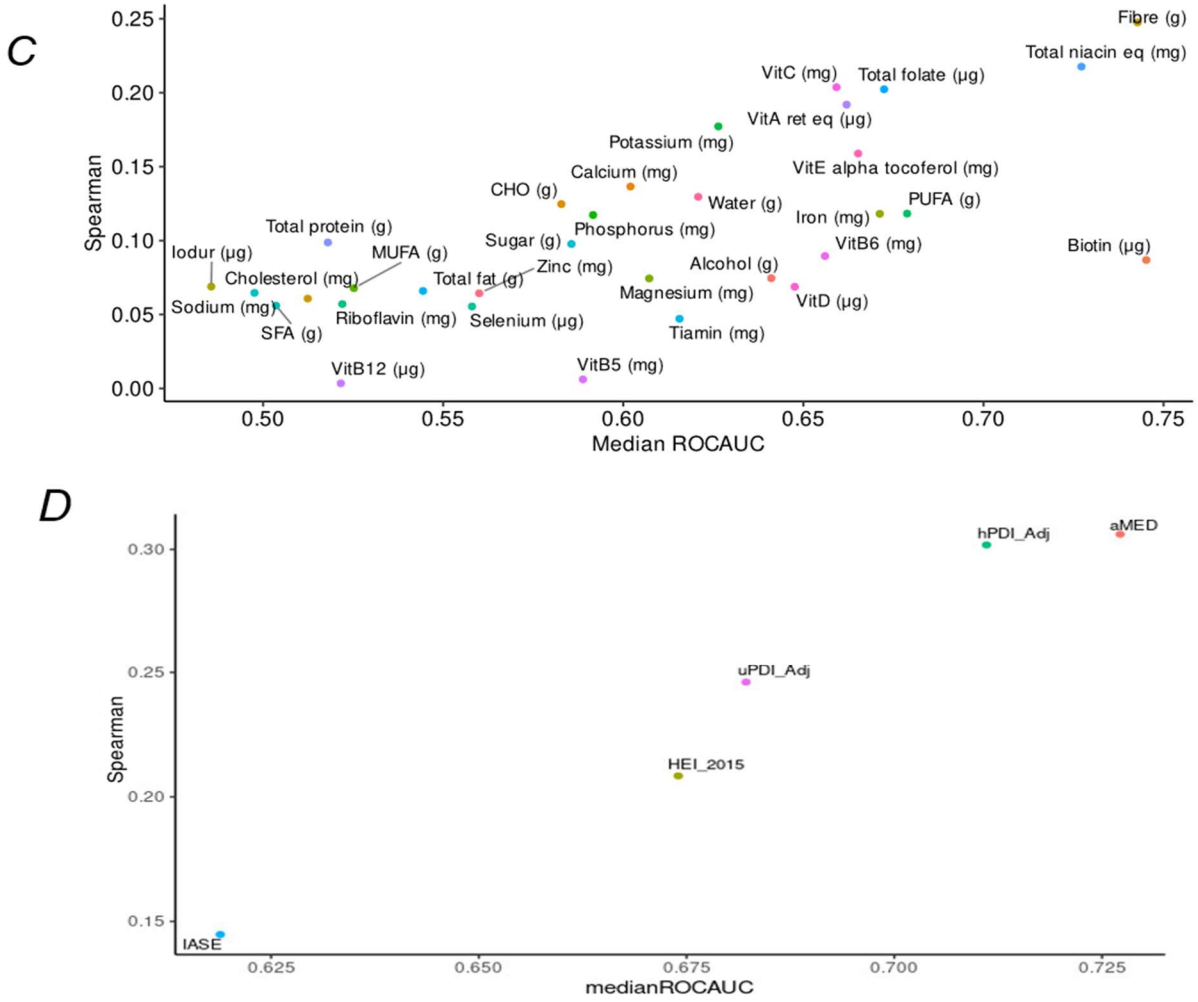


Figure 31. Prediction using machine learning technique (n= 500 baseline). Prediction of different food items (A), food groups (B), nutrients (C) and DQIs (D) using bacterial species-level genome bin (SGB)-level features information estimated by MetaPhlAn4. Y-axis and X-axis represent median Spearman's correlation and median receiver operating characteristic area under the curve (ROCAUC) from the random forest regressor and random forest classifier, respectively.

5.7 Website as contributory citizen science project

To raise awareness about the importance of the microbiome and promote a healthy lifestyle, we created the project website (https://manichanh.vhir.org/POP/en/) titled “POP Study: dietary habits and gut microbiome of the Spanish Population”. The homepage is divided into two main sections: “Study Results” and “Your Personal Results”. The website is available in three different languages including Spanish, Catalan and English; a French version is planned for the future to facilitate comparisons among French and Spanish dietary patterns.

5.7.1 *Study results*

The "Study results" section is publicly accessible and consists of four distinct parts. The first section provides an explanation of what microbiome is and why it is important. The second part briefly explores the relationship between the microbiome and diet and highlights its implications on health. The third section describes the POP Study, its objectives, and its design. Finally, a bibliography is provided to those interested in further learning (Figure 32).

Figure 32. Screenshots of the "Study results" section of the website in English version.

5.7.2 *Your personal results*

In "Your Personal Results" section, participants who donated stool samples and/or completed the sFFQ were able to access to their own results, if available. This section of the website is divided into two subsections: "Diet" and "Microbiome". In the "Diet" subsection, participants can: 1. Check their own daily intake of 19 micronutrients and four macronutrients over time sFFQ_0, sFFQ_1 and sFFQ_2), and compare them with the recommended values for the Spanish population; 2. Visualize their consumption of the 24 food groups included in the sFFQ at different timepoints and compare it with the median consumption of all participants; 3. View their scores for various DQIs and see how results compare to the median scores in POP cohort (Figure 33A, B, C). To enhance clarity, a brief explanation accompanies each result to help participants comprehend their data.

On the other side, in "Microbiome" subsection, participants can access the information from shotgun sequencing data, if available. Specifically, 1. Bacterial composition from kingdom to species level of the most abundant species characterized in the stool sample at baseline; 2. Chao1 and Shannon measures of α-diversity and the population median (Figure 33D, E); 3. The fungal composition from the kingdom to the species level, if fungal enrichment protocol was performed for this participant. The personal information can be freely downloaded in a PDF format. It is important to highlight that microbiome results were provided just for knowledge and that any clinical interpretation for participants was not given.

Finally, out of the 1,017 participants, 1,016 expressed an interest in receiving their dietary and microbiome information upon the completion of the study. As of April 22nd, 823 individual participants have accessed their results. Of these, 418 have downloaded the report in PDF format.

Figure 33. Example of the results presented to the participants in the study website. A) Estimated nutrient intake relative to the recommended guidelines by the Scientific Committee of the Spanish Agency for Food Safety and Nutrition (AESAN) and estimated energy intake from fats and carbohydrates compared to the recommended range. This graph can be viewed for any of the timepoints for which a participant has answered a sFFQ. B) Daily intake of each food group compared to the population median. This graph can be viewed for any of the timepoints for which a participant has answered a sFFQ. C) DQIs at each time point compared to the population median. Participants can access these results for aMED, HEI, hPDI and uPDI scores. D) Alpha diversity metrics for the participant's sample compared to the population median. E) Gut microbiome composition for the participant's sample, grouped at the desired taxonomic rank.

5.8 A step further: let's isolate viable bacteria

5.8.1 *Collection of healthy human gut samples (non-targeted culturomics)*

As a preliminary step, we assessed the efficiency of single-cell droplet approach by calculating the proportion of droplets that resulted in colony formation or detectable growth. The average growth yield per plate was 4.75% ± 2.13 (n= 2), values ranging from 2.69% to 8.01%, depending on the media and sample type and donor sample.

For healthy individuals (ANNEX 12), a total of 261 pure isolates were obtained across both media types (GAM and Y-LYHBHI.4 supplemented with RF) and under both liquid and solid culture conditions. Among these, 91.37% of isolates were successfully identified to the species or genus level using the updated MALDI-TOF database. The remaining 8.63% required further identification by 16S rRNA sequencing (Figure 35).

Across all processed human fecal samples, 27 distinct bacterial species, primarily obligate anaerobes, were isolated (excluding those pending 16S-based identification). Donor 1 yiealded a higher number of isolates than Donor 2. (Figure 34, Figure 35, ANNEX 13).

These isolates spanned over four phyla, predominantly Bacteroidota (55%) and Bacillota (22%), and 12 families. The most abundant families were Bacteroidaceae (33%) and Bifidobacteriaceae (19%). All species isolated had previously been cultured; no novel taxa were recovered.

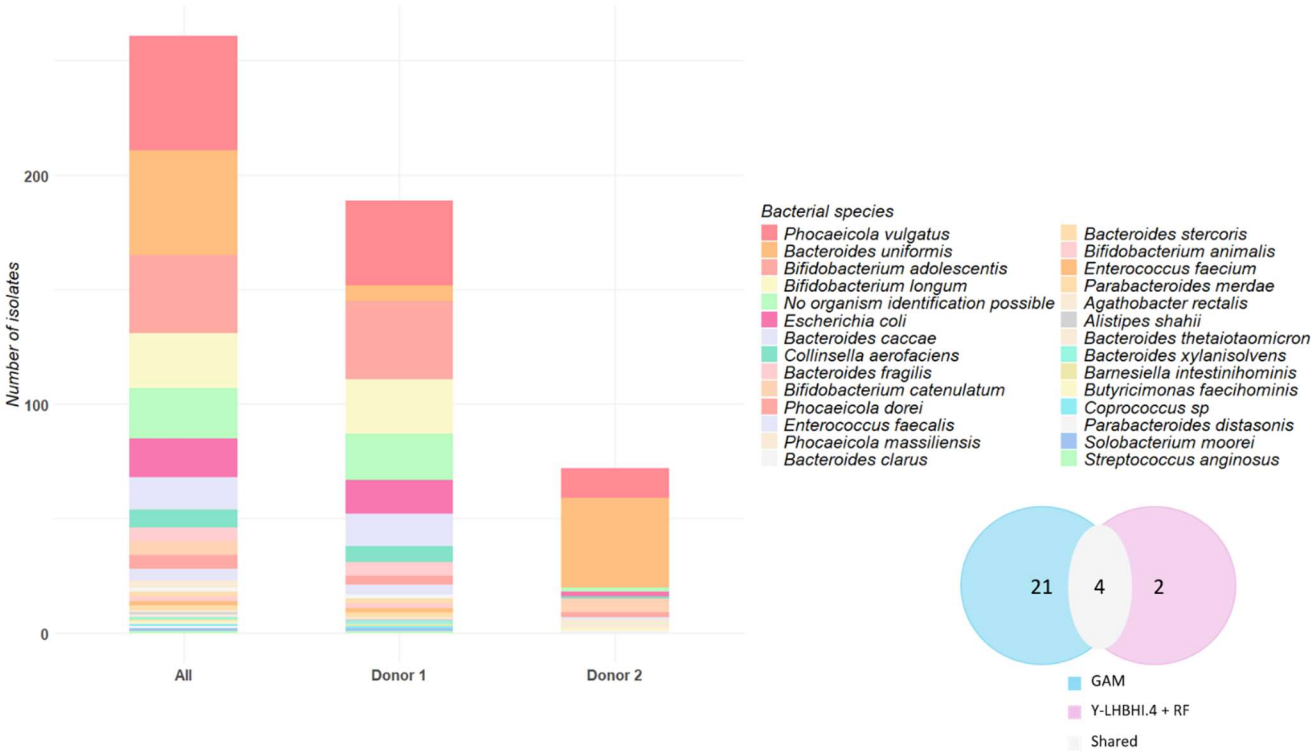


Figure 34. Number and species name of pure isolates obtained from each of the healthy donors as well as the combined number using non-targeted culturomics techique. Numbers in cycles indicate the number of different species obtained using GAM and Y-LHBHI.4 media or both.

Of the 27 bacterial species identified, 21 (78%) were exclusively isolated using GAM medium, while only two (7%) were uniquely recovered using Y-LHBHI.4 + RF. Additionally, *Bifidobacterium adolescentis*, *Bifidobacterium longum*, *Enterococcus faecalis,* and *Enterococcus faecium* were able to grow in both media types. A comparison between cultured isolates and metagenomic shotgun sequencing results is currently ongoing and was therefore not included in this thesis (Figure 34).

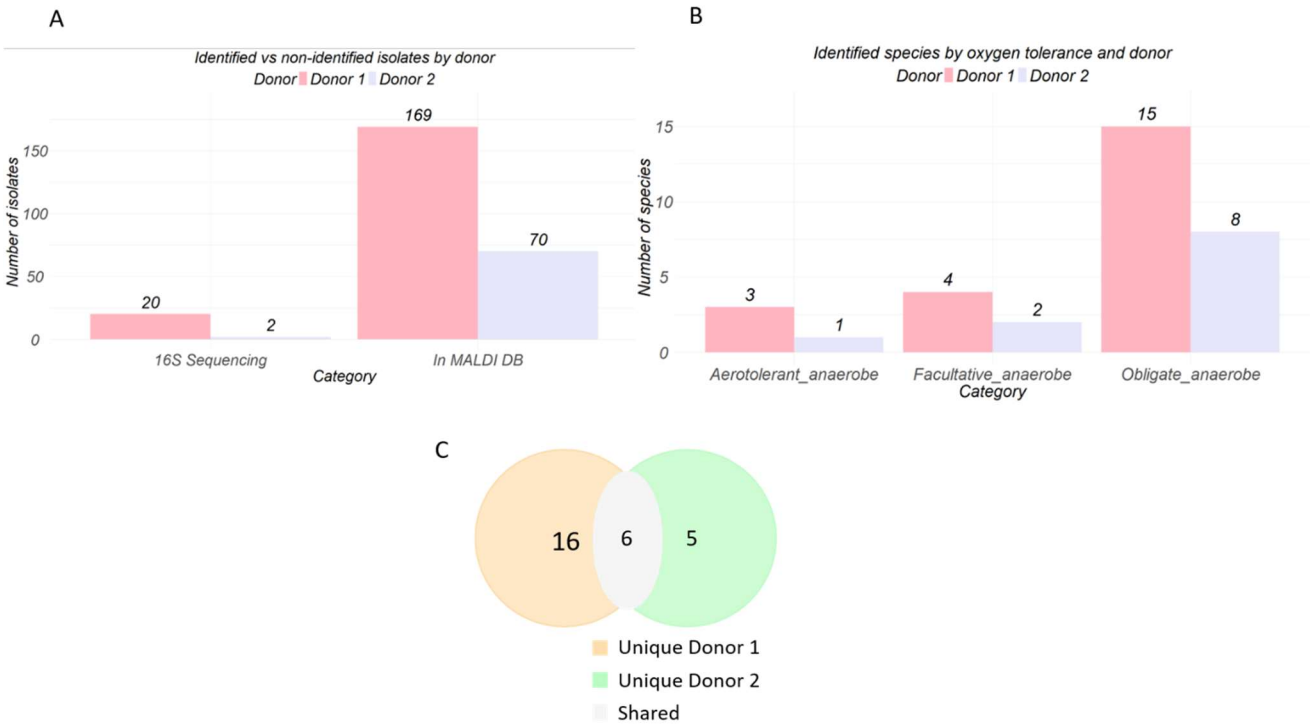


Figure 35. Number of isolated colonies per donor that were identified using MALDI-ToF updated database vs non-identified colonies subjected to 16S sequencing (A). Oxygen tolerance of identified bacterial species grouped by donor (B). Number of unique and shared bacterial species divided by donor (C).

5.8.2 *IBD-related species (targeted culturomics)*

For *E. coli*, a total of 343 colonies were obtained from CD patients (n= 2) using traditional culture methods in LB and McConkey media. From them, 40 colonies were selected for obtaining pure isolates and further subjected to species identification (Figure 36). 97.5% of these isolates were identified as "*E. coli*" when using the updated version of the MALDI-Biotyper database. Hence, just one required further 16S identification. *E. coli* concentration of IBD patients corresponded to 2,03e+05 and 1,40e+05 CFU/g of feces respectively. Comparison of isolated vs shotgun sequencing results was not possible since is still undergoing. On the contrary, no growth was obtained when targeting *F. prausnitzii* using the selected healthy donor (n= 1).

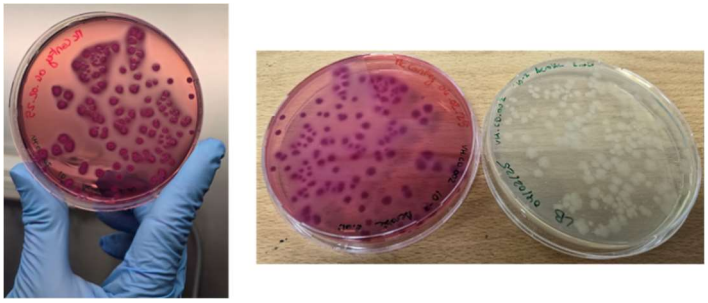


Figure 36. E.coli colonies from CD patients isolated using McConkey and LB, Miller agar plate

6 DISCUSSION

6.1 The POP Study: A relevant contribution to characterize the association between diet and microbiome

This study provides new insight into how personal traits and lifestyle factors, including diet and DQIs, can shape the human gut microbiome. It underscores the importance of adhering to both national and international guidelines to support a healthy lifestyle and, consequently, a healthier microbiome.

To explore these relationships, we conducted a longitudinal study involving 1,017 intestinally healthy volunteers from different regions of Spain. Participants provided stool samples and completed dietary questionnaires at baseline, six and twelve months. To our knowledge, this is one of the largest studies in Spain investigating diet and microbiome using shotgun metagenomic sequences, enabling analysis at the functional and species levels.

Previous research, such as the study of *Latorre-Pérez et al*. with 530 volunteers, used 16S rRNA sequencing, which lacked resolution at the species level and omitted functional data analysis (99).

Due to its Mediterranean location, Spain has traditionally followed the MedDiet, characterized by high intake of fruits, legumes, whole grains, vegetables, and nuts, as along with healthy fats from olive oil, frequent consumption of fish, moderate dairy products and fermented beverages, and low intake of red or processed meats (274).

However, adherence to the MedDiet has declined over time in Spain, with a gradual shift towards more WD patterns (275–277). In our cohort, MedDiet adherence was assessed using the aMED index, yielding a median value of 4.0 out of 9.0, confirming a significant divergence from traditional Mediterranean dietary habits.

Some prospective studies have demonstrated potential causal links between specific dietary components and non-communicable diseases (278–280). More precisely, the 2017 GBD Study emphasized the impact of 15 dietary risk factors across 195 countries, estimating their contribution to global mortality and morbidity (190).

In our Spanish cohort, only three out of twelve GBD dietary targets were met: vegetables (321.48 g/day), fruits (225.6 g/day) and fiber (27.32 g/day). Notably, vegetables and fruits were the two most consumed food groups (25-28% and 18-19% of daily intake, respectively) in our population. However, while vegetable consumption approached Spanish recommendations (300 g/day), fruit intake fell short of the 400 g/day target (281).

At the microbiome level, vegetables, especially raw and boiled leafy greens, were associated with increase α-diversity, in line with previous large cohort studies (28,205).

Similarly, fruits showed positive associations with Shannon α-diversity and was linked to an increased abundance of *Lachnospira eligens*, a butyrate-producing capable of fermenting plant pectin and stimulaing anti-inflammatory IL-10 production (282–284). This species has been related with lower weight and waist circumference (285), higher DQIs (286) and decreased abundance in several disease states (287–289).

Non-alcoholic beverages, particularly coffee, ranked third in median intake (202.03 g/day). Coffee contains polyphenols and alkaloids such as caffeine (290) that may significantly impact human gut microbiome. It was associated with ten bacterial species, most notably *L. asaccharolyticus* (formerly *Clostridium phoceensis*), and had one of the highest prediction values for dietary intake. This confirms earlier findings by *Asnicar et al.,* and *Manghi et al.,* (188,205). Additionally, *L. asaccharolyticus* has been linked with 12 coffee biomarkers including quinic acid, a major polyphenol metabolite derived from chlorogenic acid (291,292). Machine learning analysis identified coffee (caffeinated and decaffeinated) as the most predictive dietary item for microbiome composition, with an AUC > 0.8, validating further the findings by *Manghi et al*. However, the effect of coffee on the microbiome and overall health remain unclear due to mixed findings in the literature (112,186,206–208,293).

Comparison with previous Spanish studies such as ANIBES and ENALIA2 reveals similar dominant food group but with different rankings. For instance, non-alcoholic beverages were the top food group in both studies, followed by milk and dairy, vegetables, and fruits (287, 307). This shift may indicate a growing awareness of healthy eating, although full restoration of the traditional MedDiet remains a challenge. On the other side, the decline in dairy intake may be partly due to the rising popularity of plant-based beverages, now included in the "non-alcoholic beverages" category, which were less prevalent when the ANIBES and ENALIA2 studies were conducted (2013-2015).

Using individuals sFFQ items instead of broader food groups, distinct clustering patterns emerged. One cluster, associated with healthier dietary habits, included raw leafy vegetables, fresh fruit, nuts and seeds.

Nuts and seed were associated with increased α-diversity and presence of *Roserburia hominis*, consistent with previous findings (226,295) possibly due to their high content of fiber and polyphenols. Parallely, their intake correlated negatively with *F. plautii*, a flavonoid-degrading species linked to poor dietary quality, lower DQIs (99,140) and various disease outcomes such as colorectal cancer, IBD, depression and bipolar disorder. Although the underlying mechanisms remain unclear and further research is needed, it has been suggested that *F. plautii* reduce the availability of beneficial dietary flavonoids (197, 306–308).

Machine learning models identified nuts and seeds as the second most predictable dietary variable based on microbiome data (AUC= 0.76), Similarly, dark chocolate (> 70% cacao), rich in polyphenols, was associated with higher microbial diversity, species level associations were limited due to taxonomy uncertainty. Its effect on health has been extensively discussed suggesting dark chocolate intake as an effective way of reducing appetite, plasma tryglicerol levels and improving mood (237–239,254).

In contrast, a second cluster composed of less healthy, Western-style foods (confectionery, white bread and grains, sunflower oil, pastries, soft drinks, and processed foods) was associated with decreased bacterial α-diversity and negative shifts in microbiome composition. These dietary patterns have been linked in animal models to increased inflammation and risk of chronic diseases such as CKD and bone disorders (299–302).

White grain consumption was negatively correlated with *Mediterraneibacter butyricigenes*, a novel butyrate producer (303), while soft drinks positively correlated with *F. plautii*, reinforcing its association with poor adherence to the MedDiet and IBD-type microbiome alteration (99,188).

Most participants met the dietary reference intakes for macro- and micronutrients, except for CHO, calcium, vitamin D and iodine. In case of fiber, just half of the population met the recommended values. These results are in line with previous studies (98,268,304) although iodine in our data may be underestimated due to the no accounted use of iodized salt in cooking process.

Fibers play a crucial role in microbiome modulation, serving as a substrate for colonic bacteria (212). While median fiber intake met GBD targets (27.32 g/day), only 53-58% of the participants achieved this level. Suboptimal intake was associated with a decrease in α-diversity and an enrichment in *R. torques*, a mucin degrading bacterium that can compromise the gut barrier, facilitating pathogen invasion and contributing to conditions like IBD (305–308). Increase fiber intake in Spain should remain a public health priority, especially given its established role in reducing disease risk.

Vitamin D, primarily obtained through sun exposure, is paradoxically low in Spain despite ample sunlight. Vitamin D deficiency rises concern, since it is estimated that one third of the population could be at risk of its deficiency (309). Dietary sources include oily fish such as salmon and some fortified products such as margarine or milk (310). Vitamin D has been implicated in maintaining gut barrier integrity, preventing pathogen invasion (114,310,311). In our cohort, unknown *Lachnospiraceae* showed a strong positive correlation with vitamin D intake, consistent with prior findings from a year-long randomized control trial (312). Moreover, higher vitamin D intake has been associated with lower depression risk in large studies (313).

General used of dietary supplements, seems to have raised drastically during last years. In 2017, the AECOSAN reported 13.3% of general population as habitual consumers of vitamins and mineral supplements (262). Nevertheless, COVID-19 lockdown significally increased this proportion up to 21.3% (314) which seems to be gradually rising even nowadays. This fact raise concern mainly due to lack of scientific evidence regarding effectiveness of some products, despite there is a generalized idea of their good effect on health beyond their use in treating deficiencies or concrete health conditions (315)

Functional analysis identified positive associations between the L-arginine biosynthesis II and sucrose biosynthesis II pathways and the consumption of fruits, nuts and seeds, and fiber. L-arginine, a precursor for nitric oxide, homoarginine and agmatine synthesis, plays multiple physiological roles (316). Agmatine, in particular, is produced and released by colonic bacteria and may inhibit colonocyte proliferation and promote microbial diversity (317–320).

The main bacterial contributors to this pathway, were one unclassified taxon and *F. prausnitzii,* a key butyrate producer that support intestinal homeostasis, suppress inflammation, and is depleted in various diseases including IBD, IBS, type 2 diabetes, and cancer (321–327). Butyrate helps keeping the anaerobic environment in the colon by enhancing colonocyte oxygen consumption and stabilizing hypoxia-inducible factor, while its absence facilitates the buildup of potentially harmful bacteria and molecules, such as *Salmonella*, *E. coli*, and nitric oxide (NO), respectively (328).

F. prausnitzii was also identified as the second most dominant species involved in sucrose biosynthesis. Enhancement of this pathway could imply an increase in SCFAs production, although the underlying mechanisms remain unclear (329).

Another noteworthy correlation was found between item 50 of our sFFQ (coffee) and CMP-3-deoxy-D-manno-octulosonate biosynthesis pathway. The acidic sugar 3-deoxy-α-D-manno-2-octulosonate (CMP-KDO for ketodeoxyoctonate) is a key component of bacterial lipopolysaccharides (LPS). LPS are the primary constituents of the outer membrane of gram-negative bacteria, essential for maintaining the structural integrity and cell viability (330). As the previous cases, the most prevalent species remained undetermined, followed by *B. uniformis,* which has been associated with improved metabolic disfunction in murine models (331–334) and UC (335).

DQIs are commonly used to summarize diet quality and facilitate cross-country dietary comparison (98,174,177,178,182,188,336). Indices reflecting healthier diets (HEI-2015, hPDI and aMED) correlated positively with Shannon diversity, whereas uPDI_adj, which emphasizes unhealthy plant-based sources like refined grains, desserts and sugary drinks, showed a negative correlation.

Among the eight DQIs calculated, hPDI_Adj and aMED presented the highest number of bacterial correlations (14 out of 27 species) and provided the greatest predictive power using microbiome data. Observational and interventional studies have highlighted the benefits of a MedDiet in reducing CVD risk (180,337,338), improving frailty in the elderly (339) and enhancing the production of SCFAs (198). Similarly, hPDI reflects the consumption of healthy plant-based foods, while reducing animal protein intake, partially overlapping with MedDiet principles. Higher adherence to plant-based diets has been inversely associated with metabolic syndrome (177,178,182). Despite minor differences, both DQIs displayed the strongest negative correlations with *R. torques* and *F. plautii* and positive correlation with *H. parainfluenzae* or *Clostridium saccharogumia*, suggesting their potential as microbial indicators of a healthy diet. These species also showed the same tendency with typical elements of the MedDiet, including fiber, nuts, and vitamin C.

Importantly, microbiome profiling enables the stratification of healthy individuals based on their similarity to an IBD-associated dysbiotic profile. Consistent with previous findings, typical MedDiet food groups not only increased α-diversity but also appeared protective against an IBD-like microbiota, potentially by reducing "detrimental" species such as *F. plautii* and *R. gnavus*. However, further research is required to develop a comprehensive "disease score" that incorporates additional non-communicable diseases and integrates lifestyle, diet, and microbiome diversity (340,341).

6.2 Fungi microbiome and dietary patterns

Fungal microbiome populations represent a minor cell fraction of the microbiota. For example, the fungal-to-bacterial cell ratio in stool samples ranges between 10^{-9} and 10^{-4} (93). While fungal communities have traditionally been studied via ITS amplicon region, the high variability in gene copy number across species and also strains, renders it suboptimal for accurate profiling (342,343). Fungal genes comprises less than 0.08% of the total gut metagenome, making sequencing costly and technically challenging (94,344). To partially address this limitation, we employed a cost-effective enrichment protocol that successfully increased the fungal representation in shotgun sequencing and was coupled with our optimized FunOmic pipeline (93,94).

This approach succeeded in identifiying the most prevalent fungal species *Saccharomyces cerevisiae* (84.2%) and *Malassezia restricta* (34.37%), described as part of the core mycobiome previously (15,210). However, the time-intensive nature of this procedure may limit its application in large-scale population studies, requiering further optimization.

Unlike bacteria, fungal diversity showed few correlations with dietary variables, the directionality mirrored bacterial trends. Few studies have explored diet-fungi relationship, making cross-study comparisons difficult. *Shuai et al.,* reported higher fungal diversity associated with fruit intake (210) while *Sun et al.,* reported that vegetables and meat influence the gut mycobiome, albeit without specifying the direction of the effect (3). These findings suggest that healthy dietary patterns may also positively influence the fungal communities.

At the species level, *Aspergillus penicilloides* positively correlated with confectionary, foods rich in sugars such as candies and nougat. More broadly, *Aspergillus* species have been linked to total sugars and CHO intake and negatively correlated with SCFA levels (345,346).

Botryosphaeria dothidea presented the strongest negative association with red meat as a common opportunistic plant pathogen, its presence may indicate a higher intake of plant-based foods relative to animal products (347).

Finally, we hypothesized that the limited number of observed associations is partly due to the small sample size of enriched samples (n= 100), coupled with the patchy distribution of fungal species across them. Therefore, these results should be interpreted with caution.

6.3 Influence of personal traits on the Microbiome

In addition to diet, other personal characteristics significantly impact the human gut microbiome. **Aging** was associated with higher consumption of traditionally healthy food such as nuts and seeds, whole grains, fruits and fruit products. These findings were supported by two DQIs -IASE and hPDI_Adj -suggesting a diet richer in plant-based foods and lower in animal products. This aligns with two previous Spanish studies (ANIBES and *Latorre-Pérez et al*.)(99,294), that found increased vegetables, fruits, and oils intake among older individuals. Furthermore, an age-dependent increase in alpha diversity was also observed, consistent with findings from Finland, Japan, and the UK (90,188,209). However, aging does not always correlate with better diet or microbiome diversity. *Claesson et al.,* (91), compared individuals residing in long-term care facilities with community-dwelling volunteers and observed a clear clustering of microbiome profiles based on residence type. Vegetables, fruits, and meat consumption emerged as the most discriminant dietary factors. Notably, a reduction in microbial diversity among long-term care residents was associated with increased frailty measures, underscoring the critical role of diet in promoting healthy aging.

Aging was also linked to an increase in *A. muchiniphila*, known to improve glucose metabolism and metabolic health (131,151,197) though its levels are reduced in CD patients (153). Despite healthier diets, older individuals still exhibited lower levels of *Dysosmobacter welbionis*, potential due to age itself.

Dysosmobacter welbionis, a newly described butyrate producer, may offer protection against obesity (348).

Dietary patterns also varied by **gender**, with women showing better DQIs (349,350), although microbial diversity (α and β), remained largely similar between genders, with only modest compositional variations (Table 22). At the taxonomic level, *Prevotella* and *Veillonellaceae*, were found more abundant in men (100,101), possibly due to hormonal influences (102,104,105).

Transit time, estimated via stool frequency, support previous reports showing a positive association between longer transit time and α-diversity (140,142,143), possibly because it allows greater bacterial access to substrates, promoting microbial growth and diversity (140,144). Lower stool frequency (1.5 times per week) have been previously associated with increased risk of mortality (145–147) and decrease in *Blautia wexlerae*, who may play an important role against obesity and food addiction (176,351). Contrary, regular transit times (once per day or >3 times per week) promoted beneficial microorganisms like *A. muchiniphila* (140,151).

Smoking, a well-known health risk factor diseases (352–355), correlated with reduced α-diversity, poorer dietary quality (136,356), and increased alcohol consumption (357,358). Higher **BMI**, was associated with lower DQIs, greater consumption of bread and ready-to-eat meals, decreased bacterial richness and evenness (98), and higher prevalence of *R. torques*, previously linked to obesity (98,115,188).

Finally, **regional dietary habits** within Spain also influenced food choices. Partially in line with previous research claiming variations in adherence to MedDiet among Mediterranean countries (359), we demonstrated regional differences in diet and dietary patterns after classifying Spain into four regions. Specifically, the Interior region adhered more closely to the MedDiet, particularly through higher legume intake. Legumes are nutrient-dense and confer multiple health benefits. However, these regional dietary differences did not translate into significant microbiome variations.

6.4 Contributory citizens science project

Interest in diet-microbiome research is rapidly increasing. Yet, findings are often inaccessible to participants due to technical jargon and high costs. For instance, participants in the the AGP (260). Project ashad to pay around $150-200, potentially biasing samples toward wealthier, more educated populations.

The POP Study incorporated citizen science through four strategies:

First, we provided accessible explanations of microbiome science and the study's aims to inform participants.

Second, all participants received a personalized dietary report, comparing their intake to national recommendations and the other participants. Some also received simplified microbiome data with explanations, accessible via web portal or PDF format.

Third, we commited to disseminating results via open access publication (see ANNEX 14) and a free seminar to Promote public engagement and dialogue.

Fourth, Participation incurred no cost, enabling broader inclusion across socioeconomic backgrounds.

6.5 Culturomics

Non-targeted culturomics, aimed at isolating the highest number of microorganisms, has been proposed as an effective approach for obtaining viable bacterial species from human gut samples, an essential step for advancing studies on bacteria-host interactions (54,60,360,361).

In non-targeted culturomics, commercially available media are often favored due to greater reproducibility, standardization, broad applicability, and reduced preparation time and risk of contamination (362–364). Among these, GAM and BHI are commonly used for culturing anaerobic bacterial and have yielded promising results (58,60,363,365,366). In our pilot study, we selected these two media to cultivate and isolate bacteria from two frozen samples. GAM outperformed supplemented BHI, also know as L-YHBHI.4 + RF, yelding 25 (21 unique to GAM and 4 shared) out of the 27 total isolated species and achieving higher growth output (5.55% GAM vs. 3.55% L-YHBHI.4 + RF). A previous study by *Tao et al*. also found that GAM supported higher bacterial density than BHI when comparing multiple media for *ex vivo* culture of gut samples (367). Conversely, *Ito et al.* observed similar performance between the two media (362). However, due to the limited number of comparative studies between GAM and BHI, definitive conclusions remain elusive.

The number of different bacterial species isolated in the pilot study was slightly lower than those reported in earlier culturomics studies (58,60,363,364). This discrepancy may be attributable to the sample collection method. Whereas most prior studies used fresh simples (55,58,60,360,363,364,368) we employed frozen aliquots, which may negatively affect cell viability and diversity (369,370). Despite fresh samples being more optimal, they are logistically challenging to collect, especially for large-scale studies.

Nevertheless, our pilot study successfully isolated 27 different bacterial species, including several of notable relevance to health and disease. For instance, *B. animalis*, *B. adolescentis,* and *B. longum* have been suggested to be involved in SCFAs production, immune system modulation, and improved digestion (371,372). Their depletion have been proposed as biomarker for condition diseases such as IBD (308,373).

Another species of interest, *Bacteroides uniformis*, another butyrate producer, have been linked to positive outcomes in stroke recovery (206), improved metabolic function in mouse models (331–333) and proposed to have a potential therapeutic effect in UC (335). In contrary, *C. aerofaciens* have shown to decrease following Mediterranean diet intervention (339) and may play a role in obesity (374) and liver disease (375).

Other isolates such as *Parabacteroides merdae* have recently gained attention. Although some studies associate it with unhealthy diets (172,363), the evidence remains inconclusive and often contradictory (365,376), underlying the need for viable species isolation to facilitate deeper functional analysis.

Targeted culturomics focuses on isolating specific bacterial species of interest. In our pilot study, *E. coli* was successfully isolated for investigation in the context of IBD, whereas *F. prausnitzii* was not (72,73,269,323). These species are known to play critical roles in IBD pathogenesis. Studies have highlighted significant differences in *E. coli* populations between IBD patients and healthy controls, particularly the adherent-invasive *E. coli* (AIEC), which possess enhanced virulence factors, colonization rates (median 3.1 x 10^5 CFU/g of tissue in CD vs 4.8 x 10^4 CFU/g of tissue in healthy controls) and the ability to invade epithelial cells (377–379). Isolation of viable species is essential for understanding disease mechanisms and identifying thrapeutic candidates in *ex vivo* and *in vitro* models. The failure to isolate *F. prausnitzii,* an extremely oxygen-sensitive species, could be due to prolongated oxygen exposure during sample collection, potentially compromising its viability (380).

6.6 Strenghts and Limitations

6.6.1 *Strenghts*

First, at the time the sFFQ was developed, few sFFQs were freely available to assess the habitual intake of food groups and nutrients potentially relevant to the human gut microbiome in the general population. The sFFQ used here not only addresses this gap but also includes questions regarding personal traits and characteristics, offering a broader understanding of how multiple factors influence gut microbiota. Additionally, the short completion time (around 20 min), is a key advantage compared to other sFFQs used in population studies, likely encouraging better compliance and continued participation at follow-up timepoints.

Second, despite a limited sample size for metagenomic sequencing (n= 500 baseline), the observed associations between dietary components (e.g. vegetables, nuts, coffee, fruits) and microbiome - as well as with personal traïts (age, transit time) - were partially consistent with findings from larger cohorts (90,99,140,184, 28, 205,206), thereby supporting the validility of the results.

Third, the POP cohort study represents one of the few longitudinal microbiome studies with a relatively large number of participants. This design enables future analyses exploring seasonal effects on diet-microbiome interactions. Moreover, the calculation of several standarized DQIs allows comparisons with other international studies, despite the use of different dietary assesment methods across countries.

Fourth, microbiome data were obtained via shotgun metagenomic sequencing, a technique that yelds higher-resolution information compared to traditional 16S rRNA sequencing. This approach enabled taxonomic profiling down to the species level and facilitated funcional pathway analysis. Importantly, it was applied to study both bacterial and fungal microbiome communities, providing a more comprehensive view of interkingdom interactions and their relation with diet.

Fifth, one of the key features of the project was the creation of a dedicated website to raise awareness about the impact of diet and microbiome. This initiative aimed to enhance participant engagement and promote open, accessible science. It also facilitated smoother data collection and strengthened the relationships between the public and research.

Finally, the study went a step further by isolating a collection of gut bacteria using culturomics, a novel cultivation-based technique. The isolation of viable microbes opens the door to testing causal relationships through *in vivo* and *ex-vivo* models in future studies.

6.6.2 *Limitations*

Despite these strengths, the study has several limitations. One major limitation is the sample size and regional distribution. Although an effort was made to recruit a representative sampling fraction across Spanish regions, the Mediterranean area was overrepresented, which may bias results. To mitigate this, region areas were included as a covariate in the statistical models. Moreover, while the percentage of healthcare workers in the general Spanish population is estimated to be approximately 2.77%, we observed that in our cohort this percentage rose to 30%. This may suggest a higher level of awareness and motivation within this sector regarding participation in scientific research projects. However, such an overrepresentation could introduce some selection bias, potentially affecting the generalizability of the findings. The implementation of targeted recruitment strategies aimed at reaching underrepresented professional sectors—such as through community-based recruitment channels—could help improve the representativeness of the general population in future studies.

Additionally, fungal analyses were conducted with a randomly small subset (n= 100), which may explain the limited associations observed with dietary variables and interkingdom relationships. This highlights the need to improve fungal DNA enrichment protocols and bioinformatic analysis in the context of large-scale studies.

Second, sFFQ relies mainly on self-reported dietary intake, which is subjected to recall bias and misreporting. This may result in over- or underestimation of food and nutrient intakes. Notably, some nutrient estimates appeared inconsistent and should be interpreted cautiously, in line with findings from our previous study (98). Complementing dietary assessments with objective biomarkers such as urinary metabolites could be a valuable strategy, although this field is still evolving.

Third, the nutritional composition tables used may be insufficient for modern dietary analysis. They lack data on additives, cooking methods, preservatives and other food processing variables, all of which are increasingly recognized as influencial on health and microbiota composition (381).

Fourth, while some results align partially with those from other observational studies, it is important to emphasize that correlation does not imply causation. Experimental validation is needed to confirm these associations. We have begun this process by isolating viable bacteria, which are planned to be tested in explant tissue models in our laboratory.

Fifth, frozen samples collected for culturomics analysis were not suitable enough for viable species isolation. Thus, a considerable number of Bacillota was lost during the preservation and transport to INRAE facility. In the future, isolation will be carried out using fresh samples instead.

Lastly, limited funding constrained the sequencing of all collected samples at various timepoints. However, we plan to complete these analyses once sufficient financial suport becomes available.

7 CONCLUSIONS

CONCLUSIONS

Despite the limitations, the results present in this doctoral thesis offer valuable insights in how personal traits, lifestyle choices and diet exert an effect on gut microbiome, leading to the subsequent conclusions:

1. Promoting high diverse and high-quality healthy diets, rich in vegetables, fruits and fiber-rich products (legumes, whole grains and nuts and seeds) and reducing intake of less healthy food choices like white grains and bread, soft drinks or processed food could lead to an increase in bacteria diversity and a less IBD-type microbiome. Accompanied by increase of bacterial species often related with health, such as *A. muciniphila*, and detriment of species associated with poor heath outcomes such as *F. plautii* or *R. torques*.

2. Other personal traits, geography and lifestyle choices also impact gut microbiome. Among them, smoking, BMI, living in the Mediterranean region and extreme fast bowel frequencies were correlated with less bacterial diversity and, in some cases, with poorer diets. Highlighting the importance of taking all these factors into consideration when performing large-scale microbiome analysis.

3. Contributory citizen science constitutes a valid approach for engaging the general population in scientific research. This approach not only generated robust scientific data comparable to traditional clinical studies but also made possible to recruit a large cohort that would be logistically impossible through conventional research methods alone.

4. Culturomics comprise a fast and feasible approach for viable species isolation further study host-microbiome interactions, although further optimization is needed.

Conclusions

Personal traits

Gender

- ↑ *EQIs in women (except uPDI and MPDI that ↓)*
- ↑ *Ruminococcus, Allisonella and Prevotella, in men*

Age

- ↑ *bacterial and fungal α-diversity*
- ↑ *EQIs (except uPDI that ↓)*
- ↑ *Akkermansia_muciniphila in healthy aging*
- ↓ *Bifidobacterium bifidum, Dysmobacter welbionis, F.plautii*
- ↓ *IBD-type microbiome*

BMI

- ↓ *bacterial α-diversity*
- ↓ *aMED, hPDI and PDI*
- ↓ *Akkermansia muciniphila, Intestinimonas gabonensis, Intestinimonas massiliensis*
- ↑ *R. gnavus, R.torques, Coprococcus comes, Blautia welerae, Phocaeicola vulgatus*
- ↑ *IBD-type microbiome*

Transit time

- ↑ *bacterial α-diversity*
- ↑ *Akkermansia mucinphila, Intestinimonas massiliensis*
- ↓ *Blautia wexlerae*
- *CHO → Saccharolytic:Proteolytic*

Diet

White grains & bread

- ↓ *bacterial and fungal α-diversity*
- ↓ *Clostridium saccharogumia, Eubacterium eligens, Blautia stercolis*
- *Food group with highest number of associations*

Coffee

- *Highest prediction value using microbiome data*
- ↑ *Clostridium phoceensis, Massilioclostridium_coli, Dysosmobacter welbionis*
- *CMP-3-deoxy-D-manno-octulosonate biosynthesis*

Nuts and seeds

- ↑ *bacterial α-diversity*
- ↑ *Clostridium saccharogumia, Roseburia hominis*
- ↓ *F. plautii, R. torques*
- *L-arginine and sucrose biosynthesis*
- ↓ *IBD-type microbiome*

Fruits & fruit products

- ↑ *bacterial α-diversity*
- ↑ *Eubacterium eligens,*
- ↓ *F.plautii, R. torques*
- *L-arginine and sucrose biosynthesis*
- ↓ *IBD-type microbiome*

Fiber

- ↓ *R. torques*
- ↑ *H. Parainfluenzae*
- *L-arginine and sucrose biosynthesis*

Vegetables

- ↑ *bacterial and fungal α-diversity*
- ↑ *Lacnospiraceae, H.parainfluenzae*
- ↓ *Fl.plautii, R.gnavus, R.torques*
- *L-arginine and sucrose biosynthesis*
- ↓ *IBD-type microbiome*

EQIs

aMED

- ↑ *bacterial α-diversity*
- *EQI with ↑ number of correlation with bacteria*
- ↓ *F. plautii, R.torques, B.malissilensis*
- ↑ *Clostridium saccharogumia, H.parainfluenzae, A. mucinphila*
- ↑ *EQI prediction value with microbiome data*

hPDI

- ↑ *bacterial α-diversity*
- *EQI with ↑ number of correlation with bacteria*
- ↑ *Clostridium saccharogumia, Intestinimonas gabonensis*
- ↓ *R.torques, B. massiliensis, F. plautii,*

uPDI

- ↓ *bacterial and fungal α-diversity*
- ↓ *Lachnospiraceae, Blautia glucerasea*
- ↑ *Prevotellamassilia_timonensis*

HEI-2015

- ↑ *bacterial α-diversity*
- ↑ *F. prausnitzii, Blautia glucerasea*
- ↓ *IBD-type microbiome*

8 FUTURE PERSPECTIVES

FUTURE PERSPECTIVES

Our study offers new insights into the complex relationship between diet, personal traits and the human gut microbiome. However, validation of our findings in an independent cohort will be essential to assess the robustness of the associations observed. Additionally, intervention studies in animal models and/or human subjects or ex vivo experiments may be required to demonstrate causality. Bacterial isolates obtained through culturomics will be investigated using explant tissue models, and parallel studies in rats are also planned.

Longitudinal designs provide the opportunity to track the participants over time, which is particularly helpful for examining how dietary changes and seasonal variation influence bacterial and fungal microbiomes and their potential role in disease development. Due to budgetary constraints, we were only able to assess dietary changes after six and twelve months, while microbiome sequencing was not feasible at those timepoints. Nevertheless, fecal samples were collected and stored, enabling future analyses when funding permits.

As previously mentioned, fungal assesment methods require further optimization. Although our experimental enrichment protocol improved fungal recovery, its labor-intensive nature highlights the need for more efficient alternatives. One promising direction involves enhancing the bioinformatic pipeline -- an effort already underway by our bioinformatics team.

Finally, greater efforts are needed to make scientific reserach accessible and understandable to the general públic. Science relies on volunteers, and society relies on science. Promoting clearer and more comprehensive dissemination of findings can improve public perception and engagement. In addition to the development of our website, we aim to organize a free, open-access seminar where participants can ask questions and discuss current knowledge on the microbiome-diet connection.

9 BIBLIOGRAPHY

BIBLIOGRAPHY

1. Neish AS. Microbes in Gastrointestinal Health and Disease. Gastroenterology [Internet]. 2009 [cited 2023 Feb 2];136(1):65. Available from: /pmc/articles/PMC2892787/

2. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science (80-) [Internet]. 2005 Mar 25 [cited 2023 Feb 2];307(5717):1915–20. Available from: https://www.science.org/doi/10.1126/science.1104816

3. Sun Y, Zuo T, Cheung CP, Gu W, Wan Y, Zhang F, et al. Population-Level Configurations of Gut Mycobiome Across 6 Ethnicities in Urban and Rural China. Gastroenterology [Internet]. 2021;160(1):272-286.e11. Available from: https://doi.org/10.1053/j.gastro.2020.09.014

4. Cao Z, Sugimura N, Burgermeister E, Ebert MP, Zuo T, Lan P. The gut virome: A new microbiome component in health and disease. eBioMedicine [Internet]. 2022 Jul 1 [cited 2023 Feb 17];81. Available from: http://www.thelancet.com/article/S2352396422002948/fulltext

5. Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J [Internet]. 2017 Jun 1 [cited 2023 Feb 2];474(11):1823–36. Available from: /biochemj/article/474/11/1823/49429/Introduction-to-the-human-gut-microbiota

6. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14(8):1–14.

7. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med [Internet]. 2016;8(1):1–11. Available from: http://dx.doi.org/10.1186/s13073-016-0307-y

8. Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, et al. Towards the human intestinal microbiota phylogenetic core. Environ Microbiol [Internet]. 2009 Oct 1 [cited 2023 Feb 17];11(10):2574–84. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1462-2920.2009.01982.x

9. Geleijnse JM, Vermeer C, Grobbee DE, Schurgers LJ, Knapen MHJ, Van Der Meer IM, et al. Dietary Intake of Menaquinone Is Associated with a Reduced Risk of Coronary Heart Disease: The Rotterdam Study. J Nutr. 2004 Nov 1;134(11):3100–5.

10. Simes DC, Viegas CSB, Araújo N, Marreiros C. Vitamin K as a Diet Supplement with Impact in Human Health: Current Evidence in Age-Related Diseases. Nutr 2020, Vol 12, Page 138 [Internet]. 2020 Jan 3 [cited 2023 Feb 17];12(1):138. Available from: https://www.mdpi.com/2072-6643/12/1/138/htm

11. Kho ZY, Lal SK. The human gut microbiome - A potential controller of wellness and disease. Front Microbiol. 2018 Aug 14;9(AUG):1835.

12. Wilson AS, Koller KR, Ramaboli MC, Nesengani LT, Ocvirk S, Chen C, et al. HHS Public Access. 2021;65(3):723–40.

13. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: Roles of resistant starch and nonstarch polysaccharides. Physiol Rev [Internet]. 2001 [cited 2023 Feb 2];81(3):1031–64. Available from: https://journals.physiology.org/doi/10.1152/physrev.2001.81.3.1031

14. Putignani L, Del Chierico F, Petrucca A, Vernocchi P, Dallapiccola B. The human gut microbiota: a dynamic interplay with the host from birth to senescence settled during childhood. Pediatr Res 2014 761 [Internet]. 2014 Apr 14 [cited 2023 Feb 2];76(1):2–10. Available from: https://www.nature.com/articles/pr201449

15. Zhang F, Aschenbrenner D, Yoo JY, Zuo T. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. The Lancet Microbe [Internet]. 2022 Dec 1 [cited 2023 Feb 17];3(12):e969–83. Available from: http://www.thelancet.com/article/S2666524722002038/fulltext

16. Luo Y, Li J, Zhou H, Yu B, He J, Wu A, et al. The Nutritional Significance of Intestinal Fungi: Alteration of Dietary Carbohydrate Composition Triggers Colonic Fungal Community Shifts in a Pig Model. Appl Environ Microbiol. 2021;87(10):1–19.

17. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, et al. Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. PLoS One. 2013;8(6).

18. Wu X, Xia Y, He F, Zhu C, Ren W. Intestinal mycobiota in health and diseases: from a disrupted equilibrium to clinical opportunities. Microbiome 2021 91 [Internet]. 2021 Mar 14 [cited 2023 Feb 17];9(1):1–18. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-021-01024-x

19. Belvoncikova P, Splichalova P, Videnska P, Gardlik R. The Human Mycobiome: Colonization, Composition and the Role in Health and Disease. J Fungi. 2022;8(10).

20. Liang G, Bushman FD. The human virome: assembly, composition and host interactions. Nat Rev Microbiol 2021 198 [Internet]. 2021 Mar 30 [cited 2023 Feb 17];19(8):514–27. Available from: https://www.nature.com/articles/s41579-021-00536-5

21. Taylor VL, Fitzpatrick AD, Islam Z, Maxwell KL. The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence. Adv Virus Res. 2019;103:1–31.

22. Namsolleck P, Thiel R, Lawson P, Holmstrøm K, Rajilic M, Vaughan EE, et al. Molecular methods for the analysis of gut microbiota. http://dx.doi.org/101080/08910600410032367 [Internet]. 2009 Jul [cited 2023 Feb 2];16(2–3):71–85. Available from: https://www.tandfonline.com/doi/abs/10.1080/08910600410032367

23. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci Reports 2021 111 [Internet]. 2021 Feb 4 [cited 2023 Feb 2];11(1):1–10. Available from: https://www.nature.com/articles/s41598-021-82726-y

24. Sánchez-Campillo M, Pastor-Fajardo MT, Sabater-Molina M, López-Andreo MJ, Larqué E. Critical Steps for Human Gut Exfoliome RNA Profiling Analysis Using Non-Invasive Stool Samples. Ann Nutr Metab [Internet]. 2022 Apr 1 [cited 2023 Feb 2];78(2):80–90. Available from: https://pubmed.ncbi.nlm.nih.gov/34788767/

25. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev Microbiol [Internet]. 2018;16(7):410–22. Available from: http://dx.doi.org/10.1038/s41579-018-0029-9

26. Partula V, Mondot S, Torres MJ, Kesse-Guyot E, Deschasaux M, Assmann K, et al. Associations between usual diet and gut microbiota composition: Results from the Milieu Intérieur cross-sectional study. Am J Clin Nutr. 2019;109(5):1472–83.

27. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. An Open Platform for Citizen Science Microbiome Research. mSystems [Internet]. 2018;3(3):1–28. Available from: http://humanfoodproject.com/

28. Walker RL, Vlamakis H, Lee JWJ, Besse LA, Xanthakis V, Vasan RS, et al. Population study of the gut microbiome: associations with diet, lifestyle, and cardiometabolic disease. Genome Med [Internet]. 2021 Dec 1 [cited 2023 Feb 3];13(1):1–16. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-01007-5

29. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457–63.

30. Zhang Y, Chen H, Lu M, Cai J, Lu B, Luo C, et al. Habitual Diet Pattern Associations with Gut Microbiome Diversity and Composition: Results from a Chinese Adult Cohort. Nutrients. 2022;14(13).

31. Delavy M, Sertour N, Patin E, Le Chatelier E, Cole N, Dubois F, et al. Unveiling *Candida albicans*

intestinal carriage in healthy volunteers: the role of micro- and mycobiota, diet, host genetics and immune response. Gut Microbes [Internet]. 2023;15(2). Available from: https://doi.org/10.1080/19490976.2023.2287618

32. Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, et al. The gut mycobiome of the Human Microbiome Project healthy cohort. Microbiome [Internet]. 2017 Nov 25 [cited 2024 Feb 21];5(1):153. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0373-4

33. Yang RH, Su JH, Shang JJ, Wu YY, Li Y, Bao DP, et al. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. PLoS One [Internet]. 2018 Oct 1 [cited 2024 Feb 21];13(10):e0206428. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206428

34. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35(9):833–44.

35. Liu YX, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell [Internet]. 2021 May 1 [cited 2024 Jul 8];12(5):315–30. Available from: https://dx.doi.org/10.1007/s13238-020-00724-8

36. Gao B, Chi L, Zhu Y, Shi X, Tu P, Li B, et al. An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. Biomol 2021, Vol 11, Page 530 [Internet]. 2021 Apr 2 [cited 2024 Jul 8];11(4):530. Available from: https://www.mdpi.com/2218-273X/11/4/530/htm

37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods [Internet]. 2012 Apr [cited 2024 Jul 8];9(4):357. Available from: /pmc/articles/PMC3322381/

38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics [Internet]. 2014 Aug 8 [cited 2024 Jul 8];30(15):2114. Available from: /pmc/articles/PMC4103590/

39. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. bioRxiv [Internet]. 2022 Aug 22 [cited 2024 Jul 8];2022.08.22.504593. Available from: https://www.biorxiv.org/content/10.1101/2022.08.22.504593v1

40. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol [Internet]. 2019 Nov 28 [cited 2024 Jul 8];20(1):1–13. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1891-0

41. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bobakery 3. Elife. 2021;10:1–42.

42. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol. 2016;12(6):1–12.

43. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6.

44. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 2017;27(5):824–34.

45. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38(12):1–15.

46. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.

47. Patro R, Duggal G, I Love M, A Irizarry R, Kingsford C. Salmon: fast and bias-aware quantification of

transcript expression using dual-phase inference. Nat Methods. 2017;14(4):417–9.

48. Lagkouvardos I, Overmann J, Clavel T. Cultured microbes represent a substantial fraction of the human and mouse gut microbiota. Gut Microbes [Internet]. 2017 Sep 3 [cited 2024 Feb 15];8(5):493. Available from: /pmc/articles/PMC5628658/

49. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nat 2019 5687753 [Internet]. 2019 Feb 11 [cited 2024 Feb 15];568(7753):499–504. Available from: https://www.nature.com/articles/s41586-019-0965-1

50. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. BMC Biol [Internet]. 2019 Jun 12 [cited 2024 Feb 15];17(1):1–4. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0667-z

51. Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18:1185–93.

52. Cheema SK, Li R, Cameron SJS. Culturomics as a tool to better understand the human milk microbiota and host–microbiota interactions. Microbiota Host. 2023;1(1).

53. Hamad I, Ranque S, Azhar EI, Yasir M, Jiman-Fatani AA, Tissot-Dupont H, et al. Culturomics and Amplicon-based Metagenomic Approaches for the Study of Fungal Population in Human Gut Microbiota. Sci Reports 2017 71 [Internet]. 2017 Dec 1 [cited 2024 Feb 15];7(1):1–8. Available from: https://www.nature.com/articles/s41598-017-17132-4

54. Lagier J-C, Khelaifia S, Tidjani Alou M, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol [Internet]. 2016 [cited 2024 Feb 15]; Available from: www.nature.com/naturemicrobiology

55. Khelaifia S, Lagier JC, Nkamga VD, Guilhot E, Drancourt M, Raoult D. Aerobic culture of methanogenic archaea without an external source of hydrogen. Eur J Clin Microbiol Infect Dis. 2016;35(6):985–91.

56. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature [Internet]. 2016 May 4 [cited 2024 Feb 19];533(7604):543. Available from: /pmc/articles/PMC4890681/

57. Walker AW, Duncan SH, Louis P, Flint HJ. Phylogeny, culturing, and metagenomics of the human gut microbiota. Trends Microbiol [Internet]. 2014;22(5):267–74. Available from: http://dx.doi.org/10.1016/j.tim.2014.03.001

58. Huang Y, Sheth RU, Zhao S, Cohen LA, Dabaghi K, Moody T, et al. High-throughput microbial culturomics using automation and machine learning. Nat Biotechnol 2023 4110 [Internet]. 2023 Feb 20 [cited 2024 Feb 19];41(10):1424–33. Available from: https://www.nature.com/articles/s41587-023-01674-2

59. Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev [Internet]. 2015 Jan 1 [cited 2024 Feb 19];28(1):237–64. Available from: /pmc/articles/PMC4284300/

60. Afrizal A, Hitch TCA, Viehof A, Treichel N, Riedel T, Abt B, et al. Anaerobic single-cell dispensing facilitates the cultivation of human gut bacteria. Environ Microbiol. 2022;24(9):3861–81.

61. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. Int J Syst Evol Microbiol [Internet]. 2002 Nov 1 [cited 2024 Feb 19];52(Pt 6):2141–6. Available from: https://pubmed.ncbi.nlm.nih.gov/12508881/

62. Derrien M, Vaughan EE, Plugge CM, de Vos WM. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol [Internet]. 2004 Sep [cited 2024 Feb 19];54(Pt 5):1469–76. Available from: https://pubmed.ncbi.nlm.nih.gov/15388697/

63. Wan X, Yang Q, Wang X, Bai Y, Liu Z. Isolation and Cultivation of Human Gut Microorganisms: A Review. Microorg 2023 [Internet]. 2023 Apr 20 [cited 2024 Feb 19];11(4):1080. Available from: https://www.mdpi.com/2076-2607/11/4/1080/htm

64. Athanasopoulou K, Adamopoulos PG, Scorilas A. Unveiling the Human Gastrointestinal Tract Microbiome: The Past, Present, and Future of Metagenomics. Biomedicines. 2023 Mar 1;11(3).

65. Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, et al. Culturing the human microbiota and culturomics. Nat Rev Microbiol [Internet]. 2018 [cited 2024 Feb 19];16. Available from: www.nature.com/nrmicro

66. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res 2020 306 [Internet]. 2020 May 20 [cited 2023 Feb 2];30(6):492–506. Available from: https://www.nature.com/articles/s41422-020-0332-7

67. Xu F, Fu Y, Sun TY, Jiang Z, Miao Z, Shuai M, et al. The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for complex human diseases. Microbiome [Internet]. 2020 Oct 8 [cited 2023 Feb 2];8(1):1–14. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00923-9

68. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun 2019 101 [Internet]. 2019 Oct 3 [cited 2023 Feb 2];10(1):1–10. Available from: https://www.nature.com/articles/s41467-019-12476-z

69. Wilkins LJ, Monga M, Miller AW. Defining Dysbiosis for a Cluster of Chronic Diseases. Sci Reports 2019 91 [Internet]. 2019 Sep 9 [cited 2023 Feb 2];9(1):1–10. Available from: https://www.nature.com/articles/s41598-019-49452-y

70. Patangia D V., Anthony Ryan C, Dempsey E, Paul Ross R, Stanton C. Impact of antibiotics on the human microbiome and consequences for host health. Microbiology [Internet]. 2022 Feb 1 [cited 2023 Feb 2];11(1). Available from: /pmc/articles/PMC8756738/

71. De Filippo C, Di Paola M, Ramazzotti M, Albanese D, Pieraccini G, Banci E, et al. Diet, environments, and gut microbiota. A preliminary investigation in children living in rural and Urban Burkina Faso and Italy. Front Microbiol. 2017 Oct 13;8(OCT):1979.

72. Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. Gut [Internet]. 2017 May 1 [cited 2023 Feb 2];66(5):813–22. Available from: https://gut.bmj.com/content/66/5/813

73. Amos GCA, Sergaki C, Logan A, Iriarte R, Bannaga A, Chandrapalan S, et al. Exploring how microbiome signatures change across inflammatory bowel disease conditions and disease locations. Sci Rep [Internet]. 2021 Sep 21 [cited 2023 Feb 2];11(1):1–9. Available from: https://www.nature.com/articles/s41598-021-96942-z

74. Zuo T, Lu XJ, Zhang Y, Cheung CP, Lam S, Zhang F, et al. Gut mucosal virome alterations in ulcerative colitis. Gut [Internet]. 2019 Jul 1 [cited 2023 Feb 17];68(7):1169–79. Available from: https://gut.bmj.com/content/68/7/1169

75. Li X V., Leonardi I, Putzel GG, Semon A, Fiers WD, Kusakabe T, et al. Immune regulation by fungal strain diversity in inflammatory bowel disease. Nature. 2022;608(7922):E21.

76. Konturek PC, Koziel J, Dieterich W, Haziri D, Wirtz S, Glowczyk I, et al. Successful therapy of Clostridium difficile infection with fecal microbiota transplantation. J Physiol Pharmacol. 2016;67(6):859–66.

77. Freire R, Ingano L, Serena G, Cetinbas M, Anselmo A, Sapone A, et al. Human gut derived-organoids provide model to study gluten response and effects of microbiota-derived molecules in celiac disease. [cited 2023 Feb 3]; Available from: https://doi.org/10.1038/s41598-019-43426-w

78. Lindfors K, Lin J, Lee HS, Hyöty H, Nykter M, Kurppa K, et al. Metagenomics of the faecal virome

indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study. Gut [Internet]. 2020 Aug 1 [cited 2023 Feb 17];69(8):1416–22. Available from: https://gut.bmj.com/content/69/8/1416

79. Yuan X, Wang R, Han B, Sun CJ, Chen R, Wei H, et al. Functional and metabolic alterations of gut microbiota in children with new-onset type 1 diabetes. Nat Commun 2022 131 [Internet]. 2022 Oct 26 [cited 2023 Feb 3];13(1):1–16. Available from: https://www.nature.com/articles/s41467-022-33656-4

80. Liu BN, Liu XT, Liang ZH, Wang JH. Gut microbiota in obesity. World J Gastroenterol [Internet]. 2021 Jul 7 [cited 2023 Feb 3];27(25):3837. Available from: /pmc/articles/PMC8291023/

81. Yang K, Niu J, Zuo T, Sun Y, Xu Z, Tang W, et al. Alterations in the Gut Virome in Obesity and Type 2 Diabetes Mellitus. Gastroenterology [Internet]. 2021 Oct 1 [cited 2023 Feb 17];161(4):1257-1269.e13. Available from: http://www.gastrojournal.org/article/S001650852103170X/fulltext

82. Jiang S, Xie S, Lv D, Wang P, He H, Zhang T, et al. Alteration of the gut microbiota in Chinese population with chronic kidney disease. Sci Rep [Internet]. 2017 Dec 1 [cited 2023 Feb 3];7(1). Available from: /pmc/articles/PMC5460291/

83. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555(7695):210–5.

84. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe [Internet]. 2016;19(5):731–43. Available from: http://dx.doi.org/10.1016/j.chom.2016.04.017

85. Zhernakova D V., Wang D, Liu L, Andreu-Sánchez S, Zhang Y, Ruiz-Moreno AJ, et al. Host genetic regulation of human gut microbial structural variation. Nature. 2024;625(7996):813–21.

86. Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sánchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 2022;54(2):143–51.

87. Qin Y, Havulinna AS, Liu Y, Jousilahti P, Ritchie SC, Tokolyi A, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat Genet. 2022;54(2):134–42.

88. Kurilshikov A, Medina-gomez C, Bacigalupe R. Large-scale association analyses identify host factors influencing human gut microbiome composition Authors. bioRxiv. 2020;1–40.

89. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms. 2019;7(1).

90. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao JZ, et al. Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. BMC Microbiol [Internet]. 2016 May 25 [cited 2024 Apr 18];16(1):1–12. Available from: https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-016-0708-5

91. Claesson MJ, Jeffery IB, Conde S, Power SE, O'connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012;488(7410):178–84.

92. Gutierrez MW, Mercer EM, Moossavi S, Laforest-Lapointe I, Reyna ME, Becker AB, et al. Maturational patterns of the infant gut mycobiome are associated with early-life body mass index. Cell Reports Med [Internet]. 2023;4(2):100928. Available from: https://doi.org/10.1016/j.xcrm.2023.100928

93. Xie Z, Canalda-Baltrons A, d'Enfert C, Manichanh C. Shotgun metagenomics reveals interkingdom association between intestinal bacteria and fungi involving competition for nutrients. Microbiome [Internet]. 2023;11(1):1–19. Available from: https://doi.org/10.1186/s40168-023-01693-w

94. Xie Z, Manichanh C. FunOMIC: Pipeline with built-in fungal taxonomic and functional databases for human mycobiome profiling. Comput Struct Biotechnol J [Internet]. 2022;20:3685–94. Available from: https://doi.org/10.1016/j.csbj.2022.07.010

95. Kim YS. Sex and gender-related issues in biomedical science. Sci Ed. 2018;5(1):66–9.

96. Gao X, Zhang M, Xue J, Huang J, Zhuang R, Zhou X, et al. Body mass index differences in the gut microbiota are gender specific. Front Microbiol. 2018;9(JUN):1–10.

97. Sinha T, Vich Vila A, Garmaeva S, Jankipersadsing SA, Imhann F, Collij V, et al. Analysis of 1135 gut metagenomes identifies sex-specific resistome profiles. Gut Microbes. 2019;10(3):358–66.

98. Yáñez F, Soler Z, Oliero M, Xie Z, Oyarzun I, Serrano-Gómez G, et al. Integrating dietary data into microbiome studies: A step forward for nutri-metaomics. Nutrients. 2021;13(9).

99. Latorre-Pérez A, Hernández M, Iglesias JR, Morán J, Pascual J, Porcar M, et al. The Spanish gut microbiome reveals links between microorganisms and Mediterranean diet. Sci Rep [Internet]. 2021;11(1):1–12. Available from: https://doi.org/10.1038/s41598-021-01002-1

100. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. PLoS One. 2016;11(5):1–16.

101. Takagi T, Naito Y, Inoue R, Kashiwagi S, Uchiyama K, Mizushima K, et al. Differences in gut microbiota associated with age, sex, and stool consistency in healthy Japanese subjects. J Gastroenterol [Internet]. 2019;54(1):53–63. Available from: https://doi.org/10.1007/s00535-018-1488-5

102. Flores R, Shi J, Fuhrman B, Xu X, Veenstra TD, Gail MH, et al. Fecal microbial determinants of fecal and systemic estrogens and estrogen metabolites: A cross-sectional study. J Transl Med. 2012;10(1):1–11.

103. Gomez A, Luckey D, Taneja V. The gut microbiome in autoimmunity: Sex matters. Clin Immunol [Internet]. 2014;159(2):154–62. Available from: http://dx.doi.org/10.1016/j.clim.2015.04.016

104. Liu R, Zhang C, Shi Y, Zhang F, Li L, Wang X, et al. Dysbiosis of gut microbiota associated with clinical parameters in polycystic ovary syndrome. Front Microbiol. 2017;8(FEB):1–12.

105. Shin JH, Park YH, Sim M, Kim SA, Joung H, Shin DM. Serum level of sex steroid hormone is associated with diversity and profiles of human gut microbiome. Res Microbiol [Internet]. 2019;170(4–5):192–201. Available from: https://doi.org/10.1016/j.resmic.2019.03.003

106. Dahiya D, Nigam PS. Antibiotic-Therapy-Induced Gut Dysbiosis Affecting Gut Microbiota—Brain Axis and Cognition: Restoration by Intake of Probiotics and Synbiotics. Int J Mol Sci. 2023;24(4).

107. Panda S, El Khader I, Casellas F, López Vivancos J, García Cors M, Santiago A, et al. Short-term effect of antibiotics on human gut microbiota. PLoS One. 2014;9(4).

108. Bich VTN, Le NG, Barnett D, Chan J, van Best N, Tien TD, et al. Moderate and transient impact of antibiotic use on the gut microbiota in a rural Vietnamese cohort. Sci Rep [Internet]. 2022;12(1):1–12. Available from: https://doi.org/10.1038/s41598-022-24488-9

109. Rashid MU, Zaura E, Buijs MJ, Keijser BJF, Crielaard W, Nord CE, et al. Determining the long-term effect of antibiotic administration on the human normal intestinal microbiota using culture and pyrosequencing methods. Clin Infect Dis. 2015;60(Suppl 2):S77–84.

110. Anthony WE, Wang B, Sukhum K V., D'Souza AW, Hink T, Cass C, et al. Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. Cell Rep [Internet]. 2022;39(2):110649. Available from: https://doi.org/10.1016/j.celrep.2022.110649

111. John D, Michael D, Dabcheva M, Hulme E, Illanes J, Webberley T, et al. placebo-controlled study assessing the impact of probiotic supplementation on antibiotic induced changes in the gut microbiome. 2024;(March):1–11.

112. Stege PB, Hordijk J, Shetty SA, Visser M, Viveen MC, Rogers MRC, et al. Impact of long-term dietary habits on the human gut resistome in the Dutch population. Sci Reports 2022 121 [Internet]. 2022 Feb 3 [cited 2024 Apr 3];12(1):1–13. Available from: https://www.nature.com/articles/s41598-022-05817-4

113. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature. 2018;555(7698):623–8.

114. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nat Commun [Internet]. 2020;11(1):1–12. Available from: http://dx.doi.org/10.1038/s41467-019-14177-z

115. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732–9.

116. Rosenberg E, Zilber-Rosenberg I. Reconstitution and Transmission of Gut Microbiomes and Their Genes between Generations. Microorganisms [Internet]. 2022 Jan 1 [cited 2024 Jul 22];10(1). Available from: /pmc/articles/PMC8780831/

117. Dill-McFarland KA, Tang ZZ, Kemis JH, Kerby RL, Chen G, Palloni A, et al. Close social relationships correlate with human gut microbiota composition. Sci Rep [Internet]. 2019 Dec 1 [cited 2024 Jul 22];9(1). Available from: /pmc/articles/PMC6345772/

118. Brito IL, Gurry T, Zhao S, Huang K, Young SK, Shea TP, et al. Transmission of human-associated microbiota along family and social networks. Nat Microbiol 2019 46 [Internet]. 2019 Mar 25 [cited 2024 Jul 22];4(6):964–71. Available from: https://www.nature.com/articles/s41564-019-0409-6

119. Valles-Colomer M, Blanco-Míguez A, Manghi P, Asnicar F, Dubois L, Golzato D, et al. The person-to-person transmission landscape of the gut and oral microbiomes. Nature. 2023;614(7946):125–35.

120. Antunes P, Novais C, Peixe L. Food-to-Humans Bacterial Transmission. Microbiol Spectr [Internet]. 2020 Jan 16 [cited 2024 Jul 22];8(1). Available from: https://journals.asm.org/doi/10.1128/microbiolspec.mtbp-0019-2016

121. Older CE, Diesel A, Patterson AP, Meason-Smith C, Johnson TJ, Mansell J, et al. The feline skin microbiota: The bacteria inhabiting the skin of healthy and allergic cats. PLoS One [Internet]. 2017 Jun 1 [cited 2024 Jul 22];12(6):e0178555. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178555

122. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, et al. Cohabiting family members share microbiota with one another and with their dogs. Elife [Internet]. 2013 Apr 16 [cited 2024 Jul 22];2(2):458. Available from: /pmc/articles/PMC3628085/

123. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. 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 [Internet]. 2010 Aug 17 [cited 2024 Jul 23];107(33):14691–6. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1005963107

124. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. Nat Commun 2014 51 [Internet]. 2014 Apr 15 [cited 2024 Jul 23];5(1):1–12. Available from: https://www.nature.com/articles/ncomms4654

125. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature [Internet]. 2012 May 9 [cited 2024 Jul 23];486(7402):222–7. Available from: https://www.nature.com/articles/nature11053

126. Lin A, Bik EM, Costello EK, Dethlefsen L, Haque R, Relman DA, et al. Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. PLoS One [Internet]. 2013 Jan 29 [cited 2024 Jul 23];8(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23349750/

127. Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, et al. Subsistence strategies in traditional societies distinguish gut microbiomes. Nat Commun [Internet]. 2015 [cited 2024 Jul 25];6. Available from: /pmc/articles/PMC4386023/

128. Brooks CN, Wight ME, Azeez OE, Bleich RM, Zwetsloot KA. Growing old together: What we know about the influence of diet and exercise on the aging host's gut microbiome. Front Sport Act Living [Internet]. 2023;5(April):1–9. Available from: https://doi.org/10.3389/fspor.2023.1168731

129. Si J, Vázquez-Castellanos JF, Gregory AC, Decommer L, Rymenans L, Proost S, et al. Long-term life history predicts current gut microbiome in a population-based cohort study. Nat Aging. 2022;2(10):885–95.

130. Taniguchi H, Tanisawa K, Sun X, Kubo T, Hoshino Y, Hosokawa M, et al. Effects of short-term endurance exercise on gut microbiota in elderly men.pdf. Physiol Rep. 2018;6(23).

131. Munukka E, Ahtiainen JP, Puigbó P, Jalkanen S, Pahkala K, Keskitalo A, et al. Six-week endurance exercise alters gut metagenome that is not reflected in systemic metabolism in over-weight women. Front Microbiol. 2018;9(OCT).

132. Smith KS, Morris MM, Morrow CD, Novak JR, Roberts MD, Frugé AD. Associations between Changes in Fat-Free Mass, Fecal Microbe Diversity, and Mood Disturbance in Young Adults after 10-Weeks of Resistance Training. Microorganisms. 2022;10(12).

133. Cronin O, Barton W, Skuse P, Penney NC, Garcia-Perez I, Murphy EF, et al. A Prospective Metagenomic and Metabolomic Analysis of the Impact of Exercise and/or Whey Protein Supplementation on the Gut Microbiome of Sedentary Adults. mSystems. 2018;3(3).

134. Lee SH, Yun Y, Kim SJ, Lee EJ, Chang Y, Ryu S, et al. Association between Cigarette Smoking Status and Composition of Gut Microbiota: Population-Based Cross-Sectional Study. J Clin Med [Internet]. 2018 Sep 14 [cited 2023 Sep 26];7(9):282. Available from: /pmc/articles/PMC6162563/

135. Prakash A, Peters BA, Cobbs E, Beggs D, Choi H, Li H, et al. Tobacco smoking and the fecal microbiome in a large, multi-ethnic cohort. Cancer Epidemiol Biomarkers Prev. 2021;30(7):1328–35.

136. Stewart CJ, Auchtung TA, Ajami NJ, Velasquez K, Smith DP, De La Garza R, et al. Effects of tobacco smoke and electronic cigarette vapor exposure on the oral and gut microbiota in humans: A pilot study. PeerJ. 2018;2018(4):1–16.

137. Lin R, Zhang Y, Chen L, Qi Y, He J, Hu M, et al. The effects of cigarettes and alcohol on intestinal microbiota in healthy men. J Microbiol [Internet]. 2020 Nov 1 [cited 2023 Sep 26];58(11):926–37. Available from: https://pubmed.ncbi.nlm.nih.gov/32997305/

138. Rogers MAM, Greene MT, Saint S, Chenoweth CE, Malani PN, Trivedi I, et al. Higher rates of clostridium difficile infection among smokers. PLoS One. 2012;7(7).

139. Procházková N, Falony G, Dragsted LO, Licht TR, Raes J, Roager HM. Advancing human gut microbiota research by considering gut transit time. Gut. 2023;72(1):180–91.

140. Asnicar F, Leeming ER, Dimidi E, Mazidi M, Franks PW, Al Khatib H, et al. Blue poo: Impact of gut transit time on the gut microbiome using a novel marker. Gut. 2021;70(9):1665–74.

141. Törnblom H, Van Oudenhove L, Sadik R, Abrahamsson H, Tack J, Simrén M. Colonic transit time and IBS symptoms: what's the link? Am J Gastroenterol. 2012;107(5):754–60.

142. Hadizadeh F, Walter S, Belheouane M, Bonfiglio F, Heinsen FA, Andreasson A, et al. Stool frequency is associated with gut microbiota composition. Gut [Internet]. 2017 Mar 1 [cited 2024 Jul 18];66(3):559–60. Available from: https://gut.bmj.com/content/66/3/559

143. Kwon HJ, Lim JH, Kang D, Lim S, Park SJ, Kim JH. Is stool frequency associated with the richness and community composition of gut microbiota? Intest Res. 2019;17(3):419–26.

144. Müller M, Hermes GDA, Canfora EE, Smidt H, Masclee AAM, Zoetendal EXG, et al. Distal colonic transit

is linked to gut microbiota diversity and microbial fermentation in humans with slow colonic transit. Am J Physiol - Gastrointest Liver Physiol. 2020;318(2):G361–9.

145. Ma W, Li Y, Heianza Y, Staller KD, Chan AT, Rimm EB, et al. Associations of Bowel Movement Frequency with Risk of Cardiovascular Disease and Mortality among US Women. Sci Rep [Internet]. 2016 Sep 6 [cited 2024 Jul 18];6. Available from: /pmc/articles/PMC5011651/

146. Kubota Y, Iso H, Tamakoshi A. Bowel Movement Frequency, Laxative Use, and Mortality From Coronary Heart Disease and Stroke Among Japanese Men and Women: The Japan Collaborative Cohort (JACC) Study. J Epidemiol [Internet]. 2016 [cited 2024 Jul 18];26(5):242. Available from: /pmc/articles/PMC4848322/

147. Peng X, Li J, Wu Y, Dai H, Lynn HS, Zhang X, et al. Association of Stool Frequency and Consistency with the Risk of All-Cause and Cause-Specific Mortality among U.S. Adults: Results from NHANES 2005–2010. Healthc 2023, Vol 11, Page 29 [Internet]. 2022 Dec 22 [cited 2024 Jul 18];11(1):29. Available from: https://www.mdpi.com/2227-9032/11/1/29/htm

148. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut. 2016;65(1):57–62.

149. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. Gut microbiota composition associated with stool consistency. Gut [Internet]. 2016 Mar 1 [cited 2024 Jul 18];65(3):540–2. Available from: https://gut.bmj.com/content/65/3/540

150. Geerlings SY, Kostopoulos I, de Vos WM, Belzer C. *Akkermansia muciniphila* in the Human Gastrointestinal Tract: When, Where, and How? Microorg 2018, Vol 6, Page 75 [Internet]. 2018 Jul 23 [cited 2024 Jul 18];6(3):75. Available from: https://www.mdpi.com/2076-2607/6/3/75/htm

151. Dao MC, Everard A, Aron-Wisnewsky J, Sokolovska N, Prifti E, Verger EO, et al. *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: Relationship with gut microbiome richness and ecology. Gut. 2016;65(3):426–36.

152. Earley H, Lennon G, Balfe Á, Coffey JC, Winter DC, O'Connell PR. The abundance of *Akkermansia muciniphila* and its relationship with sulphated colonic mucins in health and ulcerative colitis. Sci Reports 2019 91 [Internet]. 2019 Oct 30 [cited 2024 Jul 18];9(1):1–9. Available from: https://www.nature.com/articles/s41598-019-51878-3

153. James SL, Christophersen CT, Bird AR, Conlon MA, Rosella O, Gibson PR, et al. Abnormal fibre usage in UC in remission. Gut. 2015;64(4):562–70.

154. Wang L, Tang L, Feng Y, Zhao S, Han M, Zhang C, et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurised bacterium blunts colitis associated tumourigenesis by modulation of CD8+ T cells in mice. Gut [Internet]. 2020 Nov 1 [cited 2024 Jul 18];69(11):1988–97. Available from: https://gut.bmj.com/content/69/11/1988

155. Picó C, Serra F, Rodríguez AM, Keijer J, Palou A. Biomarkers of Nutrition and Health: New Tools for New Approaches. Nutrients [Internet]. 2019 May 1 [cited 2023 Mar 24];11(5). Available from: /pmc/articles/PMC6567133/

156. Guasch-Ferre M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. Clin Chem [Internet]. 2018 Jan 1 [cited 2023 Mar 24];64(1):82–98. Available from: https://academic.oup.com/clinchem/article/64/1/82/5608703

157. Li K, Brennan L, McNulty BA, Bloomfield JF, Duff DJ, Devlin NFC, et al. Plasma fatty acid patterns reflect dietary habits and metabolic health: A cross-sectional study. Mol Nutr Food Res [Internet]. 2016 Sep 1 [cited 2023 Mar 24];60(9):2043–52. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mnfr.201500711

158. Wittenbecher C, Mühlenbruch K, Kröger J, Jacobs S, Kuxhaus O, Floegel A, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. Am

J Clin Nutr [Internet]. 2015 Jun 1 [cited 2023 Mar 24];101(6):1241–50. Available from: https://academic.oup.com/ajcn/article/101/6/1241/4564514

159. Gibbons H, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, et al. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake,. Am J Clin Nutr. 2015 Mar 1;101(3):471–7.

160. Stratakis N, Siskos AP, Papadopoulou E, Nguyen AN, Zhao Y, Margetaki K, et al. Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health. Elife. 2022;11:1–20.

161. Castellano-Escuder P, González-Domínguez R, Vaillant MF, Casas-Agustench P, Hidalgo-Liberona N, Estanyol-Torres N, et al. Assessing Adherence to Healthy Dietary Habits Through the Urinary Food Metabolome: Results From a European Two-Center Study. Front Nutr. 2022;9(June):1–11.

162. Ortega RM, Pérez-Rodrigo C, López-Sobaler AM. Dietary assessment methods: dietary records. Nutr Hosp. 2015;31:38–45.

163. Bailey RL. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. Curr Opin Biotechnol [Internet]. 2021;70:91–6. Available from: https://doi.org/10.1016/j.copbio.2021.02.007

164. Rupasinghe WS, Perera H, Wickramaratne N. A comprehensive review on dietary assessment methods in epidemiological research. J Public Heal Nutr [Internet]. 2020;3(2):204–11. Available from: https://www.alliedacademies.org/articles/%0Ahttps://www.alliedacademies.org/abstract/a-comprehensive-review-on-dietary-assessment-methods-in-epidemiological-research-12048.html

165. Fairulnizal M, Gunasegavan RDN, Khalid NM, Balasubramaniam V, Mustar S, Rashed AA. Recent techniques in nutrient analysis for food composition database. Molecules. 2020;25(19).

166. Kapsokefalou M, Roe M, Turrini A, Costa HS, Martinez-Victoria E, Marletta L, et al. Food composition at present: New challenges. Nutrients. 2019;11(8).

167. Lupiañez-Barbero A, González Blanco C, de Leiva Hidalgo A. Spanish food composition tables and databases: need for a gold standard for healthcare professionals. Endocrinol Diabetes y Nutr. 2018;65(6):361–73.

168. Gora M, Cámara M, Calderón V. A comparative analysis of the food composition tables and databases included in the EuroFIR network. Rev del Com científico AESAN. 2022;(34):103–27.

169. Bowyer RCE, Jackson MA, Pallister T, Skinner J, Spector TD, Welch AA, et al. Use of dietary indices to control for diet in human gut microbiota studies. Microbiome [Internet]. 2018 Apr 25 [cited 2023 Feb 8];6(1):77. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0455-y

170. Krebs-Smith S, Pannucci T, Subar A, Kirkpatrick S, Lerman J, Tooze J, et al. Update of the Healthy Eating Index: HEI-2015. J Acad Nutr Diet. 2018;118(9):1591–602.

171. Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB, Basiotis PP. Development and Evaluation of the Healthy Eating Index- 2005: Technical Report. Center for Nutrition Policy and Promotion, U. 2007;2600(202). Available from: http://www.cnpp.usda.gov/HealthyEatingIndex.htm.

172. Bagheri M, Shah RD, Mosley JD, Ferguson JF. A metabolome and microbiome wide association study of healthy eating index points to the mechanisms linking dietary pattern and metabolic status. Eur J Nutr [Internet]. 2021;60(8):4413–27. Available from: https://doi.org/10.1007/s00394-021-02599-9

173. Norte Navarro AI, Ortiz Moncada R. Calidad de la dieta española según el índice de alimentación saludable. Nutr Hosp. 2011;26(2):330–6.

174. Drescher LS, Thiele S, Mensink GBM. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. J Nutr. 2007;137(3):647–51.

175. Hatløy A, Torheim LE, Oshaug A. Food variety - A good indicator of nutritional adequacy of the diet? A case study from an urban area in Mali, West Africa. Eur J Clin Nutr. 1998;52(12):891–8.

176. Hosomi K, Saito M, Park J, Murakami H, Shibata N, Ando M, et al. Oral administration of *Blautia wexlerae* ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. Nat Commun. 2022;13(1):1–17.

177. Kim H, Lee K, Rebholz CM, Kim J. Plant-based diets and incident metabolic syndrome: Results from a South Korean prospective cohort study. PLoS Med [Internet]. 2020;17(11):1–18. Available from: http://dx.doi.org/10.1371/journal.pmed.1003371

178. Satija A, Bhupathiraju SN, Spiegelman D, Chiuve SE, Manson JAE, Willett W, et al. Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults. J Am Coll Cardiol. 2017;70(4):411–22.

179. Song S, Lee K, Park S, Shin N, Kim H, Kim J. Association between unhealthful plant-based diets and possible risk of dyslipidemia. Nutrients. 2021;13(12):1–12.

180. Fung TT, McCullough ML, Newby P, Manson JE, Meigs JB, Rifai N, et al. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. Am J Clin Nutr [Internet]. 2005;82(1):163–73. Available from: https://doi.org/10.1093/ajcn/82.1.163

181. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. Bmj. 1995;311(7018):1457.

182. Chen B, Zeng J, Qin M, Xu W, Zhang Z, Li X, et al. The Association Between Plant-Based Diet Indices and Obesity and Metabolic Diseases in Chinese Adults: Longitudinal Analyses From the China Health and Nutrition Survey. Front Nutr. 2022;9(June).

183. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A Short screener is valid for assessing mediterranean diet adherence among older spanish men and women. J Nutr. 2011;141(6):1140–5.

184. García-Conesa MT, Philippou E, Pafilas C, Massaro M, Quarta S, Andrade V, et al. Exploring the validity of the 14-item mediterranean diet adherence screener (Medas): A cross-national study in seven european countries around the mediterranean region. Nutrients. 2020;12(10):1–18.

185. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science (80-) [Internet]. 2016 Apr 29 [cited 2023 Feb 3];352(6285):560–4. Available from: https://www.science.org/doi/10.1126/science.aad3503

186. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science (80-) [Internet]. 2016 Apr 4 [cited 2023 Feb 3];352(6285):565. Available from: /pmc/articles/PMC5240844/

187. Zmora N, Suez J, Elinav E. You are what you eat: diet, health and the gut microbiota. Nat Rev Gastroenterol Hepatol [Internet]. 2019;16(1):35–56. Available from: http://dx.doi.org/10.1038/s41575-018-0061-2

188. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med [Internet]. 2021;27(2):321–32. Available from: http://dx.doi.org/10.1038/s41591-020-01183-8

189. Shen X, Tilves C, Kim H, Tanaka T, Spira AP, Chia CW, et al. Plant-based diets and the gut microbiome: findings from the Baltimore Longitudinal Study of Aging. Am J Clin Nutr [Internet]. 2024;119(3):628–38. Available from: https://doi.org/10.1016/j.ajcnut.2024.01.006

190. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2019;393(10184):1958–72.

191. Thiese MS. Observational and interventional study design types; an overview. Biochem Medica [Internet]. 2014 Jun 15 [cited 2023 Feb 3];24(2):199–210. Available from: https://www.biochemia-medica.com/en/journal/24/10.11613/BM.2014.022

192. Berry SE, Valdes AM, Drew DA, Asnicar F, Wolf J, Capdevila J, et al. Human Postprandial Responses to Food and Potential for Precision Nutrition. 2021;26(6):964–73.

193. Yu D, Nguyen SM, Yang Y, Xu W, Cai H, Wu J, et al. Long-term diet quality is associated with gut microbiome diversity and composition among urban Chinese adults. Am J Clin Nutr. 2021 Mar 1;113(3):684–94.

194. Statovci D, Aguilera M, MacSharry J, Melgar S. The impact of western diet and nutrients on the microbiota and immune response at mucosal interfaces. Front Immunol. 2017 Jul 28;8(JUL):838.

195. Shikany JM, Demmer RT, Johnson AJ, Fino NF, Meyer K, Ensrud KE, et al. Association of dietary patterns with the gut microbiota in older, community-dwelling men. Am J Clin Nutr. 2019;110(4):1003–14.

196. Lim MY, Hong S, Bang S-J, Chung W-H, Shin J-H, Kim J-H, et al. Gut Microbiome Structure and Association with Host Factors in a Korean Population. mSystems [Internet]. 2021 Aug 31 [cited 2023 Feb 3];6(4). Available from: https://journals.asm.org/doi/10.1128/mSystems.00179-21

197. Wu X, Unno T, Kang S, Park S. A Korean-Style Balanced Diet Has a Potential Connection with Ruminococcaceae Enterotype and Reduction of Metabolic Syndrome Incidence in Korean Adults. Nutr 2021, Vol 13, Page 495 [Internet]. 2021 Feb 3 [cited 2023 Feb 3];13(2):495. Available from: https://www.mdpi.com/2072-6643/13/2/495/htm

198. De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut [Internet]. 2016 Nov 1 [cited 2023 Feb 8];65(11):1812–21. Available from: https://gut.bmj.com/content/65/11/1812

199. Garcia-Mantrana I, Selma-Royo M, Alcantara C, Collado MC. Shifts on gut microbiota associated to mediterranean diet adherence and specific dietary intakes on general adult population. Front Microbiol. 2018 May 7;9(MAY):890.

200. Mitsou EK, Kakali A, Antonopoulou S, Mountzouris KC, Yannakoulia M, Panagiotakos DB, et al. Adherence to the Mediterranean diet is associated with the gut microbiota pattern and gastrointestinal characteristics in an adult population. Br J Nutr [Internet]. 2017 Jun 28 [cited 2023 Feb 8];117(12):1645–55. Available from: https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/adherence-to-the-mediterranean-diet-is-associated-with-the-gut-microbiota-pattern-and-gastrointestinal-characteristics-in-an-adult-population/B57ABE5976EBBFAD9E5734D81A46032F

201. Wang DD, Nguyen LH, Li Y, Yan Y, Ma W, Rinott E, et al. The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. Nat Med 2021 272 [Internet]. 2021 Feb 11 [cited 2023 Feb 8];27(2):333–43. Available from: https://www.nature.com/articles/s41591-020-01223-3

202. Gupta A, Dhakan DB, Maji A, Saxena R, P.K. VP, Mahajan S, et al. Association of *Flavonifractor plautii*, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. mSystems [Internet]. 2019 Dec 17 [cited 2023 Feb 8];4(6). Available from: https://pubmed.ncbi.nlm.nih.gov/31719139/

203. Naumovski N, Panagiotakos DB, D'Cunha NM. Untangling the 2-Way Relationship Between Red Wine Polyphenols and Gut Microbiota. Gastroenterology. 2020;158(1):48–51.

204. Le Roy CI, Wells PM, Si J, Raes J, Bell JT, Spector TD. Red Wine Consumption Associated With Increased Gut Microbiota α-Diversity in 3 Independent Cohorts. Gastroenterology [Internet]. 2020;158(1):270-272.e2. Available from: https://doi.org/10.1053/j.gastro.2019.08.024

205. Manghi P, Bhosle A, Wang K, Marconi R, Selma-Royo M, Ricci L, et al. Coffee consumption is associated with intestinal Lawsonibacter asaccharolyticus abundance and prevalence across multiple cohorts. Nat Microbiol. 2024;9(December).

206. Hammond TC, Powell E, Green SJ, Chlipala G, Frank J, Yackzan AT, et al. Functional recovery outcomes following acute stroke is associated with abundance of gut microbiota related to inflammation, butyrate and secondary bile acid. Front Rehabil Sci [Internet]. 2022 [cited 2024 Oct 1];3. Available from: /pmc/articles/PMC9644110/

207. Thirion F, Sellebjerg F, Fan Y, Lyu L, Hansen TH, Pons N, et al. The gut microbiota in multiple sclerosis varies with disease activity. Genome Med [Internet]. 2023 Dec 1 [cited 2024 Oct 1];15(1):1–17. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-022-01148-1

208. Kulecka M, Zeber-Lubecka N, Bałabas A, Czarnowski P, Bagińska K, Głowienka M, et al. Diarrheal-associated gut dysbiosis in cancer and inflammatory bowel disease patients is exacerbated by Clostridioides difficile infection. Front Cell Infect Microbiol. 2023 Jul 27;13:1190910.

209. Koponen KK, Salosensaari A, Ruuskanen MO, Havulinna AS, Männistö S, Jousilahti P, et al. Associations of healthy food choices with gut microbiota profiles. Am J Clin Nutr. 2021 Aug 1;114(2):605–16.

210. Shuai M, Fu Y, Zhong HL, Gou W, Jiang Z, Liang Y, et al. Mapping the human gut mycobiome in middle-aged and elderly adults: multiomics insights and implications for host metabolic health. Gut. 2022;71(9):1812–20.

211. Johnson AJ, Zheng JJ, Kang JW, Saboe A, Knights D, Zivkovic AM. A Guide to Diet-Microbiome Study Design. Front Nutr. 2020;7(June):1–16.

212. So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT, et al. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. Am J Clin Nutr. 2018 Jun 1;107(6):965–83.

213. Holscher HD, Gregory Caporaso J, Hooda S, Brulc JM, Fahey GC, Swanson KS. Fiber supplementation influences phylogenetic structure and functional capacity of the human intestinal microbiome: follow-up of a randomized controlled trial. Am J Clin Nutr. 2015 Jan 1;101(1):55–64.

214. Alfa MJ, Strang D, Tappia PS, Graham M, Van Domselaar G, Forbes JD, et al. A randomized trial to determine the impact of a digestion resistant starch composition on the gut microbiome in older and mid-age adults. Clin Nutr [Internet]. 2018 Jun 1 [cited 2023 Jul 6];37(3):797–807. Available from: http://www.clinicalnutritionjournal.com/article/S0261561417301164/fulltext

215. O'Callaghan A, van Sinderen D. Bifidobacteria and Their Role as Members of the Human Gut Microbiota. Front Microbiol [Internet]. 2016 [cited 2023 Jul 6];7(JUN):925. Available from: /pmc/articles/PMC4908950/

216. Clarke ST, Green-Johnson JM, Brooks SPJ, Ramdath DD, Bercik P, Avila C, et al. β2-1 Fructan supplementation alters host immune responses in a manner consistent with increased exposure to microbial components: results from a double-blinded, randomised, cross-over study in healthy adults. Br J Nutr [Internet]. 2016 May 28 [cited 2023 Jul 6];115(10):1748–59. Available from: https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/21-fructan-supplementation-alters-host-immune-responses-in-a-manner-consistent-with-increased-exposure-to-microbial-components-results-from-a-doubleblinded-randomised-crossover-study-in-healthy-adults/153AE7DD44FDF9F1AB86236C02C17E2B

217. Vanegas SM, Meydani M, Barnett JB, Goldin B, Kane A, Rasmussen H, et al. Substituting whole grains for refined grains in a 6-wk randomized trial has a modest effect on gut microbiota and immune and inflammatory markers of healthy adults, ,. Am J Clin Nutr. 2017 Mar 1;105(3):635–50.

218. Oliver A, Chase AB, Weihe C, Orchanian SB, Riedel SF, Hendrickson CL, et al. High-Fiber, Whole-Food Dietary Intervention Alters the Human Gut Microbiome but Not Fecal Short-Chain Fatty Acids. mSystems. 2021;6(2).

219. Vetrani C, Maukonen J, Bozzetto L, Della Pepa G, Vitale M, Costabile G, et al. Diets naturally rich in polyphenols and/or long-chain n-3 polyunsaturated fatty acids differently affect microbiota composition in high-cardiometabolic-risk individuals. Acta Diabetol [Internet]. 2020;57(7):853–60. Available from: https://doi.org/10.1007/s00592-020-01494-9

220. Creedon AC, Hung ES, Berry SE, Whelan K. Nuts and their effect on gut microbiota, gut function and symptoms in adults: A systematic review and meta-analysis of randomised controlled trials. Nutrients. 2020;12(8):1–21.

221. Holscher HD, Guetterman HM, Swanson KS, An R, Matthan NR, Lichtenstein AH, et al. Walnut consumption alters the gastrointestinal microbiota, microbially derived secondary bile acids, and health markers in healthy adults: A randomized controlled trial. J Nutr. 2018;148(6):861–7.

222. Bamberger C, Rossmeier A, Lechner K, Wu L, Waldmann E, Fischer S, et al. A Walnut-Enriched Diet Affects Gut Microbiome in Healthy Caucasian Subjects: A Randomized, Controlled Trial. Nutr 2018, Vol 10, Page 244 [Internet]. 2018 Feb 22 [cited 2023 Jul 7];10(2):244. Available from: https://www.mdpi.com/2072-6643/10/2/244/htm

223. Tindall AM, McLimans CJ, Petersen KS, Kris-Etherton PM, Lamendella R. Walnuts and Vegetable Oils Containing Oleic Acid Differentially Affect the Gut Microbiota and Associations with Cardiovascular Risk Factors: Follow-up of a Randomized, Controlled, Feeding Trial in Adults at Risk for Cardiovascular Disease. J Nutr [Internet]. 2020 Apr 1 [cited 2023 Jul 7];150(4):806. Available from: /pmc/articles/PMC7138683/

224. Liu Z, Lin X, Huang G, Zhang W, Rao P, Ni L. Prebiotic effects of almonds and almond skins on intestinal microbiota in healthy adult humans. Anaerobe. 2014;26(December):1–6.

225. Holscher HD, Taylor AM, Swanson KS, Novotny JA, Baer DJ. Almond consumption and processing affects the composition of the gastrointestinal microbiota of healthy adult men and women: A randomized controlled trial. Nutrients. 2018;10(2).

226. Dhillon J, Li Z, Ortiz RM. Almond Snacking for 8 wk Increases Alpha-Diversity of the Gastrointestinal Microbiome and Decreases Bacteroides fragilis Abundance Compared with an Isocaloric Snack in College Freshmen. Curr Dev Nutr. 2019 Aug 1;3(8):nzz079.

227. Rosas Jr. M, Liu C, O'Mealy C, Forsberg E, Hooshmand S, Kern M, et al. The Effects of Mixed Nut Consumption on the Microbiome in Healthy Adults. Curr Dev Nutr [Internet]. 2020 Jun 1 [cited 2023 Jul 7];4(Supplement_2):1580–1580. Available from: https://dx.doi.org/10.1093/cdn/nzaa062_037

228. Haskell-Ramsay CF, Dodd FL, Smith D, Cuthbertson L, Nelson A, Lodge JK, et al. Mixed Tree Nuts, Cognition, and Gut Microbiota: A 4-Week, Placebo-Controlled, Randomized Crossover Trial in Healthy Nonelderly Adults. J Nutr [Internet]. 2023 Jan 14 [cited 2023 Jul 7];152(12):2778–88. Available from: https://dx.doi.org/10.1093/jn/nxac228

229. Katz DL, Doughty K, Ali A. Cocoa and Chocolate in Human Health and Disease. Antioxid Redox Signal [Internet]. 2011 Nov 11 [cited 2024 Aug 8];15(10):2779. Available from: /pmc/articles/PMC4696435/

230. Hollenberg NK, Martinez G, McCullough M, Meinking T, Passan D, Preston M, et al. Aging, Acculturation, Salt Intake, and Hypertension in the Kuna of Panama. Hypertension [Internet]. 1997 Jan 1 [cited 2024 Aug 8];29(1):171–6. Available from: https://www.ahajournals.org/doi/10.1161/01.HYP.29.1.171

231. Munguía L, Gutiérrez-Salmeán G, Hernández M, Ortiz A, Sánchez ME, Nájera N, et al. Beneficial effects of a flavanol-enriched cacao beverage on anthropometric and cardiometabolic risk profile in overweight subjects. Rev Mex Cardiol [Internet]. 2015 [cited 2024 Aug 8];26(2):78–86. Available from: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-21982015000200004&lng=es&nrm=iso&tlng=en

232. García-Cordero J, Martinez A, Blanco-Valverde C, Pino A, Puertas-Martín V, San Román R, et al. Regular Consumption of Cocoa and Red Berries as a Strategy to Improve Cardiovascular Biomarkers

via Modulation of Microbiota Metabolism in Healthy Aging Adults. Nutrients [Internet]. 2023 May 1 [cited 2024 Aug 8];15(10):2299. Available from: https://www.mdpi.com/2072-6643/15/10/2299/htm

233. Lee Y, Berryman CE, West SG, Chen CYO, Blumberg JB, Lapsley KG, et al. Effects of dark chocolate and almonds on cardiovascular risk factors in overweight and obese individuals: A randomized controlled-feeding trial. J Am Heart Assoc [Internet]. 2017 Dec 1 [cited 2024 Aug 8];6(12). Available from: https://www.ahajournals.org/doi/10.1161/JAHA.116.005162

234. Mink PJ, Scrafford CG, Barraj LM, Harnack L, Hong CP, Nettleton JA, et al. Flavonoid intake and cardiovascular disease mortality: a prospective study in postmenopausal women. Am J Clin Nutr. 2007 Mar 1;85(3):895–909.

235. Sesso HD, Manson JE, Aragaki AK, Rist PM, Johnson LG, Friedenberg G, et al. Effect of cocoa flavanol supplementation for the prevention of cardiovascular disease events: the COcoa Supplement and Multivitamin Outcomes Study (COSMOS) randomized clinical trial. Am J Clin Nutr [Internet]. 2022 Jun 1 [cited 2023 Oct 5];115(6):1490. Available from: /pmc/articles/PMC9170467/

236. Leyva-Soto A, Chavez-Santoscoy RA, Lara-Jacobo LR, Chavez-Santoscoy AV, Gonzalez-Cobian LN. Daily Consumption of Chocolate Rich in Flavonoids Decreases Cellular Genotoxicity and Improves Biochemical Parameters of Lipid and Glucose Metabolism. Mol A J Synth Chem Nat Prod Chem [Internet]. 2018 Sep 1 [cited 2024 Aug 8];23(9). Available from: /pmc/articles/PMC6225491/

237. Tzounis X, Rodriguez-mateos A, Vulevic J, Gibson GR, Kwik-uribe C. Prebiotic evaluation of cocoa-derived flavanols in healthy humans by using a randomized , controlled , double-blind , crossover intervention study 1 – 3. 2011;(7):62–72.

238. Shin JH, Kim CS, Cha L, Kim S, Lee S, Chae S, et al. Consumption of 85% cocoa dark chocolate improves mood in association with gut microbial changes in healthy adults: a randomized controlled trial. J Nutr Biochem. 2022 Jan 1;99:108854.

239. Hernández-González T, González-Barrio R, Escobar C, Madrid JA, Periago MJ, Collado MC, et al. Timing of chocolate intake affects hunger, substrate oxidation, and microbiota: A randomized controlled trial. FASEB J [Internet]. 2021 Jul 1 [cited 2024 Aug 8];35(7):e21649. Available from: https://onlinelibrary.wiley.com/doi/full/10.1096/fj.202002770RR

240. Lee KW, Kim YJ, Lee HJ, Lee CY. Cocoa Has More Phenolic Phytochemicals and a Higher Antioxidant Capacity than Teas and Red Wine. J Agric Food Chem [Internet]. 2003 Dec 3 [cited 2024 Aug 9];51(25):7292–5. Available from: https://pubs.acs.org/sharingguidelines

241. Zhou Y, Zhang N, Arikawa AY, Chen C. Inhibitory Effects of Green Tea Polyphenols on Microbial Metabolism of Aromatic Amino Acids in Humans Revealed by Metabolomic Analysis. Metab 2019, Vol 9, Page 96 [Internet]. 2019 May 11 [cited 2023 Oct 5];9(5):96. Available from: https://www.mdpi.com/2218-1989/9/5/96/htm

242. Meng X, Sang S, Zhu N, Lu H, Sheng S, Lee MJ, et al. Identification and characterization of methylated and ring-fission metabolites of tea catechins formed in humans, mice, and rats. Chem Res Toxicol [Internet]. 2002 Aug [cited 2024 Aug 9];15(8):1042–50. Available from: https://pubs.acs.org/doi/full/10.1021/tx010184a

243. Dostal AM, Samavat H, Espejo L, Arikawa AY, Stendell-Hollis NR, Kurzer MS. Green Tea Extract and Catechol-O-Methyltransferase Genotype Modify Fasting Serum Insulin and Plasma Adiponectin Concentrations in a Randomized Controlled Trial of Overweight and Obese Postmenopausal Women. J Nutr. 2016 Jan 1;146(1):38–45.

244. Huang F, Zheng X, Ma X, Jiang R, Zhou W, Zhou S, et al. Theabrownin from Pu-erh tea attenuates hypercholesterolemia via modulation of gut microbiota and bile acid metabolism. Nat Commun [Internet]. 2019 Dec 1 [cited 2023 Oct 5];10(1). Available from: /pmc/articles/PMC6823360/

245. Yuan X, Long Y, Ji Z, Gao J, Fu T, Yan M, et al. Green Tea Liquid Consumption Alters the Human

Intestinal and Oral Microbiome. 2018;1800178:1–15.

246. Li A, Kou R, Liu H, Chen M, Wang J, Liu Q, et al. Multi-omics analyses reveal relationships among polyphenol-rich oolong tea consumption, gut microbiota, and metabolic profile: A pilot study. Food Chem [Internet]. 2023;426(April):136653. Available from: https://doi.org/10.1016/j.foodchem.2023.136653

247. Jin JS, Touyama M, Hisada T, Benno Y. Effects of green tea consumption on human fecal microbiota with special reference to Bifidobacterium species. Microbiol Immunol [Internet]. 2012 Nov 1 [cited 2024 Aug 9];56(11):729–39. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1348-0421.2012.00502.x

248. Farag MA, von Bergen M, Saleh BM, Homsi MN, Abd El-Al MS. How do green and black coffee brews and bioactive interaction with gut microbiome affect its health outcomes? Mining evidence from mechanistic studies, metagenomics and clinical trials. Trends Food Sci Technol [Internet]. 2021;118(PB):920–37. Available from: https://doi.org/10.1016/j.tifs.2021.11.004

249. Díaz-Rubio ME, Saura-Calixto F. Dietary fiber in brewed coffee. J Agric Food Chem. 2007;55(5):1999–2003.

250. Jaquet M, Rochat I, Moulin J, Cavin C, Bibiloni R. Impact of coffee consumption on the gut microbiota: A human volunteer study. Int J Food Microbiol [Internet]. 2009;130(2):117–21. Available from: http://dx.doi.org/10.1016/j.ijfoodmicro.2009.01.011

251. Mansour A, Mohajeri-Tehrani MR, Karimi S, Sanginabadi M, Poustchi H, Enayati S, et al. Short term effects of coffee components consumption on gut microbiota in patients with non-alcoholic fatty liver and diabetes: A pilot randomized placebo-controlled, clinical trial. EXCLI J. 2020;19:241–50.

252. Chong CW, Wong LC, Teh CSJ, Ismail NH, Chan PQ, Lim CS, et al. Coffee consumption revealed sex differences in host endogenous metabolism and gut microbiota in healthy adults. J Food Biochem. 2020;44(12):1–13.

253. Alfa MJ, Strang D, Tappia PS, Graham M, Van Domselaar G, Forbes JD, et al. A randomized trial to determine the impact of a digestion resistant starch composition on the gut microbiome in older and mid-age adults. Clin Nutr. 2018 Jun 1;37(3):797–807.

254. Wiese M, Bashmakov Y, Chalyk N, Nielsen DS, Krych Ł, Kot W, et al. Prebiotic Effect of Lycopene and Dark Chocolate on Gut Microbiome with Systemic Changes in Liver Metabolism, Skeletal Muscles and Skin in Moderately Obese Persons. Biomed Res Int. 2019;2019.

255. Ortiz-Repiso V. The 2030 Agenda and the information professionals [Internet]. Boosting the Knowledge Economy: Key Contributions from Information Services in Educational, Cultural and Corporate Environments. Chandos Publishing; 2022. 189–208 p. Available from: http://dx.doi.org/10.1016/B978-1-84334-772-9.00002-6

256. Siddiqi SM, Uscher-Pines L, Leinhos M, Dekker D, Chari R. Public Health Readiness for Citizen Science: Health Department Experiences. J Public Heal Manag Pract. 2023;29(4):464–72.

257. Bonney R, Phillips TB, Ballard HL, Enck JW. Can citizen science enhance public understanding of science? Public Underst Sci. 2016;25(1):2–16.

258. de Sherbinin A, Bowser A, Chuang TR, Cooper C, Danielsen F, Edmunds R, et al. The Critical Importance of Citizen Science Data. Front Clim. 2021;3(March):1–7.

259. Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg K V., et al. Citizen science: A developing tool for expanding science knowledge and scientific literacy. Bioscience. 2009;59(11):977–84.

260. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems [Internet]. 2018 Jun 26 [cited 2024 Feb 21];3(3). Available from: https://journals.asm.org/doi/10.1128/msystems.00031-18

261. Savio L Del, Prainsack B, Buyx A. Motivations of participants in the citizen science of microbiomics: Data from the British Gut Project. Genet Med. 2017;19(8):959–61.

262. Marcos S V, Rubio MJ, Sanchidrián FR, de Robledo D. Spanish National dietary survey in adults, elderly and pregnant women. EFSA Support Publ. 2017;13(6).

263. Comité Científico AESAN. No Title [Internet]. AESAN/BEDCA. Base de datos española de composición de alimentos v1.0. 2007. Available from: http://www.bedca.net

264. Moreiras O, Carbajal A, Cabrera L. Tabla de composición de alimentos. 19th ed. Pirámide, editor. Madrid, Spain; 2019. 1–494 p.

265. USDA. Composite of Foods Raw, Processed, Prepared. Natl Nutr Database Stand Ref. 2015;154.

266. Comité Científico AESAN. Informe del Comité Científico de la Agencia Española de Seguridad Alimentaria y Nutrición (AESAN) sobre Ingestas Nutricionales de Referencia para la población española. Rev del Com Científico la AESAN. 2019;29:43–68.

267. (US) I of M, Standing Committee on Scientific Evaluation of Dietary. Using Dietary Reference Intakes for Nutrient Assesment of Individuals. 2000.

268. Olza J, Aranceta-Bartrina J, González-Gross M, Ortega RM, Serra-Majem L, Varela-Moreiras G, et al. Reported dietary intake and food sources of zinc, selenium, and vitamins a, e and c in the spanish population: Findings from the anibes study. Nutrients. 2017;9(7).

269. Serrano-Gómez G, Mayorga L, Oyarzun I, Roca J, Borruel N, Casellas F, et al. Dysbiosis and relapse-related microbiome in inflammatory bowel disease: A shotgun metagenomic approach. Comput Struct Biotechnol J. 2021;19:6481–9.

270. Asare PT, Lee CH, Hürlimann V, Teo Y, Cuénod A, Akduman N, et al. A MALDI-TOF MS library for rapid identification of human commensal gut bacteria from the class Clostridia. Front Microbiol. 2023;14(February):1–11.

271. Lasch P, Beyer W, Bosch A, Borriss R, Drevinek M, Dupke S, et al. A MALDI-ToF mass spectrometry database for identification and classification of highly pathogenic bacteria. Sci data. 2025;12(1):187.

272. Ogier JC, Son O, Gruss A, Tailliez P, Delacroix-Buchet A. Identification of the bacterial microflora in dairy products by temporal temperature gradient gel electrophoresis. Appl Environ Microbiol. 2002;68(8):3691–701.

273. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol [Internet]. 2021;17(11):1–27. Available from: http://dx.doi.org/10.1371/journal.pcbi.1009442

274. Davis C, Bryan J, Hodgson J, Murphy K. Definition of the mediterranean diet: A literature review. Nutrients. 2015;7(11):9139–53.

275. León-Muñoz LM, Guallar-Castillón P, Graciani A, López-García E, Mesas AE, Aguilera MT, et al. Adherence to the mediterranean diet pattern has declined in Spanish adults. J Nutr. 2012;142(10):1843–50.

276. Varela-Moreiras G, Ávila JM, Cuadrado C, del Pozo S, Ruiz E, Moreiras O. Evaluation of food consumption and dietary patterns in Spain by the Food Consumption Survey: Updated information. Eur J Clin Nutr. 2010;64:S37–43.

277. Varela-Moreiras G, Ruiz E, Valero T, Ávila JM, del Pozo S. The Spanish diet: An update. Nutr Hosp. 2013;28(SUPPL.5):13–20.

278. Alamnia TT, Sargent GM, Kelly M. Dietary patterns and associations with metabolic risk factors for non-communicable disease. Sci Rep [Internet]. 2023;13(1):1–10. Available from: https://doi.org/10.1038/s41598-023-47548-0

279. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. J Nutr [Internet]. 2020;150(4):663–71. Available from: https://doi.org/10.1093/jn/nxz268

280. Micha R, Shulkin ML, Peñalvo JL, Khatibzadeh S, Singh GM, Rao M, et al. Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the nutrition and chronic diseases expert group (NutriCoDE). PLoS One. 2017;12(4):1–26.

281. Aranceta J, Blay G, Carrillo L, Fernández JM, Garaulet M, Gil A, et al. Guia de la alimentación saludable para atención primaria y colectivos ciudadanos [Internet]. 2016. Available from: file:///E:/20854194W/Downloads/Guia SENC_Capitulo 1 -WEB0 (2).pdf

282. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. Gut Microbes. 2012;3(4).

283. Chung WSF, Meijerink M, Zeuner B, Holck J, Louis P, Meyer AS, et al. Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. FEMS Microbiol Ecol. 2017;93(11):1–9.

284. Haran JP, Bucci V, Dutta P, Ward D, McCormick B. The nursing home elder microbiome stability and associations with age, frailty, nutrition and physical location. J Med Microbiol. 2018;67(1):40–51.

285. Muralidharan J, Moreno-Indias I, Bulló M, Lopez JV, Corella D, Castañer O, 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 [Internet]. 2021;114(3):1148–58. Available from: https://doi.org/10.1093/ajcn/nqab150

286. Ma E, Maskarinec G, Lim U, Boushey CJ, Wilkens LR, Setiawan VW, et al. Long-term association between diet quality and characteristics of the gut microbiome in the multiethnic cohort study. Br J Nutr. 2022;128(1):93–102.

287. Mukherjee A, Lordan C, Ross RP, Cotter PD. Gut microbes from the phylogenetically diverse genus Eubacterium and their various contributions to gut health. Gut Microbes [Internet]. 2020;12(1). Available from: https://doi.org/10.1080/19490976.2020.1802866

288. Jia W, Rajani C, Xu H, Zheng X. Gut microbiota alterations are distinct for primary colorectal cancer and hepatocellular carcinoma. Protein Cell [Internet]. 2021;12(5):374–93. Available from: https://doi.org/10.1007/s13238-020-00748-0

289. Vestergaard MV, Allin KH, Eriksen C, Zakerska-Banaszak O, Arasaradnam RP, Alam MT, et al. Gut microbiota signatures in inflammatory bowel disease. United Eur Gastroenterol J. 2024;12(1):22–33.

290. Cornelis M., Erlund I, Michelotti G., Herder C, Westerhuis J., Toumilehto J. Metabolomic response to coffee consumption application to a three-stage clinical trial. 2018.

291. Dekkers KF, Sayols-Baixeras S, Baldanzi G, Nowak C, Hammar U, Nguyen D, et al. An online atlas of human plasma metabolite signatures of gut microbiome composition. Nat Commun. 2022;13(1):1–12.

292. Wang K, Ma W, Hang D, Thompson KN, Franzosa EA, Asnicar F, et al. Interplay between coffee intake and the gut microbiome in relation to host metabolome. Gastroenterology. 2023 May;164(6):S-257-S-258.

293. Tabone M, Bressa C, García-Merino JA, Moreno-Pérez D, Van EC, Castelli FA, et al. The effect of acute moderate-intensity exercise on the serum and fecal metabolomes and the gut microbiota of cross-country endurance athletes. Sci Rep [Internet]. 2021;11(1):1–12. Available from: https://doi.org/10.1038/s41598-021-82947-1

294. Partearroyo T, Samaniego-Vaesken M de L, Ruiz E, Aranceta-Bartrina J, Gil Á, González-Gross M, et al.

Current food consumption amongst the spanish anibes study population. Nutrients. 2019;11(11):1–15.

295. Si H, Yan W, Jun S, Hongyu M, Xia Z, Kuan W, et al. Modulation of cecal microbiota and fecal metabolism in mice by walnut protein. Food Funct. 2023;15(3):1689–704.

296. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol [Internet]. 2019;4(4):623–32. Available from: http://dx.doi.org/10.1038/s41564-018-0337-x

297. Coello K, Hansen TH, Sørensen N, Munkholm K, Kessing LV, Pedersen O, et al. Gut microbiota composition in patients with newly diagnosed bipolar disorder and their unaffected first-degree relatives. Brain Behav Immun [Internet]. 2019;75(August 2018):112–8. Available from: https://doi.org/10.1016/j.bbi.2018.09.026

298. Armstrong H, Alipour M, Valcheva R, Bording-Jorgensen M, Jovel J, Zaidi D, et al. Host immunoglobulin G selectively identifies pathobionts in pediatric inflammatory bowel diseases. Microbiome. 2019;7(1):1–17.

299. Arias M, Cobo M, Jaime-Sánchez P, Pastor J, Marijuan P, Pardo J, et al. Gut microbiota and systemic inflammation changes after bread consumption: The ingredients and the processing influence. J Funct Foods [Internet]. 2017;32:98–105. Available from: http://dx.doi.org/10.1016/j.jff.2017.02.023

300. Kwon JG, Park SH, Kwak JE, Cho JH, Kim G, Lee D, et al. Mouse feeding study and microbiome analysis of sourdough bread for evaluation of its health effects. Front Microbiol. 2022;13(September):1–13.

301. Travinsky-Shmul T, Beresh O, Zaretsky J, Griess-Fishheimer S, Rozner R, Kalev-Altman R, et al. Ultra-processed food impairs bone quality, increases marrow adiposity and alters gut microbiome in mice. Foods. 2021;10(12):1–17.

302. Snelson M, Tan SM, Clarke RE, De Pasquale C, Thallas-Bonke V, Nguyen TV, et al. Processed foods drive intestinal barrier permeability and microvascular diseases. Sci Adv. 2021;7(14):1–15.

303. Kim JS, Lee KC, Suh MK, Han K Il, Eom MK, Lee JH, et al. *Mediterraneibacter butyricigenes* sp. nov., a butyrate-producing bacterium isolated from human faeces. J Microbiol. 2019;57(1):38–44.

304. Olza J, Aranceta-Bartrina J, González-Gross M, Ortega RM, Serra-Majem L, Varela-Moreiras G, et al. Reported dietary intake, disparity between the reported consumption and the level needed for adequacy and food sources of calcium, phosphorus, magnesium and vitamin D in the Spanish population: Findings from the ANIBES study. Nutrients. 2017;9(2).

305. Schaus SR, Vasconcelos Pereira Ga, Luis AS, Madlambayan Em, Terrapon N, Ostrowski MP, et al. *Ruminococcus torques* is a keystone degrader of intestinal mucin glycoprotein, releasing oligosaccharides used by Bacteroides thetaiotaomicron. bioRxiv [Internet]. 2024;1–2. Available from: https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=650377970

306. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016;167(5):1339-1353.e21.

307. Kwak MS, Cha JM, Shin HP, Jeon JW, Yoon JY. Development of a Novel Metagenomic Biomarker for Prediction of Upper Gastrointestinal Tract Involvement in Patients With Crohn's Disease. Front Microbiol. 2020;11(June):1–11.

308. Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, Rutgeerts P, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut. 2011;60(5):631–7.

309. González-Molero I, Morcillo S, Valdés S, Pérez-Valero V, Botas P, Delgado E, et al. Vitamin D deficiency in Spain: A population-based cohort study. Eur J Clin Nutr. 2011;65(3):321–8.

310. Chen TC, Chimeh F, Lu Z, Mathieu J, Person KS, Zhang A, et al. Factors that influence the cutaneous synthesis and dietary sources of vitamin D. Arch Biochem Biophys. 2007;460(2):213–7.

311. Assa A, Vong L, Pinnell LJ, Avitzur N, Johnson-Henry KC, Sherman PM. Vitamin D deficiency promotes epithelial barrier dysfunction and intestinal inflammation. J Infect Dis. 2014;210(8):1296–305.

312. Ciubotaru I, Green SJ, Kukreja S, Barengolts E. Significant differences in fecal microbiota are associated with various stages of glucose tolerance in African American male veterans. Transl Res [Internet]. 2015;166(5):401–11. Available from: http://dx.doi.org/10.1016/j.trsl.2015.06.015

313. Bertone-Johnson ER, Powers SI, Spangler L, Brunner RL, Michael YL, Larson JC, et al. Vitamin D intake from foods and supplements and depressive symptoms in a diverse population of older women. Am J Clin Nutr [Internet]. 2011;94(4):1104–12. Available from: https://doi.org/10.3945/ajcn.111.017384

314. Pérez-Rodrigo C, Citores MG, Bárbara GH, Ruiz-Litago F, Sáenz LC, Arija V, et al. Patterns of change in dietary habits and physical activity during lockdown in spain due to the covid-19 pandemic. Nutrients. 2021;13(2):1–16.

315. Wierzejska RE. Dietary supplements—for whom? The current state of knowledge about the health effects of selected supplement use. Int J Environ Res Public Health. 2021;18(17).

316. Sah N, Wu G, W.Bazer F. Amino Acids in Nutrition and Health. Amino Acids in Gene Expression, Metabolic Regulation, and Exercising Performance. In: Advances in Experimental Medicine and Biology [Internet]. Springer Nature Switzerland; 2021. Available from: http://www.springer.com/series/5584

317. Singh K, Gobert AP, Coburn LA, Barry DP, Allaman M, Asim M, et al. Dietary arginine regulates severity of experimental colitis and affects the colonic microbiome. Front Cell Infect Microbiol. 2019;9(MAR):1–15.

318. Wu M, Xiao H, Shao F, Tan B, Hu S. Arginine accelerates intestinal health through cytokines and intestinal microbiota. Int Immunopharmacol [Internet]. 2020;81(October):106029. Available from: https://doi.org/10.1016/j.intimp.2019.106029

319. Ren W, Chen S, Yin J, Duan J, Li T, Liu G, et al. Dietary arginine supplementation of mice alters the microbial population and activates intestinal innate immunity. J Nutr [Internet]. 2014;144(6):988–95. Available from: https://doi.org/10.3945/jn.114.192120

320. Mayeur C, Veuillet G, Michaud M, Raul F, Blottière HM, Blachier F. Effects of agmatine accumulation in human colon carcinoma cells on polyamine metabolism, DNA synthesis and the cell cycle. Biochim Biophys Acta - Mol Cell Res. 2005;1745(1):111–23.

321. Martin-Gallausiaux C, Marinelli L, Blottière HM, Larraufie P, Lapaque N. Conference on diet and digestive disease symposium 2: Sensing and signalling of the gut environment: Scfa: Mechanisms and functional importance in the gut. Proc Nutr Soc. 2021;80(1):37–49.

322. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A. 2008;105(43):16731–6.

323. Henry C, Bassignani A, Berland M, Langella O, Sokol H, Juste C. Modern Metaproteomics: A Unique Tool to Characterize the Active Microbiome in Health and Diseases, and Pave the Road towards New Biomarkers—Example of Crohn's Disease and Ulcerative Colitis Flare-Ups. Cells. 2022;11(8):1–15.

324. Chen H, Ou R, Tang N, Su W, Yang R, Yu X, et al. Alternation of the gut microbiota in irritable bowel syndrome: an integrated analysis based on multicenter amplicon sequencing data. J Transl Med. 2023;21(1):1–16.

325. Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, et al. Human Gut Microbiota Changes Reveal the Progression of Glucose Intolerance. PLoS One. 2013;8(8).

326. Newsome RC, Gharaibeh RZ, Pierce CM, da Silva WV, Paul S, Hogue SR, et al. Interaction of bacterial genera associated with therapeutic response to immune checkpoint PD-1 blockade in a United States cohort. Genome Med. 2022;14(1):1–15.

327. Coutzac C, Jouniaux JM, Paci A, Schmidt J, Mallardo D, Seck A, et al. Systemic short chain fatty acids limit antitumor effect of CTLA-4 blockade in hosts with cancer. Nat Commun [Internet]. 2020;11(1):1–13. Available from: http://dx.doi.org/10.1038/s41467-020-16079-x

328. Venegas DP, De La Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short chain fatty acids (SCFAs)mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Front Immunol. 2019;10(MAR).

329. Xia Y, Tan D, Akbary R, Kong J, Seviour R, Kong Y. Aqueous raw and ripe Pu-erh tea extracts alleviate obesity and alter cecal microbiota composition and function in diet-induced obese rats. Appl Microbiol Biotechnol. 2019;103(4):1823–35.

330. Heyes DJ, Levy C, Lafite P, Roberts IS, Goldrick M, Stachulski A V., et al. Structure-based mechanism of CMP-2-keto-3-deoxymanno-octulonic acid synthetase: Convergent evolution of a sugar-activating enzyme with DNA/RNA polymerases. J Biol Chem [Internet]. 2009;284(51):35514–23. Available from: http://dx.doi.org/10.1074/jbc.M109.056630

331. Romaní-Pérez M, López-Almela I, Bullich-Vilarrubias C, Evtoski Z, Benítez-Páez A, Sanz Y. *Bacteroides uniformis CECT 7771* requires adaptive immunity to improve glucose tolerance but not to prevent body weight gain in diet-induced obese mice. Microbiome. 2024;12(1):1–16.

332. López-Almela I, Romaní-Pérez M, Bullich-Vilarrubias C, Benítez-Páez A, Gómez Del Pulgar EM, Francés R, et al. *Bacteroides uniformis* combined with fiber amplifies metabolic and immune benefits in obese mice. Gut Microbes [Internet]. 2021;13(1):1–20. Available from: https://doi.org/10.1080/19490976.2020.1865706

333. Fabersani E, Portune K, Campillo I, López-Almela I, Montserrat-de la Paz S, Romaní-Pérez M, et al. *Bacteroides uniformis CECT 7771* alleviates inflammation within the gut-adipose tissue axis involving TLR5 signaling in obese mice. Sci Rep [Internet]. 2021;11(1):1–15. Available from: https://doi.org/10.1038/s41598-021-90888-y

334. Gauffin Cano P, Santacruz A, Moya Á, Sanz Y. *Bacteroides uniformis CECT 7771* ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. PLoS One. 2012;7(7).

335. Yan YT, Lei Y, Qu Y, Fan Z, Zhang T, Xu Y, et al. Bacteroides uniformis-induced perturbations in colonic microbiota and bile acid levels inhibit TH17 differentiation and ameliorate colitis developments. npj Biofilms Microbiomes. 2023;9(1):1–14.

336. Xu F, Earp JE, Vadiveloo M, Adami A, Delmonico MJ, Lofgren IE, et al. The relationships between total protein intake, protein sources, physical activity, and lean mass in a representative sample of the us adults. Nutrients. 2020;12(10):1–13.

337. Wang DD, Qi Q, Wang Z, Usyk M, Sotres-Alvarez D, Mattei J, et al. The Gut Microbiome Modifies the Association Between a Mediterranean Diet and Diabetes in USA Hispanic/ Latino Population. J Clin Endocrinol Metab [Internet]. 2022 Mar 1 [cited 2023 Feb 8];107(3):e924. Available from: /pmc/articles/PMC8851916/

338. Meslier V, Laiola M, Roager HM, De Filippis F, Roume H, Quinquis B, et al. Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. Gut. 2020;69(7):1258–68.

339. Ghosh TS, Rampelli S, Jeffery IB, Santoro A, Neto M, Capri M, et al. Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: The NU-AGE 1-year dietary intervention across five European countries. Gut. 2020;69(7):1218–28.

340. Khalili H, Håkansson N, Chan SS, Chen Y, Lochhead P, Ludvigsson JF, et al. Adherence to a Mediterranean diet is associated with a lower risk of later-onset Crohn's disease: Results from two

large prospective cohort studies. Gut. 2020;69(9):1637–44.

341. Narula N, Chang NH, Mohammad D, Wong ECL, Ananthakrishnan AN, Chan SSM, et al. Food Processing and Risk of Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. Clin Gastroenterol Hepatol [Internet]. 2023;21(10):2483-2495.e1. Available from: https://doi.org/10.1016/j.cgh.2023.01.012

342. Sharma D, Denmat SH Le, Matzke NJ, Hannan K, Hannan RD, O'Sullivan JM, et al. A new method for determining ribosomal DNA copy number shows differences between Saccharomyces cerevisiae populations. Genomics [Internet]. 2022;114(4):110430. Available from: https://doi.org/10.1016/j.ygeno.2022.110430

343. Lofgren LA, Uehling JK, Branco S, Bruns TD, Martin F, Kennedy P. Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales.pdf. Mol Ecol. 2019;(28):721–30.

344. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65.

345. García-Gamboa R, Díaz-Torres O, Senés-Guerrero C, Gradilla-Hernández MS, Moya A, Pérez-Brocal V, et al. Associations between bacterial and fungal communities in the human gut microbiota and their implications for nutritional status and body weight. Sci Rep [Internet]. 2024;14(1):1–14. Available from: https://doi.org/10.1038/s41598-024-54782-7

346. Mok K, Poolsawat T, Somnuk S, Wanikorn B, Patumcharoenpol P, Nitisinprasert S, et al. Preliminary characterization of gut mycobiome enterotypes reveals the correlation trends between host metabolic parameter and diet: a case study in the Thai Cohort. Sci Rep [Internet]. 2024;14(1):1–11. Available from: https://doi.org/10.1038/s41598-024-56585-2

347. Mondragón-Flores A, Rodríguez-Alvarado G, Gómez-Dorantes N, Guerra-Santos JJ, Fernández-Pavía SP. Botryosphaeriaceae : a complex , diverse and cosmopolitan family of fungi. Rev Mex Ciencias Agrícolas. 28699BC;12(4):643–54.

348. Le Roy T, Moens De Hase E, Van Hul M, Paquot A, Pelicaen R, Régnier M, et al. *Dysosmobacter welbionis* is a newly isolated human commensal bacterium preventing diet-induced obesity and metabolic disorders in mice. Gut. 2022;71(3):534–43.

349. Feraco A, Armani A, Amoah I, Guseva E, Camajani E, Gorini S, et al. Assessing gender differences in food preferences and physical activity: a population-based survey. Front Nutr. 2024;11(February):1–11.

350. Gil M, Rudy M, Stanisławczyk R, Duma-Kocan P, Żurek J. Gender Differences in Eating Habits of Polish Young Adults Aged 20–26. Int J Environ Res Public Health. 2022;19(22).

351. Samulenaite S, García-Blanco A, Mayneris-Perxachs J, Domingo-Rodríguez L, Cabana-Domínguez J, Fernàndez-Castillo N, et al. Gut microbiota signatures of vulnerability to food addiction in mice and humans. Gut. 2024;1–17.

352. Lubin JH, Couper D, Lutsey PL, Woodward M, Yatsuya H, Huxley RR. Risk of cardiovascular disease from cumulative cigarette use and the impact of smoking intensity. Epidemiology. 2016;27(3):395–404.

353. Chen BC, Weng MT, Chang CH, Huang LY, Wei SC. Effect of smoking on the development and outcomes of inflammatory bowel disease in Taiwan: a hospital-based cohort study. Sci Rep [Internet]. 2022;12(1):1–9. Available from: https://doi.org/10.1038/s41598-022-11860-y

354. Jacob L, Freyn M, Kalder M, Dinas K, Kostev K. Impact of tobacco smoking on the risk of developing 25 different cancers in the UK: A retrospective study of 422,010 patients followed for up to 30 years. Oncotarget. 2018;9(25):17420–9.

355. Torres J, Bonovas S, Doherty G, Kucharzik T, Gisbert JP, Raine T, et al. ECCO guidelines on therapeutics in Crohn's disease: Medical treatment. J Crohn's Colitis. 2020;14(1):4–22.

356. Zhang W, Li J, Lu S, Han N, Miao J, Zhang T, et al. Gut microbiota community characteristics and disease-related microorganism pattern in a population of healthy Chinese people. Sci Rep [Internet]. 2019;9(1):1–10. Available from: http://dx.doi.org/10.1038/s41598-018-36318-y

357. Jiang N, Lee YO, Ling PM. Association between tobacco and alcohol use among young adult bar patrons: A cross-sectional study in three cities. BMC Public Health. 2014;14(1):1–9.

358. Falk DE, Hsiao-Ye Y, Hiller-Sturmhöfel S. An Epidemiologic Analysis of Co-Occurring Alcohol and Tobacco Use and Disorders. Alcohol Res Heal. 2006;29(3):162–71.

359. Alemán JA, Rentero MPZ, Montoro-García S, Mulero J, Garrido AP, Leal M, et al. Adherence to the "Mediterranean diet" in Spain and its relationship with cardiovascular risk (DIMERICA study). Nutrients. 2016;8(11).

360. Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, et al. 1,520 Reference Genomes From Cultivated Human Gut Bacteria Enable Functional Microbiome Analyses. Nat Biotechnol [Internet]. 2019;37(2):179–85. Available from: http://dx.doi.org/10.1038/s41587-018-0008-8

361. Chang Y, Hou F, Pan Z, Huang Z, Han N, Bin L, et al. Optimization of Culturomics Strategy in Human Fecal Samples. Front Microbiol. 2019;10(December):1–11.

362. Ito T, Sekizuka T, Kishi N, Yamashita A, Kuroda M. Conventional culture methods with commercially available media unveil the presence of novel culturable bacteria. Gut Microbes. 2019;10(1):77–91.

363. Park H, Yeo S, Ryu CB. A streamlined culturomics approach for the human gut microbiota research. Sci Rep [Internet]. 2024;1–11. Available from: https://doi.org/10.21203/rs.3.rs-4388987/v1

364. Armetta J, Li SS, Vaaben TH, Vazquez-uribe R, Sommer MOA. Metagenome-guided culturomics for the targeted enrichment of gut microbes. Nat Commun [Internet]. 2025;(May 2024). Available from: http://dx.doi.org/10.1038/s41467-024-55668-y

365. Ghimire S, Roy C, Wongkuna S, Antony L, Maji A, Keena MC, et al. Identification of *Clostridioides difficile*-Inhibiting Gut Commensals Using Culturomics, Phenotyping, and Combinatorial Community Assembly. mSystems. 2020;

366. Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat Med [Internet]. 2019;25(9):1442–52. Available from: http://dx.doi.org/10.1038/s41591-019-0559-3

367. Tao X, Huang W, Pan L, Sheng L, Qin Y, Chen L, et al. Optimizing ex vivo culture conditions to study human gut microbiome. ISME Commun. 2023;3(1):1–10.

368. Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol. 2016;1(November).

369. Lau JT, Whelan FJ, Herath I, Lee CH, Collins SM, Bercik P, et al. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. Genome Med [Internet]. 2016;8(1):1–10. Available from: http://dx.doi.org/10.1186/s13073-016-0327-7

370. Bilinski J, Dziurzynski M, Grzesiowski P, Podsiadly E, Stelmaszczyk-Emmel A, Dzieciatkowski T, et al. Fresh Versus Frozen Stool for Fecal Microbiota Transplantation—Assessment by Multimethod Approach Combining Culturing, Flow Cytometry, and Next-Generation Sequencing. Front Microbiol. 2022;13(July):1–14.

371. Volokh O, Klimenko N, Berezhnaya Y, Tyakht A, Nesterova P, Popenko A, et al. Human gut microbiome response induced by fermented dairy product intake in healthy volunteers. Nutrients. 2019;11(3):3–14.

372. Le Roy CI, Kurilshikov A, Leeming ER, Visconti A, Bowyer RCE, Menni C, et al. Yoghurt consumption is associated with changes in the composition of the human gut microbiome and metabolome. BMC

Microbiol. 2022;22(1):1–13.

373. Chen H, Ou R, Tang N, Su W, Yang R, Yu X, et al. Alternation of the gut microbiota in irritable bowel syndrome: an integrated analysis based on multicenter amplicon sequencing data. J Transl Med [Internet]. 2023;21(1):1–16. Available from: https://doi.org/10.1186/s12967-023-03953-7

374. Frost F, Storck LJ, Kacprowski T, Gärtner S, Rühlemann M, Bang C, et al. A structured weight loss program increases gut microbiota phylogenetic diversity and reduces levels of Collinsella in obese type 2 diabetics: A pilot study. PLoS One. 2019;14(7):1–14.

375. Purohit A, Kandiyal B, Kumar S, Pragasam AK, Kamboj P, Talukdar D, et al. *Collinsella aerofaciens* linked with increased ethanol production and liver inflammation contribute to the pathophysiology of NAFLD. iScience [Internet]. 2024;27(2):108764. Available from: https://doi.org/10.1016/j.isci.2023.108764

376. Thomas AM, Fidelle M, Routy B, Kroemer G, Wargo JA, Segata N, et al. Gut OncoMicrobiome Signatures (GOMS) as next-generation biomarkers for cancer immunotherapy. Nat Rev Clin Oncol. 2023;20(9):583–603.

377. Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N, et al. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. Gastroenterology. 2004;127(2):412–21.

378. Lee JG, Han DS, Jo SV, Reum Lee A, Park CH, Eun CS, et al. Characteristics and pathogenic role of adherent-invasive *Escherichia coli* in inflammatory bowel disease: Potential impact on clinical outcomes. PLoS One. 2019;14(4):1–14.

379. Costa RFA, Costa RFA, Ferrari MLA, Ferrari MLA, Bringer MA, Darfeuille-Michaud A, et al. Characterization of mucosa-associated Escherichia coli strains isolated from Crohn's disease patients in Brazil. BMC Microbiol. 2020;20(1):1–11.

380. Martínez N, Hidalgo-Cantabrana C, Delgado S, Margolles A, Sánchez B. Filling the gap between collection, transport and storage of the human gut microbiota. Sci Rep. 2019;9(1):1–8.

381. Campanale C, Massarelli C, Savino I, Locaputo V, Uricchio VF. A detailed review study on potential effects of microplastics and additives of concern on human health. Int J Environ Res Public Health. 2020;17(4).

10ANNEXES

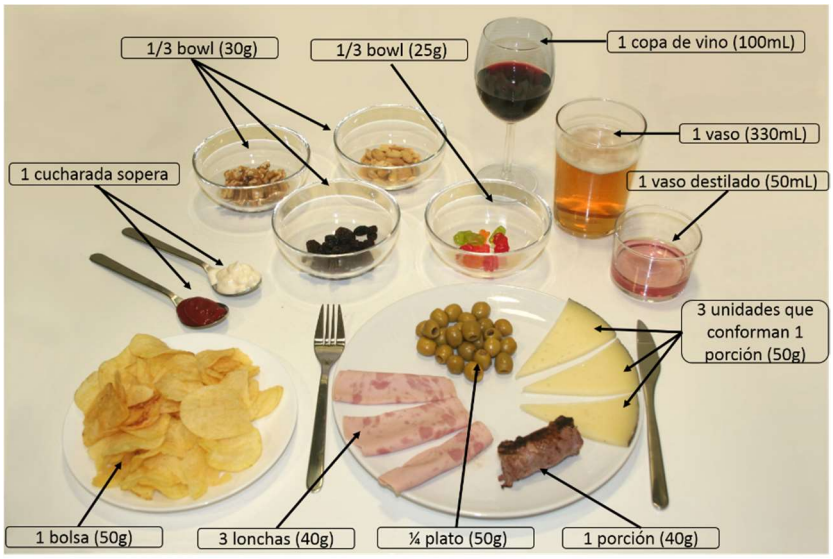
ANNEX 1. List of 58 Food items included in the sFFQ

Food ítems (g/day)
1.Raw leafy vegetables such as spinach, lettuce, endive, celery, lamb's lettuce, bean sprouts, and green beans
2. Cooked leafy vegetables such as spinach, chard, cabbage, celery, asparagus, and bean sprouts
3. Tomato
4. Onion, spring onions and leek
5. Zucchini, aubergine and cucumber
6. Carrot, pumpkin and beet
7. Bell pepper and Padrón pepper
8. Crucifers like broccoli, cauliflower, turnip, cabbage and arugula (rocket)
9. Food pickled in vinegar such as onion, gherkins, sauerkraut, capers, carrots, chives, and artichokes
10. Corn and fresh legumes such as beans, peas, and broad beans
11. Mushrooms in general
12. Potato except for potato chips and sweet potato
13. Cooked lentils, cooked kidney beans (pinto, white or black), and cooked chickpeas
14. Fresh fruit such as orange, grapefruit, banana, apple, pear, nectarine, kiwi, mandarin (2 units), strawberry (6 units), watermelon or melon (1 slice), grapes (1 cluster), and natural fruit juice
15. High-fat fruit such as avocado, olives, and coconut
16. Dried fruit such as raisins, dried figs and dried cranberries
17. Nuts and seeds such as walnuts, almonds, peanuts, hazelnuts, pistachios, pine nuts, sunflower seeds and other seeds
18. White bread such as baguette, farmhouse white bread, sliced bread, and milk bread
19. Whole wheat bread such as wholemeal baguette bread, whole meal bread, and wholemeal sliced bread
20. Breakfast cereal such as cornflakes, oatmeal, muesli and others
21. Normal/whole wheat biscuit and sponge cake
22. Cooked cereal and pasta such as noodles, macaroni, spaghetti, white rice, couscous, bulgur, and other grains
23. Cooked wholegrain cereal and pasta such as wholegrain noodles, wholegrain spaghetti, brown rice, wild rice, quinoa, and other whole grains
24. Whole milk
25. Semi-skimmed milk
26. Skimmed milk
27. Plant-based beverage and product such as almond drink, rice drink, oat drink, and soy drink
28. High-fat cheese such as cured cheese, Parmesan, Manchego, Roquefort, Gruyere, Gorgonzola, and Grana padano
29. Low-fat cheese such as Mozzarella, Buffalo, Camembert, Cheddar, goat cheese, and cottage cheese
30. Fermented dairy such as yoghurt, yoghurt drink, Greek yoghurt, and kefir
31. Dairy dessert such as tiramisu, custard, flan, and ice cream (2 scoops)
32. Chicken egg, duck egg, and quail egg
33. Fatty meat such as beef, veal, pork, wild boar, deer, lamb, and horse
34. Lean meat such as chicken, turkey, and other poultry, rabbit, hare, and kid
35. Processed meats such as salami, pork sausage, blood sausage, mortadella, fresh sausage, sobrasada (Majorcan sausage), bacon, cured ham, and boiled ham
36. Blue or high-fat fish such as anchovy, eel, elver, tuna, bonito, horse mackerel, salmon, and sardine
37. White or low-fat fish such as cod, hake, sole, monkfish, whiting and turbot

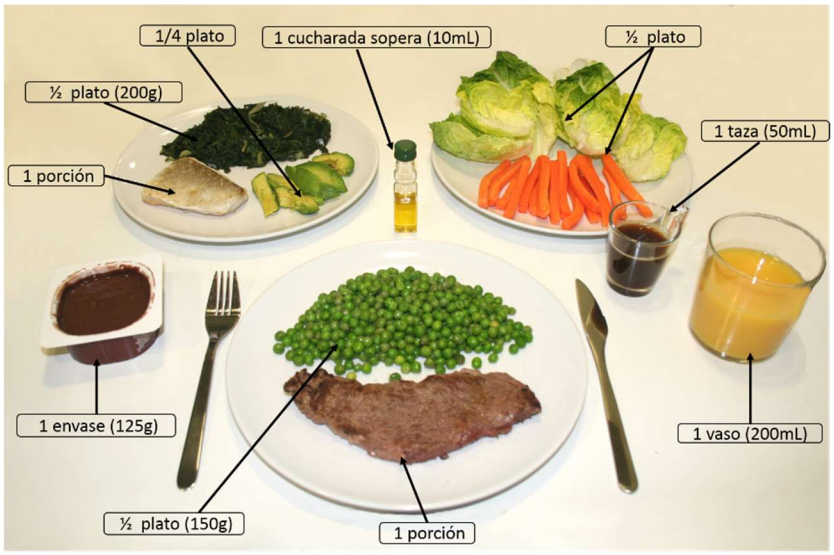
Food items (g/day)
38. Canned fish in oil such as tuna, horse mackerel, mackerel, and bonito
39. Mollusks and crustaceans such as mussel, clam, squid, octopus, cuttlefish, and prawn
40. Olive oil
41. Sunflower oil
42. Other oils such as those from corn, rapeseed, and grape seed
43. Butter and margarine
44. Pastries such as doughnut, muffins, croissants, “palmera”, churros, cakes, and puff pastry
45. Dark chocolate (> 50% cocoa) and cocoa powder
46. Confectionery such as candy bars, gummy sweets, caramels, chewing gum, nougat, and marzipan
47. Packaged tomato sauce and canned tomato
48. Other condiments such as mayonnaise, ketchup, mustard, pesto, aioli, and balsamic vinegar
49. Tea (with and without caffeine) and infusions
50. Coffee (with and without caffeine) of all kinds
51. Soft drinks such cola, diet soda, and isotonic or flavored drinks
52. Packaged fruit juices and nectar (sweetened fruit juice)
53. Wine or cava (rosé, red, vintage, must, white, muscat)
54. Beer
55. Whiskey, vodka, gin, cognac, and rum
56. Added sugar, honey, jam, and quince
57. Fried potato, nacho, salted tortilla chips, snacks, salted pretzels, potato ring crisps, twiglets, and crackers
58. Processed food such as pizza, lasagna, cannelloni, chicken nuggets, and potato omelet

ANNEX 2. Images included in the sFFQ.

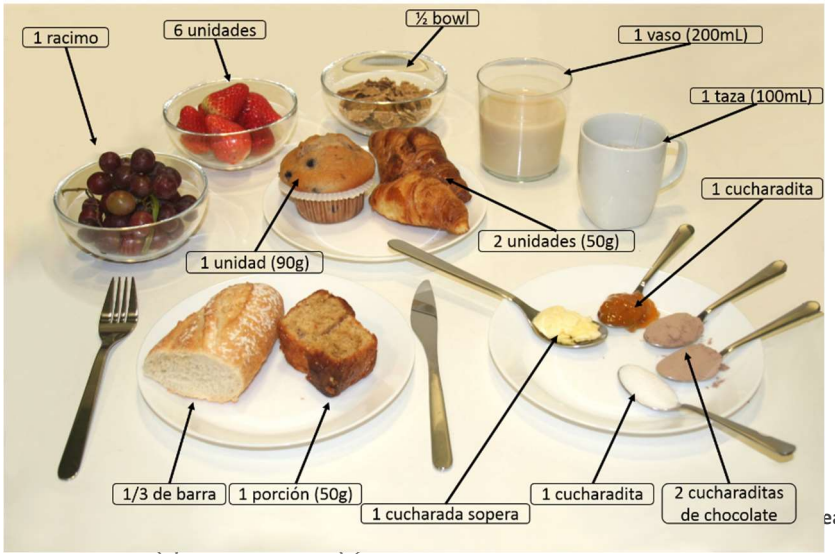
A



B



C



eakfast (A);

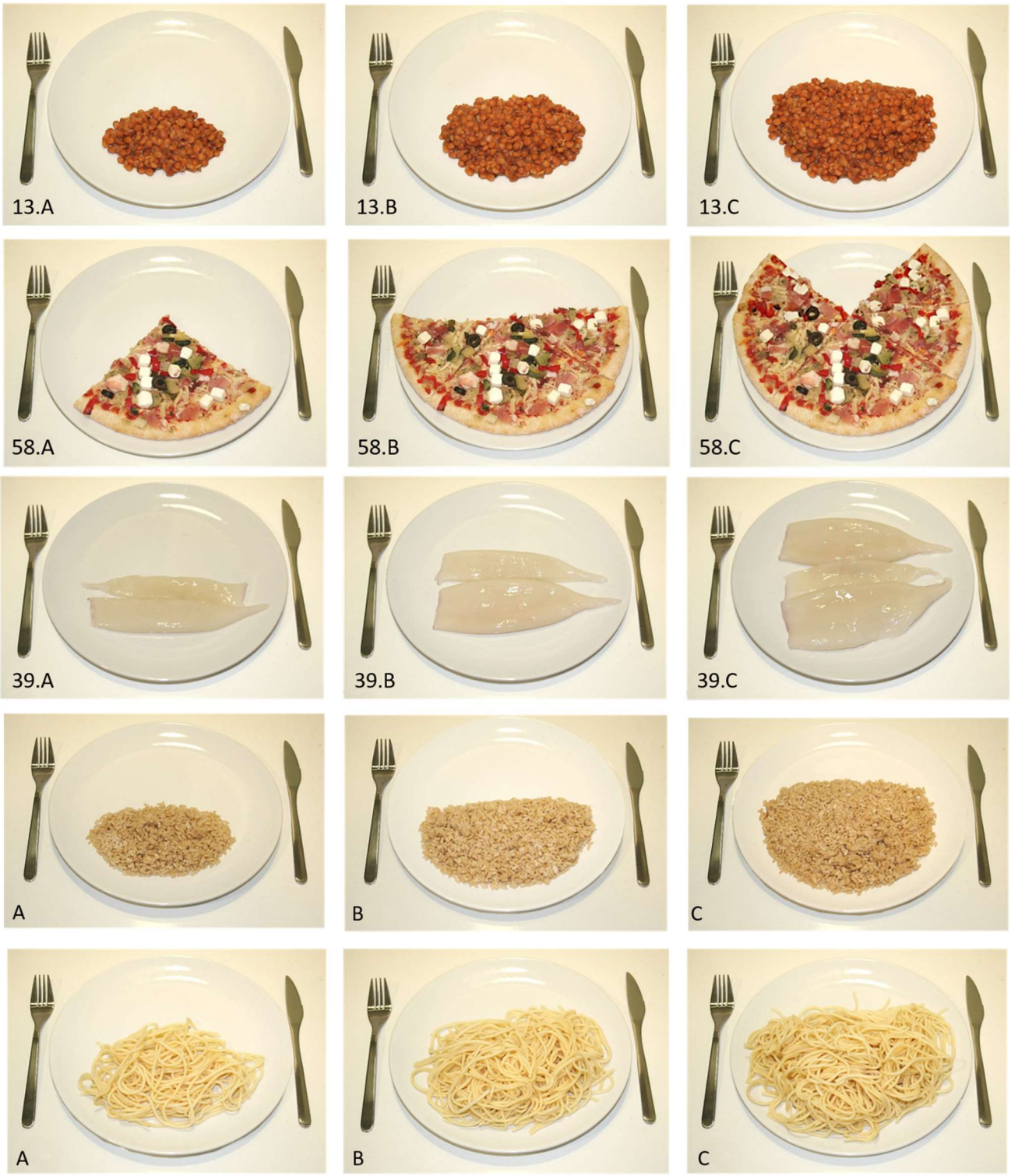


Image II. Series of photographs following ratios 1:2:3 used in the most problematic foods. Images in the middle (B) correspond to standard serving size in the sFFQ.

ANNEX 3. Old vs updated sFFQ

CUESTIONARIO DE FRECUENCIA DE CONSUMO ALIMENTARIO

Indicaciones

- El cuestionario consta de dos partes: en la primera deberá responder preguntas de información general y en la segunda deberá completar 59 preguntas a cerca de la frecuencia de consumo de alimentos del mes anterior
- La presencia de un * indica que se trata de una pregunta con respuesta obligatoria
- El tiempo de respuesta de la encuesta va a ser evaluado
- En caso de tener dudas con la ración estándar indicada, puede consultar ver imagen (alimento señalado con una flecha roja) a modo de apoyo
- Se deben responder todos los grupos de alimentos

* Datos personales

Fecha (dd/mm/aaaa)

Código de identificación

Sexo (H/M)

Fecha de nacimiento (dd/mm/aaaa)

Peso (kg)

Talla (cm)

* Número de cuestionario realizado

- 1º cuestionario
- 2º cuestionario

Si usted es de sexo femenino. ¿Esta menstruando actualmente?

- SI
- NO
- Menopausia

* ¿Cuál fue su vía de nacimiento?

- Parto Vaginal
- Cesárea
- No sabe

* ¿Usted fuma?

- ◯ SI
- ◯ NO
- ◯ En el pasado

* ¿Cuál es su grupo sanguíneo?

- ☐ O
- ☐ AB
- ☐ A
- ☐ B
- ☐ Rh(-)
- ☐ Rh (+)
- ☐ No se sabe

* ¿Usted sigue algún tipo de alimentación específica?

- ◯ NO
- ◯ SI (indique cuál)
- ◯ ¿Cuál? (especifique)

*

¿Usted consume alimentos listos para el consumo o precocinados comprados en tiendas como tortilla de patatas, pizza, lasañas, hamburguesa?

- ◯ SI
- ◯ NO

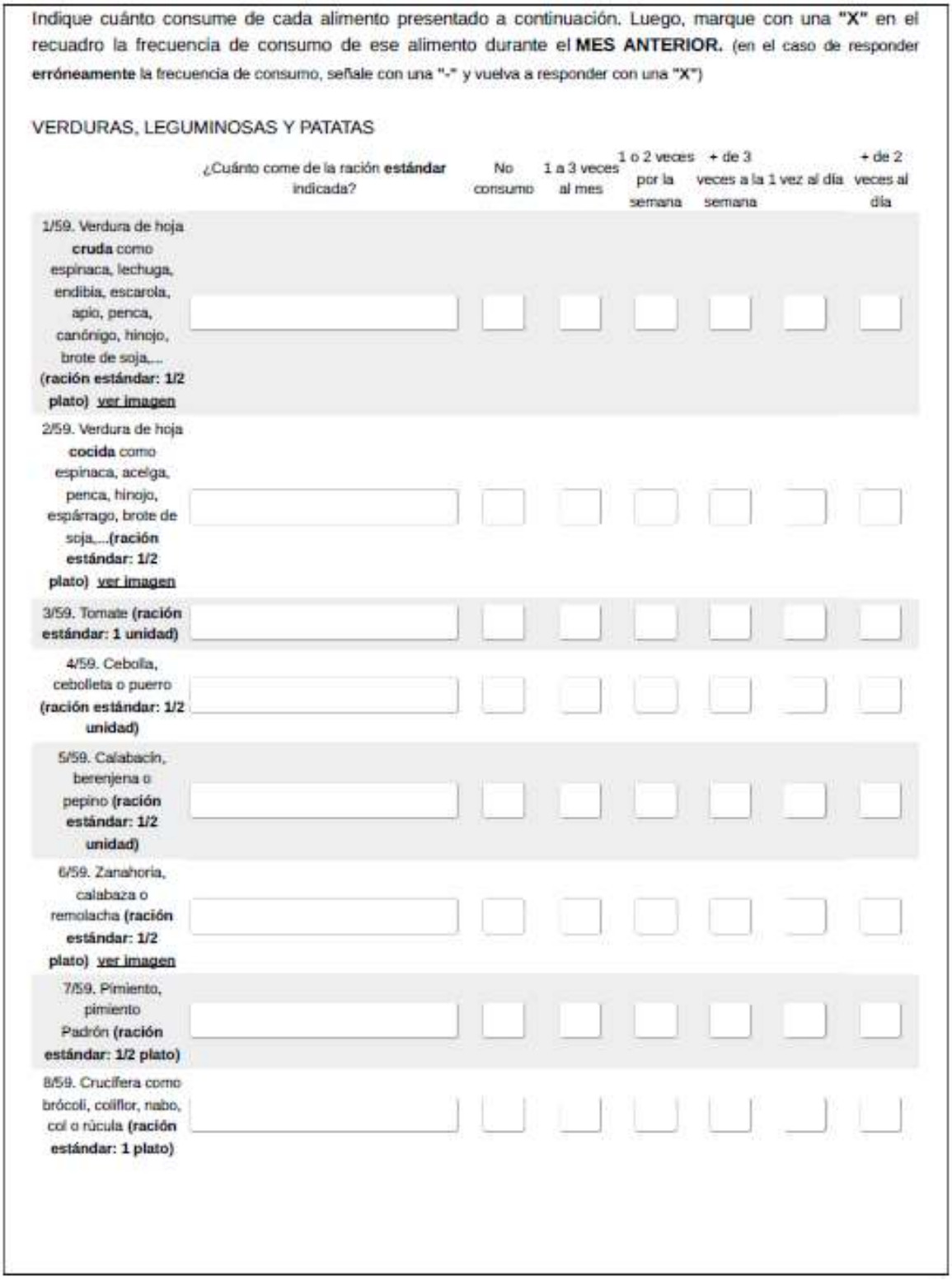
*

¿Usted consume alimentos con edulcorantes como sacarina, sucralosa, aspartamo, acesulfamo o stevia ?

- ◯ SI
- ◯ NO

*

¿Qué cantidad total de líquidos ingieres de forma aproximada a lo largo del día? Incluyendo agua, infusiones, café, leche, bebidas vegetales, zumos, refrescos, cerveza u otras bebidas alcohólicas.



Indique cuánto consume de cada alimento presentado a continuación. Luego, marque con una **"X"** en el recuadro la frecuencia de consumo de ese alimento durante el **MES ANTERIOR.** (en el caso de responder **erróneamente** la frecuencia de consumo, señale con una "-" y vuelva a responder con una "X")

VERDURAS, LEGUMINOSAS Y PATATAS

¿Cuánto come de la ración **estándar** indicada?	No consumo	1 a 3 veces al mes	1 o 2 veces por la semana	+ de 3 veces a la semana	1 vez al día	+ de 2 veces al día
1/59. Verdura de hoja **cruda** como espinaca, lechuga, endibia, escarola, apio, penca, canónigo, hinojo, brote de soja,... **(ración estándar: 1/2 plato)** ver imagen						
2/59. Verdura de hoja **cocida** como espinaca, acelga, penca, hinojo, espárrago, brote de soja,...**(ración estándar: 1/2 plato)** ver imagen						
3/59. Tomate **(ración estándar: 1 unidad)**						
4/59. Cebolla, cebolleta o puerro **(ración estándar: 1/2 unidad)**						
5/59. Calabacín, berenjena o pepino **(ración estándar: 1/2 unidad)**						
6/59. Zanahoria, calabaza o remolacha **(ración estándar: 1/2 plato)** ver imagen						
7/59. Pimiento, pimiento Padrón **(ración estándar: 1/2 plato)**						
8/59. Crucífera como brócoli, coliflor, nabo, col o rúcula **(ración estándar: 1 plato)**						

Image I. sFFQ used in the Pilot Study by means of SurveyMonkey web just available in Spanish

Manichanh Lab

Language

Manichanh Lab - Login

Please login to proceed.

User code

examplecode

Password

Login

Vall d'Hebron
Institut de Recerca

SHORT FOOD FREQUENCY QUESTIONNAIRE

Relevant information:

- The questionnaire consists of two parts: in the first you will have to answer general information questions and in the second you will have to complete 59 questions about the frequency of food consumption in the previous month
- The survey response time will be assessed
- If you have doubts regarding the indicated standard portion, you can consult the images (food indicated with a red arrow) as support

Next

Personal data:

Identification code:

tesis_user

Number of questionnaire:

1

Sex:

- ○ Male
- ○ Female
- ○ Prefer not to say

Age:

in years

Weight:

in kg

Height:

in cm

¿En qué Comunidad Autónoma reside actualmente?

If you are female, are you currently menstruating?

- ○ Yes
- ○ No
- ○ Menopause
- ○ I am not female

If you are female, are you currently pregnant?
- ○ Yes (please specify how many weeks pregnant)
- ◉ No
- ○ I am not female

How many weeks pregnant are you?

NA

What was your type of birth?
- ○ Vaginal delivery
- ○ Cesarian Section
- ○ I don't know

Do you smoke?
- ○ Yes
- ○ No
- ○ I used to

What is your blood type?
- ○ A
- ○ B
- ○ AB
- ○ O
- ○ I don't know

What is your Rh factor (rhesus)?
- ○ (+)
- ○ (-)
- ○ I don't know

Do you follow any specific type of diet?
- ○ No
- ○ Yes (please specify)

Do you eat ready-to-eat or pre-cooked store-bought foods such as pizza, lasagna, hamburgers?
- ○ Yes
- ○ No

Do you eat foods with sweeteners such as saccharin, sucralose, aspartame, acesulfame, or stevia?
- ○ Yes
- ○ No

Do you work in a healthcare field? (hospital, residence, clinic, outpatient, etc.)
- ○ Yes
- ○ No

Have you taken antibiotics in the last 2 months?
○ Yes
○ No

Do you suffer from any kind of disease?
○ No
○ Yes (please specify)

Have you suffered COVID-19?
○ Yes, once
○ Yes, more than once
○ No

Have you been diagnosed with long COVID?
○ Yes
○ No

What total amount of liquids do you drink approximately throughout the day including water, infusions, coffee, milk, vegetable drinks, juices, soft drinks, beer or other alcoholic beverages?

in L

In relation to your intestinal transit, which option is more adjusted to the frequency of your bowel movements?

Do you know how many steps have you walked on average daily the last month? If you have a smart device (mobile, watch, etc) please indicate the value that appears on it:

Steps per day (If you don't know please write 0)

At the end of the study (approximately 2-3 years), do you want to know the results of your microbiota profile?
○ Yes
○ No

Previous Next

VEGETABLES, LEGUMES AND POTATOES

Indicate how much you consume of each food presented below.
Then indicate the frequency of consumption of that food during the PREVIOUS MONTH

1/ Raw leafy vegetables such as spinach, lettuce, endive, celery, lamb's lettuce, bean sprout, green bean, ...
(standard portion: 1/2 dish)

How much of the indicated standard portion do you eat?

How often have you eaten this food during the past month?

2/ Cooked leafy vegetables such as spinach, chard, cabbage, celery, asparagus, bean sprouts
(standard portion: 1/2 dish)

How much of the indicated standard portion do you eat?

How often have you consumed this food in the past month?

3/ Tomato
(standard portion: 1 unit)

How much of the indicated standard portion do you eat?

How often have you consumed this food in the past month?

4/ Onion, spring onions or leek
(standard portion: 1/2 unit)

How much of the indicated standard portion do you eat?

How often have you consumed this food in the past month?

5/ Zucchini, aubergine or cucumber
(standard portion: 1/2 unit)

How much of the indicated standard portion do you eat?

How often have you consumed this food in the past month?

6/ Carrot, pumpkin or beet
(standard portion: 1/2 dish)

How much of the indicated standard portion do you eat?

How often have you consumed this food in the past month?

Image II. sFFQ used in the Population Study developed in our lab (English version is shown)

ANNEX 4. Median and percentiles in g/day of the different sFFQ (baseline, six and twelve months) food groups, energy and nutrients

Food group (g/day)	sFFQ1 (n= 1017)			sFFQ2 (n= 844)			sFFQ3 (n= 754)		
	P25	Median	P75	P25	Median	P75	P25	Median	P75
Alcoholic beverage	6.60	28.38	90.30	6.60	28.38	90.30	6.60	28.38	90.30
Appetizers	1.65	3.30	10.50	1.24	3.30	10.50	1.65	3.30	10.50
Biscuits breakfast cereals and cereal bars	1.65	6.45	22.50	1.65	5.28	22.50	1.65	6.30	23.10
Chocolates and derivatives	0.40	1.58	7.68	0.40	2.52	7.68	0.40	2.52	7.68
Fats and oils	10.00	17.14	30.00	10.00	13.53	30.00	10.00	15.32	30.00
Fish and shellfish	30.33	54.00	81.30	29.70	54.00	81.30	27.39	51.06	75.63
Fruit and fruit products	141.80	225.60	608.40	136.40	220.76	604.20	136.40	225.6	605.19
Legumes	19.80	41.40	57.15	19.80	41.40	63.00	16.7625	36.45	57.15
Meats and eggs	56.70	85.75	139.50	54.30	84.00	136.58	55.65	84.00	139.97
Milk and dairy products except fermented milk	24.30	104.95	221.00	25.00	107.16	226.66	25.12	110.5	228.56
Non-Alcoholic beverage	100.00	207.40	356.60	100.00	196.40	348.95	100.00	199.5	340.05
Nuts and seeds	3.15	9.60	30.00	3.15	12.60	30.00	3.15	9.60	30.00
Pastries and sweets breads	0.00	1.65	3.30	0.00	3.30	5.25	0.00	3.30	5.25
Potatoes and other tubers	9.90	31.50	48.00	9.90	31.50	48.00	9.90	31.50	31.50
Ready to eat meals	0.00	13.20	19.80	0.00	9.90	19.80	0.00	6.60	13.20
Sauces and condiments	1.32	2.76	6.84	1.32	2.76	7.89	1.32	2.76	6.11
Sausages and other meat products	1.49	4.72	9.45	1.49	4.72	9.45	1.49	4.72	9.45
Sugars and other sweets	0.00	1.32	6.30	0.00	1.05	5.99	0.00	1.32	6.40
Vegetables and vegetable products	205.50	353.35	551.20	198.52	314.99	521.61	176.17	304.28	488.37
White bread	0.00	14.70	70.00	2.31	14.70	70.00	2.31	14.70	70.00

Food group (g/day)	sFFQ1 (n= 1017)			sFFQ2 (n= 844)			sFFQ3 (n= 754)		
	P25	Median	P75	P25	Median	P75	P25	Median	P75
White grains and white pastas	5.30	16.85	25.30	5.30	16.85	25.30	5.3	16.85	25.3
Wholegrain or whole meal bread	0.00	14.70	44.80	0.00	14.70	44.80	2.31	14.70	44.80
Whole meal grains and whole meal pastas	0.00	2.64	8.40	0.00	2.64	8.40	0.00	2.64	8.40
Yogurt and fermented milk	8.25	26.25	125.00	8.25	26.25	125.00	8.25	26.25	125.00
Energy and nutrients									
Energy (kcal/day)	1366.50	1787.03	2300.91	1352.86	1722.34	2210.98	1287.42	1703.74	2232.29
Total fat (g/day)	50.83	70.17	95.39	49.55	68.02	91.18	48.91	67.46	96.89
Total protein (g/day)	62.96	80.99	102.86	61.31	78.43	100.16	58.86	77.93	98.12
Total water (g/day)	1056.24	1446.74	1874.13	1020.87	1348.21	1784.52	955.9	1326.47	1743.79
Total fiber (g/day)	19.12	27.60	39.44	18.62	27.40	37.12	18.31	26.40	37.44
Total carbohydrates (g/day)	132.21	187.50	250.34	129.29	178.12	241.90	127.98	171.93	237.90
Alcohol (g/day)	0.62	1.97	4.76	0.62	1.97	4.66	0.62	1.85	4.57
MUFA (g/day)	21.05	30.18	42.30	20.99	29.31	39.54	20.35	29.59	42.03
PUFA (g/day)	8.93	12.69	17.11	8.78	12.40	16.82	8.71	12.25	16.82
SFA (g/day)	14.67	20.57	28.44	14.66	20.32	27.75	14.47	20.35	27.73
Cholesterol (mg/day)	188.76	276.26	379.15	186.51	267.05	376.76	185.77	268.86	381.51
Vitamin A µg retinol (eq/day)	711.37	1051.82	1493.39	691.70	987.26	1436.48	638.18	961.25	1438.45
Vitamin D (µg/day)	4.00	7.36	12.59	4.07	7.67	12.40	3.79	7.32	13.05
Vitamin E (mg α tocoferol/day)	9.67	13.92	19.77	9.58	13.71	18.57	9.27	13.33	19.14
Folate total (µg/day)	331.73	471.83	660.77	325.09	457.18	627.32	305.24	437.21	639.07
Total niacin equivalent (mg/day)	45.22	82.51	153.76	46.79	87.83	152.51	45.68	81.79	152.29
Riboflavin (mg/day)	1.55	1.99	2.68	1.50	1.97	2.63	1.42	1.94	2.57
Tiamin (mg/day)	1.11	1.53	2.09	1.07	1.48	2.06	1.07	1.47	2.02

Food group (g/day)	sFFQ1 (n= 1017)			sFFQ2 (n= 844)			sFFQ3 (n= 754)		
	P25	Median	P75	P25	Median	P75	P25	Median	P75
Vitamin B12 (µg/day)	3.67	4.91	6.86	3.70	5.12	7.00	3.62	4.91	7.11
Vitamin B6 (mg/day)	1.92	2.59	3.51	1.90	2.53	3.47	1.80	2.52	3.42
Vitamin C ascorbic acid (mg/day)	142.28	231.37	354.56	133.48	228.27	335.45	131.91	227.00	343.86
Calcium (mg/day)	645.33	899.06	1221.27	640.25	872.98	1164.24	621.30	863.52	1190.04
Iron (mg/day)	11.91	15.78	21.09	11.56	15.31	20.39	11.24	14.99	20.44
Potassium (mg/day)	2827.65	3798.22	4966.56	2739.28	3686.23	4761.58	2582.40	3533.12	4760.63
Magnesium (mg/day)	357.63	472.54	618.85	344.70	461.46	602.43	334.83	463.15	612.42
Sodium (mg/day)	1106.42	1448.96	1972.86	1091.95	1427.70	1912.09	1049.72	1420.13	1885.9
Phosphorus (mg/day)	1137.54	1443.64	1832.69	1084.81	1419.78	1794.50	1052.01	1381.24	1780.49
Iodine (µg/day)	80.40	107.32	143.50	79.77	104.94	137.00	76.38	99.45	139.61
Selenium (µg/day)	55.46	74.38	97.10	54.47	72.97	95.22	53.94	71.87	96.69
Zinc (mg/day)	7.31	9.36	12.63	6.91	9.20	12.11	6.86	9.31	12.40
Sugar (g/day)	49.79	68.12	89.37	48.05	63.92	85.26	47.33	63.25	84.69

ANNEX 5. Significant association (q< 0.05) between food groups and categorical variables (Bowel frequency, gender, region areas, sweeteners intake and smoke). Output from MaAsLin2

Food group	*Variable*	*value*	*coef*	*stderr*	*N*	*p-value*	*q-value*
Sausages and other meat products	Bowel_Frequency	+ de 3 times per week	0.445466295	0.130018659	2,615	0.000622673	0.005534874
White bread	Bowel_Frequency	+ de 3 times per week	0.539029371	0.174467728	2,615	0.00202804	0.01622432
Biscuits and breakfast cereals	Bowel_Frequency	2 times per day	0.412387519	0.148223502	2,615	0.005442577	0.036794889
Vegetables	Gender	Female	0.351821619	0.067355492	2,615	2.16659E-07	4.95222E-06
Alcoholic beverage	Gender	Female	-0.948494891	0.194031532	2,615	1.19423E-06	2.20474E-05
Fats and oils	Gender	Female	0.330405412	0.069728644	2,615	2.50031E-06	4.13845E-05
Fish and shellfish	Gender	Female	0.282608847	0.076713858	2,615	0.000242781	0.002710114
Fruits and fruit products	Gender	Female	0.361588643	0.099285353	2,615	0.00028562	0.003046613
White grains	Gender	Female	-0.383607263	0.118591921	2,615	0.001260419	0.010431051
Ready to eat meals	Gender	Female	-0.42088455	0.136272857	2,615	0.002070452	0.016292079
Whole bread	Gender	Female	0.548765475	0.185811653	2,615	0.003221632	0.023790516
Non-alcoholic drinks	Gender	Female	0.31884346	0.108603992	2,615	0.003410847	0.024435919
White bread	Gender	Female	-0.505154933	0.186196312	2,615	0.006788412	0.042451216
Appetizers	Region_Areas	North Spain	-0.776274034	0.207792293	2,615	0.000198252	0.002321002
Legumes	Region_Areas	Interior	0.302846289	0.101780386	2,615	0.002999673	0.02285465
White grains	Region_Areas	North Spain	-0.540023708	0.183794968	2,615	0.003382865	0.024435919
Fats and oils	Region_Areas	North Spain	-0.301513484	0.107986136	2,615	0.005348892	0.036678114
Biscuits and breakfast cereals	Region_Areas	North Spain	0.653107292	0.237353769	2,615	0.00604543	0.040302864
White grains	Region_Areas	Interior	-0.387329081	0.142829504	2,615	0.006809883	0.042451216

Food group	*Variable*	*value*	*coef*	*stderr*	*N*	*p-value*	*q-value*
Milk and dairy	Region_Areas	Interior	0.485992005	0.182589524	2,615	0.00789879	0.047992651
Fruits and fruit products	Smoke	Yes	-0.725710208	0.151048326	2,615	1.68762E-06	2.998E-05
Alcoholic beverage	Smoke	Yes	1.20098491	0.268323248	2,615	8.01947E-06	0.000106926
Alcoholic beverage	Smoke	Former smoker	0.523539204	0.148649743	2,615	0.000436515	0.004276068
Biscuits and breakfast cereals	Smoke	Former smoker	-0.421024784	0.142236626	2,615	0.00311165	0.023337375
Ready to eat meals	Sweeteners	Yes	0.510985512	0.103763276	2,615	9.04328E-07	1.80866E-05
Sauces and condiments	Sweeteners	Yes	0.329752664	0.098289236	2,615	0.000806487	0.007033302
Sausages and other meat products	Sweeteners	Yes	0.298350927	0.102279307	2,615	0.003567423	0.025181813

ANNEX 6. Correlation values between Eating Quality Indices (DQIs), food groups, food items and numerical personal data with bacterial α- diversity (Chao1 and Shannon) using the Spearman correlation test (n= 500). Just data with FDR < 0.05 was represented.

Var1	*Var2*	*p-value*	*q-value*	*R*
Item1	Shannon	1.7626E-05	0.00204459	0.19126924
Item1	Chao1	0.00108766	0.01615781	0.14611725
Item2	Shannon	0.00098413	0.01615781	0.14736538
Item2	Chao1	0.00139291	0.01615781	0.14298807
Item3	Shannon	0.00419294	0.02702114	0.12822749
Item12	Chao1	0.00214904	0.02011404	-0.1373484
Item14	Shannon	0.00053283	0.01545214	0.15481954
Item16	Shannon	0.00510496	0.03116712	0.1254316
Item17	Shannon	0.00812632	0.04037908	0.11860345
Item18	Chao1	0.00357004	0.02436024	-0.1304733
Item18	Shannon	0.00087883	0.01615781	-0.148766
Item22	Chao1	0.00118785	0.01615781	-0.1450096
Item22	Shannon	0.00041122	0.01545214	-0.1578695
Item36	Chao1	0.00128257	0.01615781	0.14403893
Item37	Chao1	0.00623302	0.03286503	0.12253922
Item39	Chao1	0.00301825	0.02188232	0.13278239
Item41	Chao1	0.00617957	0.03286503	-0.1226652
Item41	Shannon	0.00200225	0.02011404	-0.1382826
Item44	Chao1	0.00835429	0.04037908	-0.1181865
Item45	Shannon	0.00569089	0.03286503	0.12386469
Item46	Shannon	0.00902851	0.04189229	-0.1170102
Item51	Shannon	0.0027128	0.02188232	-0.1342315

Var1	*Var2*	*p-value*	*q-value*	*R*
Item51	Chao1	0.00225416	0.02011404	-0.1367146
Item58	Shannon	0.0030038	0.02188232	-0.1328479
Item58	Chao1	0.00049553	0.01545214	-0.1556795
Fruits_and_Fruit_Products	Shannon	7.80E-05	0.00374	0.17626
VitC_mg	Shannon	0.00021	0.01359	0.16576
Age	Chao1	0.00072	0.00286	0.15129
hPDI_Adj	Shannon	0.0026	0.01838	0.13481
HEI_2015	Shannon	0.00306	0.01838	0.13258
aMED	Shannon	0.00467	0.021	0.12671
Chocolates_and_Derivatives	Shannon	0.00569	0.03035	0.12386
Age	Shannon	0.00644	0.01288	0.12206
Nuts_and_Seeds	Shannon	0.00813	0.03507	0.1186
Vegetables	Shannon	0.00877	0.03507	0.11745
Fish_and_Shellfish	Chao1	0.01313	0.04849	0.11118
BMI	Chao1	0.01971	0.01971	-0.10457
BMI	Shannon	0.01063	0.01417	-0.1145
uPDI_Adj	Chao1	0.01033	0.0372	-0.11494
Pastries_and_Sweet_Breads	Chao1	0.00835	0.03507	-0.11819
White_Bread	Chao1	0.00357	0.02142	-0.13047
Ready_To_Eat_Meals	Shannon	0.003	0.0206	-0.13285
Potatoes_and_Other_Tubercules	Chao1	0.00215	0.01719	-0.13735
uPDI_Adj	Shannon	0.00164	0.01838	-0.14089
White_Grains	Chao1	0.00119	0.0114	-0.14501
White_Bread	Shannon	0.00088	0.01055	-0.14877
Ready_To_Eat_Meals	Chao1	0.0005	0.00793	-0.15568
White_Grains	Shannon	0.00041	0.00793	-0.15787

ANNEX 7. Significant association obtained by Spearman correlations (FDR < 0.05) between age and BMI with bacterial species

Var1	*Var2*	*p-value*	*q-value*	*R*
Age	*GGB9568_SGB14980*	3.4556E-07	0.00036836	0.22621922
Age	*GGB3570_SGB4777*	5.7474E-07	0.0004595	0.22201337
Age	*GGB9677_SGB15180*	1.8959E-06	0.00121261	0.21180827
Age	*Gemmiger_SGB15295*	5.8852E-05	0.01459324	0.1791941
Age	*GGB9697_SGB15213*	6.1779E-05	0.01459324	0.17869139
Age	*GGB3617_SGB4891*	0.00010704	0.01901723	0.17290006
Age	*Desulfovibrio_fairfieldensis*	0.00027655	0.03049713	0.16243605
Age	*GGB9559_SGB14969*	0.00030699	0.03184321	0.16124584
Age	*Akkermansia_muciniphila*	0.00030867	0.03184321	0.16118334
Age	*Roseburia_sp_AM59_24XD*	0.0003308	0.03303118	0.16038953
Age	*GGB9635_SGB15106*	0.0003598	0.03303118	0.15942052
Age	*Eggerthellaceae_unclassified_SGB14322*	0.00036087	0.03303118	0.15938621
Age	*Ruminococcaceae_unclassified_SGB15260*	0.00039249	0.03303118	0.15841236
Age	*GGB9636_SGB15108*	0.00044414	0.03332468	0.15696865
Age	*GGB2983_SGB3965*	0.00045179	0.03332468	0.15676806
Age	*GGB9557_SGB14966*	0.00047158	0.03332468	0.15626405
Age	*GGB9631_SGB15085*	0.00047934	0.03332468	0.15607161
Age	*Bacilli_bacterium*	0.00076773	0.04589179	0.1504232
Age	*Blautia_sp_OF03_15BH*	0.00077491	0.04589179	0.15030957
Age	*GGB9603_SGB15035*	0.0008249	0.04644983	0.14954458
Age	*Mediterraneibacter_sp_gm002*	0.00082986	0.04644983	0.14947104
Age	*GGB9522_SGB14921*	0.00084213	0.04644983	0.1492909
Age	*Flavonifractor_plautii*	0.000747	0.04589179	-0.1507567

Var1	*Var2*	*p-value*	*q-value*	*R*
Age	*Hydrogeniiclostidium_mannosilyticum*	0.00052737	0.03588336	-0.154942
Age	*Clostridiales_bacterium*	0.00047161	0.03332468	-0.1562632
Age	*Holdemania_sp_Marseille_P2844*	0.00044068	0.03332468	-0.1570603
Age	*Enterocloster_aldensis*	0.00036894	0.03303118	-0.1591304
Age	*Pseudoflavonifractor_capillosus*	0.00027332	0.03049713	-0.1625695
Age	*Clostridium_sp_SN20*	0.00020649	0.0264143	-0.1657233
Age	*Dysosmobacter_welbionis*	0.00016089	0.02220217	-0.1684831
Age	*Clostridia_bacterium_UC5_1_1D1*	0.00015868	0.02220217	-0.1686347
Age	*Bifidobacterium_bifidum*	1.8859E-05	0.00670107	-0.190612
Age	*GGB9619_SGB15067*	6.4451E-06	0.00343525	-0.2007976
BMI	*Clostridium_fessum*	7.7009E-09	2.4628E-05	0.2553743
BMI	*Coprococcus_comes*	4.8078E-08	7.6877E-05	0.24180352
BMI	*Blautia_wexlerae*	1.5502E-05	0.00619684	0.19251148
BMI	*Roseburia_intestinalis*	6.3885E-05	0.01459324	0.17834329
BMI	*Ruminococcus_torques*	9.417E-05	0.01882226	0.17426623
BMI	*GGB9512_SGB14909*	0.00010414	0.01901723	0.17319359
BMI	*Collinsella_aerofaciens*	0.00016662	0.02220217	0.16809861
BMI	*Phocaeicola_vulgatus*	0.00039242	0.03303118	0.15841446
BMI	*Eubacterium_ramulus*	0.00045909	0.03332468	0.15657973
BMI	*Lancefieldella_parvula*	0.00056542	0.03690208	0.15411293
BMI	*Lacrimispora_celerecrescens*	0.00066977	0.04283841	0.15207999
BMI	*Dorea_sp_AF24_7LB*	0.00077364	0.04589179	0.1503295
BMI	*Clostridium_sp_AT4*	0.00088268	0.04784403	0.14871221
BMI	*GGB9608_SGB15041*	0.00084243	0.04644983	-0.1492865
BMI	*GGB6649_SGB9391*	0.00055554	0.03690208	-0.1543231
BMI	*Bacilli_bacterium*	0.00047016	0.03332468	-0.1562995
BMI	*GGB3643_SGB4948*	0.00037796	0.03303118	-0.1588504

Var1	*Var2*	*p-value*	*q-value*	*R*
BMI	*Intestinimonas_butyriciproducens*	0.00022378	0.02650548	-0.1648247
BMI	*Intestinimonas_massiliensis*	0.00021563	0.02650548	-0.16524
BMI	*Candidatus_Borkfalkia_ceftriaxoniphila*	0.00014566	0.02218212	-0.1695709
BMI	*Clostridia_bacterium*	0.00014489	0.02218212	-0.169629
BMI	*Ruminococcaceae_bacterium*	0.00012257	0.02063031	-0.1714437
BMI	*Intestinimonas_gabonensis*	7.9779E-05	0.01700898	-0.1760199
BMI	*GGB9758_SGB15368*	5.51E-05	0.01459324	-0.1798744
BMI	*Ruminococcaceae_bacterium_D5*	5.1179E-05	0.01459324	-0.1806335
BMI	*Lawsonibacter_sp_NSJ_51*	1.0304E-05	0.00470749	-0.1964108

ANNEX 8. Significant association (q< 0.05) between functional pathways and categorical variables (Bowel frequency, gender, region areas, sweeteners intake and smoke). Output from MaAsLin2

Feature	*Metadata*	*Value*	*Coef*	*Stderr*	*N*	*p-value*	*q-value*
TRNA.CHARGING.PWY..tRNA.charging	BMI	BMI	-569.85	99.08	498	1.83E-08	0.000115631
NONMEVIPP.PWY..methylerythritol.phosphate.pathway.I	BMI	BMI	-412.73	74.03	498	4.70E-08	0.000148817
PWY.3841..folate.transformations.II..plants.	BMI	BMI	-448.01	83.29	498	1.32E-07	0.000166652
PWY0.1586..peptidoglycan.maturation..meso.diaminopimelate.containing.	BMI	BMI	397.21	73.54	498	1.17E-07	0.000166652
PWY66.399..gluconeogenesis.III	BMI	BMI	-179.72	32.99	498	9.22E-08	0.000166652
PWY.7383..anaerobic.energy.metabolism..invertebrates..cytosol.	BMI	BMI	-290.96	55.43	498	2.54E-07	0.000268548
PWY.6609..adenine.and.adenosine.salvage.III	BMI	BMI	-484.62	93.74	498	3.79E-07	0.000343461
PWY.6163..chorismate.biosynthesis.from.3.dehydroquinate	BMI	BMI	-372.33	75.17	498	1.09E-06	0.000870322
PWY.7953..UDP.N.acetylmuramoyl.pentapeptide.biosynthesis.III..meso.diaminopimelate.containing.	BMI	BMI	-410.52	83.42	498	1.28E-06	0.000905666
PEPTIDOGLYCANSYN.PWY..peptidoglycan.biosynthesis.I..meso.diaminopimelate.containing.	BMI	BMI	-401.90	87.03	498	5.32E-06	0.002246376
PWY.6385..peptidoglycan.biosynthesis.III..mycobacteria.	BMI	BMI	-366.56	81.68	498	9.55E-06	0.003074491
PWY.6387..UDP.N.acetylmuramoyl.pentapeptide.biosynthesis.I..meso.diaminopimelate.containing.	BMI	BMI	-380.37	84.75	498	9.55E-06	0.003074491
PWY.6386..UDP.N.acetylmuramoyl.pentapeptide.biosynthesis.II..lysine.containing.	BMI	BMI	-371.33	84.09	498	1.31E-05	0.003619992
PWY.5686..UMP.biosynthesis.I	BMI	BMI	-356.74	82.37	498	1.90E-05	0.004154562
PWY.7790..UMP.biosynthesis.II	BMI	BMI	-356.74	82.37	498	1.90E-05	0.004154562
PWY.7791..UMP.biosynthesis.III	BMI	BMI	-356.74	82.37	498	1.90E-05	0.004154562
PWY.7199..pyrimidine.deoxyribonucleosides.salvage	BMI	BMI	-356.53	83.51	498	2.47E-05	0.005063055
PWY.6292..superpathway.of.L.cysteine.biosynthesis..mammalian.	BMI	BMI	-134.20	31.95	498	3.33E-05	0.006391338
PWY0.1296..purine.ribonucleosides.degradation	BMI	BMI	-370.59	88.43	498	3.46E-05	0.006452706
ARO.PWY..chorismate.biosynthesis.I	BMI	BMI	-260.56	62.32	498	3.61E-05	0.006532945

Feature	Metadata	Value	Coef	Stderr	N	p-value	q-value
PWY.8190..L.glutamate.degradation.XI..reductive.Stickland.reaction.	BMI	BMI	14.72	3.59	498	5.04E-05	0.008639222
ARGSYN.PWY..L.arginine.biosynthesis.I..via.L.ornithine.	BMI	BMI	-315.74	78.91	498	7.57E-05	0.012291701
ARGSYNBSUB.PWY..L.arginine.biosynthesis.II..acetyl.cycle.	BMI	BMI	-335.21	84.96	498	9.48E-05	0.015000732
PWY.822..fructan.biosynthesis	BMI	BMI	19.45	4.94	498	9.71E-05	0.015000732
PWY.5088..L.glutamate.degradation.VIII..to.propanoate.	BMI	BMI	4.10	1.04	498	0.0001	0.015224101
KETOGLUCONMET.PWY..ketogluconate.metabolism	BMI	BMI	39.70	10.52	498	0.0001	0.024464112
P125.PWY..superpathway.of..R.R..butanediol.biosynthesis	BMI	BMI	34.45	9.15	498	0.0001	0.024764044
PWY.7446..sulfoquinovose.degradation.I	BMI	BMI	6.03	1.61	498	0.0002	0.025156109
PWY.5097..L.lysine.biosynthesis.VI	BMI	BMI	-204.35	54.95	498	0.0002	0.025603777
PWY.7184..pyrimidine.deoxyribonucleotides.de.novo.biosynthesis.I	BMI	BMI	111.93	30.16	498	0.0002	0.0258482
X1CMET2.PWY..folate.transformations.III..E..coli.	BMI	BMI	-179.04	48.57	498	0.0002	0.026203758
PWY.801..homocysteine.and.cysteine.interconversion	BMI	BMI	28.49	7.75	498	0.0002	0.026759511
PWY.6293..superpathway.of.L.cysteine.biosynthesis..fungi.	BMI	BMI	38.57	10.70	498	0.0003	0.033432436
PWY.6545..pyrimidine.deoxyribonucleotides.de.novo.biosynthesis.III	BMI	BMI	103.79	29.10	498	0.0004	0.037450726
POLYISOPRENSYN.PWY..polyisoprenoid.biosynthesis..E..coli.	BMI	BMI	109.10	30.81	498	0.0004	0.039379612
PWY.2942..L.lysine.biosynthesis.III	BMI	BMI	-178.62	50.44	498	0.0004	0.039379612
PWY.6700..queuosine.biosynthesis.I..de.novo.	BMI	BMI	-338.13	96.11	498	0.0004	0.041672094
COA.PWY.1..superpathway.of.coenzyme.A.biosynthesis.III..mammals.	BMI	BMI	-233.93	66.75	498	0.0005	0.043220594
PWY.6859..all.trans.farnesol.biosynthesis	BMI	BMI	80.44	23.15	498	0.0005	0.046202201
PWY.7383..anaerobic.energy.metabolism..invertebrates..cytosol.	Bowel_movement	>3_WK	711.63	153.70	498	5.03E-06	0.002246376
PWY66.399..gluconeogenesis.III	Bowel_movement	>3_WK	424.84	91.49	498	4.72E-06	0.002246376
FUCCAT.PWY..fucose.degradation	Bowel_movement	1.5_WK	452.02	107.57	498	3.29E-05	0.006391338
PWY.6167..flavin.biosynthesis.II..archaea.	Bowel_movement	>3_WK	79.04	19.33	498	5.29E-05	0.008813521
PWY.6151..S.adenosyl.L.methionine.salvage.I	Bowel_movement	1.5_WK	-1477.86	382.44	498	0.0001	0.019287286
GLUDEG.I.PWY..GABA.shunt	Bowel_movement	1.5_WK	351.70	94.04	498	0.0002	0.025156109
PWY.5104..L.isoleucine.biosynthesis.IV	Bowel_movement	1.5_WK	159.30	42.62	498	0.0002	0.025156109
PWY.5981..CDP.diacylglycerol.biosynthesis.III	Bowel_movement	1.5_WK	530.97	141.91	498	0.0002	0.025156109

Feature	Metadata	Value	Coef	Stderr	N	p-value	q-value
PWY.7434..terminal.O.glycans.residues.modification..via.type.2.precursor.disaccharide.	Bowel_movement	1.5_WK	105.79	28.27	498	0.0002	0.025156109
PWY.5022..4.aminobutanoate.degradation.V	Bowel_movement	1.5_WK	347.97	93.40	498	0.0002	0.025603777
PWY.6317..D.galactose.degradation.I..Leloir.pathway.	Bowel_movement	>3_WK	-618.31	166.16	498	0.0002	0.025603777
FUC.RHAMCAT.PWY..superpathway.of.fucose.and.rhamnose.degradation	Bowel_movement	1.5_WK	397.73	108.70	498	0.0002	0.028142403
PWY.7383..anaerobic.energy.metabolism..invertebrates..cytosol.	Bowel_movement	1.5_WK	919.90	257.93	498	0.0004	0.037450726
PWY.6518..bile.acids.epimerization	Bowel_movement	>3_WK	6.99	1.98	498	0.0004	0.040066536
P162.PWY..L.glutamate.degradation.V..via.hydroxyglutarate.	Bowel_movement	>2_Day	24.06	6.94	498	0.0005	0.047003732
PWY.6527..stachyose.degradation	Bowel_movement	>3_WK	-733.76	212.54	498	0.0006	0.048925417
BIOTIN.BIOSYNTHESIS.PWY..biotin.biosynthesis.I	Gender	m	-422.26	111.37	498	0.0001	0.023470741
PENTOSE.P.PWY..pentose.phosphate.pathway	Gender	m	-469.46	126.66	498	0.0002	0.025908806
FASYN.ELONG.PWY..fatty.acid.elongation....saturated	Gender	m	-509.30	137.86	498	0.0002	0.026203758
PWY.6519..8.amino.7.oxononanoate.biosynthesis.I	Gender	m	-422.21	114.29	498	0.0002	0.026203758
PWY.7664..oleate.biosynthesis.IV..anaerobic.	Gender	m	-471.08	132.44	498	0.0004	0.038229832
PWY.5989..stearate.biosynthesis.II..bacteria.and.plants.	Gender	m	-414.65	118.82	498	0.0005	0.04477563
PWY.6282..palmitoleate.biosynthesis.I..from..5Z..dodec.5.enoate.	Gender	m	-433.57	124.32	498	0.0005	0.04477563
PWY.6160..3.dehydroquinate.biosynthesis.II..archaea.	Region_Areas	Islands	160.15	34.62	498	5.12E-06	0.002246376
PWY.6349..CDP.archaeol.biosynthesis	Region_Areas	Islands	150.94	32.64	498	5.16E-06	0.002246376
PWY.6350..archaetidylinositol.biosynthesis	Region_Areas	Islands	140.95	30.50	498	5.23E-06	0.002246376
PWY.6165..chorismate.biosynthesis.II..archaea.	Region_Areas	Islands	371.22	81.69	498	7.42E-06	0.002763245
PWY.7286..7..3.amino.3.carboxypropyl..wyosine.biosynthesis	Region_Areas	Islands	179.28	39.34	498	7.01E-06	0.002763245
PWY.1861..formaldehyde.assimilation.II..assimilatory.RuMP.Cycle.	Region_Areas	Islands	959.76	214.03	498	9.71E-06	0.003074491
PWY.8112..factor.420.biosynthesis.I..archaea.	Region_Areas	Islands	137.35	30.96	498	1.19E-05	0.003610002
METHANOGENESIS.PWY..methanogenesis.from.H2.and.CO2	Region_Areas	Islands	152.13	34.38	498	1.25E-05	0.003619992
PWY.5198..factor.420.biosynthesis.II..mycobacteria.	Region_Areas	Islands	138.45	31.45	498	1.39E-05	0.003685295
PWY.5209..methyl.coenzyme.M.oxidation.to.CO2	Region_Areas	Islands	123.72	28.52	498	1.83E-05	0.004154562
PWY.8113..3PG.factor.420.biosynthesis	Region_Areas	Islands	127.74	29.39	498	1.77E-05	0.004154562

Feature	*Metadata*	*Value*	*Coef*	*Stderr*	*N*	*p-value*	*q-value*
PWY.6270..isoprene.biosynthesis.I	Region_Areas	Islands	1326.38	319.91	498	4.17E-05	0.007342054
PWY.5188..tetrapyrrole.biosynthesis.I..from.glutamate.	Region_Areas	Islands	768.98	213.26	498	0.0003	0.033432436
PWY.6270..isoprene.biosynthesis.I	Season_Year	Autumn	-627.70	146.57	498	2.34E-05	0.004945873
PWY.6122..5.aminoimidazole.ribonucleotide.biosynthesis.II	Smoke	Former	620.68	162.14	498	0.0001	0.020827481
PWY.6277..superpathway.of.5.aminoimidazole.ribonucleotide.biosynthesis	Smoke	Former	620.68	162.14	498	0.0001	0.020827481
PWY.7315..dTDP.N.acetylthomosamine.biosynthesis	Smoke	YES	364.60	95.26	498	0.0001	0.020827481
PWY0.1261..anhydromuropeptides.recycling.I	Smoke	Former	-322.93	87.56	498	0.0002	0.026203758

ANNEX 9. Significant association obtained by Spearman correlations (FDR < 0.05) between DQIs, food groups, micro and macro-nutrients and items.

Var1	*Var2*	*p-value*	*q-value*	*R*
Appetizers	*Eubacterium_rectale*	8.42622E-06	0.017059168	0.198303634
Chocolates_and_Derivatives	*GGB52930_SGB73859*	3.73914E-06	0.013044843	0.205770389
Chocolates_and_Derivatives	*Clostridium_sp_AF32_12BH*	3.42191E-05	0.037519757	0.184720345
Chocolates_and_Derivatives	*GGB3478_SGB4643*	1.27109E-05	0.021389027	0.194416283
Fats_and_Oils	*GGB9644_SGB15121*	2.31809E-05	0.030675584	0.188591886
Fruits_and_Fruit_Products	*Bacilli_unclassified_SGB6473*	3.89458E-05	0.03946696	0.183416412
Fruits_and_Fruit_Products	*GGB4676_SGB6465*	1.42661E-06	0.006843467	0.214285085
Fruits_and_Fruit_Products	*GGB9758_SGB15368*	5.32511E-07	0.005108909	0.222649466
Fruits_and_Fruit_Products	*Lachnospira_eligens*	1.60431E-05	0.024626762	0.192180229
Fruits_and_Fruit_Products	*Lachnospiraceae_bacterium*	1.58896E-05	0.024626762	0.192273058
Legumes	*Haemophilus_parainfluenzae*	3.54164E-07	0.004530465	0.226017717
Meat_and_Eggs	*Dorea_formicigenerans*	1.91014E-06	0.008144855	0.211742621
Milk_and_Dairy	*Lachnospira_sp_NSJ_43*	8.47156E-06	0.017059168	-0.198253382
Nuts_and_Seeds	*Flavonifractor_plautii*	2.60805E-05	0.030796142	-0.187428556
Nuts_and_Seeds	*GGB3478_SGB4643*	8.09873E-07	0.006215934	0.219130915
Nuts_and_Seeds	*Roseburia_hominis*	4.54377E-05	0.04359289	0.181850852
Nuts_and_Seeds	*Lachnospiraceae_bacterium*	1.40747E-07	0.002700662	0.233452048
Pastries_and_Sweet_Breads	*Clostridium_saccharogumia*	1.28191E-05	0.021389027	-0.194335299
Pastries_and_Sweet_Breads	*GGB4713_SGB6526*	2.60796E-05	0.030796142	-0.187428874
Pastries_and_Sweet_Breads	*GGB9568_SGB14980*	5.58225E-06	0.016478804	-0.202122264
Pastries_and_Sweet_Breads	*Clostridium_sp_AF20_17LB*	1.27571E-06	0.006843467	-0.215250856
Ready_To_Eat_Meals	*Bacilli_bacterium*	5.42706E-05	0.049587866	-0.180030524

Var1	*Var2*	*p-value*	*q-value*	*R*
Ready_To_Eat_Meals	*Intestinimonas_gabonensis*	7.35286E-06	0.017059168	-0.199575395
Ready_To_Eat_Meals	*GGB9677_SGB15180*	1.01642E-06	0.006501037	-0.217199973
Ready_To_Eat_Meals	*GGB4584_SGB6338*	4.22723E-05	0.041595944	-0.182585749
Sugar_and_Other_Sweets	*Lactococcus_lactis*	2.6482E-05	0.030796142	-0.187277229
Vegetables	*Bacilli_unclassified_SGB6422*	3.90803E-05	0.03946696	0.183381562
Vegetables	*Clostridium_sp_AF20_17LB*	7.68432E-06	0.017059168	0.199164748
White_Bread	*Clostridium_saccharogumia*	6.85588E-06	0.017059168	-0.200225427
White_Bread	*GGB9615_SGB15053*	4.59467E-06	0.014693754	-0.203902891
White_Bread	*Blautia_stercoris*	3.74325E-05	0.03946696	-0.183816771
White_Bread	*Clostridium_sp_AF20_17LB*	2.22173E-05	0.030450351	-0.189009283
White_Bread	*Clostridium_sp_AF36_4*	2.08416E-05	0.02962288	-0.189635905
White_Bread	*GGB3490_SGB4664*	2.05852E-05	0.02962288	-0.189757025
White_Grains	*GGB9524_SGB14924*	2.74935E-05	0.031032029	-0.186905456
White_Grains	*GGB3570_SGB4777*	2.43911E-05	0.030796142	-0.188090447
White_Grains	*Mediterraneibacter_butyricigenes*	9.848E-06	0.017996521	-0.196838297
White_Grains	*GGB3490_SGB4664*	6.70978E-06	0.017059168	-0.200425082
White_Grains	*Lachnospiraceae_bacterium*	2.20925E-06	0.008478231	-0.210463714
Whole_Bread	*Lachnospiraceae_unclassified_SGB4924*	8.89054E-06	0.017059168	0.19780075
Whole_Grains	*Haemophilus_parainfluenzae*	5.41251E-05	0.049587866	0.180058178
Yogurt	*Streptococcus_thermophilus*	3.55715E-13	1.36509E-08	0.318441011
HEI_2015	*Faecalibacterium_prausnitzii*	6.62798E-05	0.027555146	0.177960539
HEI_2015	*Blautia_glucerasea*	7.93407E-05	0.029986462	0.176077926
HEI_2015	*Lachnospiraceae_unclassified_SGB4924*	2.76906E-05	0.017813831	0.186834524
HEI_2015	*Lachnospiraceae_bacterium*	0.000143502	0.042385469	0.169733715
IASE	*Prevotella_SGB1675*	8.57479E-05	0.031270908	-0.175259059
IASE	*Eubacterium_sp_AF22_8LB*	7.41872E-05	0.029096789	-0.176783125

Var1	Var2	p-value	q-value	R
MAR	Lachnospiraceae_bacterium	2.46882E-05	0.017696298	0.187970977
PDI_Adj	Dorea_formicigenerans	2.62874E-05	0.017813831	-0.187350303
PDI_Adj	Ruminococcus_torques	1.66468E-05	0.013309081	-0.191823091
hPDI_Adj	Clostridium_saccharogumia	1.46048E-06	0.003467721	0.214081836
hPDI_Adj	GGB4700_SGB6506	2.85664E-05	0.017813831	0.186525008
hPDI_Adj	Bacilli_bacterium	4.00944E-05	0.020327842	0.183122271
hPDI_Adj	Flavonifractor_plautii	0.000152928	0.042385469	-0.169039178
hPDI_Adj	Intestinimonas_gabonensis	1.37433E-06	0.003467721	0.214608134
hPDI_Adj	GGB9646_SGB15123	0.000128177	0.040991163	0.170960138
hPDI_Adj	GGB9677_SGB15180	8.78454E-06	0.010144678	0.197913312
hPDI_Adj	GGB4579_SGB6329	9.21064E-05	0.033010604	0.174501539
hPDI_Adj	GGB9758_SGB15368	0.000144821	0.042385469	0.169634026
hPDI_Adj	Blautia_sp_AF19_10LB	3.40674E-05	0.019139459	0.184764942
hPDI_Adj	Blautia_massiliensis	4.59832E-06	0.007063502	-0.203895661
hPDI_Adj	Ruminococcus_torques	1.76232E-06	0.003663335	-0.212447356
hPDI_Adj	Clostridium_sp_AF20_17LB	9.98475E-05	0.033841141	0.173643124
hPDI_Adj	Lachnospiraceae_bacterium	8.3555E-06	0.010144678	0.198382532
uPDI_Adj	Prevotellamassilia_timonensis	2.87002E-05	0.017813831	0.186478493
uPDI_Adj	GGB4700_SGB6506	0.000143928	0.042385469	-0.169701474
uPDI_Adj	GGB4585_SGB6340	0.000155412	0.042507309	-0.168862795
uPDI_Adj	GGB4603_SGB6367	5.51394E-05	0.023878811	-0.179866917
uPDI_Adj	GGB9758_SGB15368	0.000144792	0.042385469	-0.169636239
uPDI_Adj	Blautia_glucerasea	1.16469E-05	0.011273014	-0.195249341
uPDI_Adj	Lachnospiraceae_bacterium	4.75725E-06	0.007063502	-0.203586011
MDI_Adj	Dorea_formicigenerans	3.06501E-05	0.017813831	0.185823335
MDI_Adj	Ruminococcus_torques	1.93687E-05	0.01491174	0.19035194

Var1	Var2	p-value	q-value	R
PDI_Unadj	GGB9512_SGB14909	4.33478E-05	0.021454085	-0.182330372
PDI_Unadj	Dorea_sp_AF24_7LB	1.35939E-05	0.011774054	-0.193773848
PDI_Unadj	Ruminococcus_torques	3.65059E-06	0.006513113	-0.205986519
hPDI_Unadj	Bacteroides_cellulosilyticus	7.6781E-05	0.02955643	0.176422613
hPDI_Unadj	Clostridium_saccharogumia	2.17888E-05	0.016175843	0.189200414
hPDI_Unadj	GGB4700_SGB6506	1.16643E-06	0.003467721	0.216021194
hPDI_Unadj	Bacilli_bacterium	9.75019E-05	0.033779548	0.17389641
hPDI_Unadj	Flavonifractor_plautii	5.87258E-05	0.024912936	-0.179216314
hPDI_Unadj	Intestinimonas_gabonensis	1.14784E-05	0.011273014	0.195387873
hPDI_Unadj	GGB9677_SGB15180	7.0496E-08	0.001465401	0.238863229
hPDI_Unadj	GGB3570_SGB4777	0.000140958	0.042385469	0.169928591
hPDI_Unadj	Blautia_sp_AF19_10LB	3.8293E-05	0.019899896	0.183587295
hPDI_Unadj	Blautia_massiliensis	5.01729E-05	0.02317652	-0.180837212
hPDI_Unadj	Ruminococcus_torques	4.94786E-05	0.02317652	-0.180980031
hPDI_Unadj	Enterocloster_clostridioformis	0.000146893	0.042385469	-0.169479041
hPDI_Unadj	Clostridium_sp_AF20_17LB	7.25781E-05	0.029013079	0.177012804
hPDI_Unadj	Lachnospira_sp_NSJ_43	3.07452E-05	0.017813831	0.185792391
uPDI_Unadj	Prevotellamassilia_timonensis	1.24731E-05	0.011273014	0.194596528
uPDI_Unadj	GGB4603_SGB6367	7.20242E-05	0.029013079	-0.177092969
uPDI_Unadj	GGB9758_SGB15368	3.08509E-05	0.017813831	-0.185758116
uPDI_Unadj	Blautia_glucerasea	7.89925E-06	0.010144678	-0.198907397
uPDI_Unadj	GGB3571_SGB4778	0.000108097	0.035109546	-0.17279472
uPDI_Unadj	GGB3490_SGB4664	0.000151935	0.042385469	-0.169110422
uPDI_Unadj	Lachnospiraceae_bacterium	2.62019E-07	0.002624502	-0.228473239
MDI_Unadj	Akkermansia_muciniphila	0.000176108	0.046932775	-0.167488529
MDI_Unadj	GGB9512_SGB14909	1.16837E-05	0.011273014	0.195219379

Var1	Var2	p-value	q-value	R
MDI_Unadj	*Intestinimonas_gabonensis*	1.23739E-05	0.011273014	-0.194672719
MDI_Unadj	*Dorea_longicatena*	0.000107848	0.035109546	0.172819432
MDI_Unadj	*Dorea_formicigenerans*	9.47189E-05	0.033371542	0.174204488
MDI_Unadj	*Dorea_sp_AF24_7LB*	5.34402E-05	0.023635356	0.180189242
MDI_Unadj	*Ruminococcus_torques*	9.61104E-07	0.003467721	0.217677283
aMED	*Akkermansia_muciniphila*	0.000130477	0.041094225	0.17076763
aMED	*Haemophilus_parainfluenzae*	0.000100936	0.033841141	0.173527512
aMED	*Clostridium_saccharogumia*	0.000152705	0.042385469	0.169055149
aMED	*GGB4700_SGB6506*	4.90255E-05	0.02317652	0.181074252
aMED	*Bacilli_bacterium*	8.32398E-05	0.030898304	0.175572514
aMED	*Dysosmobacter_welbionis*	0.00018467	0.048591571	-0.166963861
aMED	*Flavonifractor_plautii*	1.64743E-05	0.013309081	-0.191923843
aMED	*GGB9646_SGB15123*	5.13119E-05	0.023187398	0.180606913
aMED	*GGB9677_SGB15180*	3.75991E-06	0.006513113	0.205720396
aMED	*Blautia_massiliensis*	3.70981E-05	0.019894494	-0.183907314
aMED	*Ruminococcus_torques*	3.78771E-07	0.002624502	-0.225466566
aMED	*Clostridium_sp_AF20_17LB*	1.03565E-06	0.003467721	0.217039895
aMED	*Clostridium_sp_AF36_4*	0.000173284	0.046779898	0.16766684
aMED	*Lachnospira_sp_NSJ_43*	3.73255E-05	0.019894494	0.183845658
aMED	*Coprococcus_eutactus*	1.50139E-06	0.003467721	0.213842274
aMED	*Lachnospiraceae_bacterium*	6.44717E-06	0.008934489	0.20079462
Item1	*GGB9730_SGB15291*	2.18839E-05	0.026357906	0.189157672
Item1	*Flavonifractor_plautii*	1.46516E-05	0.020588198	-0.193054647
Item1	*GGB9758_SGB15368*	6.00477E-05	0.049438112	0.178985953
Item1	*GGB3570_SGB4777*	5.3094E-08	0.000307753	0.241044787
Item1	*Lachnospiraceae_bacterium*	5.39132E-05	0.046688993	0.180098561

Var1	*Var2*	*p-value*	*q-value*	*R*
Item2	*Bacilli_unclassified_SGB6422*	7.78232E-06	0.013122681	0.199046573
Item6	*Blautia_massiliensis*	9.56712E-06	0.015844177	-0.19711106
Item6	*Ruminococcus_gnavus*	3.27775E-05	0.033776155	-0.185152102
Item8	*Haemophilus_pittmaniae*	4.77569E-05	0.04258723	0.181342511
Item8	*Ruminococcus_torques*	1.71118E-05	0.022041473	-0.191556244
Item10	*Massilimaliae_massiliensis*	5.67745E-06	0.01066726	-0.201966859
Item13	*Haemophilus_parainfluenzae*	5.2203E-06	0.010300865	0.20273715
Item14	*GGB4676_SGB6465*	4.04822E-06	0.008532719	0.205052577
Item14	*GGB9758_SGB15368*	1.3318E-06	0.003813748	0.214879683
Item14	*Lachnospira_eligens*	6.22698E-06	0.010896279	0.201115696
Item16	*Phocea_massiliensis*	1.63832E-05	0.021666744	-0.191977496
Item16	*Ruminococcus_torques*	6.06265E-06	0.010896279	-0.201362469
Item17	*Clostridium_saccharogumia*	2.46207E-05	0.028285692	0.187997996
Item17	*Flavonifractor_plautii*	1.27231E-05	0.019499585	-0.194407132
Item17	*GGB9677_SGB15180*	5.38138E-05	0.046688993	0.180117546
Item17	*GGB3478_SGB4643*	5.46192E-07	0.001809106	0.222438213
Item17	*Roseburia_hominis*	2.84835E-05	0.031077815	0.186553919
Item17	*Lachnospiraceae_bacterium*	6.21524E-08	0.000339067	0.23983518
Item18	*Bacteroides_cellulosilyticus*	2.80595E-05	0.030979655	-0.186703036
Item18	*Clostridium_saccharogumia*	5.53448E-06	0.01066726	-0.20220119
Item18	*GGB9615_SGB15053*	7.93296E-07	0.00253696	-0.219305833
Item18	*Blautia_stercoris*	6.02371E-05	0.049438112	-0.178953336
Item18	*Clostridium_sp_AF20_17LB*	2.71082E-05	0.03028995	-0.187045519
Item18	*Clostridium_sp_AF36_4*	1.15815E-05	0.018518893	-0.195302856
Item18	*GGB3490_SGB4664*	3.10278E-05	0.032699754	-0.18570101
Item18	*Lachnospiraceae_bacterium*	2.56633E-05	0.029025191	-0.187588137

Var1	Var2	p-value	q-value	R
Item19	Lachnospiraceae_unclassified_SGB4924	1.06429E-06	0.003290151	0.21680664
Item20	Massilioclostridium_coli	1.43742E-05	0.020588198	-0.193238361
Item22	GGB9524_SGB14924	7.31529E-06	0.012563602	-0.19962305
Item22	GGB9677_SGB15180	4.13249E-05	0.038628535	-0.182815852
Item22	GGB3570_SGB4777	4.39597E-06	0.008862857	-0.204305028
Item22	Mediterraneibacter_butyricigenes	2.89533E-05	0.031223127	-0.186391118
Item22	GGB3490_SGB4664	1.32673E-05	0.019845705	-0.194006756
Item22	Lachnospiraceae_bacterium	4.74823E-07	0.001693694	-0.223601735
Item24	GGB6521_SGB9212	1.64673E-05	0.021666744	0.191927975
Item24	GGB4751_SGB6580	5.58966E-05	0.047559269	0.179726299
Item24	Christensenella_sp_Marseille_P3954	2.77448E-06	0.00677134	0.208444974
Item26	Clostridium_innocuum	5.43703E-05	0.046688993	0.18001163
Item27	Intestinimonas_gabonensis	4.76337E-05	0.04258723	0.181368916
Item29	Streptococcus_thermophilus	1.48905E-06	0.004061679	0.213913918
Item29	Roseburia_sp_AF02_12	2.97529E-05	0.031716617	-0.186119729
Item29	Megasphaera_sp_BL7	2.01563E-05	0.02492453	-0.18996285
Item30	Streptococcus_thermophilus	4.06342E-14	5.766E-10	0.330328988
Item33	Dorea_formicigenerans	4.12749E-05	0.038628535	0.182828131
Item33	Coprococcus_comes	4.16516E-05	0.038628535	0.182735938
Item34	Dorea_formicigenerans	1.35703E-06	0.003813748	0.214717585
Item36	GGB9758_SGB15368	5.97785E-05	0.049438112	0.179032476
Item36	Lachnospiraceae_bacterium	5.75104E-06	0.01066726	0.201848431
Item39	GGB9512_SGB14909	3.51949E-05	0.035605735	0.184437755
Item39	Phocea_massiliensis	3.53209E-05	0.035605735	0.184401812
Item39	GGB3344_SGB4424	2.30411E-05	0.027395821	0.188651441
Item39	GGB4566_SGB6305	1.28256E-05	0.019499585	0.194330468

Var1	Var2	p-value	q-value	R
Item40	*GGB9644_SGB15121*	1.03846E-05	0.016896269	0.196337212
Item42	*GGB1228_SGB1601*	2.68855E-07	0.00113337	0.22826446
Item42	*Bacteroides_oleiciplenus*	2.62013E-14	4.85992E-10	0.332649449
Item42	*Bacteroides_sp_AF16_49*	0	0	0.466125572
Item42	*GGB6567_SGB9277*	3.60051E-05	0.035905239	0.18420869
Item42	*GGB6554_SGB9256*	3.21188E-07	0.001241151	0.226817147
Item42	*Brachyspira_aalborgi*	3.81552E-05	0.037193721	0.183623718
Item42	*Fusobacterium_gonidiaformans*	3.89014E-05	0.037193721	0.183427965
Item42	*GGB9284_SGB14237*	2.17495E-08	0.000144078	0.247779232
Item42	*Gemmiger_SGB15292*	3.28145E-06	0.007422647	0.206945028
Item42	*GGB3183_SGB4206*	2.38923E-07	0.00110791	0.229219687
Item42	*Ruminococcaceae_unclassified_SGB15257*	6.16319E-06	0.010896279	0.201210736
Item42	*GGB36331_SGB53806*	4.37299E-05	0.040154411	0.182241104
Item42	*GGB9709_SGB15238*	1.81946E-05	0.023115115	0.190960759
Item42	*GGB13472_SGB20758*	2.38923E-07	0.00110791	0.229219687
Item42	*Clostridia_unclassified_SGB20792*	2.67828E-06	0.006713208	0.208758998
Item42	*GGB4569_SGB6311*	7.18603E-11	7.40496E-07	0.286871164
Item42	*GGB3717_SGB5040*	3.81692E-06	0.008232301	0.205584552
Item42	*Mediterraneibacter_sp_NSJ_55*	2.62013E-14	4.85992E-10	0.332649449
Item42	*Roseburia_SGB4939*	2.68855E-07	0.00113337	0.22826446
Item42	*Roseburia_sp_MUC_MUC_530_WT_4D*	3.89014E-05	0.037193721	0.183427965
Item42	*GGB3751_SGB5099*	4.34783E-07	0.001612907	0.224330522
Item42	*Arthrospira_platensis*	4.9738E-14	5.766E-10	0.32926198
Item42	*GGB18384_SGB72500*	2.38923E-07	0.00110791	0.229219687
Item42	*Eggerthellaceae_unclassified_SGB63096*	4.9738E-14	5.766E-10	0.32926198
Item42	*Schaalia_turicensis*	5.08713E-05	0.044932409	0.180695424

Var1	*Var2*	*p-value*	*q-value*	*R*
Item44	*Clostridium_saccharogumia*	1.44869E-05	0.020588198	-0.193163344
Item44	*GGB4713_SGB6526*	3.2128E-06	0.007422647	-0.207134596
Item44	*GGB9568_SGB14980*	1.15761E-06	0.003463192	-0.216086402
Item44	*Blautia_sp_AF19_10LB*	1.51746E-05	0.021004767	-0.192717114
Item44	*Clostridium_sp_AF20_17LB*	5.31753E-07	0.001809106	-0.222661327
Item44	*Lachnospiraceae_bacterium*	3.18602E-05	0.033199732	-0.185436323
Item45	*GGB52930_SGB73859*	2.87456E-06	0.006835697	0.20812918
Item45	*Clostridium_sp_AF32_12BH*	3.71654E-05	0.036668043	0.183889017
Item45	*GGB3478_SGB4643*	1.27761E-05	0.019499585	0.194367402
Item50	*Clostridia_unclassified_SGB14844*	1.39709E-08	9.96686E-05	0.251047386
Item50	*GGB9494_SGB14891*	3.82299E-10	3.54552E-06	0.276067287
Item50	*Clostridia_bacterium_UC5_1_1E11*	2.83904E-08	0.000175532	0.245789219
Item50	*Clostridia_bacterium_12CBH8*	5.01923E-10	4.23176E-06	0.274262263
Item50	*Massilioclostridium_coli*	0	0	0.359032978
Item50	*GGB9619_SGB15067*	1.56142E-05	0.021295502	0.192441784
Item50	*Clostridium_phoceensis*	0	0	0.421144582
Item50	*GGB9557_SGB14966*	2.72301E-09	2.10448E-05	0.262741288
Item50	*GGB4566_SGB6305*	4.19294E-06	0.008641373	0.204734291
Item50	*GGB3570_SGB4777*	2.15562E-06	0.00555324	0.210680318
Item51	*Bacilli_unclassified_SGB6428*	2.3352E-05	0.027414085	-0.188519515
Item51	*Flavonifractor_plautii*	1.65873E-05	0.021666744	0.191857717
Item51	*GGB9677_SGB15180*	1.41945E-05	0.020588198	-0.193359209
Item51	*GGB9695_SGB15209*	1.83144E-06	0.004852897	0.212111058
Item51	*GGB4603_SGB6367*	6.0136E-05	0.049438112	-0.178970725
Item51	*GGB9758_SGB15368*	4.51931E-05	0.041091198	-0.181905868
Item57	*Eubacterium_rectale*	3.50828E-06	0.00774679	0.206344559

Var1	*Var2*	*p-value*	*q-value*	*R*
Item58	*Bacilli_bacterium*	2.47045E-05	0.028285692	-0.187964458
Item58	*Intestinimonas_gabonensis*	1.91795E-05	0.024037077	-0.190447642
Item58	*GGB9677_SGB15180*	2.95568E-07	0.001191807	-0.227494804
Item58	*GGB4584_SGB6338*	2.05377E-05	0.025061997	-0.189779602
Fibre_g	*Haemophilus_parainfluenzae*	3.13432E-06	0.025058908	0.207356121
Fibre_g	*Ruminococcus_torques*	1.27984E-08	0.000306969	-0.25168923
VitD_mcg	*Lachnospiraceae_bacterium*	2.45627E-09	0.000117827	0.263459596
Biotin_mcg	*GGB3478_SGB4643*	4.90237E-07	0.005879167	0.223336839
VitC_mg	*GGB9758_SGB15368*	1.00601E-06	0.009651636	0.2172879
VitC_mg	*Ruminococcus_torques*	4.14408E-07	0.005879167	-0.224726562

ANNEX 10. Significant association obtained by Spearman correlations (FDR < 0.05) between bacterial and fungal species

Var1	*Var2*	*p-value*	*q-value*	*R*
Candidatus_Metaruminococcus_caecorum	*Penicillium.sp.*	2.62E-10	4.7231E-07	0.58456333
Actinomyces_SGB17168	*Penicillium.brevicompactum*	9.90E-10	1.3041E-06	0.56885733
Candidatus_Protoclostridium_gallicola	*Debaryomyces.hansenii*	1.34E-08	1.1071E-05	0.53549138
Candidatus_Metaruminococcus_caecorum	*Penicillium.camemberti*	1.54E-07	7.4587E-05	0.50045167
GGB9209_SGB14148	*Vanrija.humicola*	2.26E-07	9.8318E-05	0.49462949
GGB781_SGB1024	*Neurospora.cerealis*	3.24E-07	0.00013032	0.48898201
Fournierella_massiliensis	*Cladosporium.cladosporioides*	1.17E-06	0.00034345	0.46810242
Fournierella_massiliensis	*Saccharomyces.sp.*	1.18E-06	0.00034426	0.46799392
Butyricimonas_sp_Marseille_P3923	*Bipolaris.maydis*	1.23E-06	0.00035294	0.46732267
Propionibacterium_freudenreichii	*Penicillium.sp..DTU1*	2.23E-06	0.000563	0.45710099
Candidatus_Schneewindia_gallinarum	*Aspergillus.penicilloides*	2.49E-06	0.00061288	0.45519036
Odoribacter_laneus	*Penicillium.nalgiovense*	4.43E-06	0.0009663	0.44493496
Catabacter_hongkongensis	*Bipolaris.maydis*	6.05E-06	0.00122877	0.43923145
Duodenibacillus_massiliensis	*Penicillium.nordicum*	6.61E-06	0.00131381	0.43759752
Anaerotignum_lactatifermentans	*Penicillium.nalgiovense*	6.79E-06	0.00133845	0.43709878
Bifidobacterium_animalis	*Cladosporium.cladosporioides*	7.95E-06	0.00149352	0.43417031
Anaerofustis_stercorihominis	*Penicillium.rubens*	1.29E-05	0.00213749	0.42505249
Streptococcus_mutans	*Pestalotiopsis.kenyana*	1.77E-05	0.00271614	0.41878941
Prevotella_lascolaii	*Yarrowia.lipolytica*	2.03E-05	0.00300633	0.4161654
GGB9634_SGB15099	*Penicillium.nordicum*	2.09E-05	0.00306161	0.41559062
GGB9716_SGB15269	*Penicillium.nordicum*	2.62E-05	0.00359856	0.411006
GGB4554_SGB6285	*Penicillium.roqueforti*	2.66E-05	0.00362345	0.41070525
Clostridia_unclassified_SGB4447	*Candida.albicans*	4.19E-05	0.00504945	0.40144611
Candidatus_Metaruminococcus_caecorum	*Mucor.lanceolatus*	4.33E-05	0.00516482	0.40077553

Var1	Var2	p-value	q-value	R
GGB9710_SGB15239	*Geotrichum.candidum*	4.33E-05	0.00516842	0.40073495
GGB3341_SGB4420	*Saccharomyces.cerevisiae*	5.25E-05	0.00595425	0.39672018
Streptococcus_gordonii	*Yarrowia.lipolytica*	6.25E-05	0.00673389	0.39303342
GGB1538_SGB2121	*Geotrichum.candidum*	6.26E-05	0.00674486	0.39299183
Alistipes_inops	*Penicillium.nalgiovense*	7.12E-05	0.00741321	0.3902193
Massilimicrobiota_timonensis	*Aspergillus.penicilloides*	8.43E-05	0.00840576	0.38656762
Clostridium_SGB4751	*Agaricus.bisporus*	8.58E-05	0.00847656	0.38619113
Prevotella_stercorea	*Penicillium.nordicum*	8.62E-05	0.00848638	0.38610162
GGB4768_SGB6601	*Penicillium.nordicum*	8.62E-05	0.00848638	0.38610162
GGB9636_SGB15108	*Yarrowia.lipolytica*	8.65E-05	0.00850877	0.38601211
Clostridia_bacterium_UC5_1_1E11	*Aspergillus.penicilloides*	9.06E-05	0.0087752	0.38499761
GGB9186_SGB14125	*Saccharomyces.cerevisiae*	9.14E-05	0.00881667	0.38482188
GGB9559_SGB14968	*Vanrija.humicola*	9.37E-05	0.00896194	0.38427138
Lentisphaeria_unclassified_SGB9198	*Penicillium.roqueforti*	9.39E-05	0.0089626	0.38422954
Ruminococcus_gnavus	*uncultured.Malassezia.spp..1*	9.90E-05	0.00933618	0.38305172
Bacilli_unclassified_SGB6493	*Geotrichum.candidum*	9.96E-05	0.00937177	0.38291834
Raoultibacter_massiliensis	*Pestalotiopsis.kenyana*	1.05E-04	0.00971467	0.38181742
GGB6606_SGB9340	*Penicillium.solitum*	1.18E-04	0.0105152	0.3791076
GGB3817_SGB5182	*Vanrija.humicola*	1.25E-04	0.01092909	0.37786669
GGB6522_SGB9214	*Yarrowia.lipolytica*	1.30E-04	0.0112214	0.3770611
GGB9603_SGB15035	*Agaricus.bisporus*	1.30E-04	0.01122679	0.37703872
Blautia_producta	*Penicillium.rubens*	1.34E-04	0.01146119	0.37637208
Raoultibacter_massiliensis	*Botryosphaeria.dothidea*	1.53E-04	0.01266695	0.37329122
GGB1538_SGB2121	*Yarrowia.lipolytica*	1.58E-04	0.01288991	0.37263035
Acidaminococcus_intestini	*Candida.parapsilosis*	1.59E-04	0.01294811	0.37247254
GGB9332_SGB14295	*Geotrichum.candidum*	1.79E-04	0.01408221	0.36969826
GGB9045_SGB13947	*Penicillium.roqueforti*	1.91E-04	0.01470077	0.36823123

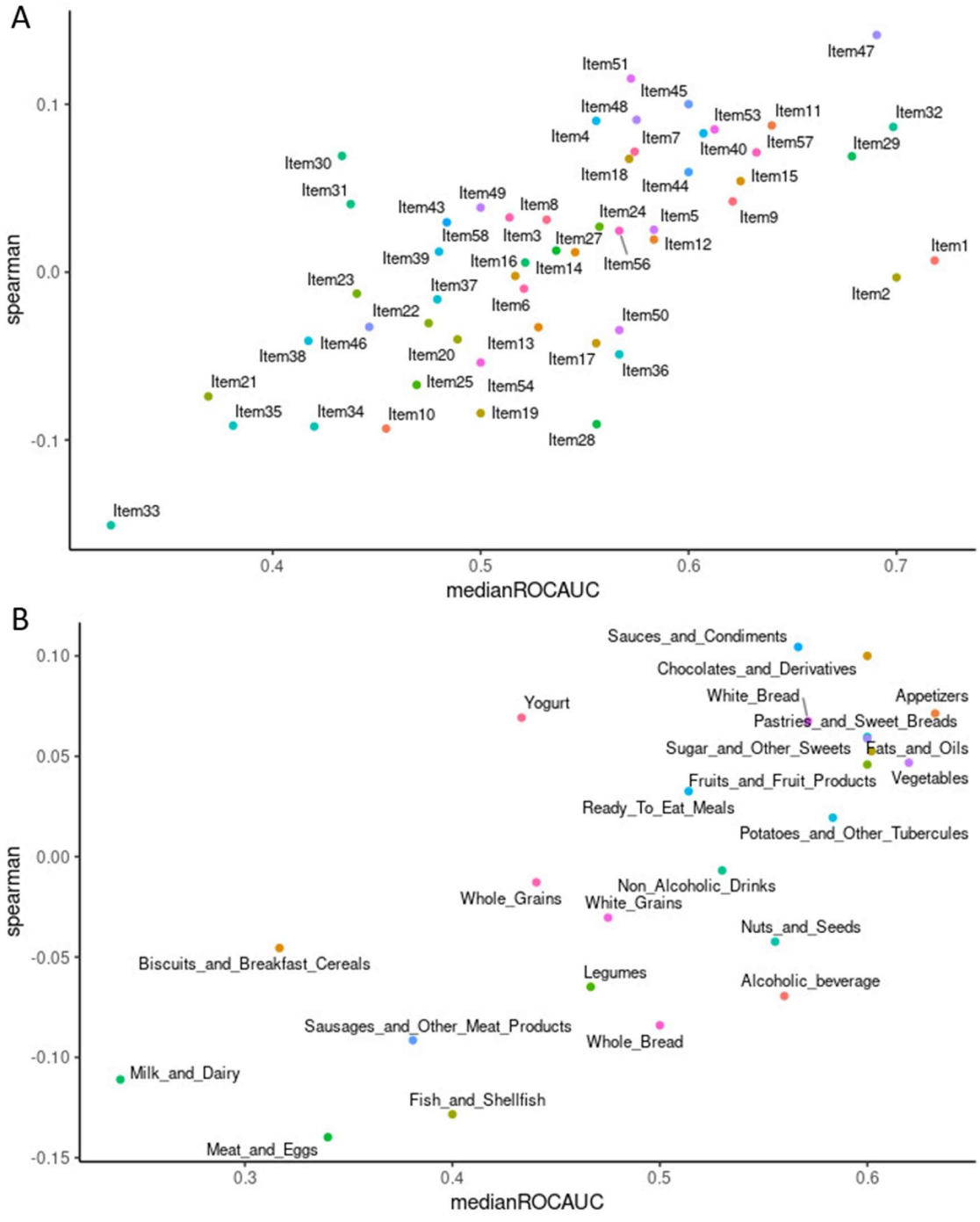
Var1	Var2	p-value	q-value	R
GGB3433_SGB4573	Penicillium.nalgiovense	1.98E-04	0.01502988	0.36739697
GGB9101_SGB14033	Bipolaris.maydis	2.05E-04	0.01529865	0.36664666
Streptococcus_infantis	Aspergillus.versicolor	2.23E-04	0.01629126	0.36459293
Blautia_hansenii	Bipolaris.maydis	2.26E-04	0.016395	0.36430592
Streptococcus_anginosus	uncultured.Malassezia.spp..1	2.31E-04	0.01658058	0.36384005
Clostridia_unclassified_SGB4372	Pichia.manshurica	2.35E-04	0.01676136	0.36338522
Candidatus_Metalachnospira_gallinarum	Saccharomyces.boulardii..nom..inval..	2.38E-04	0.01688587	0.36316157
Faecalimonas_umbilicata	Penicillium.rubens	2.41E-04	0.01699999	0.36285902
GGB4710_SGB6522	Penicillium.nalgiovense	2.58E-04	0.0178683	0.36125111
Klebsiella_pneumoniae	Saccharomyces.pastorianus	2.63E-04	0.01810537	0.36075385
Clostridia_unclassified_SGB14844	Aspergillus.penicilloides	2.66E-04	0.01819977	0.36052743
GGB5980_SGB8599	Penicillium.nalgiovense	2.88E-04	0.0191295	0.35861098
GGB29535_SGB42321	Geotrichum.candidum	2.96E-04	0.01951485	0.35792372
Bacilli_unclassified_SGB6540	Penicillium.roqueforti	3.05E-04	0.01991056	0.35723744
Streptococcus_sp_263_SSPC	Pestalotiopsis.kenyana	3.06E-04	0.01992238	0.35719522
Clostridiales_bacterium_Choco116	Aspergillus.penicilloides	3.08E-04	0.02002722	0.35700889
GGB3474_SGB4637	Malassezia.arunalokei	3.32E-04	0.02113185	0.35516444
Ruminococcus_sp_JE7A12	Penicillium.nordicum	3.34E-04	0.02115615	0.35508114
Opitutales_bacterium	Penicillium.nordicum	3.65E-04	0.02250591	0.35292483
Candidatus_Pararuminococcus_gallinarum	Meira.nashicola	3.69E-04	0.02265762	-0.3526588
GGB9291_SGB14248	Malassezia.restricta	3.73E-04	0.02282003	-0.3524049
GGB3537_SGB4727	Penicillium.nalgiovense	3.79E-04	0.02305994	0.35202385
GGB58158_SGB79798	Debaryomyces.hansenii	3.81E-04	0.02314048	0.35189221
Candidatus_Metaruminococcus_caecorum	Penicillium.paneum	3.97E-04	0.0238058	0.35086556
Clostridiaceae_unclassified_SGB4769	Malassezia.restricta	4.07E-04	0.02418078	-0.3502473
Paraprevotella_clara	Penicillium.chrysogenum	4.21E-04	0.02472814	0.34944742
GGB9494_SGB14891	Meira.nashicola	4.24E-04	0.0248644	-0.34923

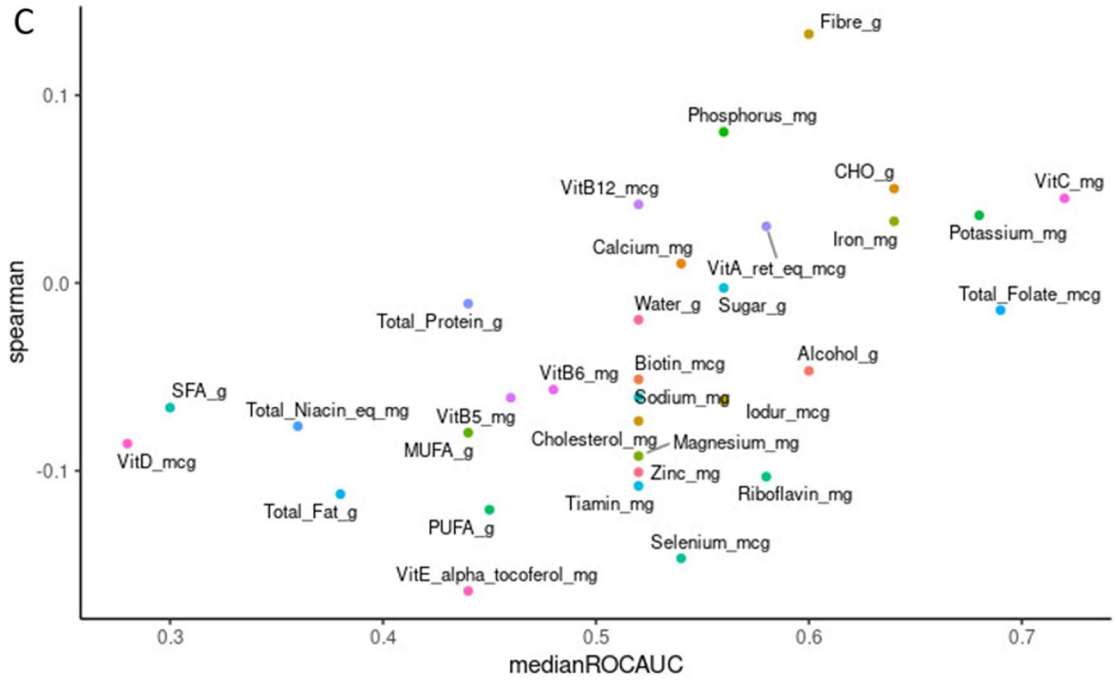
Var1	Var2	p-value	q-value	R
GGB9634_SGB15099	*Geotrichum.candidum*	4.26E-04	0.02492885	0.34911437
Anaerostipes_caccae	*Mucor.lanceolatus*	4.63E-04	0.02633483	0.34709314
Sellimonas_intestinalis	*Penicillium.brevicompactum*	4.73E-04	0.02670368	0.3465303
Clostridiales_bacterium_Choco116	*uncultured.Malassezia.spp..1*	4.75E-04	0.02673033	0.34644268
Candidatus_Metalachnospira_gallinarum	*Saccharomyces.sp.*	4.77E-04	0.02680857	0.34633427
Enterocloster_aldensis	*Penicillium.paneum*	4.80E-04	0.02694866	0.34616181
Intestinimonas_timonensis	*Saccharomyces.pastorianus*	4.85E-04	0.02710239	0.3459166
Blautia_sp_MSK_21_1	*Penicillium.roqueforti*	5.16E-04	0.02825223	0.34436269
Actinomyces_sp_ICM47	*Saccharomyces.boulardii..nom..inval..*	5.22E-04	0.02844182	0.34408138
Gordonibacter_urolithinfaciens	*Malassezia.restricta*	5.26E-04	0.02855782	0.34391861
GGB9818_SGB15459	*Penicillium.nalgiovense*	5.29E-04	0.02865344	0.34377133
GGB3226_SGB4260	*Penicillium.chrysogenum*	5.39E-04	0.02898958	0.34327053
GGB3678_SGB4991	*Penicillium.rubens*	5.42E-04	0.02909068	0.34314981
GGB9695_SGB15209	*Saccharomyces.sp.*	5.51E-04	0.02936508	0.34275222
Clostridiales_bacterium_BX7	*Geotrichum.candidum*	5.67E-04	0.02992152	0.34201318
Bacilli_unclassified_SGB6571	*Debaryomyces.hansenii*	5.79E-04	0.03037089	0.34147125
Rikenellaceae_bacterium_DSM_108975	*Penicillium.nalgiovense*	5.82E-04	0.03046042	0.34136146
Candidatus_Alangreenwoodia_gallinarii	*Neurospora.cerealis*	5.89E-04	0.03071649	0.34103982
GGB2982_SGB3964	*Malassezia.restricta*	5.95E-04	0.03090756	0.34078887
Candidatus_Aristotella_avistercoris	*Penicillium.chrysogenum*	6.13E-04	0.03150191	0.34004043
Enterocloster_lavalensis	*Penicillium.roqueforti*	6.30E-04	0.03207197	0.33933811
Sutterella_sp_AM11_39	*X.Candida..sake*	6.43E-04	0.0324481	0.33884435
Catabacter_hongkongensis	*Saccharomyces.sp.*	6.49E-04	0.03261434	0.33858704
Bifidobacterium_longum	*Candida.parapsilosis*	6.70E-04	0.03328553	0.33780299
Bacteroides_caccae	*Penicillium.sp.*	6.77E-04	0.03351285	0.33753068
GGB9694_SGB15204	*Geotrichum.candidum*	6.86E-04	0.03381687	0.33717132
Alistipes_dispar	*Penicillium.nalgiovense*	7.25E-04	0.03511491	0.335775

Var1	Var2	p-value	q-value	R
GGB3817_SGB5182	*X.Candida..sake*	7.36E-04	0.0354123	0.33538388
Mediterraneibacter_glycyrrhizinilyticus	*Penicillium.nalgiovense*	7.46E-04	0.03573754	0.33501522
Pseudoflavonifractor_sp_Marseille_P3106	*Aspergillus.versicolor*	7.74E-04	0.03655856	0.33408621
Lactonifactor_sp_BIOML_A6	*Aspergillus.versicolor*	7.76E-04	0.03660469	0.33400788
Bacteroides_caccae	*Debaryomyces.hansenii*	7.95E-04	0.03715922	0.3334008
GGB3433_SGB4573	*Penicillium.rubens*	7.99E-04	0.03731297	0.33324572
GGB6608_SGB9342	*Penicillium.nalgiovense*	8.22E-04	0.03796112	0.33252613
Streptococcus_salivarius	*Aspergillus.versicolor*	8.32E-04	0.03823596	0.33219899
Prevotella_stercorea	*Pichia.manshurica*	8.42E-04	0.03848658	0.33188317
GGB3123_SGB4135	*Penicillium.chrysogenum*	8.44E-04	0.03852986	0.33184293
Ruminococcaceae_bacterium_AM07_15	*Penicillium.nordicum*	8.62E-04	0.03906757	0.33126907
GGB3817_SGB5182	*Mucor.lanceolatus*	8.70E-04	0.0392095	0.33103817
Harryflintia_acetispora	*Mucor.lanceolatus*	8.73E-04	0.03927534	0.3309577
Clostridia_bacterium	*Candida.parapsilosis*	8.82E-04	0.03947472	-0.330696
GGB9059_SGB13976	*Geotrichum.candidum*	8.89E-04	0.0396888	0.33048113
Candidatus_Schneewindia_gallinarum	*Mucor.lanceolatus*	9.07E-04	0.04015662	0.32994844
Phocea_massiliensis	*Penicillium.nordicum*	9.57E-04	0.04161261	0.32855436
Eubacterium_sp_AF15_50	*Agaricus.bisporus*	9.70E-04	0.04195263	0.32819648
Lactococcus_lactis	*Debaryomyces.hansenii*	9.71E-04	0.04197623	0.32817143
Hafnia_paralvei	*Penicillium.nordicum*	1.04E-03	0.04395911	0.32633134
Eggerthella_lenta	*Bipolaris.maydis*	1.08E-03	0.04514226	0.32532649
GGB9667_SGB15164	*Aspergillus.versicolor*	1.09E-03	0.04532232	-0.3251674
Methanomassiliicoccales_archaeon	*Geotrichum.candidum*	1.09E-03	0.04547447	0.3250322
GGB1407_SGB1930	*Meira.nashicola*	1.11E-03	0.04605474	0.3245348
GGB3654_SGB4965	*Yarrowia.lipolytica*	1.16E-03	0.04748686	0.32343319
GGB9288_SGB14243	*Yarrowia.lipolytica*	1.16E-03	0.04752241	0.32340658
GGB9635_SGB15102	*Yarrowia.lipolytica*	1.17E-03	0.04775125	0.323232

Var1	*Var2*	*p-value*	*q-value*	*R*
Lachnoclostridium_sp_An138	*Aspergillus.penicilloides*	1.20E-03	0.0484541	0.32253196
Neobittarella_massiliensis	*Saccharomyces.sp.*	1.21E-03	0.04863447	0.32236887
Collinsella_aerofaciens	*Bipolaris.maydis*	1.21E-03	0.04873245	-0.3222973
GGB9636_SGB15107	*Geotrichum.candidum*	1.23E-03	0.04927319	0.32182723
Anaerotruncus_colihominis	*Aspergillus.penicilloides*	1.24E-03	0.04944684	0.32167774
Bifidobacterium_dentium	*Pestalotiopsis.kenyana*	1.26E-03	0.04992047	0.32127305

ANNEX 11. Prediction using machine learning technique. Prediction of different food items (A). food groups (B) and nutrients (C) using fungal species-level genome bin (SGB)-level features information estimated by MetaPhlAn4. Y-axis and X-axis represent median Spearman's correlation and median receiver operating characteristic area under the curve (ROCAUC) from the random forest regressor and random forest classifier respectively.




C
Fibre_g
Phosphorus_mg
CHO_g
VitC_mg
VitB12_mcg
Potassium_mg
Iron_mg
Calcium_mg
VitA_ret_eq_mcg
Total_Folate_mcg
Water_g
Sugar_g
Total_Protein_g
Alcohol_g
Biotin_mcg
VitB6_mg
SFA_g
Sodium_mg
Total_Niacin_eq_mg
VitB5_mg
Iodur_mcg
Cholesterol_mg
Magnesium_mg
MUFA_g
VitD_mcg
Zinc_mg
Riboflavin_mg
Tiamin_mg
Total_Fat_g
PUFA_g
Selenium_mcg
VitE_alpha_tocoferol_mg
spearman
0.1
0.0
-0.1
0.3
0.4
0.5
0.6
0.7
medianROCAUC

ANNEX 12. α-diversity measures (Shannon and Chao1), scores for hPDI and aMED indeces and IBD-similarity index scores coming from non-targeted volunteers

Donor	Gender	IBD-score	Chao1	Shannon	hPDI score	aMED score
Donor 1	Female	0.203873632	303	4.18055368	62	6
Donor 2	Female	0.209459649	287	4.353425467	60	7

ANNEX 13. List and number of bacterial isolates coming from two healthy donors. Bacteria was isolated using non-targeted method and two different media (GAM and L-YHBHI.4 supplemented with RF)

All_frozen_combined_species	Phylum	Family	Total_isolates	Donor_1_isolates	Donor_2_isolates	Media_GAM_all_isolates	Media_L. YHBHI.4_all_isolates
Agathobacter rectalis	Bacillota	Lachnospiraceae	1	1	0	1	0
Alistipes shahii	Bacteroidota	Rikenellaceae	1	1	0	1	0
Bacteroides caccae	Bacteroidota	Bacteroidaceae	14	14	0	14	0
Bacteroides clarus	Bacteroidota	Bacteroidaceae	2	2	0	2	0
Bacteroides fragilis	Bacteroidota	Bifidobacteriaceae	6	6	0	6	0
Bacteroides stercoris	Bacteroidota	Bacteroidaceae	2	2	0	2	0
Bacteroides thetaiotaomicron	Bacteroidota	Bacteroidaceae	1	0	1	1	0
Bacteroides uniformis	Bacteroidota	Bacteroidaceae	46	7	39	46	0
Bacteroides xylanisolvens	Bacteroidota	Bacteroidaceae	1	1	0	1	0
Barnesiella intestinihominis	Bacteroidota	Barnesiellaceae	1	1	0	1	0
Bifidobacterium adolescentis	Actinomycetota	Bifidobacteriaceae	34	34	0	13	21
Bifidobacterium animalis	Actinomycetota	Bifidobacteriaceae	2	2	0	0	2
Bifidobacterium catenulatum	Actinomycetota	Bifidobacteriaceae	6	0	6	0	6
Bifidobacterium longum	Actinomycetota	Bifidobacteriaceae	24	24	0	15	9
Butyricimonas faecihominis	Bacteroidota	Odoribacteraceae	1	0	1	1	0
Collinsella aerofaciens	Actinomycetota	Coriobacteriaceae	8	7	1	8	0
Coprococcus sp	Bacillota	Lachnospiraceae	1	1	0	1	0
Enterococcus faecalis	Bacillota	Enterococcaceae	5	4	1	4	1
Enterococcus faecium	Bacillota	Enterococcaceae	2	2	0	1	1
Escherichia coli	Pseudomonadota	Enterobacteriaceae	17	15	2	17	0
No organism identification possible	NA	NA	22	20	2	18	4

All_frozen_combined_species	Phylum	Family	Total_isolates	Donor_1_isolates	Donor_2_isolates	Media_GAM_all_isolates	Media_L. YHBHI.4_all_isolates
Parabacteroides distasonis	Bacteroidota	Tannerellaceae	1	0	1	1	0
Parabacteroides merdae	Bacteroidota	Tannerellaceae	2	2	0	2	0
Phocaeicola dorei	Bacteroidota	Bacteroidaceae	6	4	2	6	0
Phocaeicola massiliensis	Bacteroidota	Bacteroidaceae	3	0	3	3	0
Phocaeicola vulgatus	Bacteroidota	Bacteroidaceae	50	37	13	50	0
Solobacterium moorei	Bacillota	Erysipelotrichaceae	1	1	0	1	0
Streptococcus anginosus	Bacillota	Streptococcaceae	1	1	0	1	0

ANNEX 14. Publication related to the thesis: *Soler et al., 2025* "A contributory citizen science project reveals the impact of dietary keys to microbiome health in Spain"

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: cmanichai@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 ($q(\text{IASE}) = 0.03$; $q(\text{hPDI}) = 7{\cdot}1 \times 10^{-07}$) (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 ($q = 2{\cdot}7 \times 10^{-10}$) (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

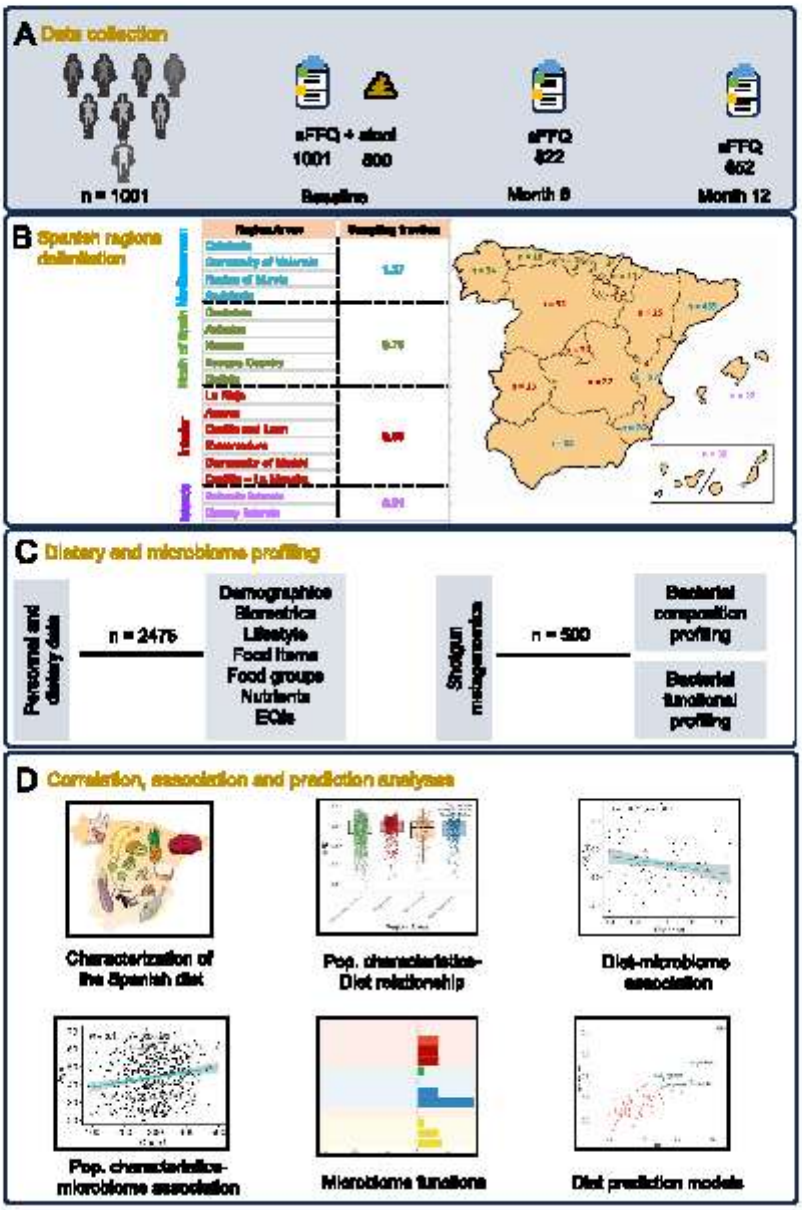


Fig. 1 | Study design. A 1001 participants reported their dietary intake and personal data through an in-house online short Food Frequency Questionnaire (sFFQ) at baseline, month six, and month 12 ($n = 2475$). Stool samples ($n = 500$) were processed at baseline for microbiome analysis. B Recruitment 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.

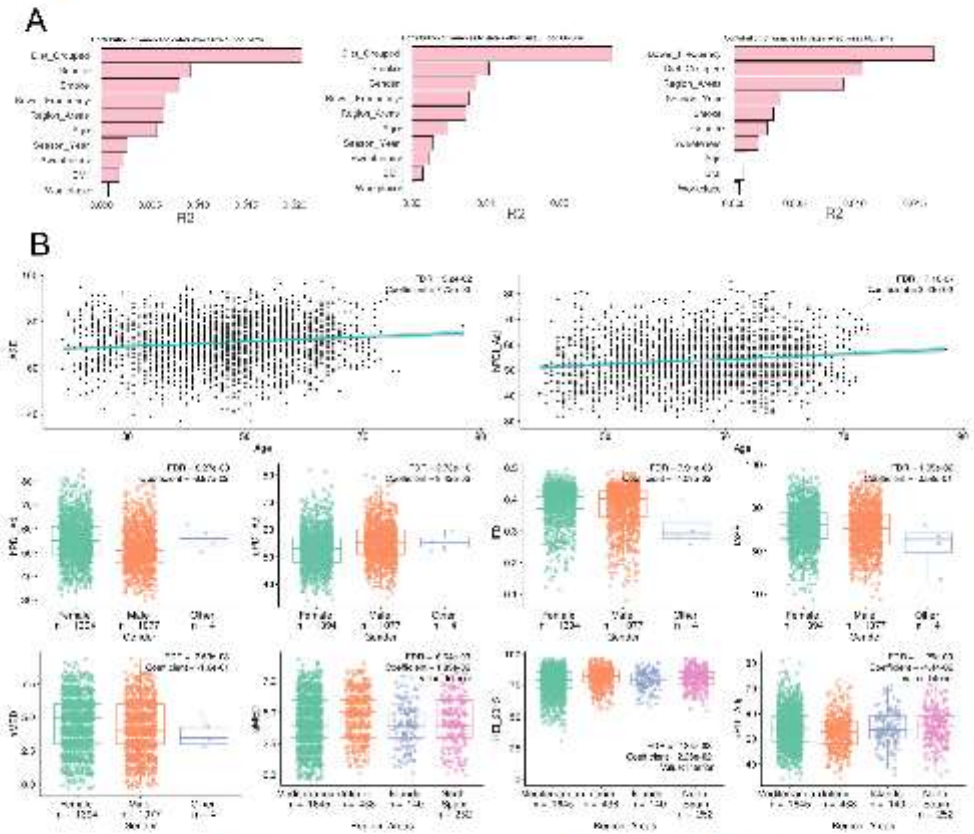


Fig. 2 | Relationship between population characteristics and dietary data. A Effect size of the population characteristics on dietary intake. The magnitude of the influence of specific characteristics on dietary intake was calculated using permutational analysis of variance (PERMANOVA), as implemented 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 MaAsLin 2 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

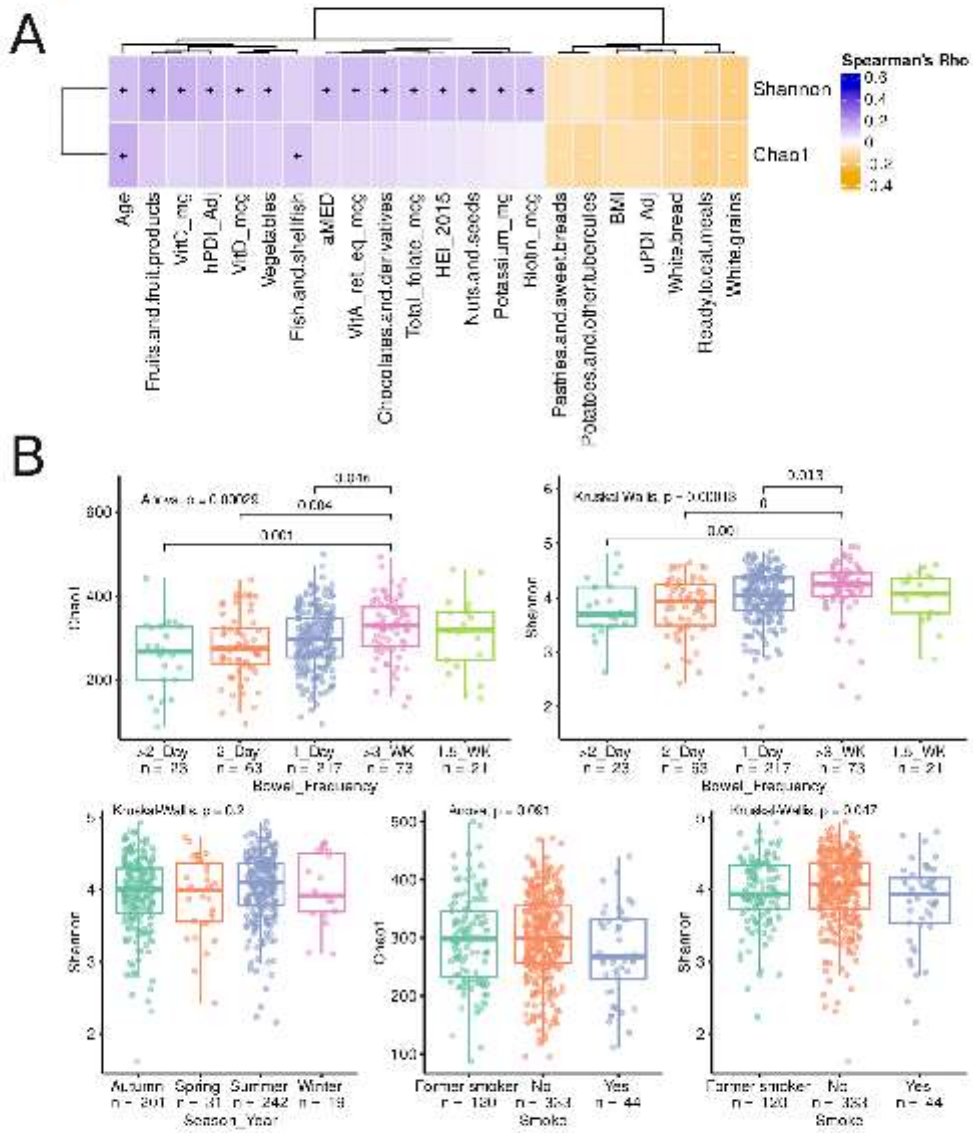


Fig. 3 | Population characteristics-microbiome alpha diversity association analysis. A Correlation between Eating Quality Indexes (EQIs), food groups, and personal data with alpha diversity (Chao1 and Shannon) using the Spearman correlation test ($n = 500$). Symbols + and − indicate significant correlations ($P < 0.05$). 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-

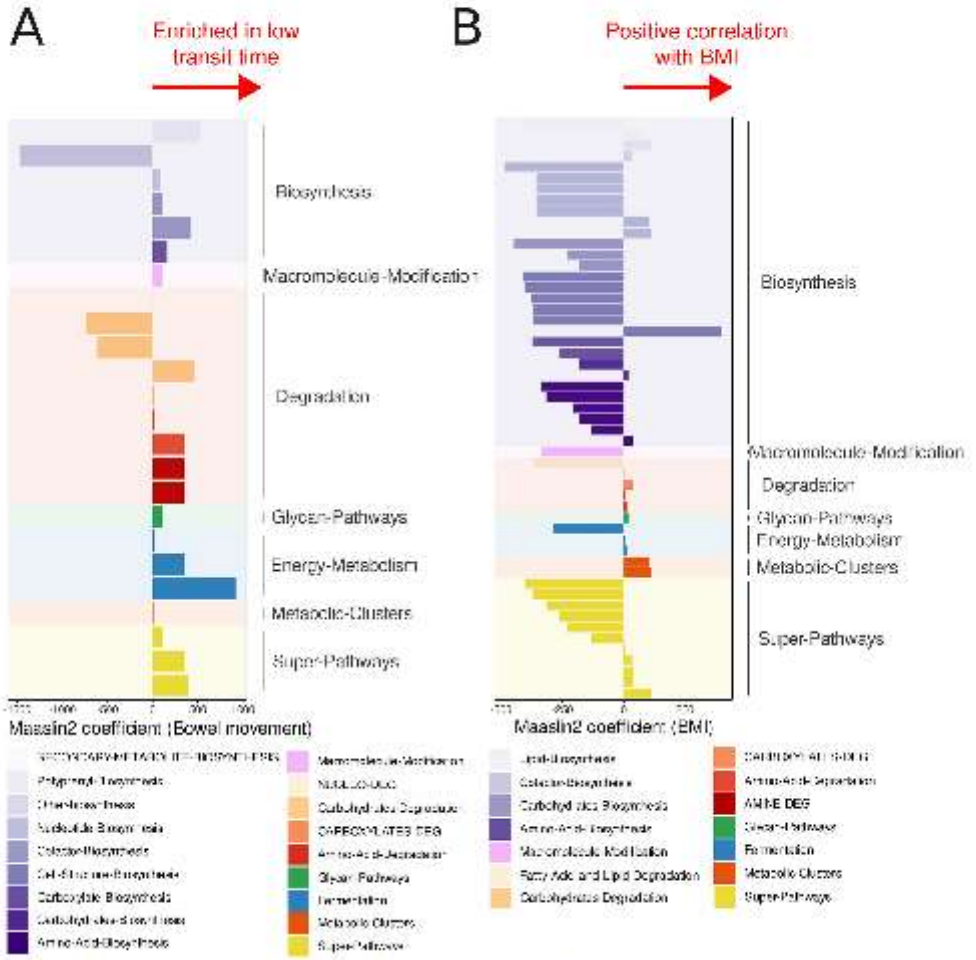


Fig. 4 | Differentially abundant metabolic pathways. Differentially abundant pathways in the microbiome of healthy individuals using to Maaslin2 depending on several conditions. Pathways were grouped according to their level 1 MetaCyc classes (broader functionality) and colored by their level 2 MetaCyc 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

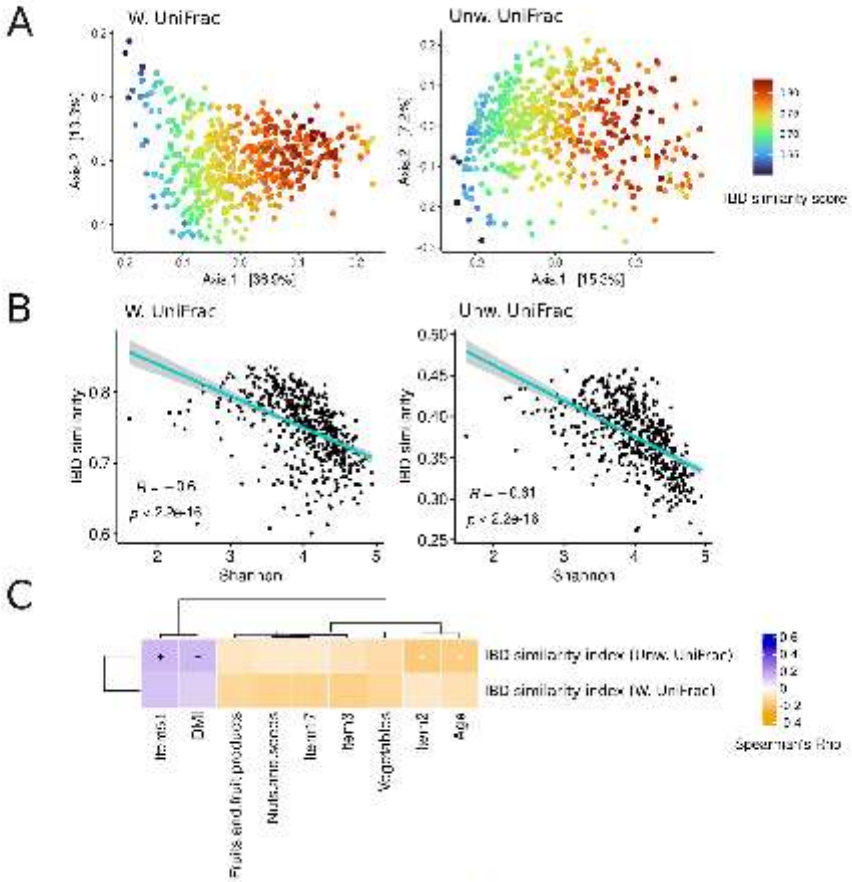


Fig. 5 | Disease similarity index and population characteristics. A Weighted (W.) and unweighted (Unw.) UniFrac distances of our cohort of healthy individuals ($n = 500$) colored by IBD-similarity score. IBD-similarity score was calculated as 1-median of a healthy sample to all samples in IBD plane ($n = 208$ 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 can not 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.[16], 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).

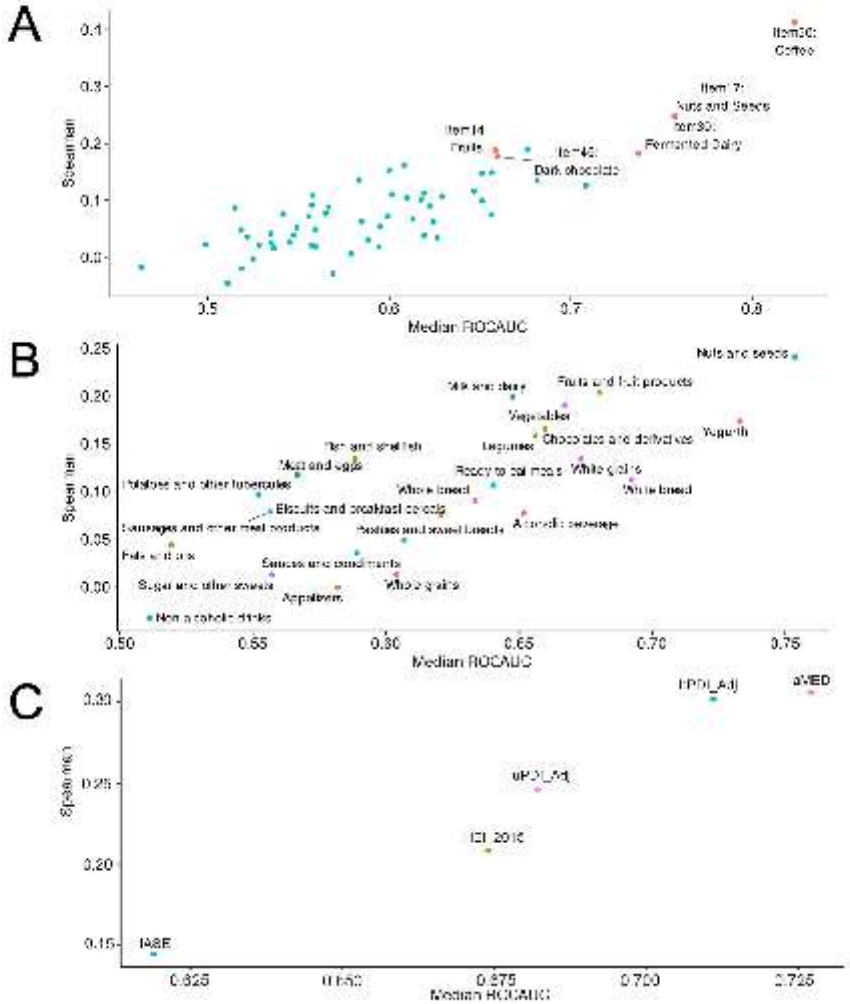


Fig. 6 | Prediction using machine learning technique. Prediction of different food items (A), food groups (B), and EQIs (C) using species-level genome bin (SGB)-level features information estimated by MetaPhlAn4. Y axis represents, for each variable, the median Spearman's correlation between observed 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[10,11]. 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[20]. Despite its benefits, adherence to the Mediterranean diet (MD) in Spain has decreased over time, shifting towards a more Western dietary pattern[21,22].

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[8]. 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[23,24]. 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[25]. 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.[8], which is based on the original Mediterranean diet scale proposed by Trichopoulou et al.[16]. 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[18]. 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[18]. 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)[19]. 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 (DIME®CA 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; PI21/00282) 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.