Universitat Autònoma de Barcelona

METAGENOMICS IN INFLAMMATORY BOWEL DISEASE

Marta Pozuelo del Río

METAGENOMICS IN INFLAMMATORY BOWEL DISEASE

A doctoral thesis presented by **Marta Pozuelo del Río** to aim for the degree of Doctor

Supervisor of the Doctoral Thesis:

Dr. Chaysavanh Manichanh

Tutor of Doctoral Thesis:

Dr. Víctor Vargas Blasco

Barcelona, September 2019

Department of Medicine

Doctoral Program in Medicine

Universitat Autònoma de Barcelona

Dr. Chaysavanh Manichanh

Principal Investigator – Microbiome Lab

Department of Physiology and Physiopathology of the Digestive Tract

Vall d'Hebron Institut de Recerca

Certify that the thesis entitled **"Metagenomics in inflammatory bowel disease"** submitted by **Marta Pozuelo del Río**, was carried out under her supervision and tutorship of **Dr. Víctor Vargas Blasco**, and authorize its presentation for the defense ahead of the corresponding tribunal.

In witness whereof, I hereby sign this document.

Barcelona, September 2019

Dr. Chaysavanh Manichanh

Front cover image: DNA sequence from *Faecalibacterium prausnitzii*, bacterium depleted in Crohn's disease

Front cover credit - Marta Vilaplana

AGRADECIMIENTOS

Por fin escribo estas palabras, significa que ya he terminado de escribir mi tesis y en breve será hora de depositarla. Han sido unos años interesantes, mejores en algunos momentos, peores en otros, pero de los cuales no me arrepiento porque por fin lo he conseguido, la tesis está terminada y en breve podré presentarla. Empecé este periodo de mi vida en un mal momento, pero gracias a la gente que me rodeaba entonces y me rodea ahora pude superarlo y avanzar en mi vida y con este libro. Probablemente lo que diga en estas páginas se quedará pequeño al lado de lo mucho que les agradezco lo que me han ayudado.

En primer lugar, gracias a la Dra. Chaysavanh Manichanh por haberme permitido realizar la tesis doctoral en tu grupo. Gracias por haberme dado la oportunidad de poder crecer profesionalmente y transmitirme tus conocimientos en el mundo de la microbiota.

Gracias a Amics del VHIR y sus donantes por haberme concedido su beca predoctoral sin la cual no habría podido llevar a cabo esta tesis. Es una realidad que, sin un soporte económico es imposible que los investigadores predoctorales nos mantengamos y menos en una ciudad como Barcelona. También querría aprovechar estas líneas para agradecer a todas aquellas personas que están luchando contra la precariedad de los investigadores predoctorales, que, aunque poco a poco, están consiguiendo que seamos un poco menos precarios.

Agradecer también su aportación en esta tesis a mi tutor el Dr. Víctor Vargas, por proporcionarme sus opiniones acerca del trabajo que estábamos realizando.

Gracias a todos los colaboradores clínicos e investigadores con los que he podido trabajar estando en este grupo, de ellos he podido aprender muchas cosas sobre las distintas enfermedades con las que he trabajado. Dr. German Soriano, Dr. Joan Verdaguer, Dra. Susana Gómez, Dra. Carmen Alonso, Dr. Fernando Azpiroz, Dr. Francisco Guarner, Dra. Virginia Robles, Estela y Victoria entre otros, gracias por aportar vuestro conocimiento y ayudarme a entender los problemas ante los que nos enfrentábamos. Pero, sobre todo, gracias a Claudia que siempre me ha ofrecido su ayuda y una sonrisa cada vez que pasaba por mi sitio, aunque también quería muchos análisis de su estudio.

Tantas horas en el VHIR no habrían sido posibles sin tener la compañía de grandes personas que he encontrado en el grupo de digestivo, tanto por parte del grupo de la Dra. Chaysavanh Manichanh como el de la Dra. María Vicario. En el segundo caso, aunque nos conocimos más tarde, las horas de la comida, chocolate y juegos de volley no habrían sido lo mismo sin vosotros, Elba, Merce, niños de máster y grado, Marina, compañera de planes malignos para sorpresas de tesis, pero sobre todo tú Cris, hermana de peluches y adicta al azúcar, has conseguido volverme una revolucionaria, y, aunque llego tarde, espero poder usarlo en el futuro en mi vida. Espero que mi precedente os sirva para conseguir lo que necesitáis.

Por mi grupo han pasado muchas personas y me quedo con un poquito de cada uno, pero no podía pasar sin agradecerle a Xavi su infinita ayuda con la informática, cuando estaba perdida siempre me ayudabas. Fran, eres una trabajadora incansable, tus palabras y ánimo siempre te ayudan a tirar para adelante. Manon, thanks for promoting Fridays of volley out of the lab. Marina y Ferrán, mis compis de ideas y cotilleos, ha sido genial compartir estos últimos meses con vosotros haciendo brain storming y pudiendo ayudaros en lo que era posible. Encarna, gracias por aportarme tu experiencia y paciencia. Guillaume, gracias por estar ahí, por ser un gracioso oculto y por poner mi cara en fotos horribles. Un gracias enorme a mi ricitos de oro, Enric, gracias por ayudarme en esta tesis con el machine learning, por darme masajes cuando me veías estresada, por ponerme la Oreja de Van Gogh para que entrase en bucle mientras trabajábamos y por dejarme escuchar uno de tus conciertos, espero ver muchos más. Estelita, gracias por enseñarme que hay que tener claras las prioridades y que, si otras cosas tardan más en llegar, ya llegarán.

Pero nada de todo esto habría sido posible sin ellos, empezaron siendo compañeros de trabajo, pasado un mes ya eran amigos, pero ahora son familia. Suchita, me dejaste tu casa cuando no podía estar sola y gracias a ti y tu boda he hecho uno de los viajes más bonitos hasta ahora, gracias por estar ahí, aunque hablemos y nos veamos poco, estoy aquí para lo que quieras. David, a ti no tengo que agradecerte nada de nada, por ser la persona más molesta que he conocido en mi vida que hasta desde Irlanda das guerra, aunque no, sabes que eso es mentira. Alba y David, “Dramas los cuatro”, aunque no lo creáis me disteis fuerza cuando no las tenía, me aguantáis los dramas en el día a día,

pero, sobre todo, nos reímos juntos, nos caemos juntos, criticamos juntos, compartimos viajes, room escapes, escapadas juntos, aunque estéis lejos, estáis ahí día tras día y eso no lo tiene todo el mundo. Si hay algo de lo que más puedo estar agradecida en este tiempo es de haberos encontrado.

La vida en el VHIR no habría sido lo mismo sin ellos, celadores, personal de limpieza, técnicos, siempre generando buen ambiente y charlando un rato por los pasillos. Sobre todo, gracias a Flor, las Montses y Vicky por ser unos soles y a Pablo, Tao y Quim por hacer más amenos los días y recogerme mis múltiples paquetes.

Gracias Giacy, que me cuidas como una hermana mayor, y Jorge, por ser los vecinos de barrio perfectos, por celebrar conmigo todos los avances y empujarme a terminar (aunque poniéndome distracciones como preparar fiestas sorpresa por medio) y por también estar conmigo cuando más lo necesitaba. Gracias a Cris y Esther, mis chicas de la uni, gracias por decir las verdades, aunque sean malas, espero que podáis disfrutar conmigo el día de la defensa y todo lo que podamos celebrar. Gracias a Mario y Manny que siempre estás pendientes de mis avances. Gracias a mis amigos de toda la vida, los Chiquilines, y a mis chicas (Paola, Laeti, Pia, Laura, Marta, Moni, Nuria, Gina) y chico (Deivid) del gimnasio que me han animado para que termine y haga el último esfuerzo, además de ayudarme a distraerme con los juegos. Por último y no por ello menos importantes, gracias a mis bachateros y salseros y con ellos, gracias al baile, han sido el aire fresco que me hacía falta, la distracción que siempre había buscado y los que consiguen sacarme de casa cuando las fuerzas flaquean, con especial mención a Marta, mi maquilladora y que me ha hecho esta portada tan preciosa, Arturo, mi pareja de coreográfico y molestador profesional y Xavi, el único salsero con el que bailo y mi arregla espaldas.

Probablemente nada habría sido posible sin él, Carlos, la persona que más horas pasa conmigo al día, que se ha hecho cargo de mis tareas de casa para que yo pudiese terminar esta tesis sin distracciones. Has estado ahí casi desde que vine a vivir a Barcelona, casi 6 años, apoyándome en los buenos y sobre todo en los malos momentos. Gracias por ser mi postdoc mientras aún eras predoc, y gracias por seguir siendo mi postdoc después. Me has ayudado profesional y personalmente y, aunque a veces sé que soy muy pesada, sigues ahí, conmigo y espero que sigas mucho más tiempo, porque

los leones y los gatos puede vivir juntos ;). Gracias por ser como eres y ser ya parte de mí. Gracias a tu familia también, Rafa, Mary y Alba y a todos los demás, por acogerme tan bien desde el minuto cero y apoyarme en este camino.

Pero gracias sobre todo a mi familia, los que siempre, desde el segundo 0 han estado ahí, los que nunca te dejan caer y siempre te apoyan. León está a 800 kilómetros, pero os siento a mi lado cada día. Gracias a mis tíos, primos y abuelos, pero en especial a ti abuelito, te fuiste hace un año, sé que te habría encantado venir a Barcelona a ver un partido del Barça y ver mi presentación de tesis, pero no puede ser, aunque tengo claro que el día de la defensa, estarás sentado en una de esas sillas aplaudiendo como el que más, aunque ninguno te podamos ver. Papá y mamá, gracias por ser mis padres, por ser exigentes, por ser vacilones, por ser controladores, pero sobre todo gracias por haberme apoyado toda mi vida, por haberme empujado a conseguir más y, aunque a veces haya hecho cosas que no os gusten (como cualquier hijo), gracias por seguir ahí siempre. Discutiremos, sí, siempre probablemente, pero la vida es así, no todo es color de rosa, lo importante es que siempre están los buenos momentos que quedan en nuestra memoria y con vosotros son infinitos. Os quiero.

Para vosotros, papá y mamá

“Life isn’t about waiting for the storm to pass, it’s about learning to dance in the rain”

Vivian Greene

INDEX

SUMMARY

The gut commensal microbiota is known to play a crucial role in maintaining intestinal homeostasis. Alterations in the microbial community composition, also known as dysbiosis, may put health status in risk and increase susceptibility to diseases. Although several diseases have been related to shifts in the gut microbiome composition, it is still uncertain whether those alterations are the cause or consequence of the disease.

Inflammatory bowel disease (IBD) is a chronic inflammatory disease with periods of active and inactive inflammation that constitutes to an important health problem. It is divided in two subtypes: Crohn's disease (CD) and ulcerative colitis (UC) that present similar symptoms but different clinical manifestations. IBD has been widely associated with an alteration of the gut microbiome composition. Nevertheless, there is no clear consensus on the microbial pattern characteristic of the disorders. Main discordances between studies are related to differences between UC and CD. Some previous publications indicate that UC microbial composition is very similar to healthy and differs from CD whereas others consider both subtypes as a unique entity and find high alterations in UC and CD microbial composition in comparison with the microbiome of healthy individuals.

The aim of this thesis was to characterize the dysbiosis in a Spanish IBD cohort to evaluate to which extend the gut microbiome composition and function could be differentiated between CD and UC and whether microbiome data could be used as diagnostic and prognostic tools. For this purpose, we analyzed fecal samples of healthy individuals, CD (affected in the ileum) and UC patients using two different methodologies: 16S rRNA gene (or 16S rDNA) and shotgun (short genomic fragments) sequencing.

As expected, we observed the presence of dysbiosis in IBD. Furthermore, we showed that microbial composition and function alterations were different for CD and UC, with greater dysbiosis in CD than in UC and with UC resembling more to a healthy state. Functional findings also confirmed this higher dysbiosis in CD than in UC and revealed

genes implicated in metabolism pathways and in immune diseases in higher abundance in CD compared with healthy individuals and UC.

Although 16S rDNA and shotgun data did not detect differences in the dysbiosis in CD and UC in a consistent manner, both methodologies allowed the classification of IBD subtypes in a similar proportion. Future studies should validate these results using other patient cohorts such as colonic CD or recently diagnosed patients before the application of these techniques as diagnostic tools in clinical practice.

RESUMEN

La microbiota intestinal desempeña un papel crucial en el manteniendo la homeostasis intesitnal. Alteraciones en la composición microbiana, también conocidas como disbisosis, pueden poner en peligro el estado de salud e incrementar el riesgo a padecer una enfermedad. Aunque muchas enfermedades se han asociado a cambios en la microbiota intestinal, todavía se desconoce si dichas alteraciones son la causa o la consecuencia de las patologías.

La enfermedad inflamatoria intestinal (EII) es una enfermedad inflamatoria crónica que se caracteriza por periodos de inflamación y constituye un problema de salud dado. La EII presenta dos subtipos: enfermedad de Crohn y colitis ulcerosa, con síntomas similares pero diferentes manifestaciones clínicas. La EII se ha relacionado ampliamente con cambios en la microbiota intestinal. A pesar de los múltiples estudios que existen, no hay un claro consenso en el perfil microbiano asociado a la enfermedad. Las principales discordancias se dan entre las diferencias asociadas a enfermedad de Crohn y la colitis ulcerosa. Algunos investigadores han demostrado que la composición microbiana en colitis ulcerosa es muy similar a la de individuos sanos y ambas difieren de la composición de enfermos de Crohn. En cambio, otros investigadores han visto que las diferencias de colitis ulcerosa y Crohn respecto a sanos son muy similares por lo que consideran ambos subtipos como una única enfermedad (EII).

El principal objetivo de esta tesis es determinar la disbiosis en una cohorte de EII española para evaluar hasta qué punto las funciones y composición microbiana difieren entre Crohn y colitis y si los datos de microbioma podrían emplearse como herramientas de diagnóstico. Para ello, analizamos muestras fecales de sanos, enfermos de Crohn y enfermos de colitis usando dos metodologías: secuenciación del gen 16SARNr (o 16S ADNr) y secuenciación por fragmentación del genoma.

Como se preveía, observamos la presencia de disbiosis en EII. Además, vimos que las alteraciones en composición microbiana y funciones eran diferentes para Crohn que para colitis, mostrando una mayor disbiosis en Crohn que en enfermos de colitis ulcerosa y con colitis mostrando un patrón muy similar a la microbiota de individuos

sanos. Los resultados funcionales encontrados en esta tesis confirman la mayor disbiosis descrita en pacientes de Crohn en comparación con pacientes de colitis ulcerosa en composición microbiana. Estos individuos presentan una mayor cantidad de genes principalmente asociados a metabolismo y enfermedades inmunes que los enfermos de colitis ulcerosa y sanos.

A pesar de que los datos de 16S ADNr y secuenciación por fragmentación no detectaron las mismas diferencias entre Crohn y colitis, ambas metodologías permitieron la clasificación de los distintos subtipos de EII con una proporción similar. Más estudios son necesarios para validar los resultados de esta tesis en otras cohortes de pacientes que incluyan Crohn localizado en colon o pacientes recién diagnosticados que no hayan sido sometidos a tratamiento antes de la aplicación de estas metodologías como herramientas diagnósticas en clínica.

LIST OF ABBREVIATIONS

A, Adenine

AI, Artificial intelligence

AUC, Area under the curve

BLAST, Basic local alignment search tool

BMI, Body mass index

C, Cytosine

CD, Crohn's disease

dNTP, Deoxyribonucleoside triphosphate

eggNOG, Evolutionary genealogy of genes: Non-supervised Orthologous Groups

EII, Enfermedad inflamatoria intestinal

EIM, Extraintestinal manifestations

FC, Fecal calprotectin

FDR, False discovery rate

FISH, Fluorescence in situ hybridization nucleotides

FMT, Fecal microbial transplantation

G, Guanine

GIT, Gastrointestinal tract

HC, Healthy controls

HMP, Human Microbiome Project

HUMAnN, HMP Unified Metabolic Analysis Network

IBD, Inflammatory Bowel Disease

IBS, Irritable Bowel Syndrome

IGC, Integrated gene catalog

KEGG, Kyoto Encyclopedia of Genes and Genomes

MEGAN, MEtaGenome Analyzer

MetaHIT, Metagenomics of the Human Intestinal Tract

MGB, Microbiota Gut Brain

MOCAT, Metagenomic Analysis Toolkit

mOTU, Metagenomic operational taxonomic unit

OTU, Operational taxonomic units

PATRIC, Pathosystems resource integration center

PCR, Polymerase chain reaction

qPCR, Quantitative polymerase chain reaction

RDP, Ribosomal database project

T, Thymine

TRUC, T-bet$^{-/-}$RAG$^{-/-}$ ulcerative colitis

UC, Ulcerative colitis

INTRODUCTION

1. Microbiota

Microbial communities, also known as microbiota, are multi-species collections of microorganisms that cohabit together in the same environment interacting with each other and are essential in the earth's ecosystem and the human body (Segata et al. 2013). In both cases, they are needed for important functions such as material degradation or nutrients absorption. Each individual of these microbial communities carries 500000 non-redundant genes (Qin et al. 2010) to develop those functions.

In nature, isolated microorganisms are rarely found as they tend to interact with each other to maintain the ecosystem. Members of microbial communities depend on each other and on the host. There are microorganism-microorganism and microorganism-host interactions in all environments. In these interactions, both members exchange molecular and genetic information (Braga, Dourado, and Araújo 2016).

One of the most crowded microbial communities is the human body. Microorganisms inhabit the human body and comprise more than 100 trillion microbes distributed along different locations of the body such as the skin, the gastrointestinal tract, the vagina or the respiratory system (Sender, Fuchs, and Milo 2016). In the human body, the densest microbial community is set in the gut (Ley et al. 2009). Gut microbial community, also known as gut microbiota, contains a quantity of microbes that accounts for more than ten-fold the number of cells in the whole body (Bäckhed et al. 2005; Turnbaugh et al. 2007) that is composed of more than 10^{13} human cells (Savage 1977).

1.1 Gut Microbiota

Gut microbiota comprises all bacteria, archaea and eukarya that coexist along all the gastrointestinal tract (GIT) starting from the oral cavity, through the esophagus, the stomach, small and large intestine until the rectum. Gut microbiota was firstly reported by the father of the microscope, Antonie van Leeuwenhoek, who described the presence of "strange little animals" using a microscope (Toledo-Pereyra 2009) his own stool and characterized the known species *Giardia spp*. in presence of diarrhea (Rajilić-Stojanović and de Vos 2014).

The composition of the intestinal microbiota in the GIT sections depends on the oxygen availability, pH, temperature and tissue structure (Booijink et al. 2010; Jenkinson and Lamont 2005). The vast majority of gut microbiota inhabits the human colon due to favorable conditions (Ley, Peterson, and Gordon 2006; Turnbaugh et al. 2007).

The gut microbiota is not yet completely characterized since it is composed by bacterial species still unknown that has not been identified. A recent database published called Culturable Genome Reference (CGR) has shown a better resolution in the description of the gut microbiome being able to identify approximately 70% of the microbial sequences in a sample (Zou et al. 2019). Intestinal microbiota mainly consists of two bacterial phyla independently of the geographical location (**Figure 1**), Bacteroidetes and Firmicutes, which contribute for more than 90% of microorganisms present in the gut (Ley et al. 2009; Qin et al. 2010; Wexler and Goodman 2017). Less dominant phyla include Proteobacteria, Actinobacteria, Fusobacteria and Verrucomicrobia phyla (Eckburg et al. 2005; Hayashi, Sakamoto, and Benno 2002). Although there is a quite stable gut microbiota composition at phylum level specially in middle-aged individuals, at lower taxonomical levels there are greater changes depending on age, way of delivery, medication, use of antibiotics, diet, physical activity and genetic host factors (Odamaki et al. 2016; Palmer et al. 2007; Shin et al. 2016; de Souza and Fiocchi 2016; Yatsunenko et al. 2012).

Viruses, archaea and eukaryotes (mainly fungi) also play an important role in the GIT. Gut viruses, known as gut virome, target eukaryotic host cells, bacteria or even archaea but it is mainly composed of bacteria-infecting phages also called bacteriophages. Those bacteriophages show higher abundance in early stages of life what has been associated with the continuous reshaping of gut microbiota in the first years (Lim et al. 2015). Gut virome is a novel field compared with gut bacteriome but, although little is known of these microorganisms, microscopy technologies have shown that the predominant viral order is Caudovirales that encompasses most of the known phages (Carding, Davis, and Hoyles 2017). Nowadays, the development of new techniques is improving the characterization of the gut virome (Milani et al. 2017).

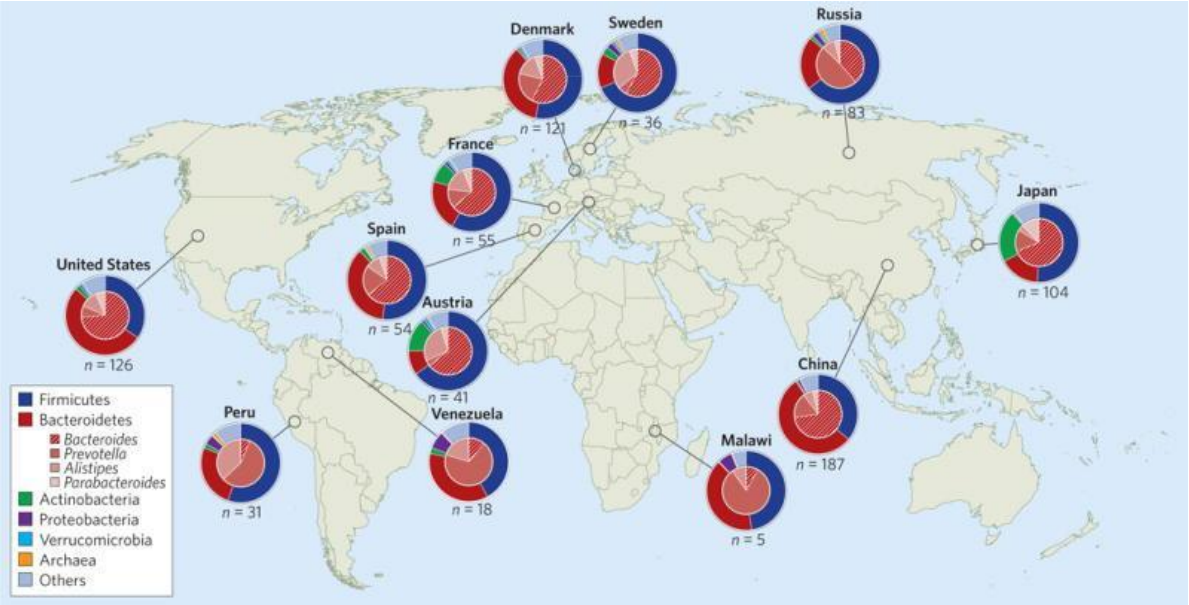


Figure 1. Global distribution of human gut microbiota abundances at phylum and genus level (From Wexler et al., Nature Microbiology 2017). Each pie represents the composition at phylum and genus level from the different countries and sample sizes. Bacteroides and Firmicutes account for more than 90% of the microbial composition in almost all geographical localizations.

Gut mycobiome accounts for approximately between 0.03 and 0.1% of the gut microbiota and is mainly composed of *Candida sp.* according to cultivation and non-cultivation techniques (Z. K. Wang et al. 2014). However, there is a lot of taxa variability among studies showing instability in time and between individuals (Hillman et al. 2017). As for gut virome, functions of the gut mycobiome are not well understood but it has been seen that they have an impact in the gastrointestinal function and contribute to digestive diseases (Z. K. Wang et al. 2014).

Archaea also contributes to a small proportion to the gut microbiota. Despite its low relative proportion, the most important Archaea genera in the gut, *Methanobrevibacter*, plays an important role in the GIT by its contribution to the production of methane. In the recent years, advances in culture independent techniques have allowed the identification of high diversity of archaea in the gut microbiota (Gaci et al. 2014; Hillman et al. 2017).

Gut microbiota can be considered as an additional organ that communicates with the host and is able to develop functions that the host cannot perform by itself. It is involved

in energy consumption, storage and chemical transformation among others (Bäckhed et al. 2005). Moreover, it is involved in health and disease status due to its influence in nutrition, protection against pathogens and immune function of the host (Maslowski and MacKay 2011; O'Hara and Shanahan 2006). It plays an important role in the digestion of specific food substances that otherwise could not be degraded such as dietary carbohydrates (fiber) (Marchesi et al. 2016). Gut microbiota helps the host to absorb and metabolize these nutrients and, in return, it lives and replicates in a nutrient enriched niche provided by the host (Hooper, Midtvedt, and Gordon 2002). In the last few years, interest in studies relating gut microbiota to brain functions, known as microbiota-gut-brain (MGB) axis, has strongly grown suggesting that it could influence stress response, behavior, cognition and brain chemistry and development (Cryan et al. 2012; Scriven et al. 2018; Zhu et al. 2017). Interactions in this MGB axis are bidirectional meaning that not only the gut microbiota can affect brain function, but also the brain can affect the gut microbiota indirectly by modifying their niche with alterations in intestinal motility or directly by secreting molecules in the gut (Carabotti et al. 2015; Rhee, Pothoulakis, and Mayer 2009).

1.2 Microbiota in health and disease

Interest in identifying the role of the gut microbiota in the health of the host has grown over the last decade in different fields. Furthermore, current high-throughput sequencing technologies provide information about gut microbiota composition and functions with a substantial decrease in prices and higher resolution than some years ago. Considering all these improvements of the techniques and the increasing interest in the gut microbiota in the last years, the development of important research projects with big cohorts in this field has grown.

1.1.1 International projects

In the United States in the 2000s, the Human Microbiome Project (HMP) was designed to deeply characterize and understand how microbiome affects human health. This

project consists of two phases. Firstly, HMP tended to describe human microbiota from different body sites (nasal passages, oral cavity, skin, gastrointestinal tract, and urogenital tract) from 300 healthy subjects trying to determine if there was a healthy core microbiota. Secondly, the ongoing project of the HMP involves 3 different cohorts (pregnancy and preterm birth, onset of inflammatory bowel disease (IBD) and onset of type 2 diabetes) to investigate how the human microbiome can contribute to disease status (https://hmpdacc.org/, (Turnbaugh et al. 2007)).

Meanwhile, in Europe, another project was gaining importance. Metagenomics of the Human Intestinal Tract (METAHIT) project focused on analyzing only microbial communities from the gut. This community contains high concentrations of microbes resulting in an ecosystem of high interest to study. The aim of this project is to investigate how microorganisms in the gut can contribute to health status. (http://www.metahit.eu/index.php?id=234, (Qin et al. 2010)).

1.1.2 Dysbiosis

High diversity and richness are commonly associated with health status. When a disbalance, called dysbisosis, appears in the microbial community composition and/or function of an individual, it can lead to a disease status (Ni et al. 2018)). Many studies investigating diseases in Western countries have associated dysbiosis with environmental factors and Western way of life in which the use of antibiotics is well established (Mosca, Leclerc, and Hugot 2016). Intestinal disorders such as IBD or Irritable Bowel Syndrome (IBS) have been clearly linked with alterations in the gut microbiota (Gevers 2015; Hall et al. 2017; Moustafa et al. 2018; Pascal et al. 2017; Pozuelo et al. 2015; Sokol et al. 2017) but also non-intestinal disorders such as metabolic, autoimmune and neurological disorders (Clarke et al. 2012; Qin et al. 2012; La Rosa et al. 2018; Santiago et al. 2016).

Whether gut microbiota is the cause or consequence of the disorders is still unresolved. In the specific case of obesity, animal models have demonstrated that gut microbiota play an important role in the disease. Ridaura *et al.* transplanted fecal microbiota from adult twins discordant for obesity into germ free mice that were fed with a low-fat

mouse diet. This study resulted in the transmission of high body and fat mass together with a metabolism typical from an obese mouse (Ridaura et al. 2013). In the case of digestive disorders such as IBD, animal models have also shown the implication of gut microbiota in the development of the illness. Ni *et al.* showed in mice models of spontaneous colitis, T-bet$^{-/-}$RAG$^{-/-}$ ulcerative colitis (TRUC) mice, that germ free animals did not develop ulcerative colitis. However, colitis requires not only the presence of microorganisms but also an inflamed environment (Ni et al. 2017). Nevertheless, to elucidate whether the gut microbiota acts as a cause or a consequence of different disorders, more studies are needed.

2. Analyses of the gut microbiome

Fecal and intestinal tissue samples could be used a proxy to characterize the gut microbiome. Collection of fecal samples is a non-invasive technique and represents the easiest way to obtain samples. The main drawback of fecal samples is that they may not be representative of all the gut microbial composition. According to previous studies, it has been shown that fecal microbiota differs from microbiota adhered to the mucosal surface with less Proteobacteria (Carstens et al. 2018; Durbán et al. 2011). On the other hand, tissue samples comprise rectum or ileal biopsies collected after colonoscopies or colonic resections and effluent or mucosal samples obtained after bowel transplantations (Booijink et al. 2010; Gevers 2015). Collecting these types of samples involves invasive techniques for the host and need a bowel preparation before intervention except for the rectum biopsies (Gevers 2015). Due to the difficulty to obtain biopsies from the small intestine, microbiota from this section of the GIT remains poorly characterized (Booijink et al. 2010).

Studies of the gut microbiota have evolved over the years. Traditionally, those studies used culture methods, however, although this technique is still applied, it is used for microbiota interactions and not for characterization. New techniques have appeared to overcome the limitations of culture. These new approaches focus on bacterial DNA present in samples to determine microbial composition and potential functions.

2.1 Culture dependent techniques

Gut microbiota has been traditionally characterized using culture techniques that are based in the capacity of microorganisms to grow and replicate in an environment with optimal conditions and nutrients (Rajilić-Stojanović and de Vos 2014). In the 1970s Moore and Holdeman demonstrated that the gut microbiota was composed by at least 50% of strictly anaerobic bacteria (Moore and Holdeman 1974). This discovery made researchers to be very careful when working with culture methods to maintain an oxygen-deprived environment.

In the 1990s-2000s, high-throughput technologies showed that between 60% to 80% of the gut microbiota had not grown in culture (Hooper et al. 2002). The reason that was proposed to explain this lack of growth was the inadequate environmental conditions and absence of suitable nutrients and medium (Cryan et al. 2012; Hayashi et al. 2002; Langendijk et al. 1995; Suau et al. 1999). Researchers working with culture had to take several points into account when interpreting their results; 1) the number of phenotypic characteristics and biochemical reactions to test was limited, 2) the number of microorganism that could grow in culture was restricted and 3) microorganisms that grow in culture could act differently in medium than in their natural niche as they might be involved in co-operational networks with other members of the gut microbiota. Consequently, nowadays, culture techniques are rarely used for microbial characterization, but they are still used for antibiotic resistance studies, species isolation, growth factors production or gut microbiota dynamics among others (D'hoe et al. 2018; Rashid et al. 2015; Zhang et al. 2015).

2.2 Culture independent techniques

New methods have appeared to overcome the limitations generated by culture techniques. Currently, most common methods for microbiome characterization are Next Generation Sequencing Techniques (Clooney et al. 2016). The cost of these

methodologies has decreased in the last few years boosting microbiome studies (J Gregory Caporaso et al. 2012) (**Figure 2**).

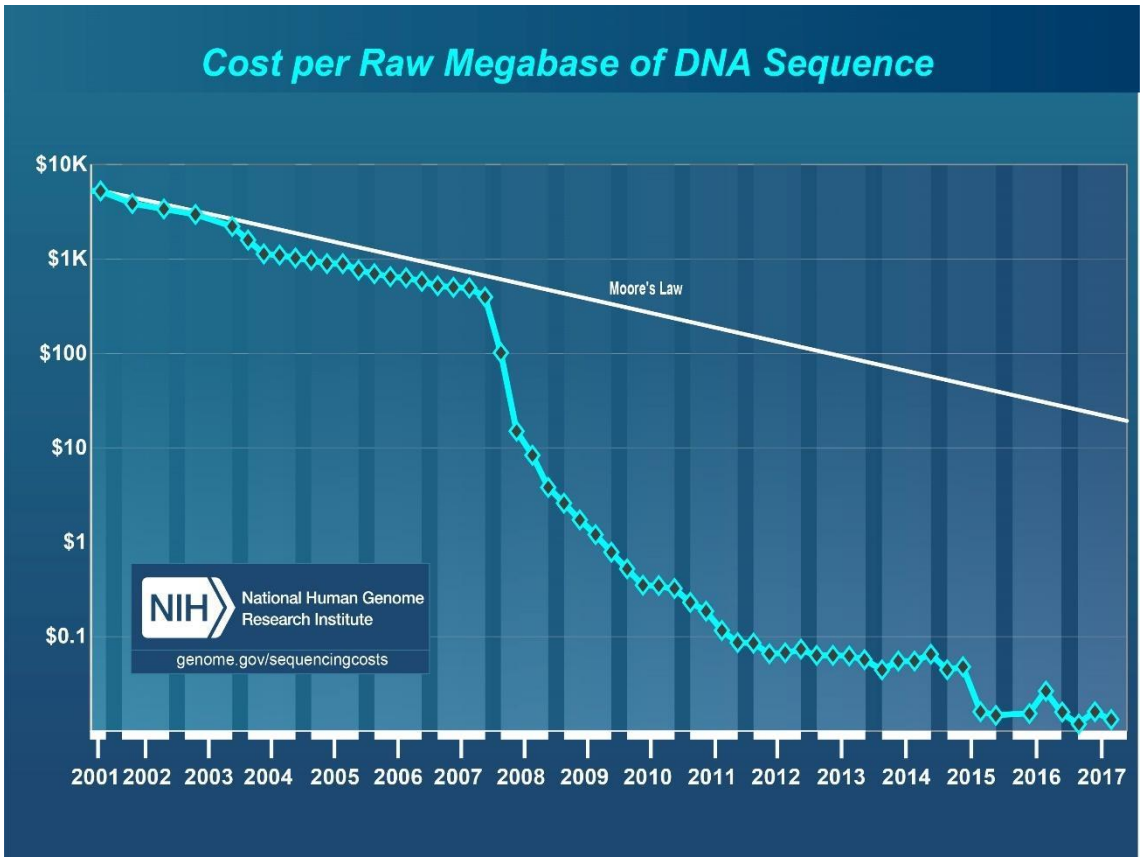


Figure 2. Sequencing cost over the last decade. (From https://www.genome.gov/sequencingcosts). Cost of high throughput sequencing methodologies have exponentially decreased in the last few years contributing to a raise in the studies of microbiome.

Next generation sequencing techniques used in gut microbiome analyses are targeted-sequencing and shotgun sequencing on microbial DNA or RNA. In most studies, targeted-sequencing approaches focus on bacterial 16S rDNA gene that is used to describe taxonomical composition in samples. On the other hand, DNA shotgun sequencing methods identify potential genes to make an exhaustive analysis and observe not only the taxonomical composition but also gene composition (Hillmann et al. 2018; Jovel et al. 2016). This approach allows the identification of gut microbial

potential functions in the GIT, however, not all these genes are active at the time of sampling. To avoid this problem and determine which functions are expressed, we use metatranscriptomics. Metatranscriptomics is based in RNA shotgun sequencing to identify active genes and quantify their expression at sampling time. There are other 'omics' that are also used to study the human gut microbiome (Segata et al. 2013). Metaproteomics studies all proteins present in a sample. Metabolomics is responsible for identifying metabolites, chemical substances present in the samples (**Figure 3**). These metabolites can be associated with the gut microbiota in two different ways: on the one hand, the presence of one metabolite can promote the growth of specific species whereas on the other hand that species can produce that metabolite (Franzosa et al. 2019).

2.2.1 Targeted-sequencing methods

Targeted-sequencing methods involve the amplification of marker genes by Polymerase Chain Reaction (PCR) to taxonomically describe a microbial community. Marker genes are orthologous groups of genes with known localization that are used to differentiate between taxonomical groups and its heritability is easy to follow. The selection of marker genes is a critical decision as the precision of microbial detection depends on their capacity to differentiate members of a microbial community. Marker genes vary depending on the microorganisms that are been analyzed. In the case of bacteria and archaea, 16SrDNA is a widely marker gene used over the last 30-40 years (Woese and Fox 1977) and has been described as a good source of information for bacterial evolution together with 18rDNA gene for fungi (Fox, Pechman, and Woese 2018).

16S rDNA gene is a highly conserved gene of about 1500 nucleotides (nts) that is only present in the small subunit of bacterial and archaeal ribosomes. 16S rDNA gene contains several conserved regions that flank nine hyper-variable regions (V1-V9) that differ between species. Conserved regions allow the use of universal primers to detect all bacteria present in a sample whereas hypervariable regions determine which bacteria is being detected (**Figure 4**).

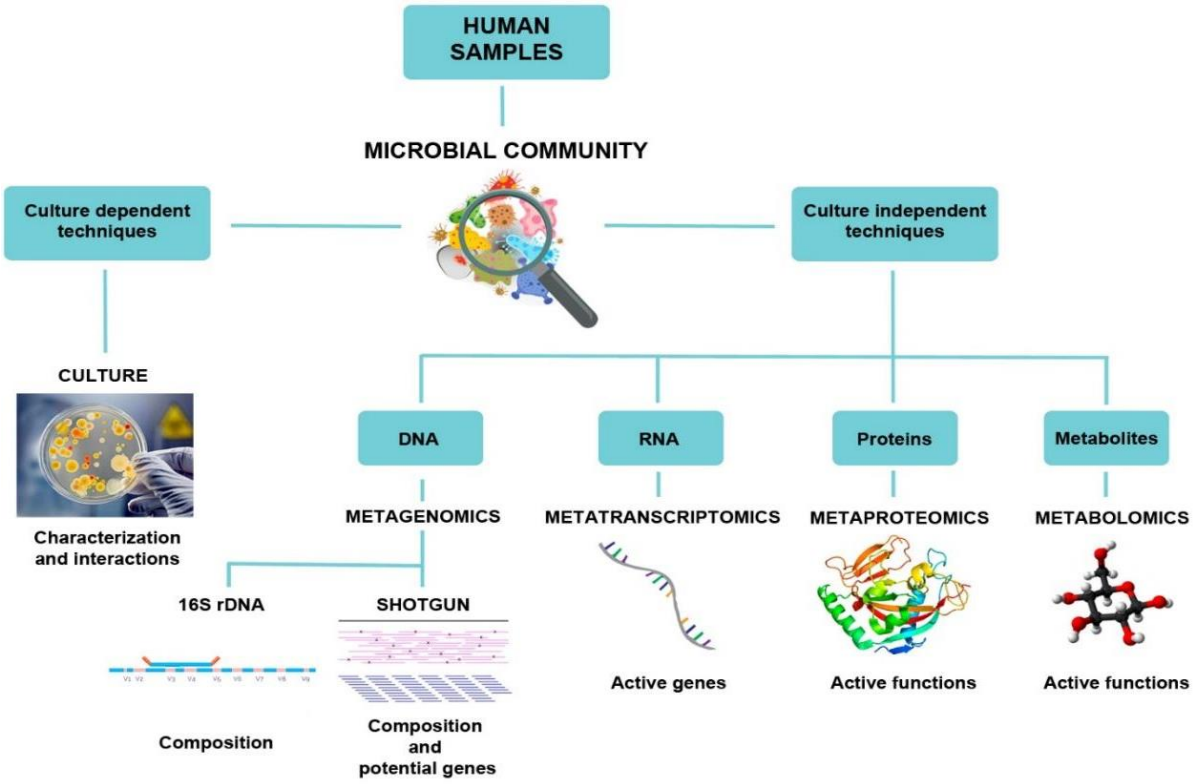


Figure 3. Different methods to study the microbiota: identification techniques for gut microbiota have evolved in the last few years to improve the detection of gut microbiota. Culture techniques can only provide information of approximately 20% to 50% (Wexler and Goodman 2017) of the gut microbiota but they are still in use to characterize bacteria and determine possible interactions between them. To improve this percentages of detection, new techniques have been developed based on DNA (metagenomics), RNA (metatranscriptomics), proteins (metaproteomics) and metabolites (metabolomics). In metagenomics, researchers use targeted or non-targeted methodologies to determine composition or both composition and potential genes, respectively. Metatranscriptomics, metaproteomics and metabolomics identify which genes and functions are active at the time of sampling but not the genetic potential of the sample.

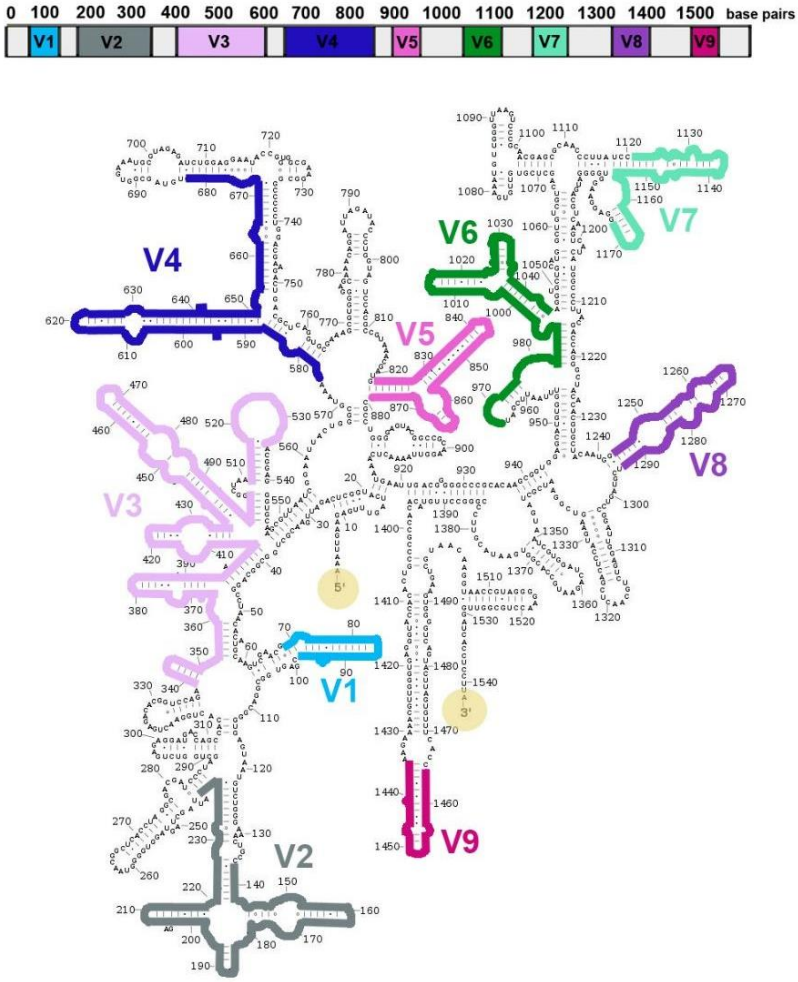


Figure 4. *E. coli* 16S rDNA gene structure. Conserved regions are represented in light grey and variable regions in different colors. Figure adapted from http://rna.ucsc.edu/rnacenter/xrna/xrna_gallery.html. Conserved regions are present in all bacteria and allow the design of universal primers that amplify all bacteria present in a sample. Hypervariable regions differ between bacterial species and allow the classification of microbiota in different phylogenetic categories.

In the last 30 years, 16SrDNA sequencing has become the prevalent mechanism used to analyze microbiota from the GIT and other environments. The main issue to consider is which hypervariable region is most appropriate to obtain a taxonomic profile closest to the reality. Studies have shown that V4 region in different environments contributes to a better approximation of the microbial composition in a sample (Zhang et al. 2018; Zhao et al. 2013). Sequencing with universal primers results in millions of sequences that correspond to bacteria and archaea in microbial communities. All these 16S rDNA sequences are then stored in several databases such as GreenGenes

(http://greengenes.secondgenome.com/), SILVA Ribosomal RNA database (https://www.arb-silva.de/) or Ribosomal Database Project (RDP) (https://rdp.cme.msu.edu/ databases.

Sequencing techniques have evolved in the last years showing important advances in sequencing platforms. Nowadays, most researchers in the microbiome filed use the Illumina MiSeq or HiSeq technology that provides millions of raw reads after quality filtering than older platforms such as Sanger, PacBio, Ion Torrent PGM or Roche 454 GS FLX Titanium platforms (Allali et al. 2017). In addition, several tools for analyzing 16S rDNA sequences have been developed such as QIIME (Quantitative Insights Into Microbial Ecology) (Navas-Molina et al. 2013), mothur (Schloss et al. 2009), UPARSE (Edgar 2013), DADA2 (Callahan et al. 2016), MetaAmp (Dong et al. 2017) or ANCHOR (Gonzalez, Pitre, and Brereton 2019). In general, most of these tools are based in the clustering of raw sequences into operational taxonomic units (OTUs) following different clustering algorithms. They classify each OTU as a bacterial species and generate abundances tables that are used to compute alpha (species present in an ecosystem) and beta (species differences between ecosystems) diversity and differential bacterial abundances.

Targeted approaches are also used not only for identifying, but also for quantifying microbial DNA. Techniques such as real-time quantitative PCR (qPCR) are based on the amplification of a targeted marker gene. Depending on the primers used for the qPCR, this technique allows the detection and quantification of total bacteria, in the case of using universal primers, or of a specific group of species if primers specific to a certain group are utilized (Bartosch et al. 2004; Carey et al. 2007).

2.2.2 Shotgun sequencing

Shotgun sequences of the genomic DNA present in a sample rather than focusing on a specific gene is an approach based on the fragmentation of all the DNA extracted and the sequencing of these small fragments that will be later assembled to build bigger genomic fragments or entire bacterial genomes. Shotgun sequencing allows the identification of most entities present in the microbiome (bacteria, archaea, eukaryotes

and viruses) and does not focus only on a specific group as targeted-sequencing. Another advantage of shotgun sequencing is that this technique reaches deeper taxonomic levels providing information at species level whereas 16S rDNA sequencing only reaches genera level with high precision. Moreover, shotgun sequencing provides not only taxonomical information but also allows functional identification. The main drawbacks of shotgun sequencing are its elevated cost in money and time for running the sequence analyses, exceeding by far those of 16S rDNA sequencing (Barko et al. 2018) (**Figure 5**).

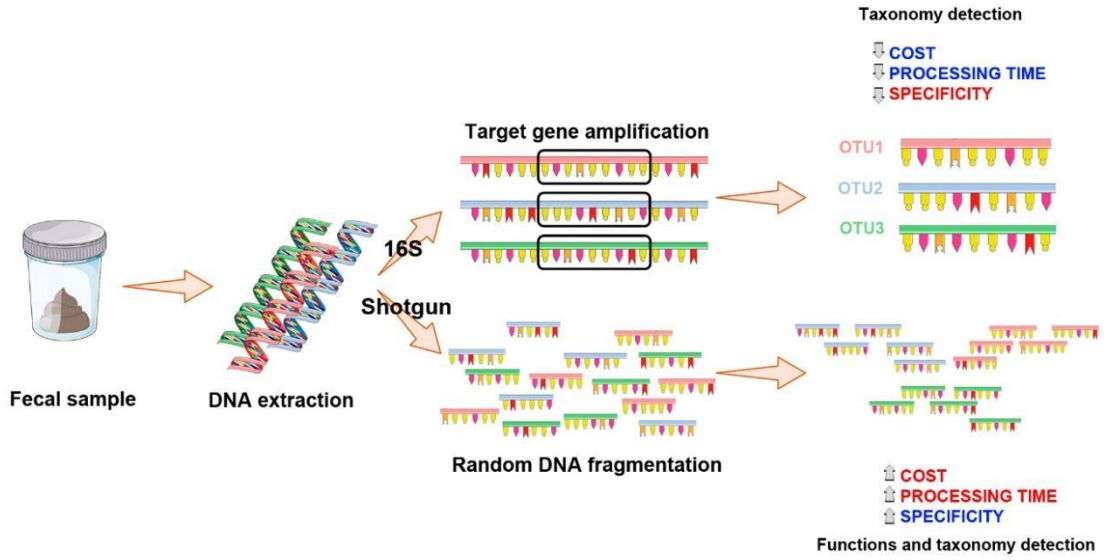


Figure 5. Comparison between 16S rDNA targeted sequencing and shotgun sequencing of a fecal sample. Metagenomic approaches focus their analyses on DNA extracted from a sample. There are two basic methods: 1) Targeted gene amplification that focalizes in just one gene, in this case, 16S rRNA gene which is adequate to identify bacteria and 2) Shotgun sequencing which is based in the fragmentation of all the DNA present in a sample and assembly of these fragments to reconstruct genomic fragments or whole genomes. Although the cost in money and processing time together with storage space of 16S target gene amplification constitute the pros of this technique, the possibility to reach species level in the analyses is very restricted. Furthermore, 16S gene amplification methodology can only give compositional information and make a prediction of possible functions associated to microorganisms detected. Shotgun sequencing also allows the identification of other entities different from bacteria that are not restricted to the 16S rRNA gene sequencing.

Similar to the bioinformatics workflow used in the 16S rRNA analysis, several pipelines have been developed for the analyses of bacterial sequences from shotgun sequencing data mainly coming from Next generation sequencing (NGS) platforms (Lam et al. 2011). HUMAnN (HMP Unified Metabolic Analysis Network) (Abubucker et al. 2012; Franzosa et al. 2018), MOCAT (Metagenomic Analysis Toolkit) (Kultima et al. 2012, 2016), MG-RAST (Keegan, Glass, and Meyer 2016) or MEGAN (MEtaGenome Analyzer) (Huson et al. 2007) are pipelines widely used for the metagenomic analyses starting from a quality control of usually paired-end raw reads and returning abundance tables of genes annotated with functional databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al. 2016) or eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (Huerta-Cepas et al. 2016).

3. IBD

Intestinal bowel disease (IBD) is an immune mediated chronic inflammatory disease which causes a substantial economic burden on the healthcare system. IBD can affect different sections of the gastrointestinal tract causing ulceration of the intestinal mucosa and encompasses two main subtypes with different clinical manifestations: Crohn's disease (CD) and ulcerative colitis (UC). IBD usually appears during young adulthood and is characterized by periods of inactive (remission) and active inflammation (relapse) (Dalal and Chang 2014).

CD can affect any section of the GIT from the mouth to the anus, but it is commonly present in the ileocecal section (**Figure 6**). Inflammation in CD can appear asymmetrically and segmented in all intestinal layers so there can be several transmural inflamed areas separated by healthy portions (Laass, Roggenbuck, and Conrad 2014). CD patients are classified according to the Montreal classification considering the age of diagnosis and the localization and behavior of the disease in five subtypes: ileocolitis (affects the ileum and the colon), ileitis (only in the ileum), gastroduodenal Crohn's disease (affects the stomach and the duodenum), jejunoileitis (jejunum) and Crohn's colitis (only affects the colon) (Satsangi et al. 2006; Silverberg et al. 2005). Relapse

periods in CD occur in most CD patients with percentages of 53%, 85% and 90% at 1, 5 and 10 years, respectively (Aniwan, Park, and Loftus 2017).

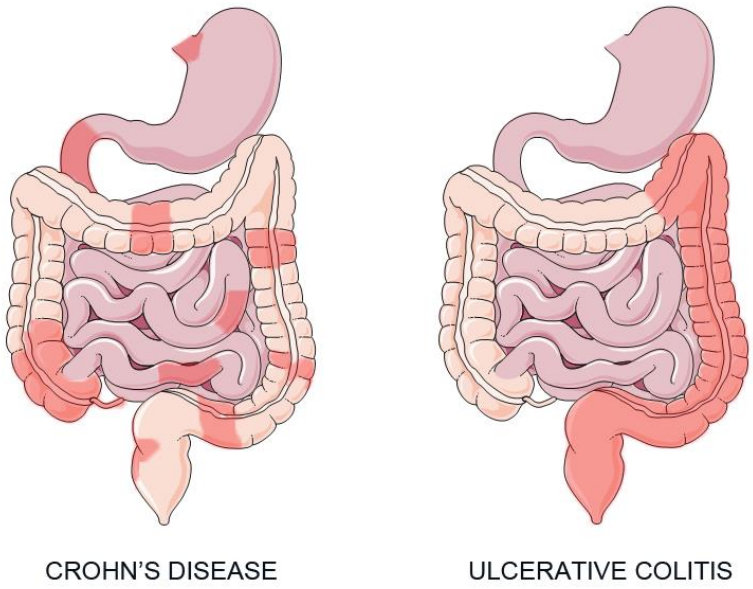


Figure 6. Inflammation areas in Crohn's diasease (CD) and ulcerative colitis (UC). IBD is characterized by periods of active (relapse) and inactive (remission) gastrointestinal inflammation. CD patients can present inflammation in any section of the gastrointestinal tract from the mouth to the anus whereas patients with UC only present inflammation in the colon region.

UC consists on localized and symmetric inflammation in the colon and affects in 95% of the cases the rectum (**Figure 6**). UC affects equally women and men and may appear at any age. Inflammation appears in the mucosal layer of the colon starting in the rectum and expanding in circles (Dalal and Chang 2014). Depending on the extent and severity of the inflammation, we can distinguish four types of UC: proctitis (limited to the anus and rectum), proctosigmoiditis (affecting the rectum and the sigmoid colon), left-sided colitis (begins in the rectum and extends to the splenic flexure) and pan-ulcerative colitis (affects the entire colon). Many patients present long periods of remission but the probability of not suffering a relapse in two years is close to 20%, while this percentage is reduced to 5% in periods of 10 years (Ghosh, Shand, and Ferguson 2000).

3.1 Prevalence and incidence

The incidence of both subtypes of IBD, CD and UC, has increased in the last few years in the Western areas including North America, Europe, Australia and New Zealand with high incidence of 10 to 30 cases per 100,000 individuals and a prevalence of 0.5% of the total population (Aniwan and Park 2017). Newly developing industrialized areas in which IBD had been rarely described such as India or Thailand incidence has importantly grown with values of 1.4 per 100,000 in 2011 (Cosnes and Cortot 2011; Kaplan 2015; Loddo and Romano 2015) (**Figure 7**).

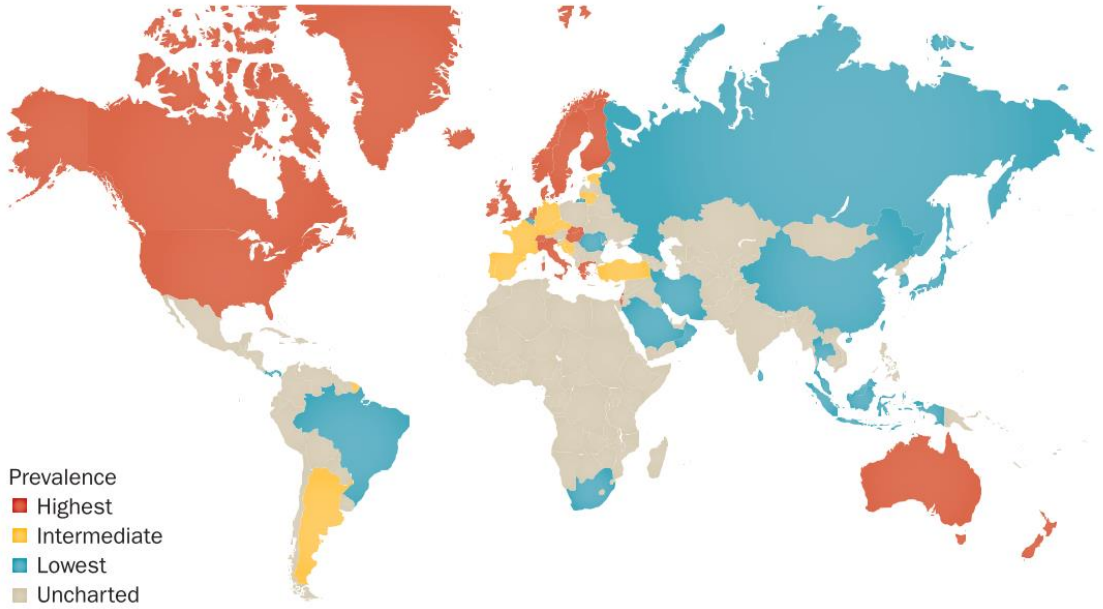


Figure 7. IBD incidence map. (From Kaplan et al., Nature Reviews Gastroenterology and Hepatology 2015). Low or blue areas corresponds to annual incidence less than 4 cases per 10^5 individuals, intermediate or yellow values correspond to incidence of 5-10 IBD cases per 10^5 individuals and finally, red or high areas refer to IBD incidence of more than 10 cases per 10^5 individuals. More industrialized areas such as North America, Europe or Australia show high and medium incidence of IBD. However, newly developing countries such as South Africa, India or China still present low number of IBD cases but they have been increasing for the last few years.

Considering both subtypes of IBD, both prevalence and incidence are higher in UC than in CD in Western and emerging populations (Ananthakrishnan 2015; Aniwan and Park

2017; Molodecky et al. 2012). Although IBD is still rare in Africa and South America, some cases have been reported in Nigeria showing a small increase in incidence (Ukwenya et al. 2011). These new diagnosed IBD patients coincide with the creation of a unit of gastroenterology equipped with diagnostic tools, suggesting that the growth in IBD cases may not be due to an increase in the number of patients but to a better diagnosis.

3.2 Etiology

The etiology of IBD is related to several factors that cannot individually explain the origin of the disease (**Figure 8**). Genetic susceptibility, stress, pollution, smoking, physical activity, medication together with other extrinsic factors cooperate together in the onset and progression of the gastrointestinal disorder (Abraham, Ahmed, and Ali 2017; Aniwan et al. 2017).

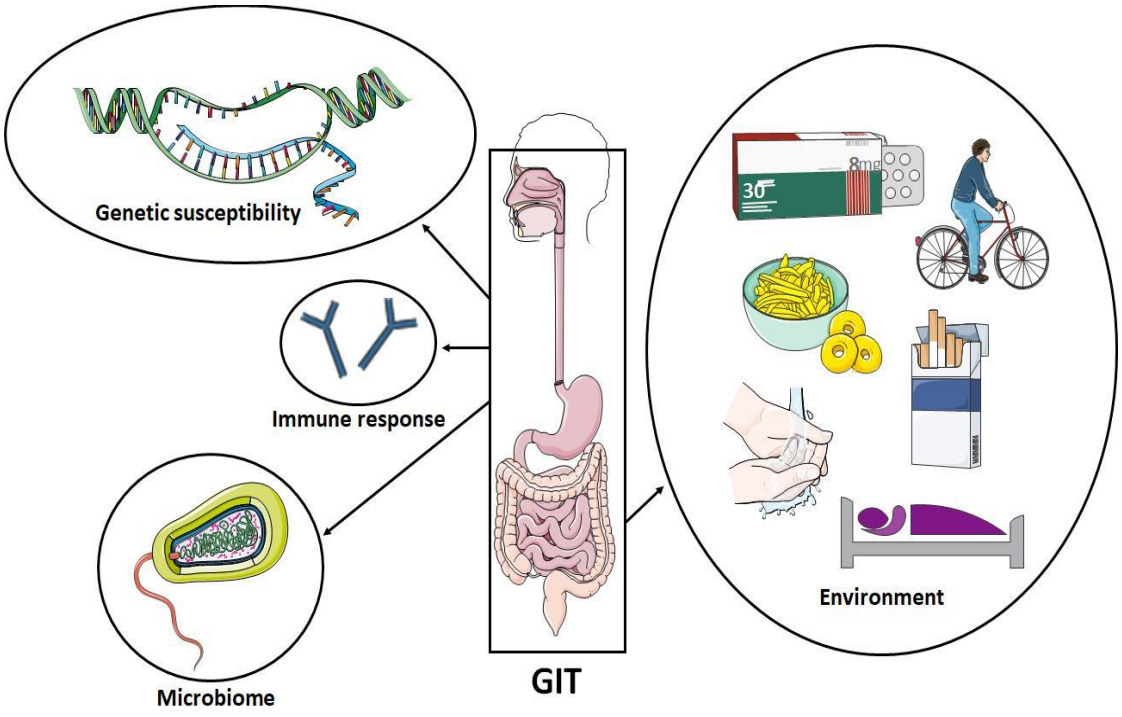


Figure 8. Etiology of IBD. The mechanism of IBD is not completely understood. There is a combination of several factors: genetics, environment, immunity and microbiome. They all contribute to the development of the disease. Among the environmental factors, smoking habit, stress or diet have been seen to have an important effect on the disease.

Over the years, different genetic factors have been associated with IBD. There are more than 200 risk loci associated to IBD (Momozawa et al. 2018). A great majority of the detected loci are linked to IBD while only a few of them are specific to UC or CD (Abraham et al. 2017; Liu et al. 2015). The high number of common loci indicates that they share common inflammatory pathways. However, specific loci can explain clinical, endoscopic and histologic differences found in patients (Ramos and Papadakis 2019).

Smoking is the most widely studied extrinsic factor in IBD. Active smoking is positively associated with CD. Conversely, it may act as a protection in UC (Harries, A. D., Baird, A. & Rhodes 1982; Mahid et al. 2006). Nevertheless, Jones *et al.* studied passive smoking and did not find any correlation with IBD development in children that were exposed to tobacco during their childhood (Jones et al. 2008).

Dietary factors have also been associated with IBD. Typical diets from Western countries characterized by low quantities of fiber, high in animal fat and sugar increase the risk of IBD (Ananthakrishnan 2015; Lewis and Abreu 2017). Researchers have shown that these diets are related to a decrease in microbial diversity or dysbiosis, which promotes the susceptibility to IBD (Chiba, Nakane, and Komatsu 2019).

Stress and depression are important disorders that affect Western populations and have been also studied as risk factors for IBD. Several studies have shown that IBD patients have a higher tendency to suffer from depression (Nowakowski, Chrobak, and Dudek 2016; Walker et al. 2008) which correlates with inflammation (Mittermaier et al. 2004). Moreover, depression can also worsen the development of the disease. However, the use of antidepressants that reduce the production of inflammatory cytokines have an impact in reducing the possibilities to develop IBD (Frolkis et al. 2018) In the case of stress, different studies in animal models have demonstrated that stress has a relation in both development and reactivation of the inflammation (Singh, Graff, and Bernstein 2009).

3.3 Symptomatology, diagnosis and treatment

IBD symptoms include diarrhea and abdominal pain together with blood in feces that can generate anemia leading to chronic fatigue in patients (Gasche et al. 2004). IBD is considered as a systemic disorder because its symptoms do not appear only in the intestine but also in other parts of the body. IBD can present dermatological, oral, ophthalmological and muscoeskeletal manifestations among others so it can affect every single organ in the body. The presence of extraintestinal manifestations (EIM) differ between adult and pediatric patients (Jang, Kang, and Choe 2019). It is important to distinguish which are those EIM to select the most appropriate treatment.

Furthermore, in many cases, IBD patients suffer from anxiety and bad mood contributing to disease worsening, which may lead to flares (Jordan, Hayee, and Chalder 2018). IBD symptoms constitute an important burden for the patients. They limit working capacities and welfare of the affected population constituting an important handicap for their life and driving to important indirect health-care costs. On the other hand, IBD also generates high direct health-care costs including hospitalization, surgeries, medical care or medicines (Kaplan 2015).

IBD diagnosis requires several tests and procedures such as blood tests and imagine and endoscopic procedures. Endoscopy is crucial to differentiate IBD from other disorders and to discriminate the two main subtypes of IBD. Endoscopies are also involved in the follow up of disease activity and treatment response. The main drawback of endoscopies is the invasive feature of the technique that also requires a bowel preparation (Spiceland and Lodhia 2018).

Current therapeutics in IBD focus on improving the quality of life of patients and lead to clinical remission of the disease. In most serious cases, surgery is even considered. In the past few years, treatments for IBD have not suffered many modifications and their use varies according to the type and severity of the disease. Oral aminosalicylates such as sulfasalazine or mesalazine are topical anti-inflammatory drugs that reduce gut inflammation when they are in contact with the mucosa. Both corticosteroids and immunomodulators acts as regulators of the immune system response reducing cellular production of inflammatory substance and reducing immune system activity,

respectively, to reduce intestinal inflammation (Abraham et al. 2017; Aniwan and Park 2017).

The implication of the gut microbiota in IBD development has led to the use of targeted therapeutic methods suitable for its modification. Clinicians usually prescribe antibiotics in IBD during infectious complications (Sartor 2004). Antibiotics such as metronidazole and ciprofloxacin have also shown good effectiveness as primary treatment for Crohn's disease. Metronidazole together with ornidazole contribute to prevent post-surgical relapse of the disease (Hammer 2011). Antibiotics also showed high effectiveness for treating pouchitis (Cohen et al. 2019). In general, the use of antibiotics focuses on decreasing the invasion of aggressive bacteria and reducing their proliferation in the lumen.

Probiotics are also used as a treatment for IBD. Probiotics are viable bacteria that have beneficial activities for the gut and have three possible ways of action in the intestine: alteration of immunoregulation, inhibition of pathobionts or strengthening the function of the mucosal and epithelial barrier (Hammer 2011). Several clinical trials have shown that probiotics are useful for treating active IBD and for preventing inflammation periods showing good results in pouchitis and UC more than in CD (Sartor 2004).

Additionally, not only the function of probiotics has been studied in IBD but also the function of prebiotics. Prebiotics are substances that can be ingested and favor the growth of beneficial microorganisms. Small studies with prebiotics have been done on UC patients showing a small decrease of inflammation in cases of mild to moderate UC (Mitsuyamat, Toyonaga, and Sata 2002).

Another alternative treatment is fecal microbial transplantation (FMT). The success of the treatment of *Clostridium difficile* infection with FMT has driven to the study of FMT as a possible therapeutic strategy for IBD. Several clinical trials have shown that the efficacy of this treatment for IBD is more modest than for *Clostridium difficile* infection and that it may depend on many variables including the composition of the stool donor. The importance of the diversity in the donor sample is leading to the concept of the "FMT super-donor" that would provide better results in FMT than others (Wilson et al. 2019). A recent study compiled all the studies that have been made in FMT and IBD

(Paramsothy et al. 2017). Although response of IBD patients to FMT is variable, summarizing the 53 studies, Paramsothy and colleagues showed that FMT was more efficient in CD patients than in UC when inducing a remission. They also showed that FMT with oral administration by infusion improved remission better than lower gastrointestinal tract administration in UC (Paramsothy et al. 2017)

3.4 Physiopathology

Numerous genomic studies have identified human candidate genes involved in the development of inflammatory processes (Momozawa et al. 2018). Among all these genes, the most promising ones encode proteins for the recognition and the presentation of the bacterial antigens and the coordination of innate and adaptative responses of immune system (Cho 2008). These findings suggest that the immune response to pathogens, especially bacterial, would be disrupted in IBD. Moreover, many environmental risk factors (smoking, appendicitis, etc.) as previously explained, may also participate in this immune dysfunction (Legaki and Gazouli 2016).

The intestinal barrier protects the body from potential threats. It consists of a bacterial biofilm, a layer of mucus and the intestinal epithelium, where specific cells providing the innate immune defenses (dendritic cells, Paneth cells, macrophages and neutrophils) are found. In IBD, each of these defenses is altered. The number of globet cells secreting the mucins, that which constitute the protective mucus of the intestinal epithelium, is decreased in IBD. In addition, the use of DNA chips has revealed a decrease in the expression of genes encoding mucins in the ileum and colon of patients (Moehle et al. 2006). The cohesion of the intestinal mucosa is ensured by the cellular junctions of the epithelial cells of the intestine. However, the proteins forming the tight junctions of the enterocytes (occluding, cadherins and catenins) have been found in decreased amount in IBD (Hill et al. 2004). Epithelial cells are also the first line of defense against invasion by pathogenic organisms, they can identify the bacterial pathogenic components by their receptors to extracellular bacterial peptides TLR (toll-like receptor) and intracellular NOD2 (or NOD2/CARD15 for nucleotide-binding oligomerization domain/caspase-activating recruitment domain 15) (Ramos and Papadakis 2019). Then,

they produce antimicrobial peptides (β-defensins HBD) and express MHC molecules to initiate the adaptative immune response of the mucosa. In IBD patients these signaling pathways are impaired and induce altered microbial peptide production and innate immune response (**Figure 9**).

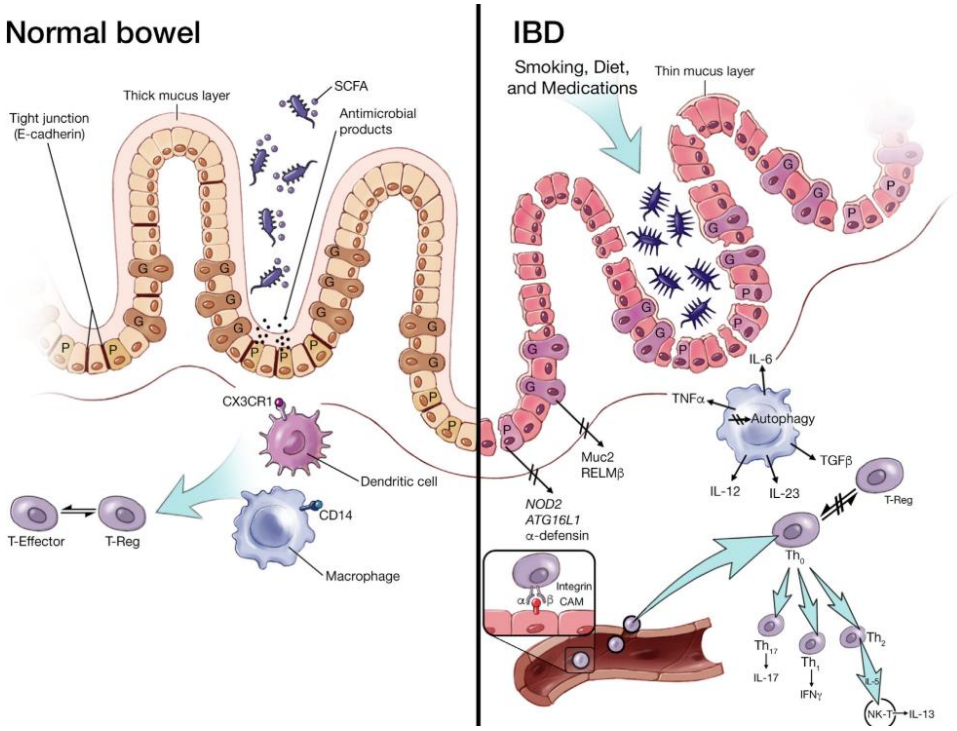


Figure 9. Gut physiology in a healthy and an IBD conditions. (From Ramos *et al.*, Mayo clinic proceedings 2019). IBD is characterized by and abnormal activity of the immune response and genetic expression combined with environmental factors. In IBD, the mucosal layer is formed by epithelial cells with disrupted activity due to the decrease presence in tight junctions, to the reduced production of antimicrobial products. In IBD there is an uncontrolled activation of different types of T-cells that migrate to inflamed tissues.

Abnormalities in the innate immune response disrupt antigen recognition and presentation of the effector cells. When IBD is active, there is an imbalance between the number of effector T cells (Th) and regulatory T cells (Treg). In Crohn's disease, Th1 lymphocytes, characterized by a high production of IL-2, IFNγ (interferon) and Th17, that predominate (Bamias et al. 2003; Fujino et al. 2003). Conversely, in UC, patients' mucosa

is infiltrated mainly by atypical Th2 lymphocyte that characterize the production of IL-5, IL-13 and TGFβ (transforming growth factor) (Targan and Karp 2005) (**Figure 9**).

4. IBD and gut microbiota

Although the origin of IBD is not completely clear, several studies, both in human and animal models, have demonstrated a strong correlation between gut microbiota and the onset and progression of IBD. Nowadays, the most recent model suggests that IBD is the result of an inappropriate immune response to microbiota in individuals with genetic susceptibility (Abraham et al. 2017; Sartor 2006; Weingarden and Vaughn 2017). CD and UC share several common clinical signs and symptoms. However, at the microbiome level, dysbiosis appears to differentiate the two IBD subtypes (Andoh et al. 2011; de Souza and Fiocchi 2016).

First evidences showing that gut microbiota was related to IBD appeared in animal models of colitis which revealed that shifts in the composition of *Lactobacillus sp*. together with other factors preceded colitis development. Colitis attenuated after driving *Lactobacillus sp*. to normal concentrations (Madsen et al. 1999).

As mentioned above, culture-based methods constitute a bottleneck in the study of gut microbiota. Most commensal bacteria do not grow in culture so genomic approaches have been an important advance in the study of dysbiosis in CD and UC overcoming the limitation of culture techniques. First approaches to identify the association between IBD and gut microbiota used classical techniques such as single strand conformation polymorphism together with cloning experiments and qPCR or temporal temperature gradient gel electrophoresis of 16S rDNA. In these studies, investigators found an association between CD and an alteration in the gut microbial composition showing higher quantities of Enterobacteria in CD than in healthy (Seksik et al. 2003). They also observed a reduced diversity in UC compared with healthy controls showing a percentage of dysbiosis higher in comparison with CD (Ott et al. 2004).

New molecular techniques such as DNA amplification, cloning and sequencing of 16S rRNA gene applied in the early 2000s in intestinal microbiota studies already

demonstrated the advantages of those techniques over culture identifying more that 50% of novel species (Hayashi et al. 2002; Manichanh et al. 2006) The arrival of metagenomics has boosted research associating microbiome alterations and IBD (Seksik et al. 2003; Tamboli 2003). Many studies have used targeted metagenomics with 16S gene whereas the use of shotgun sequencing techniques is less common but increases each day.

Manichanh *et a*l. published one of the first fecal metagenomic library based on the cloning of bacterial genomic fragments of 40 kb obtained from 6 healthy controls and 6 CD patients. Then, using this library, they screened and sequenced the 16S rRNA gene and showed a reduced diversity in CD patients compared to healthy controls, in particular in the Firmicutes phylum (Manichanh et al. 2006). More concretely, *Clostridium leptum*, from Clostiridiales order, was found in less proportion in CD patients than in healthy controls, which validated a previous study using amplified and cloned rRNA gene libraries in CD patients and healthy volunteers (Mangin et al. 2004). In this line, other studies identified a reduction of Clostridiales such as *Faecalibacterium prausnitzii* in ileal (Baumgart et al. 2007) and in colonic (Martinez-Medina et al. 2006) mucosal tissue of patients with CD. This result suggests that a deficiency of the bacterium may be a marker of CD independently of the localization of the disease.

Active and inactive periods of IBD may be associated with differential gut microbial abundance. Sokol *et al.* studied these shifts with real-time qPCR targeting the 16S rRNA gene in fecal samples from a cohort of CD patients in remission and relapse, UC patients in remission and relapse and patients with infectious colitis (n = 57) and healthy controls (n = 27). *Clostridium leptum* (with *Faecalibacterium prausnitzii* as its main representative), *Clostridium coccoides* and *Bifidobacterium* were significantly less abundant in active IBD (UC and CD) and infectious colitis compared with healthy controls again confirming a dysbiosis in CD patients and confirming previous studies described (Sokol et al. 2009).

Until now, most studies have analyzed association between gut microbiota and IBD without segregating the disease into two subtypes. However, Gophn and colleagues used cloning and sequencing techniques of the 16S rRNA gene in healthy, UC and CD individuals and showed that there were not differences between healthy volunteers and

UC patients whereas CD patients presented higher quantities of Proteobacteria and Bacteroidetes and lower abundances of Clostridia (Gophna et al. 2006). Moreover, the authors took inflamed and non-inflamed mucosal tissues from the same IBD participant and showed that there were no significant differences in the gut microbiota of both areas. This finding implies that microbial community composition may not be the direct cause of inflammation what agrees with a more recent study by Forbes and colleagues (Forbes, Van Domselaar, and Bernstein 2016). However, other studies have shown microbial differences between inflamed and non-inflamed tissues (Walker et al. 2011).

IBD symptoms in children vary extensively and may consist of minor extra intestinal manifestations, which makes the diagnosis very difficult. Some studies of gut microbiota in IBD pediatrics' stool and mucosal samples have been implemented to find changes in microbial composition that could help in the challenging and late diagnosis. Papa *et al.* developed a tool for diagnosis named SLiME, that was able to differentiate between pediatric patients with IBD from other patients with similar symptoms using fecal samples with a sensitivity of 80.3% and a specificity of 69.7% but was not sufficient to replace endoscopy (Papa et al. 2012). Moreover, analyzing 16S rRNA pyrosequencing data, they saw that alterations in active IBD included increased proportions of Enterobacteriales and reduction of *Subdogranulum* or *Butyricicoccus* both from the Clostridiales order. This result is in accordance with studies in adults. Segregating the two subtypes of IBD in mucosal tissue of another children cohort has demonstrated that reductions in bacterial alpha diversity only varied in CD and not in UC. However, and disagreeing with previous studies, increment on *Faecalibacterium prausnitzii* was described for CD (Hansen, Richard K. Russell, et al. 2012). Using Illumina MiSeq 16S rRNA sequencing, a more recent technology, in mucosal and stool samples collected from a pediatric CD cohort, defined dysbiosis associated with CD. The alteration was characterized by an increased abundance of *Enterobacteriaceae*, *Pasteurellaceae*, *Veilloneallaceae* and *Fusobacteriaceae* families and decreased proportions of Erysipelotrichales, Bacteroidales and Clostridiales (Gevers 2015). For a deeper taxonomical classification of the gut microbiota, Gevers *et al.* selected a subset of stool samples and applied shotgun sequencing with Illumina HiSeq2000 platform. Species found increased in CD corresponded to bacteria known to contribute to dysbiosis in IBD

residing in the mucosa such as *Escherichia coli*, or *Fusobacterium nucleatum* (Mirsepasi-Lauridsen et al. 2019; Palmela et al. 2018; Strauss et al. 2011) or in the oral microbiota as *Haemophilus* or *Veillonella* (Kim et al. 2013). Enrichment of *Ruminococcus gnavus* in IBD, especially in active IBD, has also been described (Hall et al. 2017). As previously described, many studies have been performed in pediatric IBD cohorts showing shared and unshared results. Although there is an agreement in some bacteria associated to IBD in those studies, there is also a lack of consistency between results what could be associated to the unstable microbial composition at early ages (Odamaki et al. 2016).

As previously mentioned, higher resolution techniques are needed for a deeper characterization of the composition of gut microbiota, but also for determining potential functions of those microorganisms. Shotgun sequencing together with metabolomic technique have recently described clear differences between healthy controls and CD patients whereas UC patients tend to show fewer differences with the healthy volunteers metabolically, functionally and compositionally (Franzosa et al. 2019).

Although gut microbiota in IBD has been widely studied, other microorganisms different from bacteria have not been extensively characterized. Analyses of 16S rRNA sequences do not provide information about neither viruses nor fungi so other techniques such as 18S rRNA sequencing or shotgun sequencing are necessary to fulfill this objective.

Recent studies in enteric virome in IBD samples have shown alterations in both forms of the disorder with an important increase of Caudovirales bacteriophages (Norman et al. 2015) whereas other studies reported higher proportions of Caudovirales in CD compared to UC but both higher than in healthy volunteers in a pediatric cohort (Fernandes et al. 2019).

Mycobiome, or fungal microbiome, in IBD has been suspected to contribute to its pathogenesis for many years. One example of this hypothesis is driven by Card9 or Dectin-1 genes among others which are IBD-associated genes and are related to immune response to fungi infection (Richard et al. 2015). High-throughput technologies have been applied to study the fungal content in gut microbiota in IBD patients showing an important fungal dysbiosis in IBD. Sokol *el al*. found a reduction of fungal diversity in UC and non-ileal CD patients defined by higher values of the ratio Basidiomycota-

Ascomycota (Sokol et al. 2017). They also showed that IBD and more concretely IBD in flare was associated with a reduction of *Saccharomyces cerevisiae* previously shown to reduce colitis in mice (Sivignon et al. 2015) and with an increase in *Candida albicans*.

In summary, both targeted and non-targeted metagenomic approaches have contributed to elucidate whether gut microbiota differs between healthy controls and IBD patients. In general, patterns of dysbiosis follow a similar trend in most of the studies, however, there are still discrepancies. The proposal of decreased diversity in IBD patients is a common tendency whereas there are clear doubts regarding the differences between CD and UC. Butyrate producing bacteria such as *Faecalibacterium prausnitzii* or *Clostridium leptum*, both from Firmicutes phylum, which are essential in the gut microbiota of human are reduced in IBD patients reducing the quantity of one of the main butyrate sources in the gut (Kumari, Ahuja, and Paul 2013; Manichanh et al. 2006; W. Wang et al. 2014) (**Table 1**).

Table 1. Summary of previous findings in microbial composition and functionality in IBD

Paper	Methodology	Disorder of the samples	Type of samples	Increased compared with HC	Decreased compared with HC
Mangin et al 2004	16S rRNA cloning & sequencing	CD	Stool	*Escherichia coli* *Bacteroides vulgatus*	*Clostridium leptum*
Gophna *et al* 2006	16S rRNA cloning & sequencing	UC & CD	Mucosa	Proteobacteria (CD) Bacteroidetes (CD)	Clostridia (CD)
Manichanh *et al* 2006	16S rRNA cloning & sequencing	CD	Stool		Firmicutes Clostridiales *Clostridium leptum* *Clostridium coccoides*
Martínez-Medina *et al* 2006	16S rRNA PCR DGGE	UC & CD	Mucosa	*Clostridium spp* (CD) *Escherichia coli* (CD) *Ruminococcus torques* (CD)	*Faecalibacterium prausnitzii* (CD)
Baugmart *et al* 2007	16S rDNA QPCR, sequencing & FISH	CD	Mucosa	*Escherichia coli*	Clostridiales
Sokol *et al* 2009	16S rRNA QPCR	UC & CD	Stool		Firmicutes *Clostridium leptum* *Clostridium coccoides* *Faecalibacterium prausnitzii* Bifidobacteria
Papa *et al* 2012	16S rRNA sequencing	UC & CD	Stool	*Escherichia-Shigella* (active disease)	Porphyromonadaceae Rikenellaceae

Paper	Methodology	Disorder of the samples	Type of samples	Increased compared with HC	Decreased compared with HC
				Coryneobacteriaceae (active disease)	
Kumain *et al* 2013	16S rRNA FISH	UC	Stool		*Clostridium coccoides* *Clostridium leptum*
Gevers *et al* 2014	16S rRNA sequencing (Illumina) Shotgun metagenomics sequencing	CD	Stool & mucosa	*Enterobacteriaceae* *Pasteurellacaea* *Veillonellaceae* *Fusobacteriaceae*	Erysipelotrichales Bacteroidales Clostridiales
Sokol *et al* 2016	16S rRNA sequencing ITS2 sequencing	UC & CD	Stool	Basidiomycota/Ascomycota ratio *Candida albicans*	*Saccharomyces cerevisiae*
Moustafa *et al* 2018	Shotgun metagenomics sequencing	UC & CD	Stool	Proteobacteria	Bacteroidetes Firmicutes
Franzosa *et al* 2018	Shotgun metagenomics sequencing Metabolomics	UC & CD	Stool	Metabolites associated to *Escherichia coli* Oxidative stress associated enzymes	

HYPOTHESIS

Gut microbiota along with host genetics, host immunity and environmental factors plays an important role in the development and perpetuation of IBD. Although previous studies have demonstrated the presence of dysbiosis in CD and UC patients compared to healthy controls, no clear consensus on the microbial profiles characteristic of the diseases have been proposed. Moreover, many studies have disagreed in the level of dysbiosis in UC compared to healthy state and have reported main differences between healthy controls and IBD but not between UC and CD.

In this thesis, we first hypothesize that different methodologies (16S rRNA and DNA shotgun sequencing) used to characterize these alterations should be consistent and equivalent. Then, we hypothesize that taxonomic and functional alterations occur in the gut microbiota between healthy controls and IBD patients and that differences could be found between CD and UC patients. Moreover, we believe that the differences between CD, UC and healthy would be such that we can design a diagnostic and prognostic tool for at least one of the two IBD subtypes. Finally, we believe that the compositional and functional analyses of fecal samples at the DNA level would provide an understanding of the physiopathology of IBD and its subtypes.

OBJECTIVES

MAIN OBJECTIVE

The main objective of this thesis was to describe differences in microbiome alterations (dysbiosis) between Crohn's disease and ulcerative colitis. To characterize the microbiome composition in large set of fecal samples (more than 2000), we used the 16S rRNA gene sequencing technique and to recover functional understanding of the gut microbiome in a subset of the large cohort, we used a more expensive technique, the DNA shotgun (short DNA fragments) sequencing technique.

SECONDARY OBJECTIVES

This study aims to:

a) Characterize the microbial composition and functions in fecal samples that differentiates IBD from healthy individuals with 16S rRNA data.
b) Compare 16S rDNA data with shotgun data and determine whether both techniques are consistent for IBD.
c) Define the fecal microbiome associated to severity of the disease.
d) Develop an algorithm to classify IBD samples based on 16S rDNA data and another one based on shotgun data.

METHODS

1. Ethics statement

In this study, we analyzed our own unpublished data and validated our results with other unpublished as well as published data. Our unpublished data were obtained from two different cohorts: Spanish IBD, Spanish UC and Belgian CD cohorts. Published data from different cohorts were also recovered from our own server or through public databases: our Spanish IBS cohorts (NCBI-SRA accession number: PRJNA268708) (Pozuelo et al. 2015), UK healthy twin (European Nucleotide Archive (ENA) accession numbers: ERP006339 and ERP006342) (Goodrich et al. 2014), French IBD (Sokol et al. 2017) and German anorexia (Misra and Klibanski 2016) cohorts) for 16S rDNA sequencing and American IBD cohort (Franzosa et al. 2019) for metagenomics. Protocols for the unpublished Spanish IBD, Spanish UC and the Belgian CD cohorts were submitted and approved by the local Ethical Committee of the University of Vall d'Hebron in Barcelona (Spain) and by the University Hospital Gasthuisbert in Leuven (Belgian), respectively. All participants gave written informed consent for their participation in the study.

2. Study design

In the Spanish IBD cohort, we recruited 34 patients with CD and 33 patients with UC for a follow-up study of one year, 65 healthy relatives (36 and 29 CD and UC relatives, respectively) with a follow-up of 3 months and 40 healthy non-related controls without follow-up. Inclusion criteria for patients were UC and CD diagnosis confirmed by histology and endoscopy, clinical remission for at least 3 months defined by the validated CD activity index (CDAI) for CD and the colitis activity index (CAI) for UC, a stable maintenance therapy (amino-salicylates, azathioprine or no drug) and previous history of at least three clinical recurrences in the past 5 years. In the case of healthy controls (HC), the inclusion criteria consisted of not having previous history of chronic disease. We collected clinical parameters (tobacco use, medical treatment) and diagnostic criteria (location and behavior of CD or extension of UC) at inclusion and during the follow-up. Clinical recurrence was defined by a value of 4 or higher for CAI and higher than 150 for CDAI. Exclusion criteria for this study were pregnancy or breast-

feeding, severe concomitant disease involving the liver, heart, lungs or kidneys and treatments with antibiotics in the last previous 8 weeks.

Patients were subjected to a follow-up of one year during which they gave samples every 3 months until relapse or until the end of the year whereas controls had a follow-up of 3 months giving one sample at the beginning and one at the end of the period. We collected a total of 419 fecal samples from 132 participants (65 HC and 67 IBD patients) at different time points for microbiome analyses. The 419 stool samples also include samples from CD and UC patients that showed flare during the study and provided a fecal at the time of recurrence.

On the one hand, all these 419 samples were analyzed for 16S rRNA method for taxonomical profiling. On the other hand, we selected a subset of samples for shotgun sequencing to generate a functional profiling. We selected a total of 178 samples that included 62 samples of healthy CD and UC-relatives at baseline and 126 samples from patients with IBD from a total of 64 patients (34 UC and 30 CD).

In the Belgian cohort, we included 55 patients with CD undergoing curative ileocecal resection at the University Hospital Leuven. We collected a total of 195 fecal samples at four time points before surgery and during postoperative follow-up (baseline, 1, 3 and 6 months after surgery) for microbiome analyses.

We used the Spanish UC cohort as validation of our results, and it enrolled long remission UC patients who provided samples at baseline and after taking probiotics. However, we only analyzed basal samples to avoid a possible effect of the probiotic in gut microbiota.

Published cohorts are described in the chapter **Population description** from the **Results** section.

3. Sample collection

After deposition of fecal matter, volunteers homogenized their feces with a spatula and immediately froze them in their home freezer at -20ºC. They then brought the frozen samples in a freezer pack to the laboratory where we stored them at -80ºC until further

processing. We performed aliquots of 250mg on solid CO_2 (dry ice) to maintain the frozen status of the sample to avoid the degradation of nucleic acids.

4. DNA extraction from fecal samples

Before starting the extraction procedure, we weighted 800mg of 0.1mm Zirconia/Silica beads, previously sterilized with UV, in tubes. We added in each tube 250µl of 4M guanidine thiocyanate, 40µl of 10% N-lauroylsacorsine and 500µl of 5% N-lauroylsarcosine and incubated at 70°C for 1 hour to chemically lyse the samples. To ensure the cell wall disruption of gram-positive bacteria and avoid introducing a bias in the recovery of all bacteria (Santiago et al. 2014), we performed a mechanical disruption using a Beadbeater (Biospec Products). We added Poly Vinyl Poly-Pyrrolidone (PVPP) in multiple washing steps to precipitate and discard aromatic molecules such as aromatic molecules nucleic debris, cellular debris or proteins. To clear lysates, we performed an enzymatic digestion of RNA. Resulting DNA from previous step was precipitated and ethanol-purified. We resuspended pure DNA in 200µl Tris-EDTA buffer (**Figure 10**).

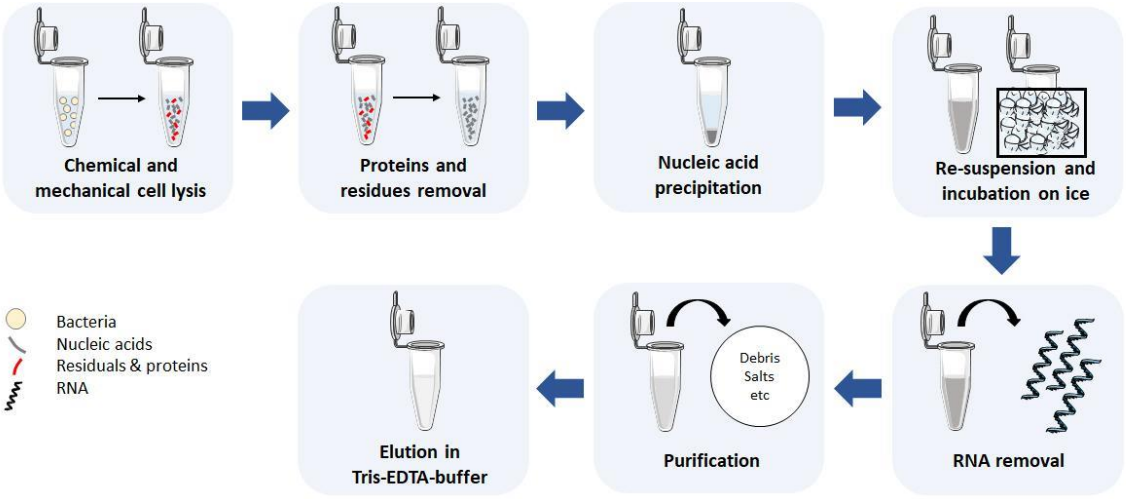


Figure 10. Extraction method. All the steps performed in the manual protocol for the extraction of DNA are summarized to better understand the process.

5. 16S rRNA gene amplification

In order to profile the microbiome composition, we amplified by PCR the hyper variable region (V4) of bacterial and archaeal 16S rRNA gene. The 5' ends of the forward (V4F_515_19) and reverse (V4R_806_20) primers targeting the V4 region of 16S gene were tagged with specific sequences for Illumina® MiSeq Technology (**Table 2**). 12 base-paired Golay codes were specified downstream of the reverse primer sequence (VV4R_806_20) to allow multiplex identification of individual samples (J. Gregory Caporaso et al. 2012; Navas-Molina et al. 2013).

Table 2. Primers used for 16S rRNA gene amplification of Illumina MiSeq sequencing

PRIMER TYPE	SEQUENCE 5' → 3' ILLUMINA FLOWCELL – BARCODE – ADAPTER – LINKER – V4 REGION
Forward	AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
Reverse	CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXXXAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT

We then ran standard PCR using 0.75 units of *Taq* polymerase (AmpliTaq Gold, Life Technologies®) and 20pmol/µl of the forward and reverse primers (IDT Technologies®) in a final volume of 50µl in a Mastercycler gradient (Eppendorf®) at 94°C for 3 minutes, followed by 35 cycles of 94°C for 45 seconds, 50°C for 60 seconds 72°C for 90 seconds and finally, a cycle of 72°C for 10 minutes.

6. Agarose gel and purification

We performed a 1% agarose gel stained with ethidium bromide and ran it in 1x Acetate EDTA (TAE) buffer. 5µl of PCR product were mixed with 6x loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in water) and placed in wells of solidified agarose gel along with 100bp DNA Molecular Weight Marker XIV (Roche®) at about 90-100V for 35-45 minutes to visualize the amplicon bands in a Gel Doc XR+

system (Bio-Rad®). We confirmed PCR amplification by the appearance of amplicon bands. Absence of bands could be explained by the presence too little genomic template DNA in the sample or by the presence of PCR inhibitors present in the sample. In the latter, we diluted the genomic DNA as an attempt to get rid of inhibitors.

After the amplification of the targeted gene was confirmed, we purified the corresponding PCR products using the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions and further quantified them using a NanoDrop ND-1000 Spectophotometer (Nucliber®).

7. Illumina sequencing

The sequencing process is described in Illumina website (https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf) and consists of:

"Sequencing templates are immobilized on a proprietary flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface bound template and low non-specific binding of fluorescently labelled nucleotides. Solid-phase amplification creates up to 1,000 identical copies of each single template molecule in close proximity (diameter of one micron or less)." "Sequencing by synthesis technology uses four fluorescently labelled nucleotides to sequence the tens of millions of clusters on the flow cell surface in parallel. During each sequencing cycle, a single labelled deoxyribonucleoside triphosphate (dNTP) is added to the nucleic acid chain. The nucleotide label serves as a terminator for polymerization, so after each dNTP incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. Since all four reversible terminator-bound dNTPs (A, C, T, G) are present as single, separate molecules, natural competition minimizes incorporation bias. Base calls are made directly from signal intensity measurements during each cycle which greatly reduces raw error rates" (**Figure 11**).

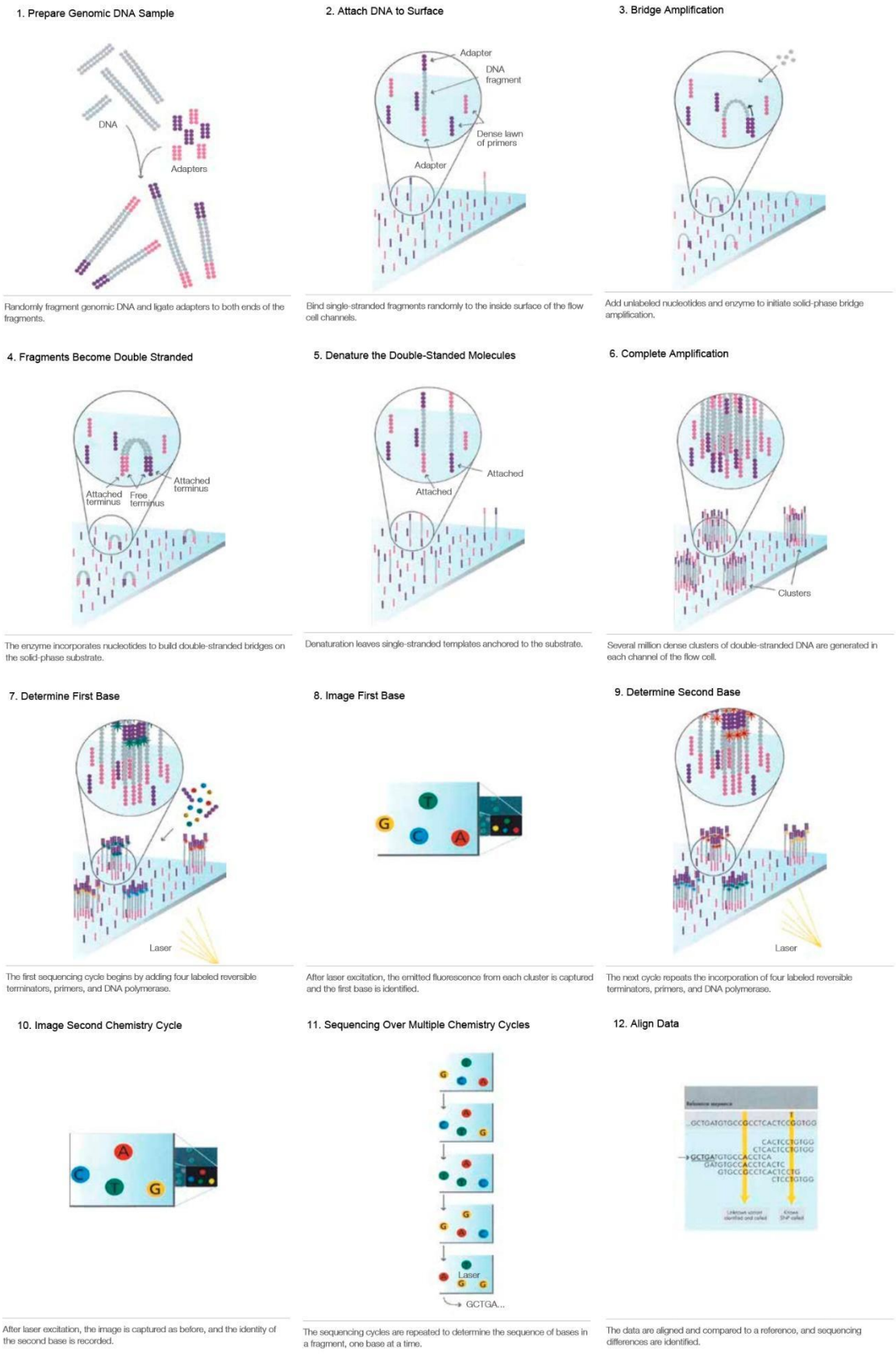


Figure 11: Illumina sequencing technology (Adapted from www.illumina.com)

7.1 16S rRNA sequencing

For 16S rRNA amplicons, we prepared pools of equally concentrated samples (240ng of DNA per sample) that were later diluted to 2nM. Then, denatured templates were further diluted to 5ρM and subsequently combined to give an 85% 16S rRNA gene amplicon library and 15% PhiX control pool. We performed sequencing with Illumina MiSeq technology, as described in section Illumina sequencing, at the genomics core of the Autonomous University of Barcelona (UAB, Spain). We received single end sequences that were later analyzed.

7.2 DNA shotgun sequencing (or metagenomic sequencing)

Extracted genomic DNA were sent to Beijing Genomics Institute (BGI) where the shotgun pair-end sequencing was performed with Illumina HiSeq technology following Illumina standards.

8. Database preparation for 16S rRNA analyses

We performed 16S rRNA analyses combining the Greengenes (version gg_13_8) and the PATRIC (Pathosystems Resource Integration Center) databases. The Greengenes database contains 16S rRNA gene annotated sequences from Bacteria and Archaea and the PATRIC database is composed of all known pathogens, many of which are not present in Greengenes.

To combine the Greengenes release gg_13_8 and the PATRIC databases, we extracted the 16S rRNA sequences and their taxonomical annotation from the annotated genomes in PATRIC avoiding different strains of the same bacterial species. We formatted both, sequences and taxonomical annotation from PATRIC, to QIIME compatible files. Finally, we combined the obtained PATRIC files with the Greengenes database into a single database. Due to annotation differences in both databases, we found repetitions in some genera. To fix this problem, we changed taxonomical annotation for those repeated genera using PATRIC annotation as it was the most recent one.

9. 16S rRNA sequences analyses

9.1 Automatization of the analyses

We developed an automatization system, written in python (v.2.7.6), to run the command lines proposed by the QIIME pipeline (v.1.9.1) for all the steps needed for the 16S rRNA sequence analyses. For analyzing and organizing the data, the script calls QIIME (v.1.9.1), biom package and a perl script. The automatization script requires as input single-end raw reads fastq files from the Illumina MiSeq sequencing platform and a metadata file with variables for correlating clinical data with microbiome data and for comparing groups of subjects. Using just a single command, the script finally returns bacterial abundance tables, alpha and beta diversity results, PCoA and rarefaction figures and basic statistical tests for differential abundance presence between several groups. The python script we implemented could work in parallel using all the available CPUs of the computer at the same time reducing the time of analysis. All the steps included in the automatization of the analysis are described below (**Figure 12**).

9.2 Upstream analyses

The first step of the analysis consisted of assessing the quality of the sequences provided by the sequencing platform using the FastQC software. Then, we verified whether the metadata contained all necessary information such as sample identifiers, barcodes, primer sequences, time points, sample status, information about symptoms, treatments and other clinical information of interest.

We performed all the remaining steps of the upstream analyses using the QIIME (v.1.9.1) pipeline following the guidelines proposed by Navas and colleagues (Navas-Molina et al. 2013). We made a demultiplexing step to remove barcodes together with the linker primer sequence from all raw sequences and assign each read back to its sample with the correspondence barcode-sampleID. We also performed a quality filtering step in which we removed sequences with a quality Phred score smaller than 20, a threshold commonly used in the microbiome field.

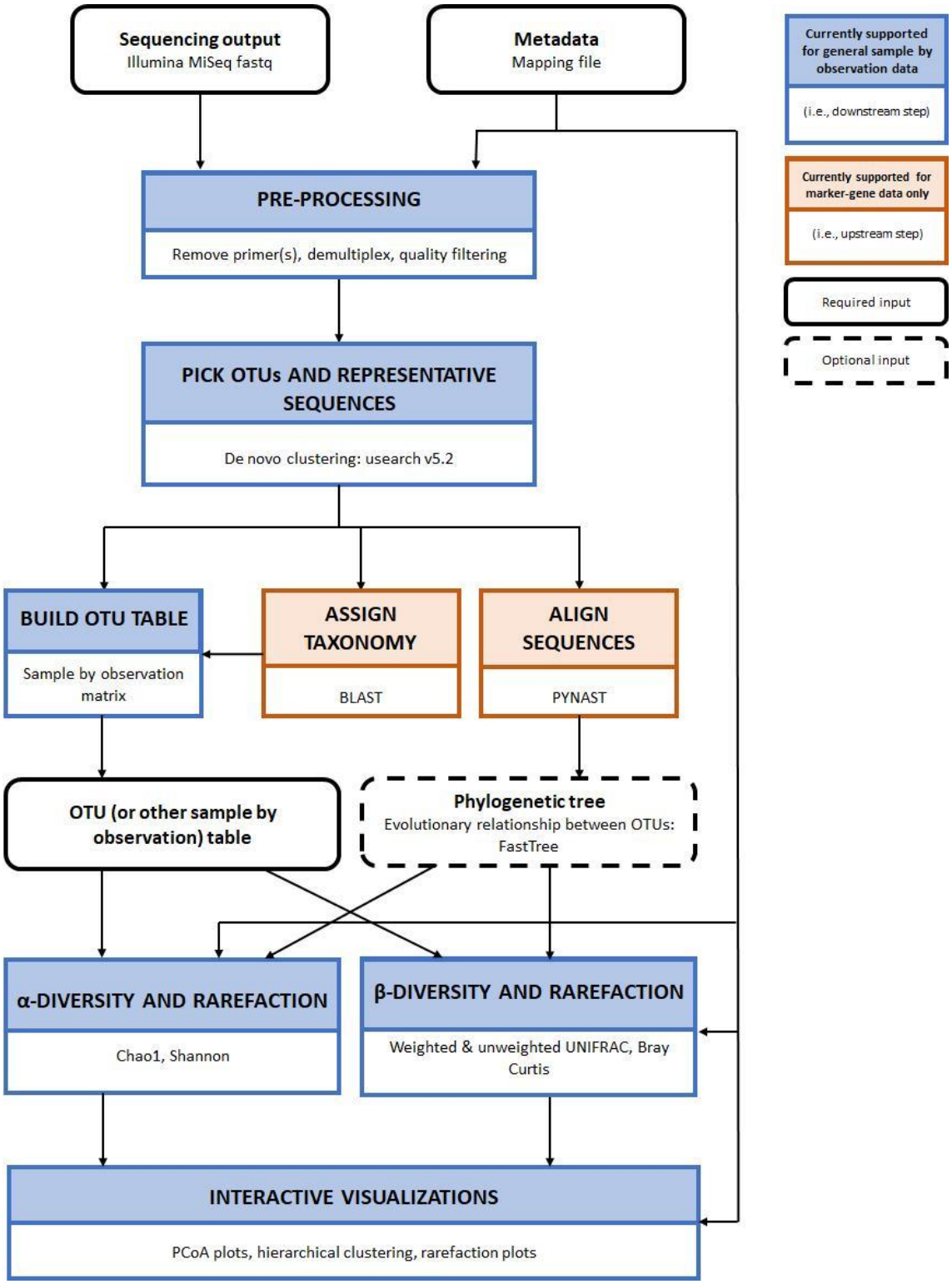


Figure 12. QIIME 1.9 workflow. (Adapted from (Navas-Molina et al. 2013)). The upstream steps include all processes that lead to the obtention of an OTU table and a phylogenetic tree. They include the preprocessing of raw reads, OTU clustering, taxonomical assignment and phylogenetic tree construction. Downstream processes encompass all the steps used for the interpretation of the results with alpha and beta diversities and visualizations.

In order to cluster sequences into Operational Taxonomic Units (OTUs) also named molecular species based on sequence similarity of 97%, we utilized the USEARCH v5.2 clustering tool, a *de novo* method (Edgar 2010). During the clustering step, we also removed chimera sequences with UCHIME, following Edgar recommendations (Edgar et al. 2011).

We assigned taxonomy using the Basic Local Alignment Search Tool (BLAST) algorithm using our combination of Greengenes (gg_13_8 release) and PATRIC databases generated as previously described.

We aligned representative sequences using PyNAST against the Greengenes template alignment. We need this step to build phylogenetic tree using FastTree.

Finally, we generated an OTU table that contains all the OTUs with their predicted taxonomy and abundances for each sample.

9.3 Downstream analyses

We performed downstream analyses including diversity analyses and statistical tests with QIIME 1.9.1 and R programming with 3.4.3 version for 16S rRNA sequences. We utilized biom program to convert tables between biom and txt format. We assigned numbers to unknown species to avoid collapsing possible novel species into one. Finally, we summarized the OTU table into different taxonomic levels from phylum to species.

9.3.1 Diversity

We rarefied OTU tables to perform diversity analyses to normalize and overcome cases in which samples have different number of sequences. In this case, we rarefied the OTU table at 6760 sequences per sample and were able to keep 2045 samples, which accounted for 115.5 million of reads used in further analyses.

To estimate microbial richness and evenness of sample, also known as alpha diversity, we calculated Chao1 and Shannon indexes respectively using QIIME (Chao et al. 2006; Hughes et al. 2001).

Between-samples diversity or beta diversity was computed using the weighted and unweighted UniFrac and the Bray Curtis methods to generate distance matrices that were later utilized for clustering samples in hierarchical cluster trees with Unweighted Pair Group Method with Arithmetic mean (UPGMA) and Principal Coordinate Analyses representations (PCoA) using QIIME.

10. Shotgun sequencing analyses

10.1 Upstream analyses

We used HUMAnN2 (Franzosa et al. 2018) pipeline for the analyses of shotgun sequencing. HUMAnN2 evolved from HUMAnN pipeline (Abubucker et al. 2012) to improve accuracy and time of analyses. First, we performed a quality filtering of the raw reads and human sequences removal with FASTX toolkit. We then combined the files high quality sequences of both pairs into a single file because HUMAnN2 does not consider that sequencing provided paired reads.

We then introduced the unique fastq file of high-quality reads into HUMAnN2 pipeline that performs the following four steps: 1. Identification of known species using marker genes to obtain a compositional profiling and reduce databases in the following steps with MetaPhlAn2 tool (Truong et al. 2015) and the ChocoPhlAn pangenome database (Huang et al. 2014); 2. Nucleotide-mapping of all raw reads against the identified species pangenomes with Bowtie2; 3. Translated-search of those reads that have not mapped with known species pangenomes against a protein database, UniRef90 using DIAMOND; 4. Gene family and pathway abundances profiling, HUMAnN2 provides a stratification of the contribution of each bacteria to the specific gene family or pathway (**Figure 13**). During the Bowtie2 step two samples from CD patients provided errors, so we removed them and worked with 176 samples instead of 178.

We annotated the results obtained from UniRef90 with KEGG (functional database) (Kanehisa et al. 2016) using HUMAnN2 resources. In summary, we obtained taxonomical abundance tables from MetaPhlAn2, pathway abundances tables with MetaCyc

identifiers, functional abundance table from KEGG and gene abundance tables with UniRef50 information.

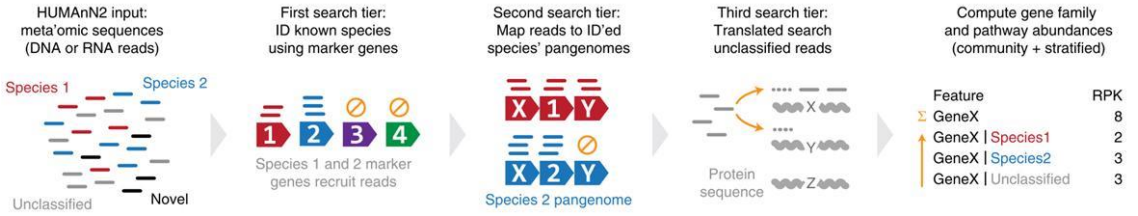


Figure 13. HUMANn2 workflow. (From Franzosa et al. 2018) HUMANn2 makes a first mapping against marker genes to detect species pangenomes. It constructs a personalized database with the pangenomes identified in the previous step and performs a nucleotide-search to find potential genes. Finally, those reads that have not mapped are tested against a protein database (UniRef90). HUMANn2 gives a genes abundance table stratified by species contribution.

10.2 Downstream analyses

We processed abundance tables with R packages. We computed alpha diversity Shannon and chao1 indexes with *vegan* and *fossil* packages, respectively. For beta diversity, we applied Bray-Curtis index using *vegan* package.

11.Statistics for diversity, differential abundance and correlation analyses

For each bacterium, we checked normality of data distribution with Shapiro-Wilk test (Shapiro and Wilk 1965). As bacterial abundances showed non-parametric distribution we performed the Kruskal Wallis one way analysis of variance (Kruskal and Wallis 1952) for comparisons between more than two groups and independent data (i.e. CD, UC and HC at basal timepoint), Friedman test (Friedman 1937) for more than three groups for paired data (i.e. CD samples from same individuals at different timepoints), Mann-Whitney U test (Mann and Whitney 1947) for two groups for independent data (i.e. remission and relapse at last timepoint of CD patients) and Wilcoxon signed rank test (Wilcoxon 1945) for two groups for paired data (i.e. basal timepoint in remission and

relapse sample from the same patient). In all these analyses, we corrected for multiple comparisons using the false discovery rate (FDR) multiple testing correction considering corrected p-values lower than 0.05 significant.

Alpha diversity showed normal distribution, therefore comparisons between groups were done with Student's t-test and Bonferroni correction for multiple comparisons. We used Adonis test to compare between groups for beta diversity. We also compared beta diversity obtained from basal samples to other timepoints using the mixed-design ANOVA model, a repeated measures analysis of variance.

To detect associations between bacteria and clinical information, we computed Spearman's rank correlation coefficient.

12. IBD classifier

We created a classifier to discriminate between CD and non-CD samples. We used artificial intelligence (AI) that is currently being used across a number of sectors, including healthcare and bioinformatics. We compared three state-of-the-art machine learning algorithms, random forest (Breiman 2001), AdaBoost (Kégl 2013) and XGBoost (Chen and Guestrin 2016) which are easy to interpret to obtain the best classifier. These algorithms learn from existing data and use their acquired knowledge to classify new samples. We used these methods to create an application that can determine whether a specific sample with a known microbial composition corresponds to a CD patient or not. For the implementation, we used python 2.7 and sklearn package.

For 16S rRNA data, we trained our AIs with a set of samples from the Spanish IBD cohort. We performed a grid search over reasonable ranges in order to tune the parameters of each algorithm. We have measured the goodness of their performance as Area Under the Curve (AUC) that is the standard metric for classification problems. Moreover, we performed 5-fold-cross validations to assess overfitting. We used the best performing algorithm to explore the features (i.e. the species) that better discriminate between CD and non-CD. We created a classifier to discriminate between CD and non-CD samples.

In metagenomics, we trained the algorithms with data from our Spanish IBD cohort and tested them with the American cohorts already described.

13. Fecal calprotectin assay

Fecal calprotectin (FC) is a marker of intestinal inflammation. We measured this protein in a subset of the Spanish cohort with a commercial ELISA (Calprest; Europistal SpA, Trieste, Italy) following the manufacturer's instructions. We read optical densities at 405nm with the microplate ELISA reader (Multiskan EX; Thermo Electron Coporation, Helsinki, Finland). We tested samples in duplicate and results were computed from a standard curve and expressed as µg/g stool.

RESULTS

1. Population description

For our discovery cohort we enrolled 172 participants (40 healthy controls non-related to the patients (HnR), 34 patients with CD, 33 patients with UC and 36 and 29 healthy relatives (HRs) of the CD (HR(CD)) and UC (HR(UC)) patients, respectively). We characterized their gut microbial communities in a longitudinal study. HRs were first-degree relatives from which we did not have information on whether they were living in the same house of the patient at the period of sampling.

Number of samples per individual differed in the three groups of the study. Non-related healthy controls (HnR) provided a single time point fecal sample whereas HRs provided two samples 3-months apart. In the case of patients with CD and UC, they provided fecal samples 3-months apart over a 1-year follow-up. When patients with IBD developed a recurrence, they provided a fecal sample just after the onset and stopped giving fecal samples for the study. During the 1-year follow-up, 13 patients with CD (38%) and 18 patients with UC (54%) developed recurrence. In total, we collected 419 fecal samples for microbiome analyses (**Table 3**).

Table 3. Summary table of discovery cohort. Spanish IBD cohort of 419 fecal samples from healthy controls, CD and UC patients.

	Basal	Interval	Last TP	Last TP Remission	Last TP Relapse
HnR	40	-	-	-	-
HR	65	-	57	-	-
CD	34	73	-	20	13
UC	32	52	-	14	18

To validate our results of 16S rDNA sequencing for CD we used a CD Belgian cohort. With the collaboration of Professor Severine Vermeire, we recruited 55 CD patients who underwent ileocecal resection at the University Hospital Leuven. They provided samples before surgery and during a postoperative follow-up at 1, 3 and 6 months after surgery.

In total, we collected 195 fecal samples (**Table 4**) that were extracted and analyzed with the 16S rRNA protocol described in the **Methods** section.

Table 4. Summary of all cohorts used for 16S and shotgun analyses.

Cohort	Number of samples	Baseline	Other TimePoints
Discovery cohort: IBD Spain			
HnR	40	40	
HR (CD)	64	36	28
HR (UC)	58	29	29
CD	140	34	106
UC	116	32	84
Validation cohort			
CD Belgium			
CD	195	55	140
UC Spain			
UC	41	41	
IBS Spain			
IBS	202	125	77
IBD France			
HC	38	38	
CD	146	146	
UC	86	86	
Healthy UK			
HC	1041	1041	
German anorexia			
HC	59	59	
Anorexia	99	99	
American IBD cohort			
HC	34	34	
CD	68	68	
UC	53	53	
Dutch IBD cohort			
HC	22	22	
CD	20	20	
UC	23	23	

In the case of UC, we used a Spanish validation cohort that consisted of 41 UC patients (**Table 4**) who provided samples at baseline and after taking probiotics. However, we only analyzed basal samples to avoid a possible effect of the probiotic in gut microbiota. These samples were also extracted and analyzed with the 16S rDNA sequencing protocol described in the **Methods** section.

We performed characteristics comparisons between the CD Spanish and Belgian cohorts and between the UC Spanish discovery and validation cohorts with the chi square (χ^2) test for categorical variables and the t-test for continuous variables. We set significant differences with p-values smaller than 0.05. In the discovery cohort, patients with CD showed inflammation mostly in the ileum (L1, 35%) and in the ileocolon (L3, 64,7%) (**Table 5**). In UC, the distribution of disease behavior at sampling was: proctitis (E1, 27.3%), left-sided colitis (E2, 33.3%) and pancolitis (E3, 39.4%) **(Table 6**).

Previous studies have demonstrated that there is an association between smoking habit and IBD (Thomas et al. 2000). In this regard, we analyzed the connection between smoking habit and severity of the disease (remission (REM) and recurrence (REL)) using the χ^2 test. We did not find any link between being smoker or ex-smoker and disease severity.

To validate our discovery, together with the Belgian CD and Spanish UC cohorts, we also used already published IBD and non-IBD cohorts (**Table 4**). We analyzed sequences obtained from an IBD cohort enrolled in France. This cohort consisted of 38 HC individuals, 146 CD and 86 UC patients (Sokol et al. 2017). Each patient provided one fecal sample. The differences with our IBD cohorts were that the V3-V5 region (instead of the V4) of the 16S rRNA gene was used for PCR amplification and sequences were generated using the Ion Torrent sequencing platform (instead of a MiSeq platform).

We also used sequence data from 3 non-IBD cohorts of V4 16S rDNA MiSeq sequencing: an IBS cohort already published by our group (Pozuelo et al. 2015) that consisted of 202 fecal samples from 125 IBS patients who provided two fecal samples 3-months apart; an anorexia German cohort of 99 patients and 59 healthy individuals (Misra and Klibanski 2016) and a British cohort of 1041 fecal samples from 977 healthy twins (Goodrich et al. 2014).

Table 5. Baseline clinical characteristics of the patients with CD and UC

Baseline clinical characteristics	CD Spanish cohort (n = 34)	CD Belgian cohort (n = 55)	Comparison between cohorts (p value)
Male/female (%)	13/21 (38.2/6.7)	29/26 (52.8/47.2)	0.201
Median (IQR) age at surgery (years) or at sample collection	34 (18-58)	41.3 (26.5-52.9)	0.141
Median duration of disease (IQR) at surgery (years) or at sampling	6.5 (0-28)	15.7 (4.1-27.1)	0.0002
Maximum disease location (Montreal classification)			0.682
L1 ileal (%)	12 (35)	18 (34)	
L2 colonic (%)	0 (0)	0 (0)	
L3 ileocolonic (%)	22 (64.7)	35 (66)	
L4 isolated upper disease (%)	2 (5.8)	2 (3.8)	
Disease behaviour at surgery (Montreal classification)			0.009
B1 non-stricturing, non-penetrating (%)	3 (8.8)	2 (3.8)	
B2 stricturing (%)	22 (64.7)	21 (39.6)	
B3 penetrating (%)	5 (14.7)	30 (56.6)	
p perianal disease (%)	3 (8.8)	15 (28.3)	
Active smoking at surgery (%)	10 (29.4)	16 (30.2)	0.012
Medication at surgery or at sampling			
Mesalamine–sulfasalazine (%)	4 (11.8)	4 (7.5)	0.012
Corticosteroids (%)	2 (2.9)	10 (18.9)	0.183
Immunosuppressants (%)	14 (41.1)	12 (22.6)	0.087
Anti-TNF (%)	12 (23.5)	7 (13.2)	0.023
Antibiotics (%)	0 (0)	9 (16.9)	0.033
Methotrexate	1 (2.9)		
Other	10 (29.4)		
None	1 (2.9)		

Table 6. Baseline clinical characteristics of the patients with CD and UC

	UC Spanish cohort 1 (n=33)	UC Spanish cohort 2 (n=41)	Comparison between cohorts (p value)
Male/female (%)	9/24 (27.2/72.7)	17/24 (41.4/58.5)	0.595
Median (IQR) age at sample collection	43 (24-62)	43 (24-68)	0.500
Median duration of disease (IQR) at sampling	9 (1-23)	10 (1-34)	0.392
Disease behavior at sampling			0.208
E1 proctitis	9 (27.3)	18 (43.9)	
E2 left sided colitis	11 (33.3)	10 (24.4)	
E3 pancolitis	13 (39.4)	13 (31.7)	
Medication at sampling			
Mesalamine (%)	11 (24)	26 (63.4)	0.021
Corticosteroids (%)	2 (6)	0	0.617
Immunosuppressants (%)	8 (24)	0	0.026
Other	2 (6)	3 (7.3)	0.708
None		2 (4.8)	

For DNA shotgun data, we used a subset of the discovery cohort of 178 samples composed by 64, 30 and 32 samples of healthy controls, CD and UC patients respectively at basal time and 21 (11 remission (REM), 10 relapse (REL)) and 31 (14 REM, 17 REL) samples at the timepoint of the follow-up for CD and UC patients, respectively (**Table 7**). We compared our results with a published cohort of 155 American individuals (68 CD patients, 53 UC patients and 34 non-IBD controls) and 65 samples from Dutch participants (20 CD patients, 23 UC patients and 22 healthy controls) (**Table 4**) (Franzosa et al. 2019).

Table 7. Sample size per group in shotgun sequencing analyses

	Basal	Last TP Remission	Last TP Relapse
HC	64	-	-
CD	30	11	10
UC	32	14	17

To measure the IBD activity in the discovery cohort, we used the fecal calprotectin (FC), a marker of inflammation. The effectiveness of this marker was assessed on a subset of 174 fecal samples from the discovery cohort provided by 122 participants. We measured FC at baseline and either after 1-year in remission or at recurrence for patients with CD and UC and at baseline for healthy individuals. In the case of remission, FC was significantly higher than in the HRs but lower than in the cases of recurrence (**Figure 14**). Calprotectin concentration did not differ between CD and UC patients, either in remission nor in relapse. With these results, we concluded that FC is an inadequate maker to differentiate UC from CD.

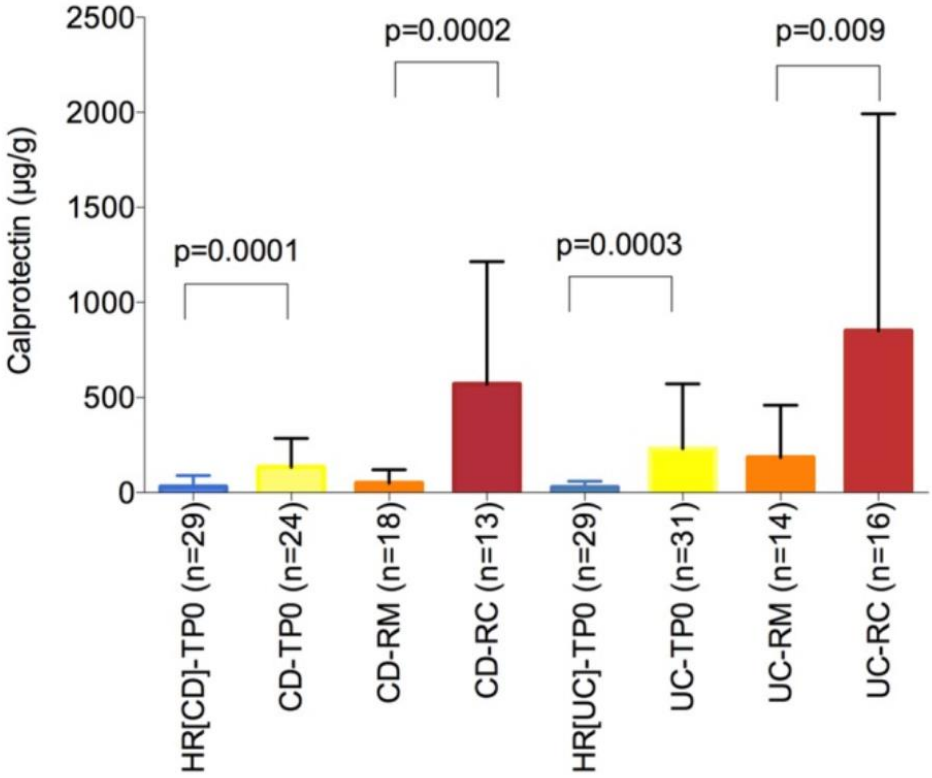


Figure 14. Calprotectin biomarker of inflammation. Calprotectin was measured in the stool of healthy relatives of CD (HR[CD]) and UC (HR[UC]) patients, and in the stool of patients with CD and UC at baseline (TP0) and after 1-year in remission (RM) and at recurrence (RC). The Mann-Whitney test was used to compare differences between groups

2. Compositional differences in gut microbiota in IBD patients using 16S rRNA sequencing technique

We analyzed a total of 2056 fecal samples from the discovery (Spanish IBD cohort) and the validation cohorts (Belgian CD, Spanish UC, Spanish IBS, German anorexia and British healthy cohorts), using the 16S rRNA gene analyzed by the bioinformatics tool QIIME 1.9. From the 2056 samples, after quality filtering we obtained a total of 115.5 million of high-quality sequences with a number of reads ranging from 1 to 223,896 per sample. We obtained the OTU table and rarefied at 6760 reads per sample, removing samples with less than 6750 reads and keeping 2029 samples for further analyses. In these analyses we could not include the French cohort due to technical differences, but we analyzed it separately obtaining a total of 8.5 million high-quality sequences for 232 patients with IBD (146 CD and 86 UC) and 38 healthy controls (HC) following the protocol described in the section.

2.1 CD is more dysbiotic than UC

To evaluate the stability of the fecal microbiome of patients with UC and CD, we analyzed the IBD Spanish cohort at five timepoints (baseline and 3, 6, 9, 12), using the weighted UniFrac distance, a metric used for comparing microbial community composition between samples. The higher the UniFrac index obtained, the higher the distance between samples. Patients with CD, but not with UC, showed higher UniFrac distances between time-point samples compared with their HRs (Mann Whitney test, $p = 0.01$) over the one-year follow-up. This result indicates that there is a higher instability in CD microbiome compared with healthy controls. Contrarily, UC patients presented a more stable microbiome composition over time even compared with their healthy relatives (Mann-Whitney test, $p = 0.015$). Furthermore, we also compared the UniFrac distances obtained between baseline samples and the rest of samples collected at later time points using a mixed-design ANOVA model, a repeated measures analysis of variance. The results also showed that the microbiome of patients with CD was

significantly more unstable than that of patients suffering from UC (mixed ANOVA, $p < 0.001$) along time (**Figure 15**).

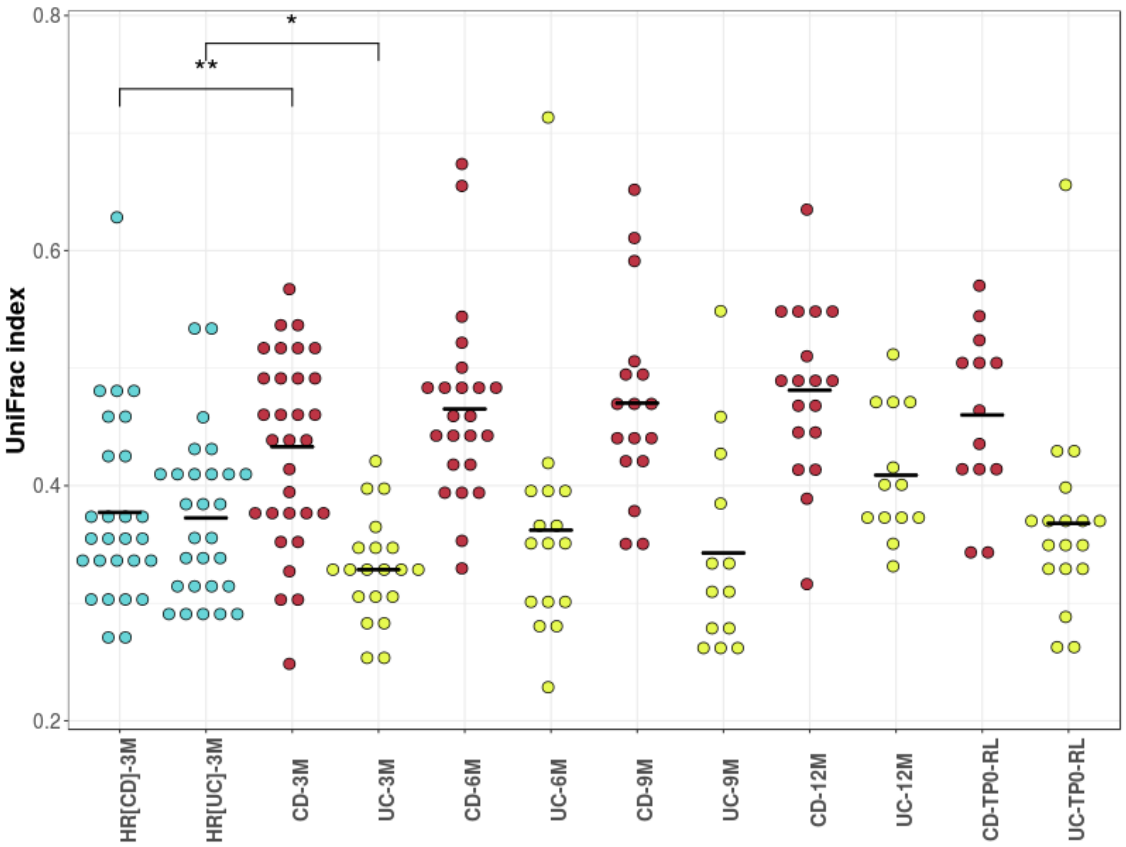


Figure 15. Microbiome stability. Unweighted UniFrac distances were calculated between different time periods for healthy relatives HR(CD) (relatives of patients with CD), HR(UC) (relatives of patients with UC) and patients with CD and UC (3M, 3 months; 6M, 6 months; 9M, 9 months; 12M, 12 months). CD-RC and UC-RC refer to samples collected during recurrence of the disease. At 3-month interval, patients with CD and UC presented significant differences in their UniFrac indexes compared with their HR (Mann-Whitney U test, *p = 0.01). We compared the UniFrac indexes obtained between samples collected at baseline and each other time points using the mixed-design ANOVA model and found that the microbiome of patients with CD was significantly more unstable than that of patients with UC (mixed-ANOVA, $p < 0.001$).

We compared the microbial composition between the groups using a multivariate analysis of variance (NPMANOVA test) on distance matrices (Bray-Curtis, weighted and unweighted UniFrac). Overall, the microbial community of the two groups of controls (relatives (HR) and non-relatives (HnR)) were not significantly different from each other ($p = 0.239$ for weighted UniFrac distance and $p = 0.134$ for Bray-Curtis distance). Nonetheless, we found differences in two genera: *Collinsella* was more abundant in healthy IBD relatives than non-relatives (Mann-Whitney test, FDR < 0.0001) whereas an unknown Peptostreptococcaceae was more abundant in HnR than in HRs (FDR < 0.0001).

Considering all samples, the microbiome of patients with CD differed significantly from that of healthy controls (relatives and non-relatives (Healthy or HC)) (NPMANOVA test; FDR < 0.0015) based on both UniFrac (weighted and unweighted) and Bray-Curtis distnces. However, the microbiome of patients with UC differed from that of HC only based on weighted UniFrac distance metrics (FDR = 0.009) and not on Bray-Curtis and unweighed UniFrac distances. These results validate the lower degree of dysbiosis of the UC microbiome compared with CD. Patients with CD and UC also showed a significant difference in their global microbiome composition (NPMANOVA test, FDR = 0.0015 for weighted UniFrac distances and FDR = 0.003 for unweighted UniFrac and Bray Curtis distances) (**Figure 16**).

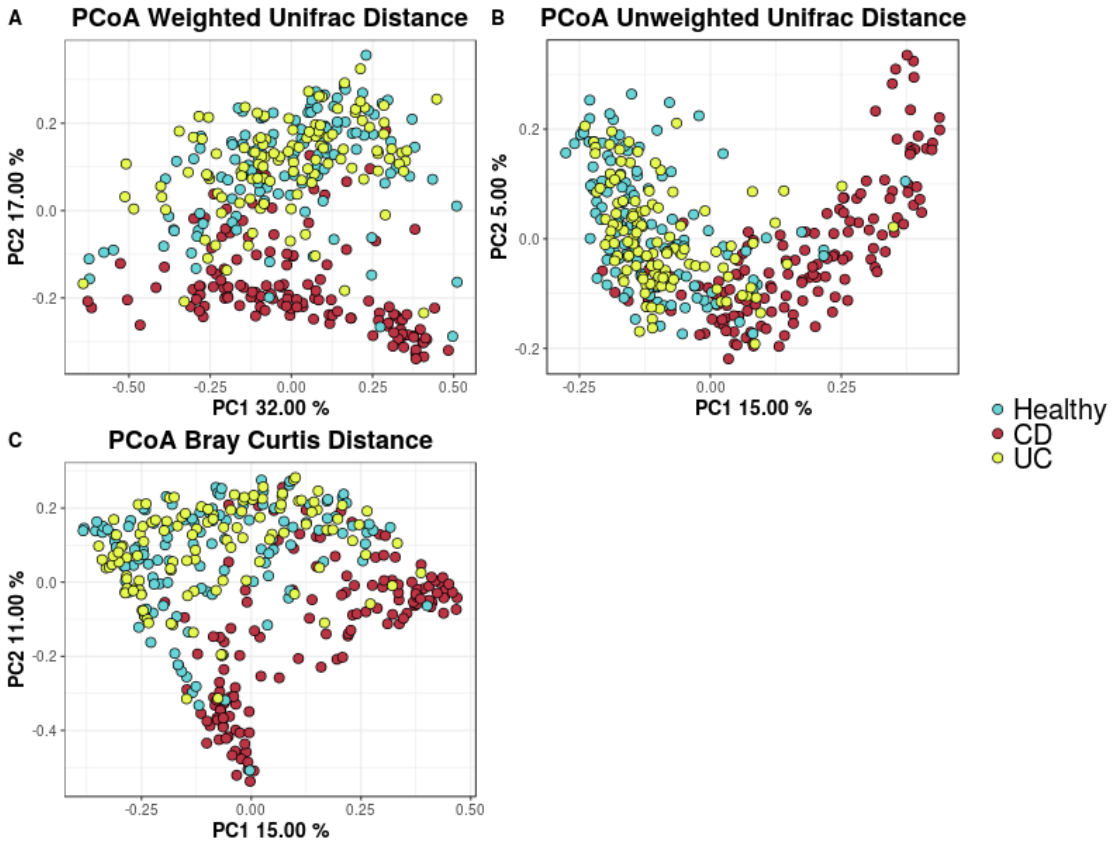


Figure 16. Dysbiosis in patients with IBD. Microbiome clustering based on unweighted (A) and weighted (B) UniFrac distances and Bray-curtis distances (C) Principal Coordinate Analysis. Significant differences were observed between healthy (combining HnR, healthy relatives HR(CD) and HR(UC)) and patients with CD (NPMANOVA test; FDR = 0.0015 for weighted UniFrac distances and FDR = 0.003 for unweighted and Bray-Curtis distances). Only differences with weighted UniFrac distances were observed between UC patients and healthy individuals (FDR = 0.009).

To determine the alpha-diversity of each group, we computed Chao1 and Shannon indexes. The Chao1 index estimates the richness whereas the Shannon index estimates richness and evenness. CD patients showed lower microbial diversity compared with the two control groups (t-test, FDR < 2 x 10^{-6}) for both indexes. We did not find differences between UC and control groups (t-test, FDR = 0.44 and FDR = 0.2 for Shannon and Chao1 indexes respectively) (**Figure 17**). We did not find differences between remission and relapse last timepoints of each subtype of IBD nor between basal and last timepoints. These findings indicate that alpha-diversity may not be a useful biomarker of IBD severity.

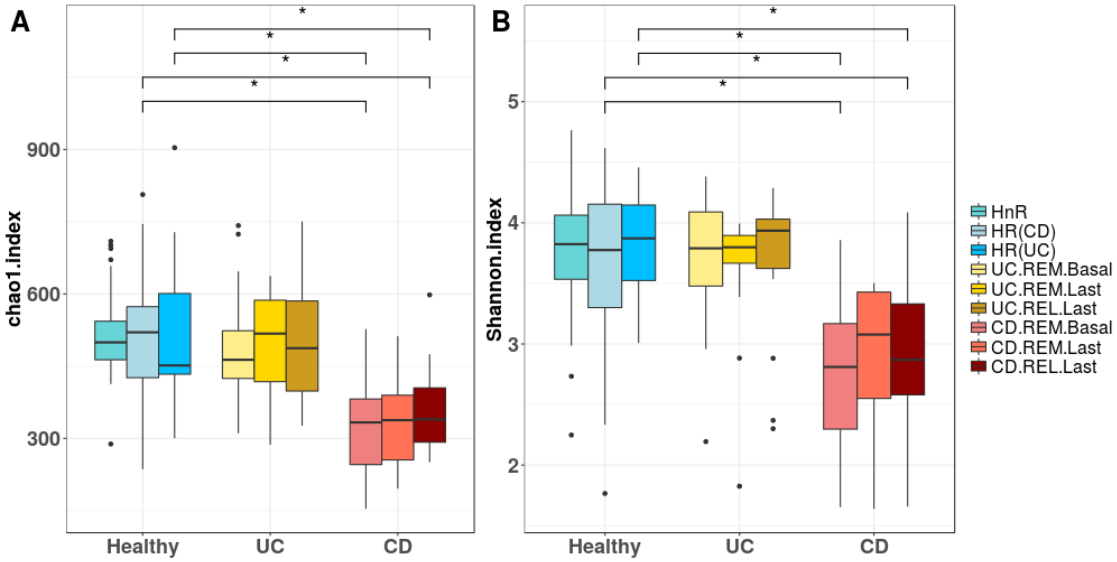


Figure 17. Dysbiosis in patients with IBD. Microbial richness was calculated based on the Chao1 index (A) and microbial richness and evenness on the Shannon index (B). Using the Student's test, the microbiome of patients with CD presented significantly lower richness and evenness than healthy controls (HnR, HR(CD) and HR (UC)) and UC patients, but patients in remission and in recurrence (CD.REM.Last vs CD.REL.Last and UC.REM.Last vs UC.REL.Last) did not present significant differences *$p < 0.05$.

To determine associations between microbiome and IBD subtypes, we used OTU (Operational Taxonomic Unit, or molecular species) table, which displays the relative abundance of each OTU for each sample and performed the analyses at different taxonomical levels from phylum to genera. We identified 7 phyla that significantly differed between HC, UC and CD (Kruskal-Wallis test, FDR < 0.05). CD showed higher abundance of Proteobacteria and Fusobacteria (FDR = 4.56×10^{-6} and FDR = 0.0003 respectively) and lower abundance of Euryarchaeota, Lentisphaerae, Tenericutes and Verrucomicrobia (FDR = 0.0006, FDR = 0.005, FDR = 1.4×10^{-5} and FDR = 0.03, respectively) compared with HC. In the case of UC, only two phyla (Lentisphaerae and Verrucomicrobia) have a higher proportion in HC (FDR = 0.02 and FDR = 0.001, respectively). Four phyla were significantly different between UC and CD. Fusobacteria and Proteobacteria (FDR = 5.81×10^{-5} and FDR = 6.44×10^{-6}, respectively) were more frequent in CD whereas Actinobacteria and Tenericutes were in higher relative abundance in UC (FDR = 0.002 and FDR = 0.003, respectively) (**Figure 18**).

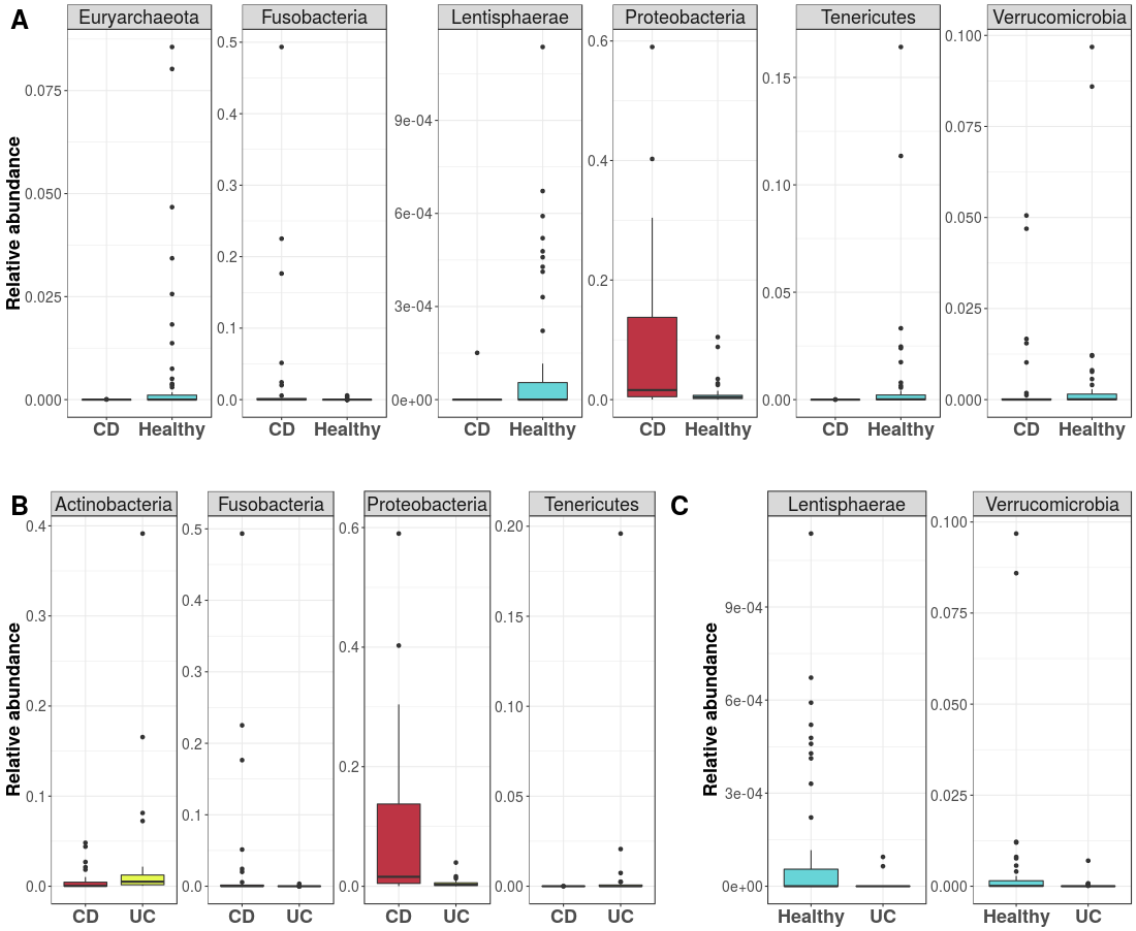


Figure 18. Dysbiosis in patients with IBD at phylum level. Taxonomic differences were detected at phylum level between HC and CD (A), CD and UC (B) and UC and HC (C) using pairwise Mann-Whitney tests (corrected p values; false discovery rate < 0.05) on significantly different phyla from Kruskal-Wallis tests.

We reported eleven genera enriched in patients with CD compared with HC whereas 29 genera were found in higher abundance in HC. Only 8 genera were in higher abundance and two in lower abundance in patients with UC compared with HC. Alteration of CD and UC with respect to healthy controls accounted for 40 genera versus 10, respectively (**Figure 19**). These results also suggest that dysbiosis is greater in CD than in UC patients at different taxonomical levels.

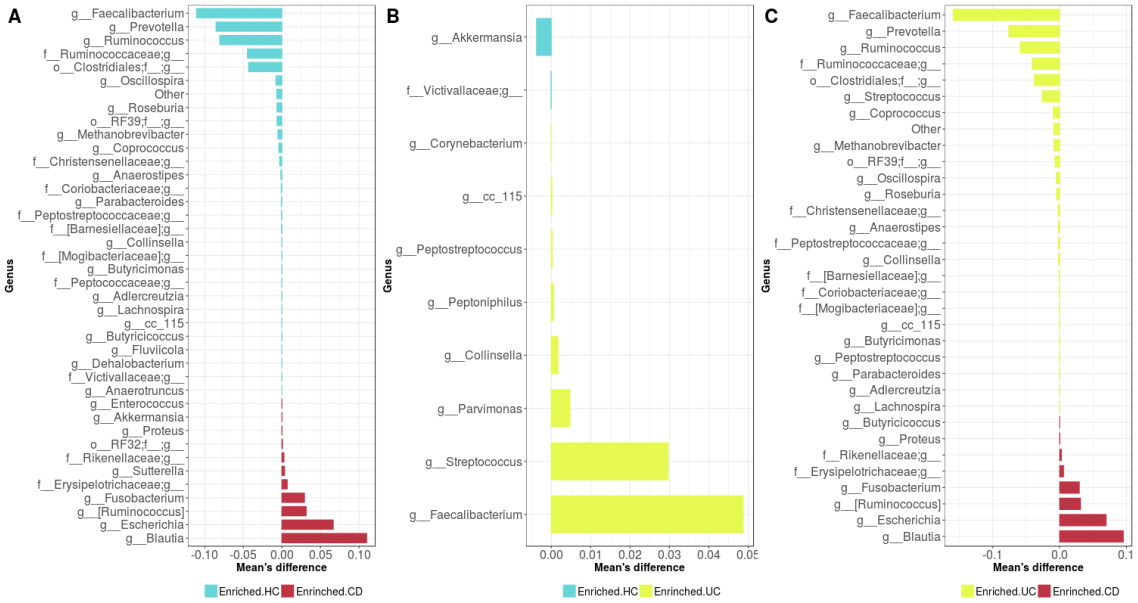


Figure 19. Dysbiosis in patients with IBD at genus level. Taxonomic differences were detected at genus level between HC and CD (A), CD and UC (B) and HC and UC (C) pairwise Mann-Whitney tests (corrected p values; false discovery rate < 0.05) on significantly different genera from Kruskal-Wallis tests.

In order to find microbial signatures of recurrence, we compared the fecal microbiome in samples of UC and CD patients at recurrence of the illness with those of patients that remained in remission after 1 year of follow-up using the non-parametric Kruskal-Wallis test. We did not find significant differences between these two groups in any of the IBD subtypes. This result suggests that either the recurrence status is not associated with additional alterations of the microbial composition or that the 16S rRNA data are insufficient to detect differences between recurrence and remission status.

Moreover, in order to determine the predictive value of recurrence in these patients, we used the Spanish IBD and the Belgian CD cohorts. For the Spanish IBD cohort, we compared baseline fecal samples of those patients who developed recurrence later (n = 13 for CD and n = 18 for UC) with basal samples of those who remained in remission after 1 year of follow-up (n = 21 for CD and n = 15 for UC). We did not find predictive biomarkers of recurrence either for CD nor UC using Kruskal-Wallis test. In the case of the Belgian cohort, composed by CD patients with more severe conditions such that they required an intestinal resection, we evaluated the predictive value of recurrence

performing Kruskal-Wallis test on the fecal samples that were collected before surgery comparing patients based on their Rutgeerts scores obtained 6 months after surgery. The results showed that patients who developed a recurrence after the surgery (Rurgeerts score of i3 and i4, n = 28) harbored a higher relative abundance of *Streptococcus* (p = 0.002, FDR = 0.17) than those who remained in remission (Rutgeerts score of i0 and i1, n = 26) at baseline. These results suggest that the relative abundance of *Streptococcus* before surgery could be a good predictive value of CD recurrence.

2.2 Relation between microbiome, smoking habit and clinical data

We studied the association between the relative abundance of all the groups of bacteria and the smoking habit using the Mann-Whitney test. CD patients contained an unknown genus of the Peptostreptococcaceae family that was in a higher proportion in smokers (FDR = 0.006) whereas non-smokers showed higher quantities of *Eggerthella lenta* (**Table 8**).

In UC, smokers presented greater abundance of *Butyricimonas*, *Prevotella* and Veillonellaceae (FDR < 0.04). On the other hand, non-smokers showed higher proportion of Clostridiaceae and *Bifidobacterium adolescentis* (FDR < 0.03).

We investigated a potential link between the localization of CD and extension of UC (obtained by Montreal classification) with the relative abundance of determined bacterial species, as suggested by Vermeire and colleagues (Vermeire, Van Assche, and Rutgeerts 2006). Mann-Whitney test identified more presence of *Enterococcus faecalis* and an unknown species of the family Erysipelotrichaceae in stool samples when the disease was localized in the ileum instead of in the ileocolon. We used the Kruskal-Wallis test to correlate disease behavior with microbial community composition and found that proctitis presented association with higher proportions of an unknown Clostridiales, *Clostridium*, an unknown Peptrostreptococcaceae and Mogibacteriaceae (FDR < 0.05) in stool (**Table 9**, **Table 10**). We did not find any type of correlation between any microbial groups and treatment.

Table 8. Relative abundance of microbial genera associated with smoking habit in UC and CD. Mann-Whitney test (FDR < 0.05) for UC (*Prevotella*, *Butyricimonas*, *Bifidobacterium adolescentis*, Clostridiaceae and Vellionellaceae) and CD (unknown Peptostreptococcaceae, *Eggerthella lenta*).

	UC		CD	
SMOKING HABIT	**SMOKERS**	**N-SMOKERS**	**SMOKERS**	**N-SMOKERS**
Unknown Peptostreptococcaceae	0.000000	0.000081	0.000275	0.000009
Prevotella	0.133849	0.046244	0.000403	0.020871
Butyricimonas	0.000126	0.000242	0.000037	0.000079
Eggerthella lenta	0.000202	0.000073	0.000018	0.000070
Bifidobacterium adolescentis	0.002345	0.036023	0.001375	0.005323
Clostridiaceae	0.001361	0.007342	0.032509	0.002131
Velllionellaceae	0.009328	0.004139	0.000843	0.002920

Table 9. Relative abundance of microbial genera associated with localization of CD. Mann-Whitney test (FDR < 0.05).

SITE OF DISEASE - CD	**L1**	**L3**	
Unknown Erysipelotrichaceae	8.633E-03	2.244E-02	
Enterococcus faecalis	2.017E-04	4.034E-05	

Table 10. Relative abundance of microbial genera associated with extension of UC. Kruskal-Wallis test (FDR < 0.05).

SITE OF DISEASE - UC	**E1**	**E2**	**E3**
Mogibacteriaceae	0.00235	0.00114	0.00071
Unknown Clostridiales	0.06815	0.05187	0.02447
Clostridium	0.00903	0.00051	0.00062
Unknown Peptostreptococcaceae	0.00020	0.000000	0.000002

2.3 Microbial marker discovery

Our comparisons of the microbiome composition of CD, UC and HC revealed that *Faecalibacterium* together with an unknown genus of Peptostreptococcaceae, *Anaerostipes*, *Methanobrevibacter* and an unknown genus of Christensenellaceae were more abundant in HC and UC and almost absent or did not appear in CD. In the case of *Fusobacterium* and *Escherichia*, we found higher abundance of these genera in CD patients, but it was almost absent in UC patients and HC. Finally, *Collinsella* was found mostly in UC cases so it allowed the discrimination between UC and CD. Based on these findings, we developed an easy-to-use algorithm to discriminate CD and non-CD samples based on eight genera with differences between CD and UC and between CD and HC identified previously. We developed our algorithm in basal samples from HnR, HR, CD and UC (40, 65, 34 and 33 respectively) individuals from the Spanish IBD cohort. The algorithm retains samples that do not contain *Faecalibacterium* or unknown genus of Peptostreptococcaceae, *Anaerostipes* and unknown genus of Christensenellaceae or contains *Fusobacterium* and *Escherichia* but not *Collinsella* and *Methanobrevibacter* (**Figure 20**).

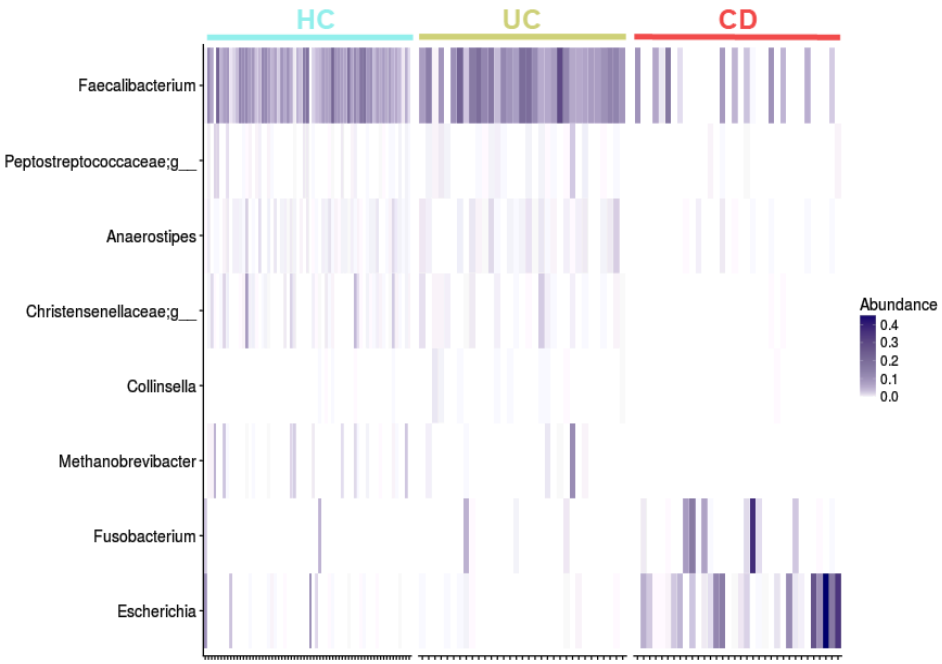


Figure 20. Microbial marker discovery. Eight bacterial genera showed potential to discriminate between HC (unrelated HnR and related HR) and patients with CD and UC in the discovery cohort: 65 HC, 33 UC patients and 34 CD patients.

Table 11. Contingency table for the discovery cohort of the algorithm.

		Conditions		
		CD	non-CD	Total
Predicted	CD	81 **(76.42%)**	17	98
	non-CD	25	141 **(87.94%)**	166
	Total	106	158	264

We validated the algorithm using different digestive disorders cohorts. To evaluate the sensitivity of the markers, we utilized the CD Belgian cohort of 54 CD patients recruited at the University Hospital Leuven (Belgian CD cohort). We applied our algorithm to a total of 193 fecal samples and obtained an overall sensitivity of 69.43%. To analyze the specificity of the biomarker proposed, we used the Spanish UC cohort already described composed by 41 UC patients that were enrolled at the University Hospital Vall d'Hebron and provided one sample before probiotic intake. We tested our algorithm proposed and obtained a specificity of 90.24%. We also tested the specificity of the technique on three non-IBD published cohorts of digestive disorders and healthy individuals. We used the already described Spanish IBS, the German anorexia and the British (twin pairs) cohorts. IBS is a disease that shares symptoms with CD, which may include abdominal pain, cramps, constipation and diarrhea. Due to their common characteristics, a biological marker could be useful to distinguish both diseases and avoid unnecessary endoscopies. We applied our algorithm to 202 fecal samples collected from 125 IBS individuals. Out of 202 samples with IBS we could identify 22 as CD samples what corresponds to only a 10.89% of false positives and a specificity of 89.11%. Anorexia samples are part of a study that was designed to detect dysbiosis in patients with anorexia compared with HC and evaluate the changes in microbiome after a weight gain in the same patients (Mack et al. 2016). As demonstrated in this previous study, anorexia is associated with shifts in the composition of the gut microbiota. To evaluate whether dysbiosis is similar in anorexia and CD, we tested our algorithm in these samples. For

this purpose, we analyzed a total of 99 stool samples from anorexia patients from which the algorithm detected 2 samples as being CD showing therefore a specificity of 97.98%. We finally tested the algorithm with 1016 and 59 healthy control samples from the British and German anorexia cohorts, respectively. Regarding the healthy twin pairs, it was originally designed to evaluate how the host genetic variation shapes the gut microbiome. Our algorithm falsely identified as CD only 77 samples out of the 1016 throwing a specificity of 92.42%. Overall, in the validation cohort, we analyzed a total of 193 CD samples and 1417 non-IBD samples throwing a sensitivity of 69.43% and a specificity of 92.45% (**Table 12**).

Table 12. Confusion table for the validation cohort. It includes the Belgian CD cohort, the Spanish UC cohort, the Healthy UK cohort, the Spanish IBS cohort and the German anorexia cohort.

		Conditions		
		CD	non-CD	Total
Predicted	CD	134 **(69.43%)**	107	241
	non-CD	59	1310 **(92.45%)**	1369
	Total	193	1417	1610

Figure 21 shows the profile of the 8 microbial species that were included in the algorithm and used in the validation dataset of 1610 fecal samples from different diseases: HC, CD, UC, IBS and anorexia. This heatmap clearly confirms that CD is characterized by a different abundance profile of the eight microbial biomarkers compared with the other disease groups. This result is also confirmed by a separate clustering based on the unweighted UniFrac PCoA representation (**Figure 22**).

Previous tests described in this study were done on 16S rDNA MiSeq sequencing of the fragment V4. To evaluate the reproducibility of the algorithm with other techniques

such as the V3-V5 variable region of the 16S rRNA gene (instead of V4) and the Ion Torrent sequencing platform (instead of Illumina MiSeq), we collaborated with the Professor Harry Sokol who provided us with a French IBD cohort. We thus tested the accuracy of the algorithm on 232 stool samples (146 CD, 86 UC and 38 HC). Our algorithm showed a sensitivity of 60% and a specificity of 94.8% in this data. Moreover, we noticed that this dataset did not carry any sequences belonging to the genus *Collinsella* and a very low abundance of *Methanobrevibacter*, which in our algorithm allow the differentiation between UC and CD, which can explain the decrease in sensitivity. The low sensitivity of the algorithm in this cohort could be either due to a different geographical sampling of more probably to different technical approaches. The good sensitivity obtained with the Belgian cohort (69.43%) points to a technical issue.

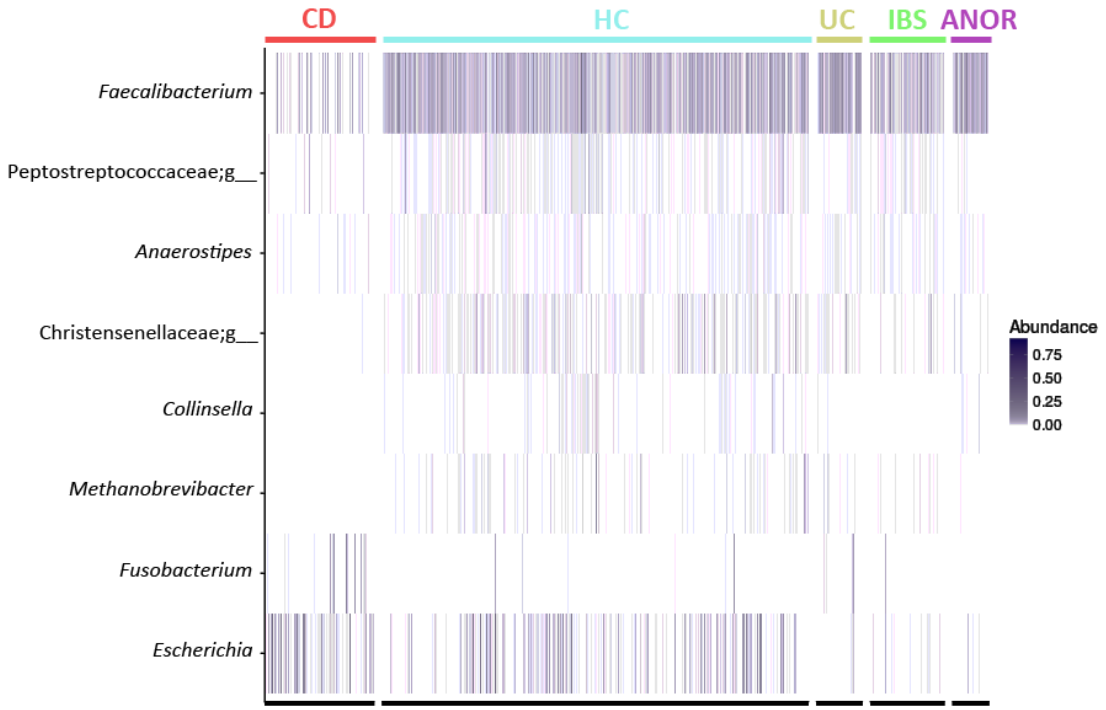


Figure 21. Microbial marker validation. Eight bacterial genera showed potential to discriminate between CD and non-CD in the validation cohort of 1610 samples: HC (n = 1075), CD (n = 193), UC (n = 41), IBS (n = 202) and anorexia (n = 99). Each blue bar represents the presence of each microbial group for each subject. Participants in each group are underlined with a specific color code (blue = all healthy controls, red = CD, yellow = UC, green = IBS and purple = anorexia). The plot was performed using an R script on relative abundance of the eight bacterial genera. The gradient color for the bars corresponds to white = absent, clear blue = low abundance and dark blue = high abundance.

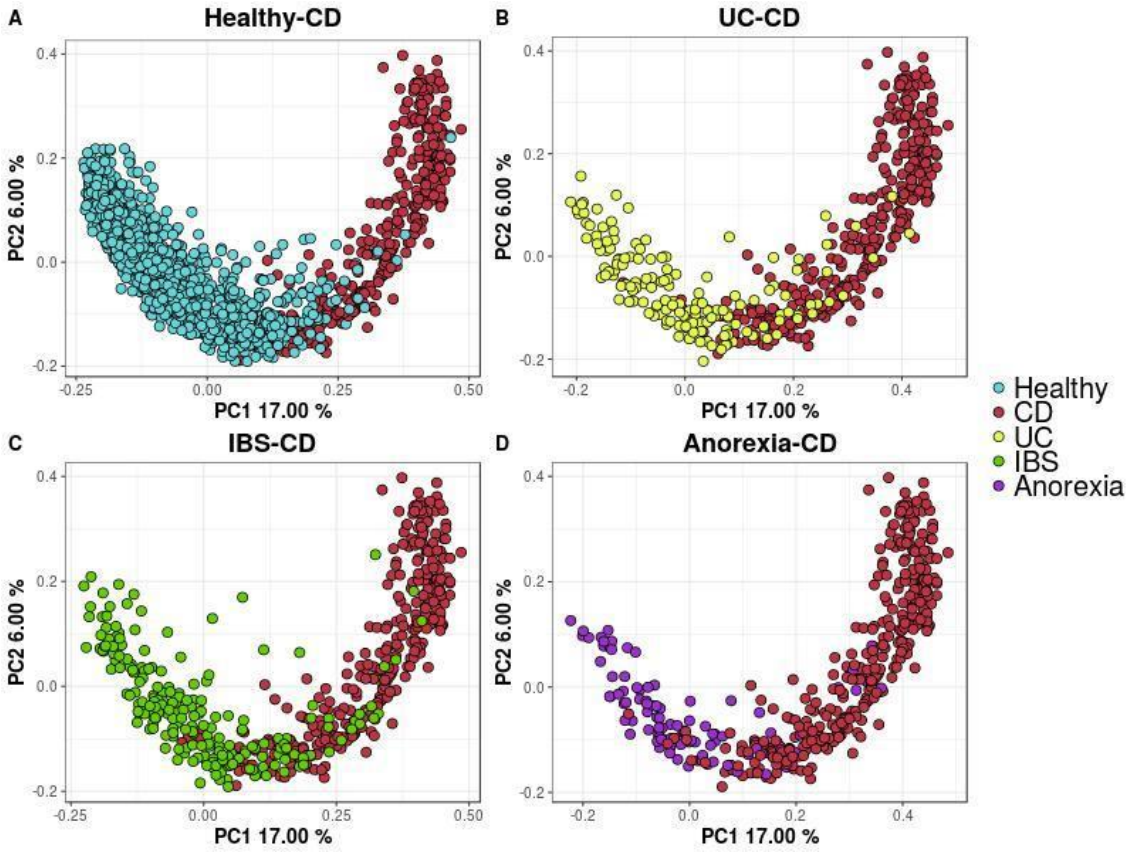


Figure 22. Unweigthed UniFrac Principal Coordinate Analysis representation of the various groups of subjects. Significant differences were found between CD and HC, UC, IBS and anorexia (NPMANOVA test, $p < 0.001$). NPMANOVA is a non-parametric multivariate test of variance.

Table 13. Contingency table in the validation cohort from France

		Conditions		
		CD	non-CD	Total
Predicted	CD	88 **(60.27%)**	30	118
	non-CD	58	94 **(75.81%)**	152
	Total	146	124	270

2.4 IBD classifier

As we have already demonstrated, differences in microbial composition can be used to classify IBD patients according to their phenotype. We decided to use artificial intelligence (AI) to obtain a classifier with a better performance of this classification on our samples. To obtain the best classifier, we trained three AI algorithms: Random Forest (RF), AdaBoost and XGBoost in OTU abundance tables. We evaluated the performance of the classifier with a five-cross validation in the Spanish IBD cohort and between cohorts training the classifier in the Spanish IBD cohort and validating on the independently Belgium CD, Spanish UC, German Anorexia, Spanish IBS and UK cohorts. Although all algorithms performed better than the microbial marker previously proposed to distinguish between CD and non-CD samples, we obtained the best results with RF model with an area under de curve (AUC) of 0.96 (**Figure 23 A**) for our cross validation showing sensitivity and specificity values of 80% and 93.9%, respectively (**Table 14**). The classifier established a combination of *Faecalibacterium* and *Eubacterium* as the best features to distinguish CD samples from non-CD samples. The independent validation with the CD Belgian, UC Spanish and non-IBD cohorts used provided an AUC of **0.99** (**Figure 23 B**) and even higher values than the five-cross validation for sensitivity and specificity of **94.8%** and **99.5%** (**Table 15**).

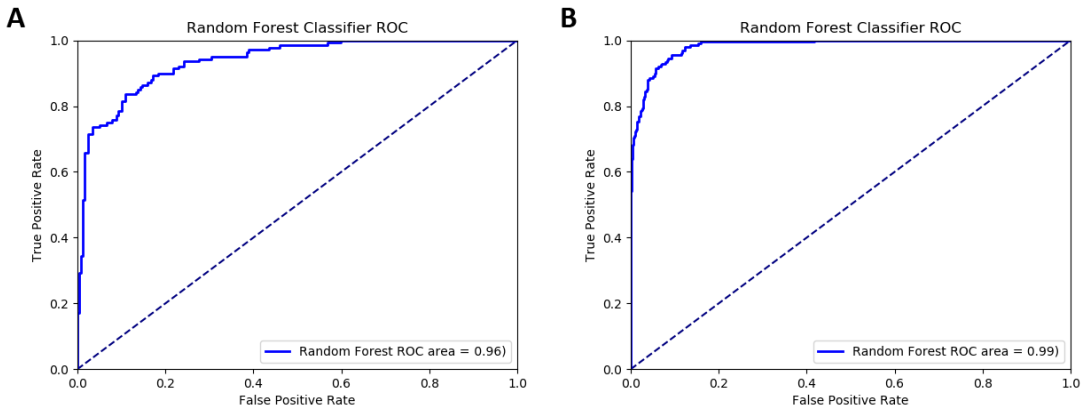


Figure 23. ROC curves for cross-validation and independent validation of Random Forest classifier.

Table 14. RF classifier for predicting CD (five-cross-validation).

		Conditions		
		CD	non-CD	Total
Predicted	CD	112 **(80%)**	17	129
	non-CD	28	262 **(93.9%)**	290
	Total	140	279	419

Table 15. RF classifier for predicting CD (independent validation).

		Conditions		
		CD	non-CD	Total
Predicted	CD	183 **(94.8%)**	7	**190**
	non-CD	10	1410 **(99.5%)**	**1420**
	Total	**193**	**1417**	**1610**

3. Compositional and functional microbiome in IBD using DNA shotgun sequencing (or shotgun metagenomics)

We hypothesized that using both compositional and functional information from shotgun data (sequences of random DNA fragments) will provide us more insights in the pathophysiology of IBD than using 16S data.

We randomly selected a subset of 178 samples (out of 379 samples from UC, CD and related healthy individuals) from the Spanish IBD cohort to perform DNA shotgun sequencing (**Table 7**). The random selection of a subset of the cohort was motivated by the high sequencing const and more time-consuming analyses of the shotgun data compared to the 16S rRNA data.

We obtained paired-end fastq files per sample with a mean of 34,564,409.47± 12,059,962.07 paired reads that we analyzed using the bioinformatics pipeline HUMAnN2 (Franzosa et al. 2018) following the standard process proposed by their developers and described in the **Methods** section of the present work.

3.1 Dysbiosis in IBD using DNA shotgun sequencing

As we did with 16S data, we investigated the differences between CD, UC, HC and evaluated the stability in time for both remission and relapse status.

We used the Bray-Curtis index, a metric to compare microbial composition between two different communities, on strain level abundance tables. Here, we did not use UniFrac distance as we did for 16S because it requires phylogenetic information, that we cannot generate using shotgun sequences since they could not be aligned being random DNA sequences. Using a multivariate analysis of variance on Bray-Curtis distance matrices, the NPMANOVA test, we saw that patients with CD at all timepoints significantly differed from UC patients and healthy controls (NPMANOVA test; p = 0.0015) (**Figure 24 A**). Nevertheless, UC patients did not significantly differ from healthy controls. We also compared the two different status of the disease, remission and relapse, in both

subtypes of IBD. Neither in UC nor in CD we were able to significantly differentiate samples between inactive or active disease status (**Figure 24 B**).

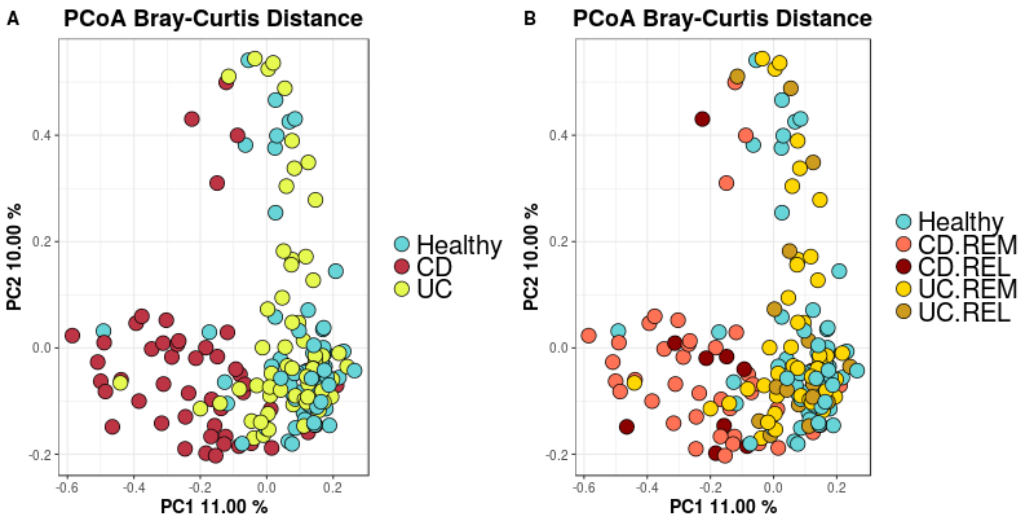


Figure 24. Dysbiosis in patients with IBD with shotgun data. PCoA computed from Bray-Curtis distance matrices from strain abundance tables of IBD patients and healthy individuals. Gut microbiome separates samples of CD patients with those of UC patients and HC (A). However, it is not able to evaluate the disease severity (B).

We computed the mean Bray-Curtis distance between each patient sample and all healthy samples to validate the differences seen in Figure PCOA. We saw that CD samples showed higher distances with healthy controls than UC, but these differences were not significantly different (Error! Reference source not found.). We did not observe significant differences between remission and relapse status. So far, taking into account these observations at the composition level using shotgun data, we validated our previous findings using 16S data.

We also used the Bray-Curtis index to assess the stability over time for CD and UC, but we could not do it for healthy controls because for economical limitation only the baseline sample for each individual was sequenced. We did not identify higher instability in CD samples than in UC patients; activeness of the disease neither contributes to higher instability. This lack of differences between CD and UC contrast with what we found using 16S data, which could be attributed to a smaller sample size.

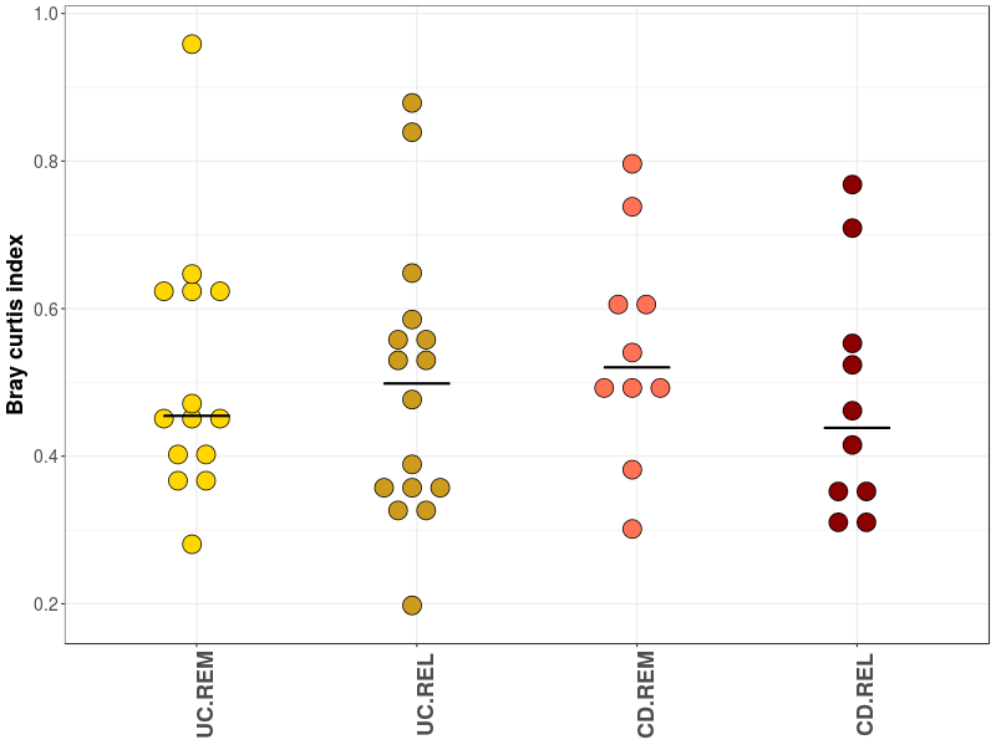


Figure 25. Stability in time associated with activity of the disease. Bray-curtis beta diversity distance between basal and last timepoints splitted by activeness of the disease, show that there are not significant differences between CD and UC, nor between remission and recurrence status.

We calculated alpha diversity with Shannon and Chao1 indexes to determine microbial diversity per group. Considering samples at all timepoints, CD showed lower α diversity than healthy controls for both indexes (t-test, FDR < 10^{-7}), however, UC did not differ from HC (**Figure 26**). This result is in accordance with 16S rDNA sequencing.

We compared alpha diversity between different timepoints and disease's activeness to determine whether disease's activeness was associated with the microbial diversity in CD and UC patients. We only found a significant decrease of chao1 alpha diversity index in the last timepoint of UC between remission and relapse status (t-test, FDR = 0.03). This decrease in alpha diversity in active disease of UC was not significant for Shannon diversity index (**Figure 27**), which suggests that only richness and not evenness was associated with the disease severity.

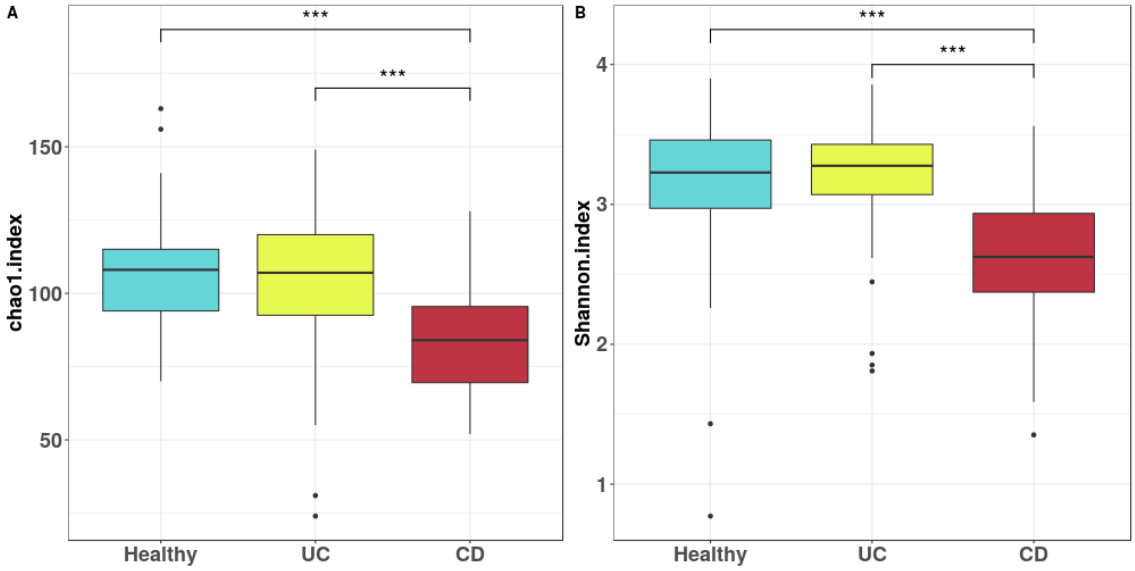


Figure 26. Alpha diversity in the different subtypes of IBD. (A) Chao1 and (B) Shannon indexes for alpha diveristy showed a decrease in alpha diversity for CD samples in comparison with UC patients and healthy individuals (t-test, FDR < 10^{7}).

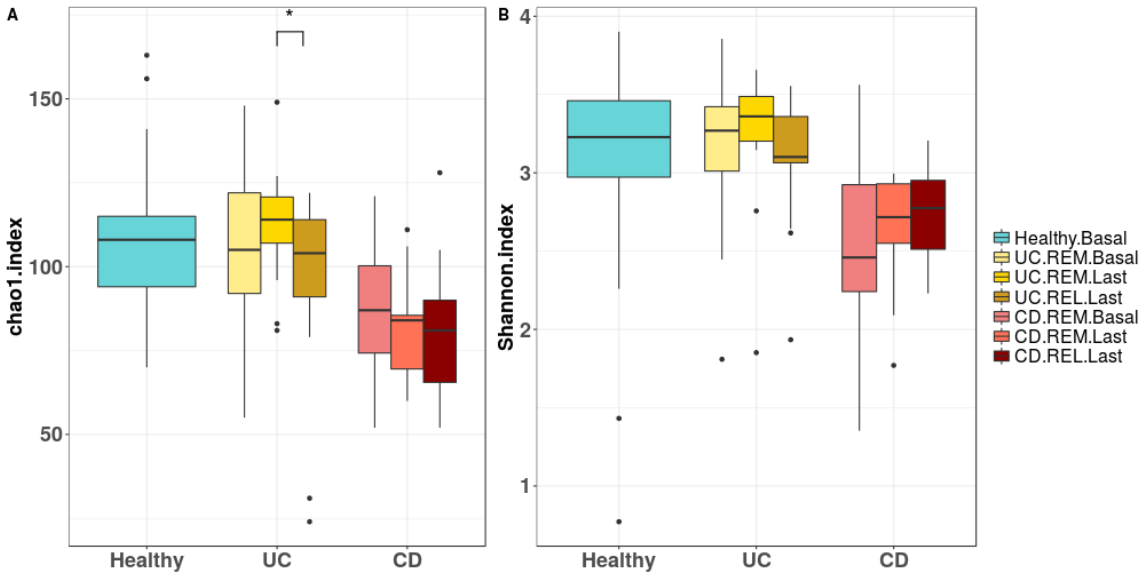


Figure 27. Alpha diversity in activeness and time of the disease. (A) Chao1 and (B) shannon indexes for alpha diversity did not presented significant differences (t-test, FDR < 0.05) between basal and last timepoints nor in activeness of the disease. Only for chao1 index we identified a difference between remission and relapse last timepoints for UC with an FDR = 0.03.

We identified 5 phyla that were significantly different between HC, CD and UC: Proteobacteria (Kruskal-Wallis test, FDR = 0.0001), Ascomycota (Fungi, FDR = 0.0001), Euryarchaeota (FDR = 0.006), Fusobacterium (FDR = 0.006) and Verrucomicrobia (FDR = 0.01). We performed pairwise Mann-Whitney tests to identify the groups that were implicated in this variation shown in the tests. Proteobacteria, Fusobacteria and Ascomycota were more frequent in CD patients than in healthy controls whereas Euryarchaeota was more present in healthy individuals. The main differences between CD and UC came from Ascomycota and Proteobacteria showing both higher abundance in CD patients than in UC. Finally, UC presented more Verrucomicrobia than healthy individuals (**Figure 28**).

We determined differences at genus level (Kruskal-Wallis test, FDR < 0.05) and identified that healthy individuals were enriched in 23 genera and depleted in 14 genera in comparison with CD. Only 8 genera were different between UC and HC with four of them enriched in each of the groups. Finally, CD presented a lower abundance in 19 genera and higher abundance for 12 genera in comparison with UC (**Figure 29**). Comparing these results with 16S, only 3 out of the 8 genera found in metagenomic sequencing for the comparison between UC and HC matched with the genera identified in 16S rDNA analyses and only 10 and 8 of the genera significantly different between CD and HC and between CD and UC, respectively, in shotgun sequencing were in common with 16S rDNA sequencing data (**Table 16**). Nevertheless, although there were some discordant bacteria between 16S and shotgun data, results in both techniques determined that the dysbiosis found is higher for CD than for UC. These differences could be due to the sample size or the database used to identify bacteria in each approach.

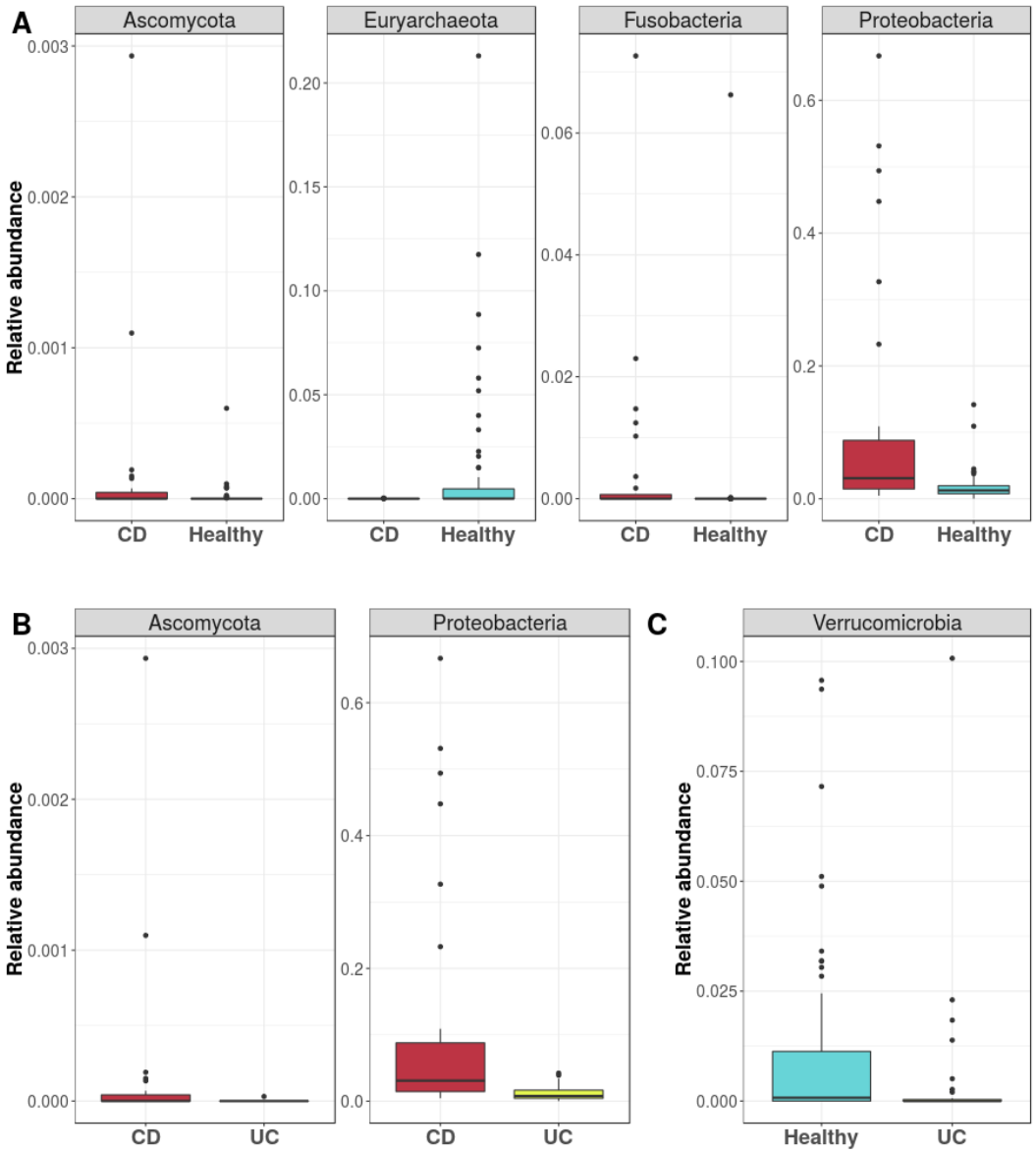


Figure 28. Relative abundance in phylum. Significantly different phyla between the three groups of the study (Kruskal-Wallis test, FDR < 0.05). (A) CD samples were enriched in 3 genera (Ascomycota, Fusobacteria and Proteobacteria) while they were depleted in Euryarchaeota. (B) UC and CD only differed in two phyla, Proteobacteria and Ascomycota, both more abundant in CD. (C) HC showed higher abundance of Verrucomicrobia than UC.

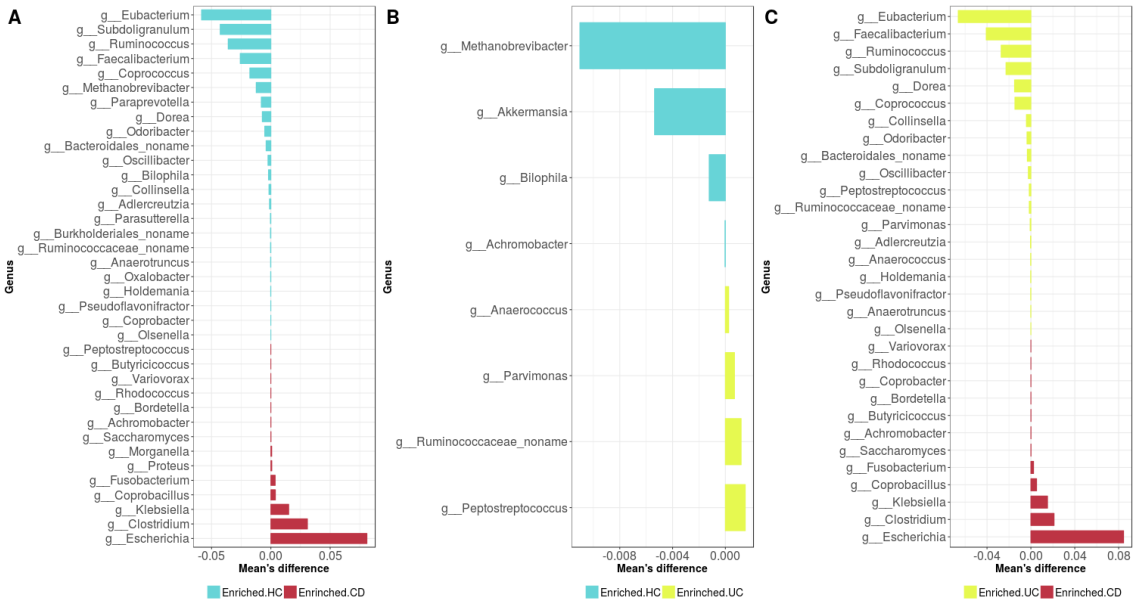


Figure 29. Genera enrichment barplot. Significantly different genera between the three groups of the study (CD, UC and HC) (Kruskal-Wallis test, FDR < 0.05) represented by the difference of means. (A) 14 genera were more frequent in CD samples while they were depleted in 23 compared with HC. (B) UC showed higher abundance in 4 genera and lower abundance in another 4 genera in comparison with CD. (C) CD was enriched in 12 genera whereas it was decreased in 19 genera in comparison with UC.

Table 16. Common significantly different genera identified between 16S rDNA sequencing and shotgun sequencing per group of comparisons

CD.HC	UC.HC	CD.UC
Ruminococcus	*Akkermansia*	*Faecalibacterium*
Faecalibacterium	*Parvimonas*	*Ruminococcus*
Coprococcus	*Peptostreptococcus*	*Collinsella*
Methanobrevibacter		*Coprococcus*
Collinsella		Unknown Ruminococcaceae
Aldercreutzia		*Aldercreutzia*
Unknown Ruminococcaceae		*Fusobacterium*
Butyricicoccus		*Escherichia*
Fusobacterium		
Escherichia		

In contrast with 16S rDNA sequencing, shotgun sequencing not only detects significant differences at genus level as shown in **Figure 28** and **Figure 29**, but also allows a deeper taxonomical analysis and provides information at species and strain levels with high resolution. However, strains are not as well characterized as species in databases, more than 50% of strains identified in our samples, corresponded to unclassified strains from known species. We detected 101 species that were significantly different between the three groups (Kruskal-Wallis, FDR < 0.05). Among them, 89 species differed between healthy and CD patients at baseline, 57 of these species were enriched in HC whereas the 32 remaining species were more present in CD. In contrast, we found only 24 significant species that were different between UC and HC, 9 were enriched in HC and 15 were more frequent in UC. Finally, we found 66 species that were significantly different between UC and CD, 45 showed higher abundance in UC whereas 21 were more frequent in CD (**Table 17**).

We checked changes in time for each bacteria and activeness of the disease, but we did not find any significant species that was associated with remission or recurrence status nor the time between the first and the last samples as previously shown with beta diversity with general microbiota.

Table 17. List of species that were differentially abundant between HC, UC and CD. FDR column corresponds to FDR correction for Kruskal-Wallis test between the three groups. Values represent the difference in abundance for each species between both groups. In red, species enriched in CD, in blue, species enriched in HC and in yellow, species enriched in UC. Blank spaces indicate that those species are not significantly different in that comparison.

Species	FDR	CD vs HC	CD vs UC	UC vs HC
s__Lachnospiraceae_bacterium_9_1_43BFAA	2.54678E-07	0.001525658	0.001629714	
s__Clostridium_clostridioforme	4.60138E-07	0.012749108	0.01300207	
s__Clostridium_ramosum	2.00743E-06	0.001327658	0.001394516	
s__Coprobacillus_unclassified	3.66713E-06	0.004801805	0.005198389	
s__Eubacterium_ventriosum	3.66713E-06	-0.003530593	-0.004889155	
s__Ruminococcus_gnavus	8.20768E-06	0.017618718	0.01756636	
s__Ruminococcaceae_bacterium_D16	1.67509E-05	-0.000290564	-0.001486498	0.001195934
s__Pseudoflavonifractor_capillosus	4.16229E-05	-8.53865E-05	-0.000208171	
s__Alistipes_shahii	8.05112E-05	-0.004464754	-0.001461131	-0.003003623
s__Subdoligranulum_unclassified	8.05112E-05	-0.042337421	-0.02218407	
s__Eubacterium_eligens	0.000104283	-0.007543927	-0.017331719	0.009787792
s__Lachnospiraceae_bacterium_4_1_37FAA	0.00016019	0.000397235	0.000404848	
s__Parvimonas_micra	0.000262846		-7.02313E-05	6.55047E-05
s__Peptostreptococcus_stomatis	0.000313722	-2.47781E-05	-0.0004286	0.000403822
s__Adlercreutzia_equolifaciens	0.000313722	-0.001174414	-0.000392149	
s__Fusobacterium_nucleatum	0.000336969	0.003857455	0.003862917	
s__Eubacterium_siraeum	0.000336969	-0.007805327	-0.009864828	
s__Roseburia_inulinivorans	0.000336969	-0.006035777	-0.005310004	
s__Bacteroidales_bacterium_ph8	0.000406363	-0.004015543	-0.00327627	
s__Ruminococcus_obeum	0.00043402	-0.002686472	-0.006572371	
s__Variovorax_unclassified	0.000471729	7.45198E-06	8.13479E-06	
s__Anaerostipes_unclassified	0.000539078	0.002213111	0.002261963	

Species	FDR	CD vs HC	CD vs UC	UC vs HC
s__Bordetella_unclassified	0.000569039	3.23361E-05	5.23033E-05	
s__Dorea_formicigenerans	0.000647955	-0.001769741	-0.008699336	0.006929595
s__Escherichia_coli	0.000708232	0.07085613	0.074064211	
s__Faecalibacterium_prausnitzii	0.000729273	-0.025833881	-0.040705261	
__Parabacteroides_goldsteinii	0.001099025	-0.001153757	-0.000622036	
s__Subdoligranulum_variabile	0.001099025	-2.53017E-05	-2.89235E-05	
s__Odoribacter_splanchnicus	0.00118829	-0.004845641	-0.003296253	
s__Ruminococcus_albus	0.001362225	-1.96697E-05	-1.08213E-05	
s__Parvimonas_unclassified	0.00136389		-0.000634735	0.000623867
s__Lachnospiraceae_bacterium_6_1_63FAA	0.001669619	0.000585901	0.000613071	
s__Alistipes_finegoldii	0.001825518	-0.001266845		-0.001413508
s__Eubacterium_rectale	0.001825518	-0.030301502	-0.019291129	
s__Ruminococcus_lactaris	0.001877581	-0.005010391	-0.005723355	
s__Alistipes_senegalensis	0.002610484	-0.00031664	-0.000170841	
s__Blautia_producta	0.00264439	0.007258318		
s__Rhodococcus_erythropolis	0.00306888	2.70748E-05	2.78467E-05	
s__Achromobacter_unclassified	0.003081414	6.45324E-05	8.55433E-05	-2.10109E-05
s__Eubacterium_ramulus	0.003081414	-0.001812833	-0.003292219	
s__Anaerotruncus_unclassified	0.00317855	-0.000228705	-0.000192458	
s__Ruminococcus_bromii	0.003357483	-0.032889746		-0.021373316
s__Peptostreptococcaceae_noname_unclassified	0.004173372	-0.000306776	-0.000239539	
s__Saccharomyces_cerevisiae	0.00451695	0.000110133	0.0001238	
s__Roseburia_hominis	0.004584258	-0.001013656	-0.001857237	
s__Bacteroides_caccae	0.004685566	-0.006322043	-0.013718996	
s__Bacteroides_xylanisolvens	0.005510412	-0.000790266		-0.000622941
s__Alistipes_indistinctus	0.005510412	-0.000397129	-0.000856244	
s__Lachnospiraceae_bacterium_5_1_63FAA	0.006918481	-0.00095075	-0.001959342	

Species	FDR	CD vs HC	CD vs UC	UC vs HC
s__Oxalobacter_formigenes	0.006992486	-0.000186619		
s__Dorea_longicatena	0.006992486	-0.005629347	-0.006502623	
s__Anaerococcus_vaginalis	0.008369524		-0.000152576	0.00014908
s__Coprobacter_fastidiosus	0.008369524	-6.25428E-05	4.37994E-05	
s__Bilophila_wadsworthia	0.011325959	-0.000285789		
s__Coprococcus_catus	0.013085713	-0.00124144	-0.001455202	
s__Paraprevotella_xylaniphila	0.01335504	-0.000357591	-0.000505613	
s__Eubacterium_hallii	0.01335504	-0.007464821	-0.011277468	
s__Finegoldia_magna	0.01335504	8.13333E-06		
s__Dorea_unclassified	0.01381762	0.000159123	0.000280934	
s__Morganella_morganii	0.01549547	0.00071114		
s__Bilophila_unclassified	0.01569034	-0.001724154		-0.001203461
s__Desulfovibrio_desulfuricans	0.016901457	-6.73845E-05		-5.84234E-05
s__Methanobrevibacter_smithii	0.016901457	-0.011754081		-0.010442614
s__Olsenella_unclassified	0.017571427	-3.26896E-06	-3.68783E-05	
s__Proteus_mirabilis	0.018459753	0.000778491		
s__Paraprevotella_clara	0.018459753	-0.002334047	-0.000977086	
s__Anaerostipes_caccae	0.018459753	1.5298E-05		
s__Peptoniphilus_lacrimalis	0.018638608			0.000241972
s__Lachnospiraceae_bacterium_ICM7	0.018638608			0.000001325
s__Methanobrevibacter_unclassified	0.019306682	-0.000632284		
s__Parabacteroides_merdae	0.019581838	-0.004258583		
s__Peptostreptococcus_anaerobius	0.019619006	1.85833E-05		0.001094678
s__Clostridium_sp_KLE_1755	0.020689721		-0.001498609	0.001113731
s__Coprococcus_sp_ART55_1	0.020752101	-0.012601066		
s__Collinsella_aerofaciens	0.023570962	-0.001700846	-0.003362834	
s__Bacteroides_nordii	0.024159051	-0.000107369	-0.000295606	

Species	FDR	CD vs HC	CD vs UC	UC vs HC
s__Alistipes_putredinis	0.024454209	-0.014168744		
s__Coprococcus_comes	0.026261761	-0.002839866	-0.007371941	
s__Anaerococcus_obesiensis	0.026607962		-0.00010535	0.000100277
s__Erysipelotrichaceae_bacterium_2_2_44A	0.026607962	8.79025E-05		
s__Akkermansia_muciniphila	0.027442387			-0.005366661
s__Klebsiella_pneumoniae	0.030269454	0.014472602	0.014544053	
s__Butyricicoccus_pullicaecorum	0.033538803	2.04958E-06	9.87933E-05	
s__Clostridium_bolteae	0.035699622	0.014056539		
s__Bifidobacterium_bifidum	0.036084092			0.00514523
s__Blautia_hydrogenotrophica	0.036084092		-0.002298767	0.00309153
s__Blautia_hansenii	0.036084092	0.000192299	0.000217321	
s__Lachnospiraceae_bacterium_3_1_46FAA	0.036734636	-0.000285182		
s__Streptococcus_australis	0.03694745	-5.31611E-05	-0.00026791	
s__Paraprevotella_unclassified	0.03694745	-0.00539116		
s__Ruminococcus_flavefaciens	0.03694745	-7.60323E-06	-6.97042E-06	
s__Parasutterella_excrementihominis	0.037185353	-0.000414709		
s__Burkholderiales_bacterium_1_1_47	0.037185353	-0.000351933		
s__Clostridium_citroniae	0.037185353	0.0002662		
s__Rhodococcus_qingshengii	0.041900334	2.63333E-06		
s__Enterococcus_avium	0.041900334	0.00001985		
s__Ruminococcus_callidus	0.042303763	-0.001779332		
s__Achromobacter_xylosoxidans	0.042625745		1.59033E-05	
s__Aggregatibacter_segnis	0.042705548			4.80938E-05
s__Clostridiales_bacterium_1_7_47FAA	0.046184373	0.000127018		
s__Lachnospiraceae_bacterium_5_1_57FAA	0.047560693	0.000952761		

3.2 Relation between shotgun microbiome data and smoking and clinical data and patients' characteristics

As smoking habit has been related to IBD subtypes in literature, we used the Kruskal-Wallis test to determine if our data validate this association between gut microbiome and smoking habit for both IBD subtypes combined. We did not identify any bacteria that was associated with any of the three possible status of the smoking habit: Ex-smoker, non-smoker and active smoker. We also performed the Mann-Whitney test considering ex-smokers as non-smokers and compared between non-smokers and active-smokers, but we did not find any bacteria that was associated with tobacco use. We separated the analyses by IBD subtypes. In this case, we find one bacterium that was positively associated with smoking habit in UC, *Actinomyces graevenitzii* (FDR = 0.03), however, in CD we did not identify any bacterium. These findings confirm somehow the differences previously reported between CD and UC for smoking habit.

We then tested association between disease localization and microbiome composition. We did not find any bacteria associated with UC extension. For CD localization, we could only test the types L1 (ileal CD) and L3 (ileocolonic CD) with Mann-Whitney test due to a lack of samples from other localizations (L2 (colonic), and L4 (isolated upper disease)). We could not find significant differences, but we identified a tendency (Mann-Whitney test, FDR = 0.16) in the orders of Pasteurellales and Bacteroidales to be more present in L3 than in L1. Finally, we also checked behavior of UC patients in B2 (stricturing) and B3 (penetrating) groups and found a trend towards a higher proportion of the order Deinococcales in B3 than in B2.

We performed Spearman correlation to determine whether there were any bacteria associated with the continuous values of fecal calprotectin, an inflammation marker. Although we identified 3 species (*Peptoniphilus harei* (r = 0.35)*, Anaerococcus obesiensi* (r = 0.37) and *Anaerococcus vaginalis (*r = 0.39)*)* that were moderately positively correlated with calprotectin. The samples they presented a lot of values close to 0 so these results should be considered with caution. BMI was also associated with different species. On the one hand it was negatively correlated with unclassified species of *Eggerthella* (r = -0.48) and *Flavonifractor plautii* (r = -0.34), and on the other hand, it was

found positively associated with 2 species (an unclassified species of *Paraprevotella* (r = 0.44) and *Paraprevotella clara* (r = 0.44) (**Figure 30**).

We did not perform any type of association with treatment due to the small sample size of each group.

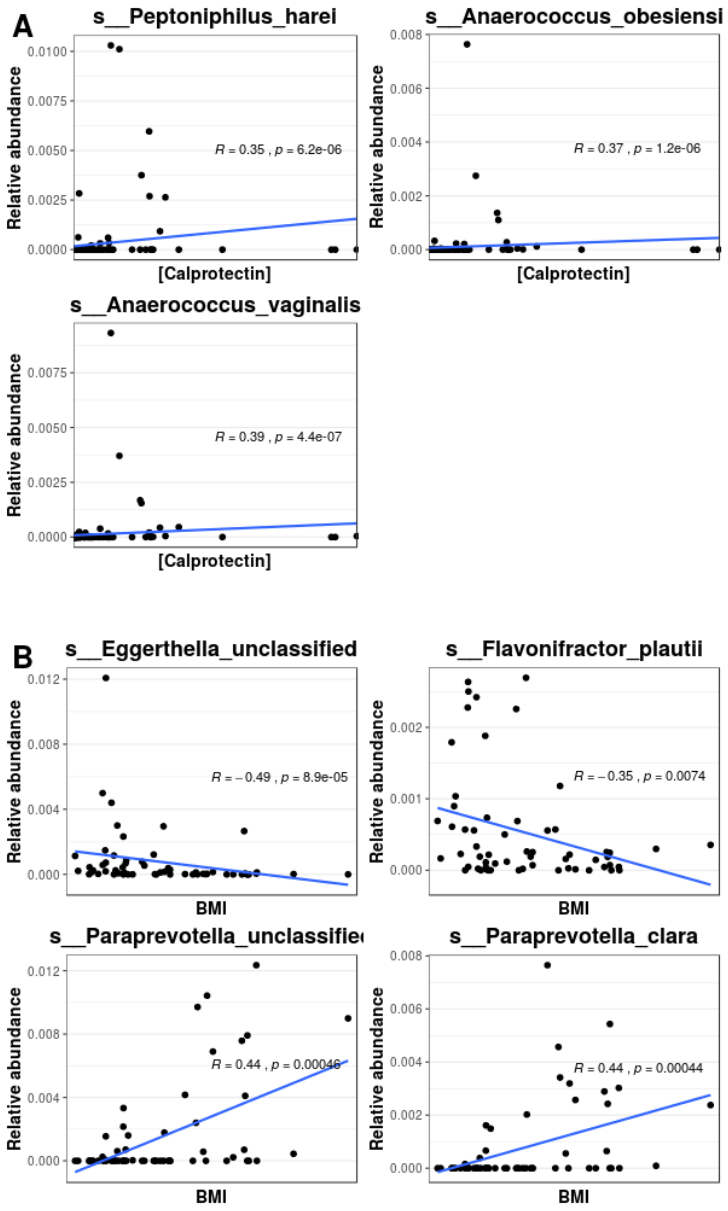


Figure 30. Microbial correlation with biological and clinical parameters. Spearman correlation with continuous variables calprotectin concentration (A) and BMI (B). Calprotectin was moderately positively correlated with 3 species but most of the correlation was driven by the high quantity of 0 values. BMI was moderately negatively correlated with 2 species and positively associated with other 2 species.

3.3 Shotgun sequencing and 16S sequencing

Although global microbiota composition estimated with 16S rDNA and shotgun sequencing clearly separate Crohn's disease patients from ulcerative colitis patients and healthy individuals (**Figure 31**), not all the individual bacteria implicated were the same for both methodologies. Thus, we sought to estimate how much the microbial frequencies were equivalent between the two techniques. We identified from phylum to genus levels which microbial annotations were common between the two methodologies. Our analyses showed that the proportion of common identities decreased for more specific taxa (**Table 18**). At genus level, in 16S rRNA data, common microorganisms comprised a mean relative abundance of 0.74 ± 0.16 per sample and in shotgun data they contributed to a mean of 0.72 ± 0.14, what suggested that a high proportion of microbial composition is detected in the same manner. To further evaluate this finding, we performed Pearson correlation analyses between common entities in 16S rDNA and shotgun sequencing for all samples at all taxonomical levels, from phylum to genus. We determined that correlation values were higher than 0.9 for nearly 70% of the samples and over 0.75 for 86% of the samples at phylum level. However, correlation was worse at lower taxonomical levels, and, at genus correlation values were over 0.9 for nearly 30% of the samples and over 0.75 for 60% of the samples (**Figure 32**).

The differences in taxonomic annotations between both techniques could be attributed to the different databases used. Indeed, each approach required a specific database for identifying microorganisms. Thus, we checked if phylum identified only in shotgun sequencing data were also present in our 16S data, which is a combination of Greengenes and PATRIC databases. Among the shotgun data, we can detect virus and fungi phyla, which cannot be detected using 16S, as we originally PCR amplify bacterial DNA. For the Bacteria kingdom, 3 out of the 4 phyla identified with shotgun sequencing were present in 16S rDNA database. This result suggests that 30% of the microbial composition that corresponds to unshared microorganisms may be influenced by the sequencing technique.

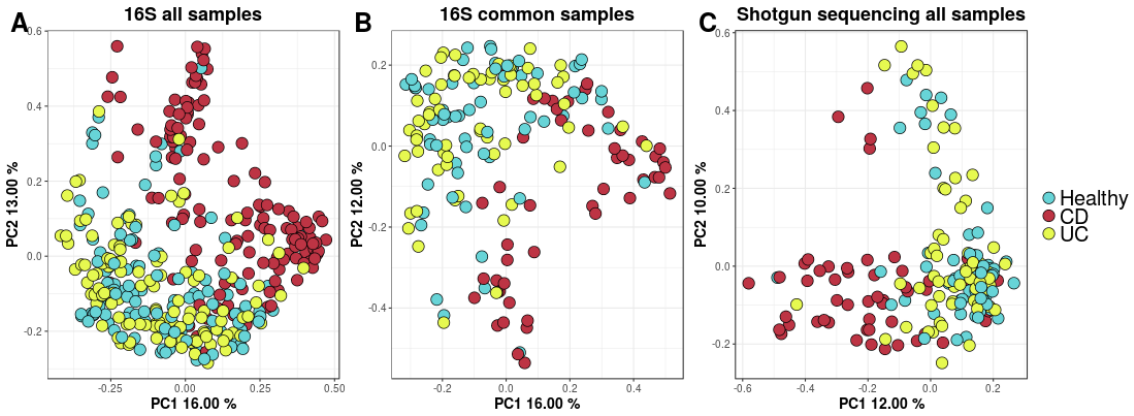


Figure 31. Dysbiosis in IBD in different techniques. PCoA of Bray-Curtis beta diversity distance computed from (A) OTU table of all samples analyzed using 16S rDNA data, (B) OTU table from commonly selected samples between 16S rDNA and shotgun sequencing analyzed with 16S rDNA data and (C) Strain abundance table from all samples analyzed using shotgun sequencing. In all cases we found significant differences between CD and UC (NPMANOVA, FDR = 0.0015) and CD and HC (NPMANOVA, FDR = 0.0015) but not between UC and HC

.

Table 18. Common and technique-specific microorganisms at different taxonomical levels between 16S rDNA and shotgun sequencing.

	PHYLUM	CLASS	ORDER	FAMILY	GENUS
COMMON	9	16	23	35	53
ONLY.16S	4	10	23	47	103
ONLY.SHOTGUN	6	18	18	56	147

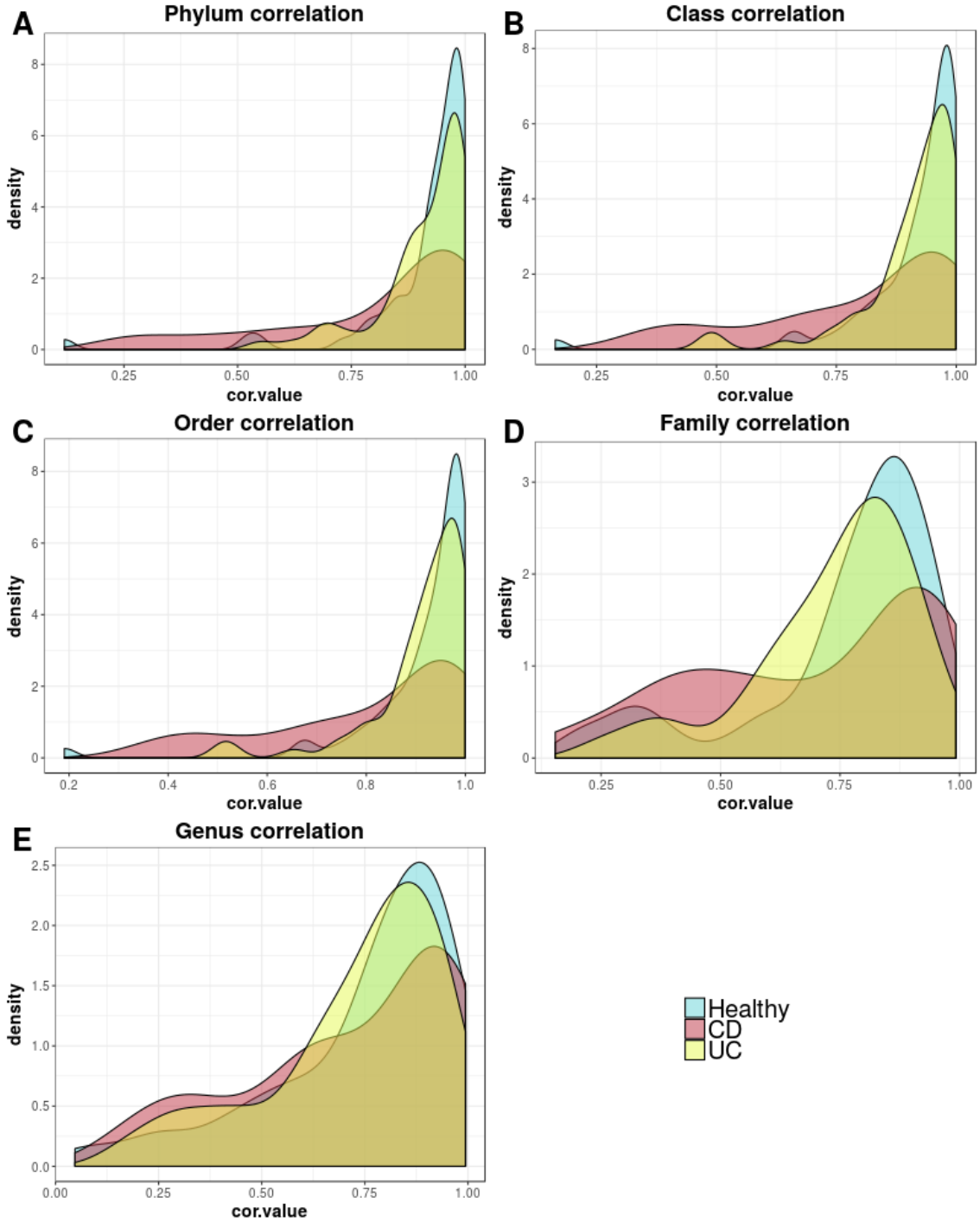


Figure 32. Correlation of common bacteria between 16S rDNA and shotgun sequencing. Density plots of Pearson correlation values of abundance of common bacteria between 16S rDNA sequencing and shotgun sequencing for (A) phylum, (B) class, (C) order, (D) family and (E) genus.

3.4 Comparison with another recent shotgun data study

Another recent IBD-shotgun metagenomic study was published by the groups of Drs. Huttenhower and Ramnik (Franzosa et al. 2019). We aimed to compare their findings with ours. To do so, we combined our dataset with this already published cohort of IBD (Franzosa et al. 2019), which contained American members enrolled in PRISM (the Prospective Registry in IBD Study at MGH) and Dutch individuals. We used the species abundance table obtained with HUMAnN2 from stool samples of 155 American individuals (68 CD patients, 53 UC patients and 34 non-IBD controls) and 65 samples from Dutch participants (20 CD patients, 23 UC patients and 22 healthy controls). We combined our species abundance table obtained from our cohort of 178 samples (64 HC, 51 CD and 63 UC) and computed Bray-Curtis beta diversity distance and a multivariate analysis of variance on distance matrices with the NPMANOVA tests.

We analyzed differences between the three groups of the study, HC, CD and UC and between studies. In both types of comparison, we detected significantly different microbiome (NPMANOVA test, FDR = 0.001 for all tests). Increasing sample size, we still detected differences between CD and healthy individuals and CD and UC patients as previously shown for the Spanish cohort driven mainly by the first coordinate of the PCoA (**Figure 33 A**). However, we also found differences between HC and UC that we could not detect in the Spanish cohort. Moreover, we detected significant differences between cohorts, in this case, mainly driven by the second coordinate as shown in the PCoA representation (NPMANOVA, FDR = 0.001) (**Figure 33 B**).

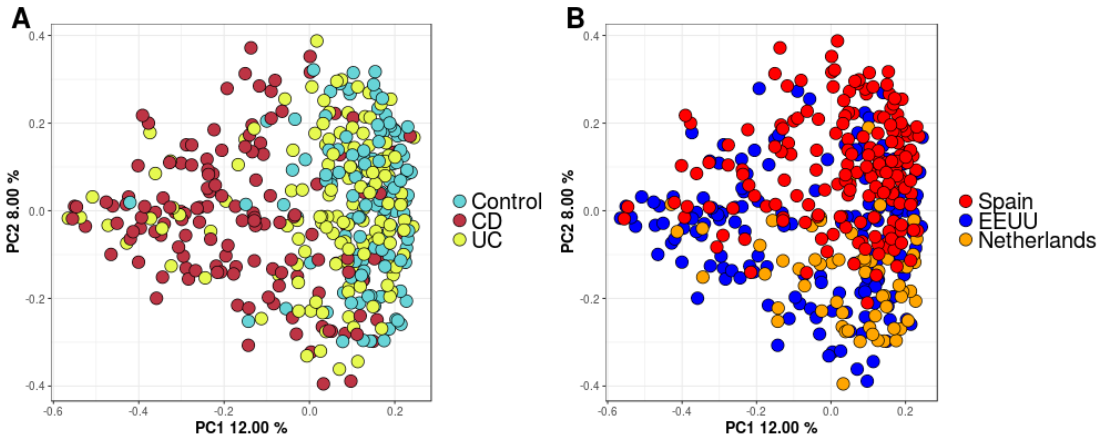


Figure 33. Dysbiosis in Spanish, American and Dutch cohorts combined. Differences were found between HC, CD and UC (NPMANOVA, FDR = 0.001) (A) and between the three different cohorts (NPMANOVA, FDR = 0.001) (B).

3.5 Validation of microbial marker in shotgun sequencing

In the previous chapter, we developed a microbial signature to identify CD samples based on 16S data. We tested whether we could apply the same biomarkers on shotgun data. To do so, we selected the genera included in our proposed signature (*Faecalibacterium*, *Anaerostipes*, *Collinsella*, *Escherichia*, *Fusobacterium,* unknown Peptostreptococcaceae and *Methanobrevibacter* (**Figure 34 A**)) in the genus abundance table from the shotgun data. Notice that one of the genera included in the signature was not detected in shotgun data (unknown Christensenellaceae) (**Figure 34 B**).

We were able to detect only 20 out of the 51 CD samples showing a sensitivity of 39.22% whereas we identified as CD a total of 20 UC and HC samples out of a total of 127 showing a specificity of 84.25% (**Table 19**). This result showed that the biomarkers proposed for 16S rRNA gene was not valid for shotgun data. Five of those genera were detected significantly different also in shotgun sequencing between the three groups (**Figure 29**), however, unknown Peptostreptococcaceae and *Anaerostipes* were not in different abundance.

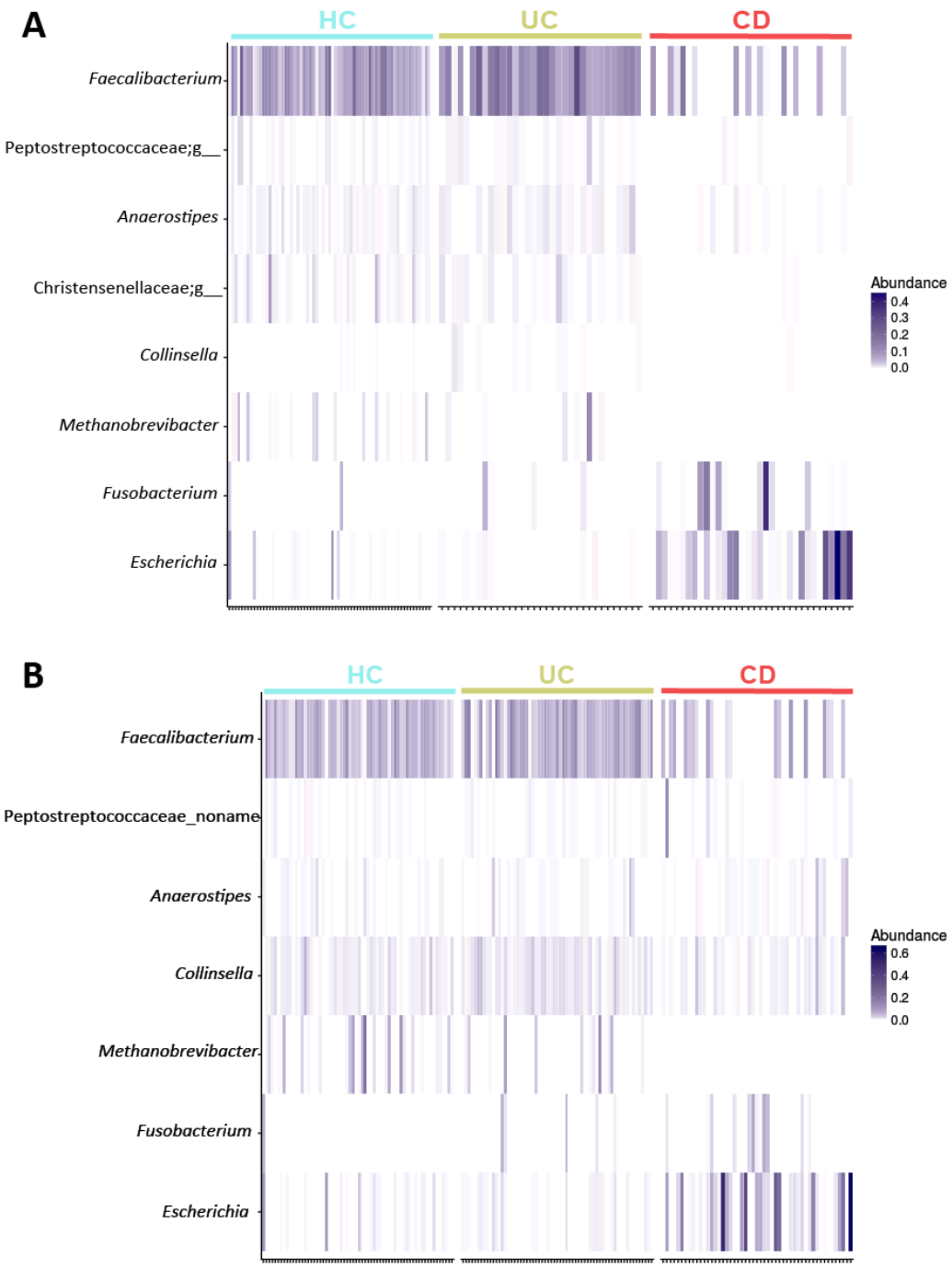


Figure 34. Microbial biomarker proposed to detect CD samples. Heatmaps that represents the relative abundance of genera implicated in the microbial marker we proposed for 16S rDNA data (A) applied in shotgun data (B). *Faecalibacterium, Collinsella* and *Methanobrevibacter* are more frequent in HC and UC than in CD whereas *Fusobacterium* and *Escherichia* have higher abundances in CD.

Table 19. Specificity and sensitivity of the 16S rDNA microbial marker on shotgun data.

		Conditions		
		CD	non-CD	Total
Predicted	CD	20 **(39.22%)**	20	40
	non-CD	31	107 **(84.25%)**	138
	Total	51	127	178

3.6 IBD classifier in shotgun sequencing

As we did with 16S rDNA sequencing data, we tried to improve the performance of the microbial marker with machine learning techniques. We trained three different methods: Random Forest (RF), AdaBoost and XGBoost with the 178 samples analyzed with shotgun sequencing from our Spanish IBD cohort. Then, we validated it with an independently cohort recently published dataset from American and Dutch IBD patients and healthy individuals. As it happened for 16S rDNA sequencing, the method that best performed to classify CD patients versus non-CD was the RF with an AUC of 0.93 (**Figure 35 A**). Our 5-cross validation resulted in a sensitivity of 72.5% and a specificity of 95.2% (**Table 20**). The independent validation provided an AUC of 0.75 (**Figure 35 B**) with sensitivity and specificity of 63.6% and 87.1% respectively (**Table 21**). In this classifier, among all the features (species) used, *Faecalibacterium prausnitzii* was the main bacteria involved in the classification in CD and non-CD samples.

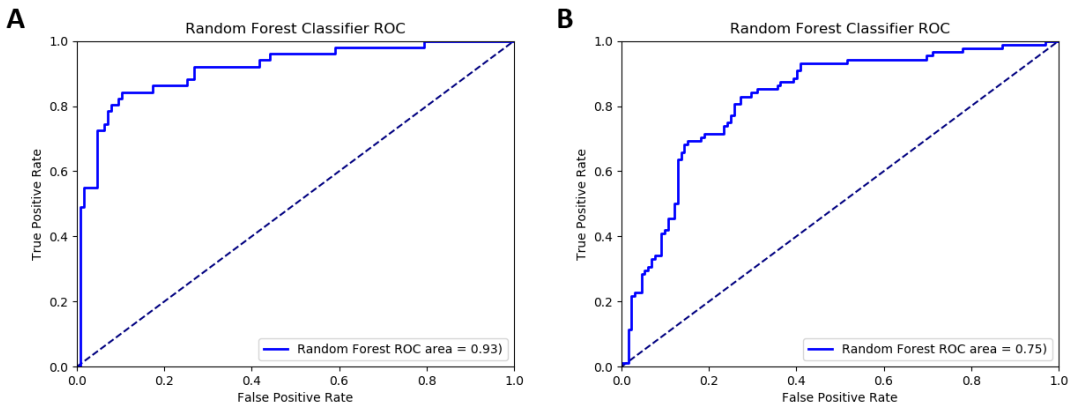


Figure 35. ROC curves for five-cross-validation (A) and independent validation (B) of Random Forest classifier in shotgun sequencing.

Table 20. RF classifier for predicting CD (five-cross-validation) for shotgun sequencing.

		Conditions		
		CD	non-CD	Total
Predicted	CD	37 **(72.5%)**	6	43
	non-CD	14	121 **(95.2%)**	135
	Total	51	127	178

Table 21. RF classifier for predicting CD (independent validation) for shotgun sequencing.

		Conditions		
		CD	non-CD	Total
Predicted	CD	56 **(63.6%)**	17	73
	non-CD	32	115 **(87.1%)**	147
	Total	88	132	220

3.7 Functional alterations in IBD

Next, we use the shotgun data to investigate the functional differences between IBD subtypes. In this step, two samples failed quality control, so we worked with 176 samples. Using the HUMAnN2 pipeline, we obtained an abundance table of genes per sample with annotation using the UniRef90 database, a functional database (see the **Methods** section page 53). We identified a mean of 32.72%±0.08 of unmapped reads. We computed Bray-Curtis distances to determine whether bacterial genes could also differentiate between CD, UC and HC. We determined that, as we described with taxonomical composition, genes can also differentiate between HC, UC and CD. We corroborated our previous findings identifying significant differences using the NPMANOVA test (FDR = 0.0015) between CD and HC, CD and UC but not between UC and HC (**Figure 36 A**).

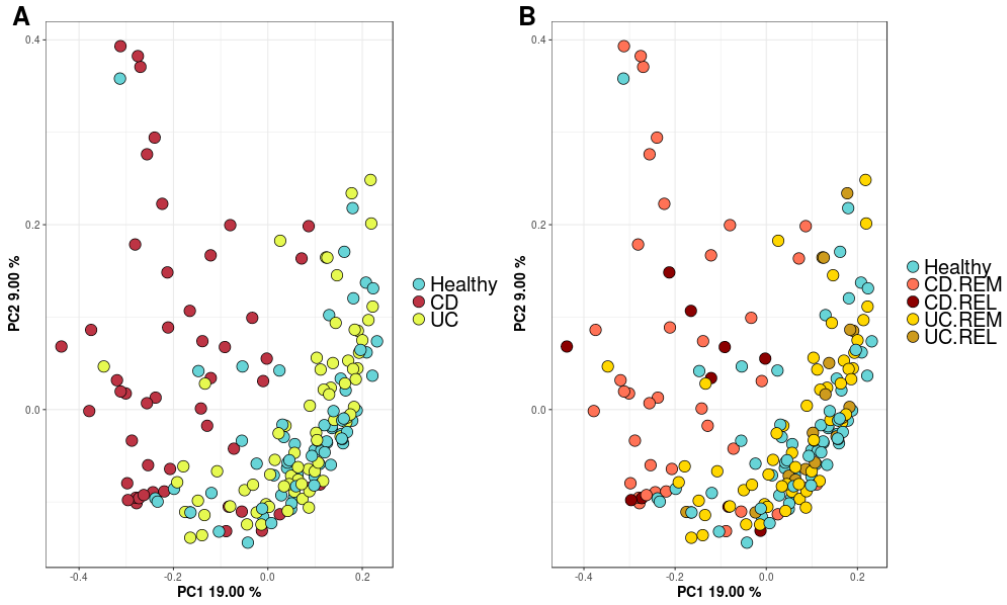


Figure 36. PCoA representation of Bray-Curtis distances in UniRef90 genes abundance table. (A) We found significant differences with genes between HC and CD and between UC and CD (NPMANOVA test, FDR = 0.0015) but not between UC and HC (FDR = 0.564). (B) As identified for taxonomical composition tests, we did not find significant differences between remission and relapse status for any of the subtypes of IBD.

We assessed microbiome changes according to disease activity (remission versus recurrence) using the Bray-Curtis distances and the NPMANOVA test and we did not find any significant differences between remission and recurrence in the disease for any of the IBD subtypes (**Figure 36 B**).

Counting the number of non-redundant detected genes has been shown to be useful for the investigation of microbiome alteration associated with obesity (Le Chatelier et al. 2013), where low gene count was linked to obesity. Thus, we determined the distribution of gene counts per group of study, HC, UC and CD. We saw that healthy and UC patient samples had more genes identified than CD patients (**Figure 37**). This result is in line with the results of alpha diversity analyses that showed higher diversity for UC and healthy than for CD patients.

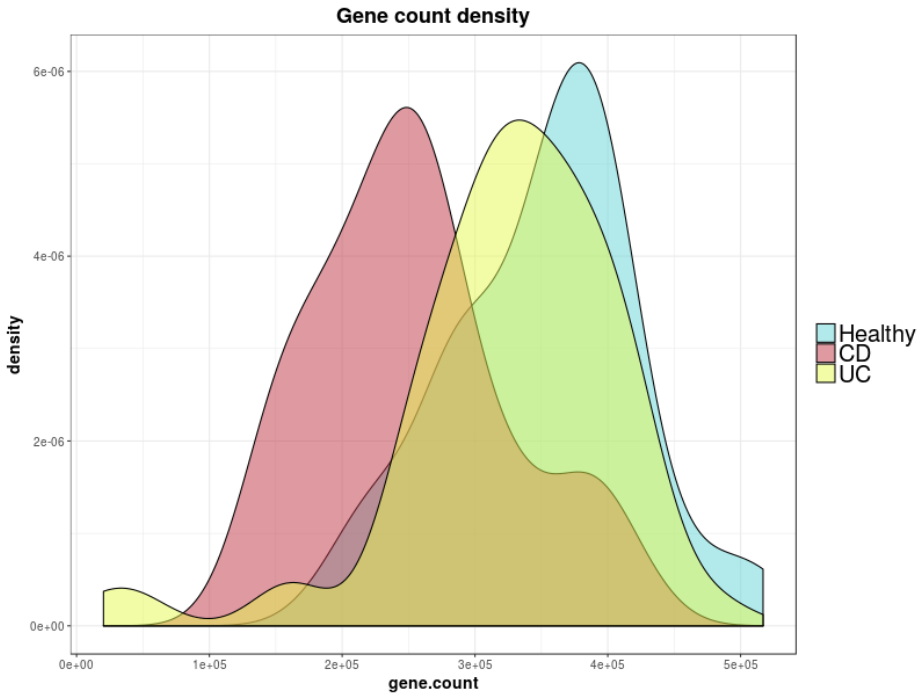


Figure 37. Gene count distribution of HC, UC and CD samples. CD samples had a smaller number of genes identified with UniRef90 database than UC and healthy individuals.

We annotated the UniRef90 identified clusters with KEGG Orthologous ids (KO ids) and functional categories. Among the 67.20% of reads that mapped with genes, we were only able to annotate a mean of 10.86% reads with KO ids that constitute a mean of

7.31% of the total number of reads. With this amount of information, we individually analyzed different levels of functional categories between HC, UC and CD with Kruskal-Wallis test. Reads that did not mapped with any gene accounted for the most significant differences between groups (FDR = 1.75 x 10^{-6}). They were more present in healthy and UC than in CD patients. This result may be related to the higher number of genes and greater diversity in those groups. On the one hand, we did not identify any functional category significantly different between HC and UC. On the other hand, genes associated with human diseases, metabolism, organismal systems, genetic information processing, and environmental information processing were significantly different between HC and CD and between CD and UC, except for genetic and environmental information processing for the latter comparison (**Figure 38**). Deeper levels of functional categories showed a gain of 25 functional groups out of 48 in CD in comparison with HC, and of 24 groups in comparison with UC (**Figure 39**). Most of the groups related to metabolism were affected in CD, only lipid, xenobiotics and nucleotide metabolism were not altered in CD.

We then assessed differences in functional categories between remission and relapse status of both IBD subtypes, UC and CD. We did not find significant differences for any functional category between both groups of disease activity.

We analyzed differences in pathways between CD, UC and HC groups with samples from basal timepoint. For this purpose, HUMAnN2 generated pathway abundance tables annotated with the MetaCyc database. We identified 180 pathways that were significantly different between the three groups (Kruskal-Wallis test, FDR < 0.05). Among these 180 differentially abundant pathways, 173 differed between HC and CD, 152 between CD and UC and only 11 were different between UC and HC (**Table 22**).

Different bacteria contribute to these pathways. In some cases, just one species is responsible for the specific pathway, however, in other pathways, several bacteria are involved. HUMAnN2 stratifies the contribution of each species identified to each pathway. With this information we observed that when the pathway was more frequent in CD, the main contributor to this pathway was *Escherichia coli* whereas if the pathway was more present in UC or HC, the main contributor was *Faecalibacterium prausnitzii* (**Figure 40**). There were other species also implicated in the differential presence of

pathways, but it is important to comment that a high percentage of the reads were associated to unclassified bacteria.

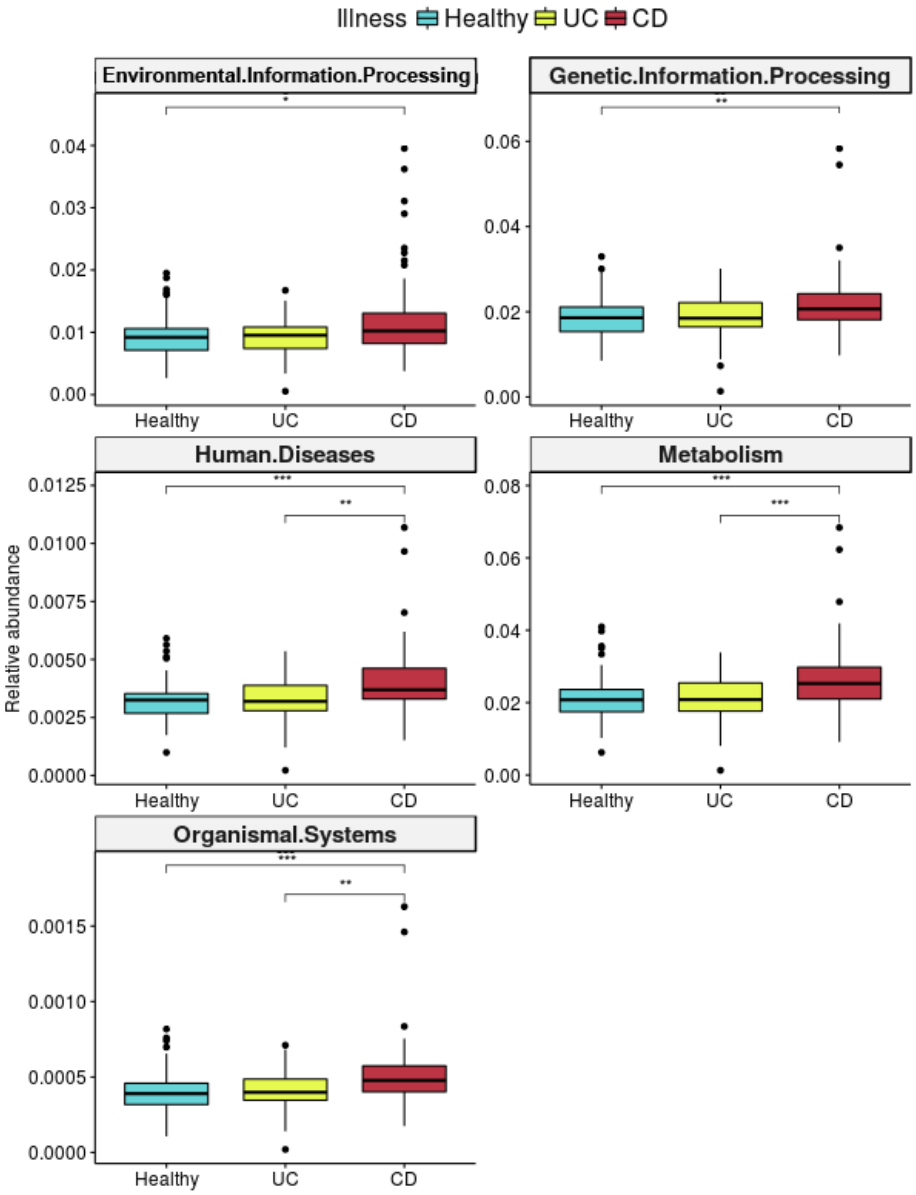


Figure 38. Most general level of functional categories differentially significant between UC, CD and HC. Significant differences were found between CD patients and UC and HC individuals (Kruskal-Wallis test, FDR < 0.05). Only genetic and environmental information processing were not significantly different between UC and CD.

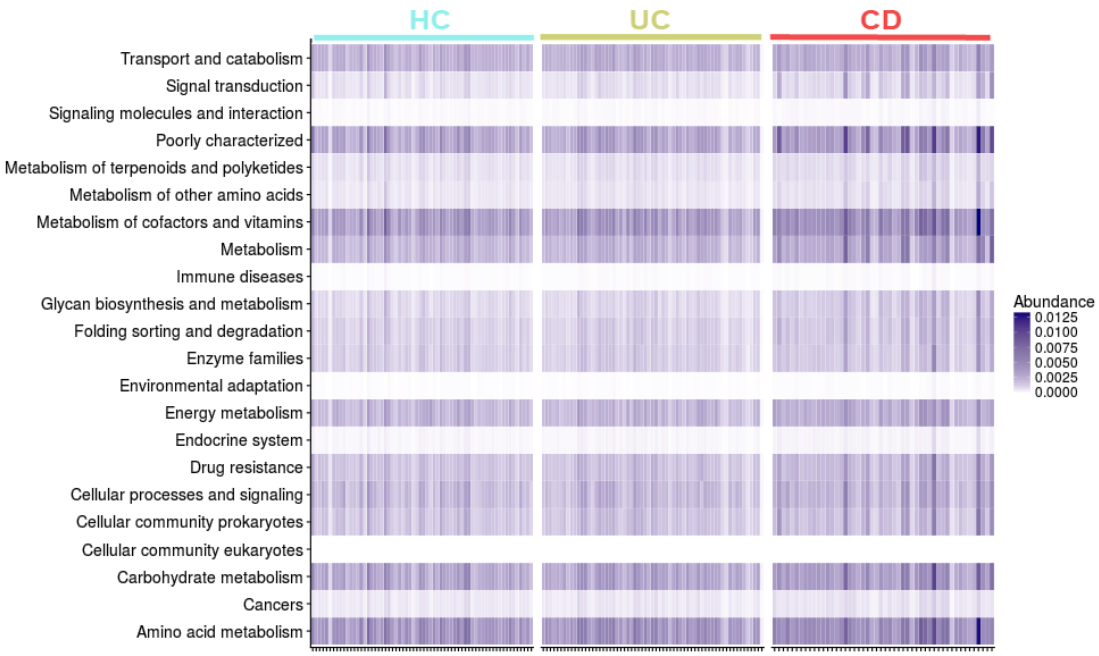


Figure 39. Functional categories significantly different between HC, UC and CD. Kruskal-Wallis test (FDR < 0.05) identified significant differences between CD and UC and CD and HC. No significant differences were identified between UC and HC for the 7.31% of reads that were assigned to KO ids.

Table 22. Significant pathways between HC, UC and CD. Kruskal-Wallis test, 21 lower FDR shown. FDR column corresponds to FDR correction for Kruskal-Wallis test between the three groups. Values represent the difference in abundance for each pathway between both groups. In red, pathways enriched in CD, in blue, pathways enriched in HC and in yellow, pathways enriched in UC. Blank spaces indicate that those pathwats are not significantly different in that comparison.

	FDR	CD vs HC	CD vs UC	UC vs HC
PWY 6590: superpathway of Clostridium acetobutylicum acidogenic fermentation	3,90744E-07	-1,99557E-05	-1,03473E-05	-9,60837E-06
CENTFERM PWY: pyruvate fermentation to butanoate	3,90744E-07	-1,63419E-05	-8,28179E-06	-8,06015E-06
ORNDEG PWY: superpathway of ornithine degradation	1,13339E-05	1,45146E-05	1,56961E-05	
PWY 5177: glutaryl CoA degradation	5,18047E-05	-3,76187E-05	-4,55392E-05	
ARGDEG PWY: superpathway of L arginine: putrescine: and 4 aminobutanoate degradation	9,72264E-05	1,48189E-05	1,59594E-05	
ORNARGDEG PWY: superpathway of L arginine and L ornithine degradation	9,72264E-05	1,48189E-05	1,59594E-05	
PWY 6891: thiazole biosynthesis II: Bacillus	0,000109226	7,14261E-06	7,73566E-06	
AST PWY: L arginine degradation II: AST pathway	0,000109226	1,42823E-05	1,5198E-05	
PWY 5138: unsaturated: even numbered fatty acid: beta: oxidation	0,000109226	3,55989E-05	3,91158E-05	
PWY 6895: superpathway of thiamin diphosphate biosynthesis II	0,000126441	5,3217E-05	6,29532E-05	
PWY 7315: dTDP N acetylthomosamine biosynthesis	0,000177144	-9,74453E-06	-1,13916E-05	
PWY 241: C4 photosynthetic carbon assimilation cycle: NADP ME type	0,000177144	1,44685E-05	1,56279E-05	
PWY0 1338: polymyxin resistance	0,000177144	3,64182E-05	3,93648E-05	
PWY4FS 7: phosphatidylglycerol biosynthesis I: plastidic	0,000178079	6,76858E-05	5,68549E-05	
PWY4FS 8: phosphatidylglycerol biosynthesis II: non plastidic	0,000178079	6,76856E-05	5,68555E-05	
ENTBACSYN PWY: enterobactin biosynthesis	0,000197023	4,65383E-05	5,05745E-05	
PWY0 1533: methylphosphonate degradation I	0,000271262	2,00153E-05	2,14029E-05	
PWY 2723: trehalose degradation V	0,000313081	-4,03135E-05	-4,99447E-05	
PWY 7456: mannan degradation	0,000313081	1,99064E-05	2,13335E-05	
ECASYN PWY: enterobacterial common antigen biosynthesis	0,000313081	1,98453E-05	2,15688E-05	
PWY 7409: phospholipid remodeling: phosphatidylethanolamine: yeast	0,000313081	2,26582E-05	2,40709E-05	

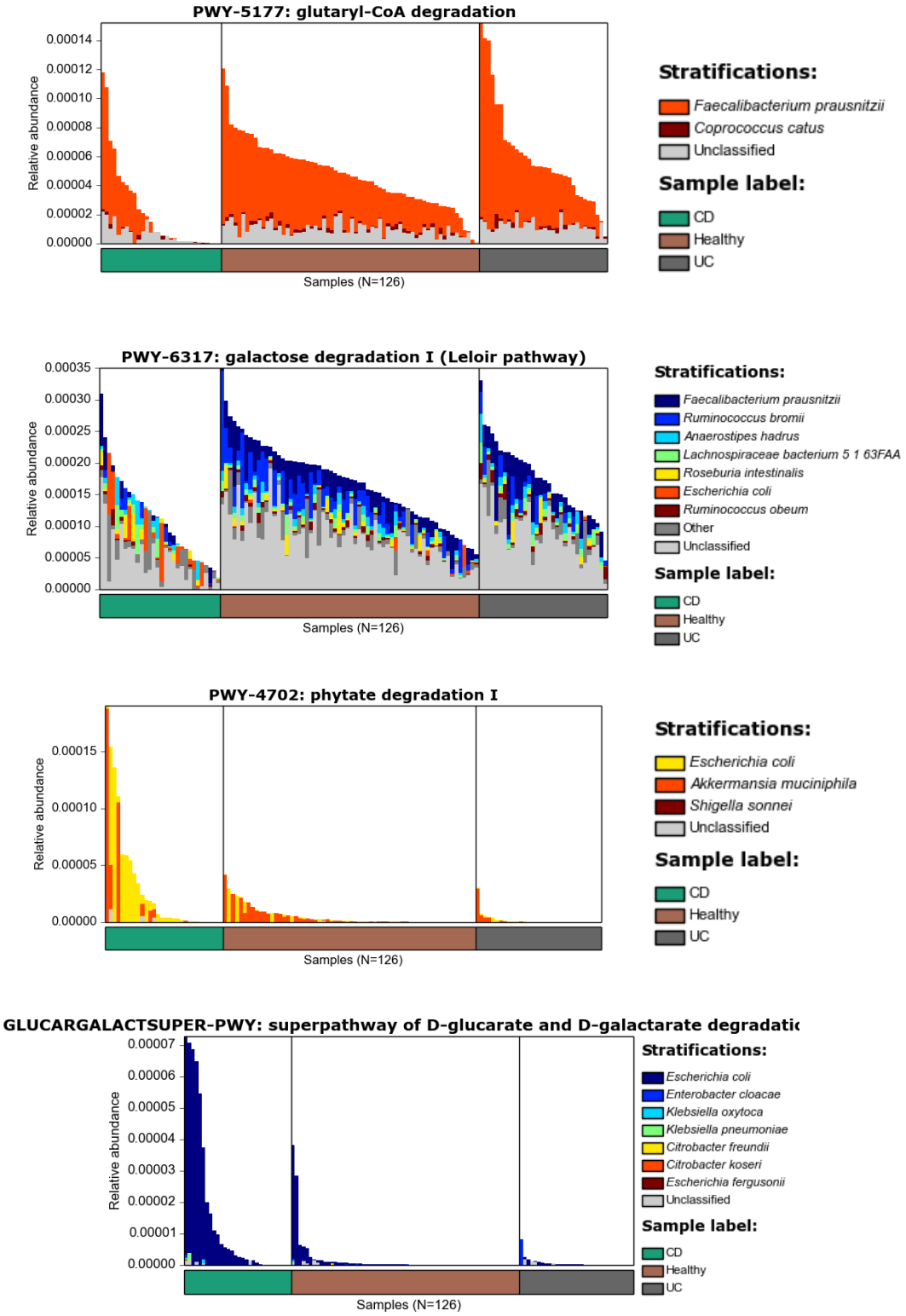


Figure 40. Examples of significantly abundant pathways between HC, UC and CD and their species stratification. The first two figures represent pathways that were more frequent in UC and HC, the main driver of these pathways is *F. prausnitzii* whereas in the last two figures, the main driver is *E. coli* and are more present in CD than in UC or HC. Phytate degradation pathway also shows the difference between HC and UC mainly driven by *A. muciniphila*.

DISCUSSION

This doctoral thesis represents one of the biggest studies of IBD patients that combines 16S rDNA and shotgun sequencing data. In the following lines, I will discuss the main topics addressed that include: microbial dysbiosis in IBD, microbiota as a diagnostic tool and the comparison between 16S and shotgun sequencing and functional analysis of the microbiome in IBD.

Dysbiosis in IBD

In this study, we reported IBD-associated changes in fecal samples, confirming, clarifying and complementing previous studies. We confirmed dysbiosis in IBD samples and we found differences between IBD subtypes. Although UC and CD share many epidemiologic, immunologic, therapeutic and clinical features and have been studied together along the years, we have confirmed that they are two distinct subtypes at the microbiome level as demonstrated in other studies in IBD adult cohorts (Andoh et al. 2011; Forbes et al. 2018).

Our results showed that CD had lower diversity than UC and healthy whereas UC presented similar alpha diversity than healthy individuals, which agrees with previous studies in adult and some pediatric cohorts (Forbes et al. 2018; Hansen, Richard K Russell, et al. 2012; Shah et al. 2016). However, the difference between IBD subtypes is still a subject of debate, as recent pediatric studies have found few differences between UC and CD and consider both subtypes of IBD as a unique disease (Knoll et al. 2016; Malham et al. 2019; Michail et al. 2012).

We found ten genera consistent between 16S rDNA and shotgun sequencing to suffer changes in CD with respect to healthy. Our findings suggest that the microbiome shifts in CD are more associated with a loss of commensal symbionts (beneficial microorganisms) than with a gain of pathobionts. Beneficial microorganisms in the IBD context include butyrate producing bacteria, such as *Faecalibacterium* (*F. praunitzii*) (Khan et al. 2012), *Methanobrevibacter* (*M. smithii*), *Butyricicoccus* (*B. pullicaecorum*) or *Coprococcus*; many species of *Ruminococcus* (except for *Ruminococcus gnavus* that was enriched in CD patients) that participate in the degradation of cellulose; *Aldercreutzia* (*A. equolifaciens*), a isoflavone degrader that produces equo, a substance with reported

beneficial properties (Sheflin et al. 2017); and *Collinsella* (*C. aerofaciens*), a carbohydrate degrader. All these genera may interact with the gut immune system to maintain homeostasis. Our findings replicate previous results regarding butyrate-producing bacteria (Eeckhaut et al. 2013; Takahashi et al. 2016) and *Ruminococcus* (Hall et al. 2017; Takahashi et al. 2016) species. Pathobionts or potential pathogenic microorganisms include *Fusobacterium (F. nucleatum)* or *Escherichia (E. coli)*. *Fusobacterium* has been associated with infections (Huggan and Murdoch 2008) and colorectal cancer (Kostic et al. 2012; Leung, Tsoi, and Yu 2015) whereas *Escherichia* has been related to IBD (Darfeuille-Michaud et al. 2004; Wright et al. 2015). Moreover, shotgun sequencing identified *Saccharomyces cerevisiae* (fungi) as more frequent in CD than in healthy controls, in accordance with previous microbiome studies based on antibodies (Vermeire et al. 2001). This finding contradicts previous research by Sokol and colleagues who showed a decrease of *S. cerevisiae* in CD patients in comparison with healthy and UC individuals using quantitative real-time PCR (qPCR) (Sokol et al. 2017). Therefore, further studies are needed to elucidate the effects of IBD on fungi.

Our analyses revealed that dysbiosis in UC was much lower than in CD patients. The relative abundance of only three genera were consistently altered in UC in both technical approaches. Dysbiosis in UC is probably driven, by the species *Akkermansia muciniphila*, more frequent in healthy individuals and *Peptostreptococcus anaerobius* and two species of *Parvimonas*, in higher abundance in UC patients, as determined by shotgun sequencing. Our finding on the alteration of *Akkermansia muciniphila* agrees with previous studies which also showed higher abundances of this species in healthy individuals (Malham et al. 2019; Shah et al. 2016). This bacterium works as a mucin degrader (Derrien et al. 2004) in the human gut and has been widely studied. It constitutes one of the main symbionts that live in the human gut increasing the function of the intestinal barrier (de Vos 2017). Plovier and colleagues showed that *A. muciniphila* contributed to the recovery of health status in obese and diabetic mice, showing its capacity for indicating a health metabolism (Plovier et al. 2017). In accordance with our study, Masoodi and colleagues described an increase of *Parvimonas* together with a decrease of Verrucomicrobiales (*A. muciniphila* phylum) in UC patients (Masoodi et al. 2019). Finally, *Peptostreptococcus anaerobius* has been recurrently found more

frequent in patients with UC and colorectal cancer which agrees with our findings (Fite et al. 2013; Furrie et al. 2004; Yang and Jobin 2017).

Alterations in microbial composition between CD and UC follow a similar pattern than between healthy individuals and CD: CD is depleted of commensal symbionts (*Faecalibacterium*, *Copropococcus*, *Ruminococcus*, *Collinsella* and *Aldercreuitzia*) and enriched of few pathobionts (*Escherichia* and *Fusobacterium*) in comparison with UC. The overall results confirm higher dysbiosis in CD than UC and similar microbial composition between UC patients and healthy individuals with small shifts in microbial composition as previously described (Gophna et al. 2006).

Besides the identified common bacteria between both methodologies, other microorganisms had different frequencies between the three groups of the study (HC, UC and CD) in one of the techniques. For example, shotgun data revealed a lower abundance of *Eubacterium* and *Subdoligranulum* in CD samples compared to UC and HC, in agreement with previous published works (Kaakoush et al. 2012a; Takahashi et al. 2016). These genera were not either differentially detected nor even detected, respectively, by 16S rDNA data. Shotgun sequencing unveiled *Bifidobacterium bifidum* as being more frequent in UC patients in comparison with healthy subjects, however, previous studies detected another species of *Bifidobacterium*, *Bifidobacterium breve,* in higher abundance in UC patients (Franzosa et al. 2019). This result could suggest that *Bifidobacterium* genus are altered in UC (Forbes et al. 2018) although we did not find this difference with 16S rDNA sequencing. Therefore, further studies are required to clarify these findings. In the case of 16S rDNA data, for which we only worked at genus taxonomical level due to the lack of accuracy for the use of 16S sequences at the species level, we observed that *Prevotella* and *Oscillospira* were enriched in HC and UC compared to CD in agreement with previous studies (Douglas et al. 2018; Kaakoush et al. 2012b; Masoodi et al. 2019). However, these genera were not differentially detected nor even detected, respectively, using shotgun data.

Functional profiling with KEGG modules of IBD stool samples confirmed greater dysbiosis in CD than UC in comparison with healthy individuals. Genes implicated in this dysbiosis are mainly associated with a diverse variety of metabolic pathways (such as carbohydrates, amino acids or terpenoids), cellular processes and immune diseases.

Such genes were enriched in CD samples in comparison with healthy subjects or UC patients. Changes in metabolism of cysteine and nucleotides in IBD have been already described showing more abundance in UC and CD than in healthy individuals (Morgan et al. 2012), however, differences between subtypes of IBD, have not been proposed yet. Very few pathways were altered between UC and healthy individuals, this was not the case of CD what again confirms the higher dysbiosis in CD. Interestingly, the annotation of genes with the MetaCyc pathways database allowed us to detect higher abundance of genes involved in metabolic pathways driven by *E. coli* in CD compared to healthy controls and UC, whereas, genes involved in pathways driven by *F. prausnitzii* were more abundant in healthy subjects and UC compared to CD. This is the case of the glutaryl-CoA degradation pathway (**Figure 40**). Glutaryl-CoA is a molecule implicated in the pathway from glutarate to butyrate. A lack of genes able to degrade this molecule, will end in a reduction of butyrate production (Vital, Howe, and Tiedje 2014). This result confirms our results suggesting a lack of butyrate-producing bacteria in CD patients. On the other hand, *E. coli* mainly drives the superpathway of D-glucarate and D-galactarate that is more present in CD patients than in UC and healthy subjects. This pathway is part of the carbohydrate metabolism and as hypothesized by Miele and colleagues, IBD patients poorly absorb polysaccharides which may contribute to increase of specific bacteria, such as *E.coli*, producing malabsorption in the intestine. In this regards, high carbohydrate diets are restricted for IBD patients to avoid the overgrowth of these type of bacteria (Miele et al. 2018). Among the pathways altered in UC patients in comparison with healthy individuals, only one (phytate degradation) out of the eleven detected, was driven by differences in *A. muciniphila*, the main species altered in UC patients. The rest of the pathways were associated to unclassified bacteria so further analyses are required to figure out which are the species implicated in the differences between healthy individuals and UC patients.

The main limitations of functional and pathway profiling are the lack of annotation of genes leading to a possible bias in the interpretation of the results (Li et al. 2014). In our study, the information given is based only on 7% of the reads that could be converted from genes to functions and pathways, so a great majority of reads could not be mapped to a known gene. Moreover, the vast majority of genes were not assigned to known

metabolic pathways (41.7% of the total annotated functions), and so the relative abundance of pathway-associated functions was quite small, biasing even more the analysis at the metabolic level (Li et al. 2014). Therefore, interpretation and conclusion on any results of the analyses should be taken with caution. Future work should focus on incrementing the functional annotation of genes in these databases.

Dysbiosis described in this study could be associated with the differences in the level of inflammation that characterize each IBD subtype. Stronger alterations of microbial composition in CD could be explained by the more severe degree of a transmural inflammation detected in CD patients in comparison with UC in which the inflammation only affects the mucosa. Finally, both microbiome composition and functional analyses pointed out to a greater dysbiosis in CD compared to healthy subjects and UC patients.

Microbiota as a diagnostic tool

Differences in microbial composition between IBD subtypes opens the door to use microbiota as a diagnostic tool for IBD patients. Based on this idea, Guo and colleges proposed a combination of *F. prausnitzii* and *F. nucleatum* as a microbial biomarker to diagnose CD patients (Guo et al. 2019). We propose a more complex combination of microbial species based on absence or presence of eight genera to differentiate CD from non-CD using fecal samples. Our algorithm showed a good performance in the discovery cohort, specially discriminating non-IBD samples. The evaluation of the effectiveness of the microbial marker classifying CD in the independent validation cohort worsened, but it was more efficient identifying non-CD samples. These findings determine that our algorithm is very precise when correctly discarding non-CD samples. We also tested the performance of our microbial marker using the French IBD cohort. In this case, we obtained lower values of sensitivity and specificity.

The slightly lower sensitivity of the Belgian cohort could be influenced by the use of all samples in the algorithm, independently of time, and by the different conditions of this cohort in comparison with the discovery cohort: Belgian CD patients suffer a more severe status of the disease and were subjected to an intervention during sample collection. Moreover, in the Belgium cohort, all the samples with CD who took

antibiotics were detected by the algorithm, suggesting that antibiotics intake prior to sampling did not affect detection by the algorithm. The worse performance of the microbial marker in the French cohort could be explained by a difference in the methodological approach or by the geographical difference between cohorts. The low accuracy obtained with the French cohort may point to a limitation of this method as a diagnostic tool, as the use of the microbial biomarker we are proposing could be restricted to the use of a similar methodology for microbial analyses (the V4 fragment of 16S rRNA gene with Illumina MiSeq). In summary, the use of the microbial marker is subjected to the cohort heterogeneity and depends on microbial extraction and sequencing approaches.

These results suggest that the microbial marker proposed for IBD is an easy-to-use mechanism to classify CD versus non-CD of 16S rDNA (V4 fragment) sequencing fecal samples from Illumina MiSeq. However, further experimental designs could be proposed to evaluate the extent to which the method used here could be implemented in a laboratory.

To evaluate another molecular technique to classify CD samples, we tested the performance of our microbial signature in samples analyzed with shotgun data. We obtained a sensitivity of 39% and a specificity of 84%, indicating that an algorithm proposed for 16S rDNA sequencing may not be appropriate for shotgun sequencing. A possible reason is that each technique detects a different subset of genera. For instance, unknown genus of Christensenellaceae, a genus detected in 16S and included in the microbial marker, was not even detected in shotgun sequencing.

To improve the performance of the classification, we developed two new classifiers using machine learning: one for 16S rDNA sequencing and another for shotgun sequencing. Our classifiers clearly improved the sensitivity and specificity of the biomarker providing a good diagnostic tool for CD. The 16S rDNA classifier performed even better when applying in other cohorts so this classifier could be tested to be published as a diagnostic tool. The shotgun classifier showed lower performance when applied in other cohorts, probably due to geographical differences, as suggested in our comparison between the Spanish IBD cohort and the American and Dutch IBD cohorts. Moreover, the lower performance of shotgun data classification may be consequence

of overfitting due to the low number of samples, reducing accuracy when testing other cohorts. Previous studies by Papa and colleagues, provided the SLiME classifier that was trained on an IBD pediatric 16S rDNA dataset and was able to separate IBD samples from non-IBD samples (Papa et al. 2012). Franzosa and colleagues developed a classifier for IBD for shotgun sequencing with random forest model. Their random forest model provided similar accuracy than our tool and their classifier was able to differentiate between UC, CD and healthy individuals depending on microbiome composition and/or metabolites (Franzosa et al. 2019). Franzosa's classifier also worked worse with their Dutch validation cohort, supporting cohort heterogeneity as a limitation of IBD classifiers.

Microbial changes in basal samples from patients that will later develop a recurrence of the disease could be used to create an algorithm that predicts a recurrence in time. Sokol and colleagues described differences in microbiome between active (recurrence) and non-active (remission) IBD with less load of *F. prausnitzii* in recurrence fecal samples (Sokol et al. 2009). In this line, other studies proposed differences between both status of the disease with lower proportions of *Subdoligranulum* or *Butyricicoccus* in active IBD (Papa et al. 2012). However, we did not find differences between both groups of basal samples. Moreover, we did not detect variations in the gut microbiome with time and disease severity in any IBD subtype using either 16S rDNA or shotgun data. Therefore, we were not able to develop a recurrence predictor based on microbiota changes using our cohort and methodology. Future studies should investigate further other statistical approaches using machine learning tools to analyze these data such as a combination of Random Forest and C5.0, method to visualize the classification decision process.

The microbial biomarker and the machine learning classifier can lead to new non-invasive diagnostic tools which may be valuable to assess patients with non-specific signs and symptoms suggestive of IBD, thereby facilitating clinical decision-making when the diagnosis of CD is initially uncertain. Indeed, these tools could be combined with either imaging techniques or calprotectin data to confirm diagnosis. Moreover, situations in which the diagnosis is uncertain between UC and CD (localized in the colon) even with colonoscopy, these tools could help to elucidate whether the sample corresponds to one or another subtype of IBD.

Comparison of 16S and shotgun sequencing

Nowadays, many investigators are sceptic about the use of 16S rDNA. This thesis can provide guidelines to choose between 16S rDNA sequencing and shotgun sequencing, an unresolved topic in metagenomic studies. We compared microbial taxonomic assignment from phylum to genera between both techniques and determined that although only 17.5% of genera detected were common between both methods, however, these common genera constituted more than 70% of the bacterial abundance of reads assigned. Moreover, abundances of common bacteria in both methodologies were highly correlated. The main microbial species implicated in the differences between CD, UC and healthy subjects that have been widely described, corresponded to bacteria that were detected in both techniques.

Both methodologies can be used indistinctly and efficiently to distinguish CD from non-CD, demonstrating that both techniques may be adequate diagnosis tests. However, the cost of 16S rDNA sequencing is much lower than the cost of shotgun sequencing and the analyses are less time-consuming so, if the effectiveness of the diagnostic tool is similar, 16S rDNA will allow the analyses of much more samples in less time and with lower cost. Nevertheless, if the objective of the study focuses on detecting a determined species or function, shotgun sequencing would be more appropriate. Moreover, if researchers are searching for fungus or viruses, shotgun sequencing should be chosen as it allows the detection of these microorganisms, whereas 16S rDNA sequencing is specific for bacteria.

The characterization of dysbiosis in IBD has shown that although main bacteria implicated in differences between groups are common between shotgun and 16S data, other microorganisms were only identified by one or the other technique. At this point, we need to deal with two type of results: bacteria detected only in one technique and bacteria that was significantly different in one technique but not in the other. The first case could be explained by differences in databases such as annotation or lack of determined species whereas the second problem suggests that working with only one gene instead of working with more than one marker gene as it is the case of MetaPhlAn for taxonomic assignation in shotgun sequencing, could produce differences in

abundances detected for each bacteria, especially for those that do not differ so much between groups.

Limitations and strengths

This study has several strengths but also limitations. We have worked with a large IBD cohort comparing two different approaches, shotgun and 16S rDNA sequencing. Until now, studies combining shotgun and 16S rDNA sequencing techniques have focused their efforts on finding compositional dysbiosis with 16S rDNA and functional dysbiosis with shotgun sequencing, however, to our knowledge, studies have not provided the comparison of both techniques.

Although we are working with a large cohort, we only analyzed CD patients with the disease localized in the ileum or ileo-caecal section of the gut. To complete this study, the cohort should include patients with CD localized only in colon to confirm that the results are extensive to all CD subtypes. Moreover, our cohort lacks IBD adult patients recently diagnosed, treatment free. Many of the previous studies of IBD are based on pediatric cohorts of recently diagnosed IBD patients, however, to the best of our knowledge, none has been performed in adults.

Validation cohorts used in this thesis by the 16S rDNA approach constitute another drawback. Only the French cohort is equivalent to our cohort as it contains healthy, UC and CD individuals. However, the samples were processed with a different technique than ours, adding another potential source of heterogeneity besides geography. The other cohorts used only included CD, UC or healthy individuals, so differences between patient's status is confounded by the cohort. This limitation arose from the lack of publicly available cohorts with similar characteristics to ours, which challenged the comparison of the results between studies. This problem demonstrates that uploading data and clear metadata is very important to improve research, because comparisons between cohorts can bring more knowledge about specific diseases. Not only, providing data is important, but also standardize techniques of analyses to avoid confounding factors as much as possible.

A general limitation of most microbial studies is the use of relative abundances of bacteria instead of considering absolute values associated with microbial load. Vandeputte and colleagues studied how significant differences changed between relative and absolute counts, comparing qPCR and flow cytometric loads, between healthy and CD individuals. They showed that *Bacteroides* was significantly different in relative counts but not in absolute values whereas *Prevotella* followed the inverse association (Vandeputte et al. 2017) suggesting that depending on the method used, interpretation and conclusions could highly differ.

CONCLUSIONS

The results of the present doctoral thesis, despite the limitations, offer valuable insights into the involvement of the microbiota in inflammatory bowel disease, leading to the following conclusions:

1. Alterations of gut microbiota are associated with IBD showing greater dysbiosis in Crohn's disease than in ulcerative colitis at the metagenomic level. Crohn's disease is characterized by a loss of beneficial bacteria such as *Faecalibacterium prausnitzii*, butyrate-producing bacteria, than by an increment of pathobionts such as *Escherichia coli* and *Fusobacterium nucleatum* whereas ulcerative colitis is more similar to healthy individuals with small shifts in the microbial community.

2. Microbial composition can distinguish between Crohn's disease and ulcerative colitis which could be used as a diagnostic tool in the future to avoid invasive techniques. However, the use of the microbial signature and classifiers proposed should be further evaluated on different techniques of fecal sample processing in order to standardize and generalize this approach.

3. Shotgun and 16S rDNA sequencing techniques provide similar results for classifying IBD subtypes, however, depending on the objective and the budget, one method may be more appropriate than the other. Shotgun sequencing appears to be more appropriate for microbial identification at strain or species levels and functional profiling whereas 16S rDNA data, a much cheaper technique, performs efficiently the IBD subtype classification.

4. Functional analyses in IBD provide similar conclusions than microbial compositional analyses for both subtypes of the disease. Shifts in genes and functions mainly occur in CD whereas UC is similar to healthy individuals. Those changes appear in pathways detected in bacteria mainly altered in CD and are associated principally to metabolism.

5. Functional profiling is still poorly characterized in all metagenomic analyses and we must be cautious about the interpretation of the results. There may be other

interesting between-group differences that are hidden in the proportion of unknown genes, which might reveal themselves in future re-analyses, strengthen by improved gene annotation.

6. Altogether, our study validates and clarifies previous works as for the alteration of the gut microbial community of IBD patients and provide new insights in the distinction between Crohn's disease and ulcerative colitis at the taxonomic and functional levels. Further studies will be needed to validate our findings and our microbial marker to apply it in clinical practice

FUTURE LINES

Future efforts in IBD should be focused on finding the main differences between inflammation status using other information sources. We did not find differences in microbial composition nor in genes composition, however, we are blind about the genes that were expressed at the time of sampling. Metatranscriptomics, proteomics and metabolomics could be used to elucidate genes, pathways and metabolites that could be influencing the inflammation in the gut.

To apply microbiome as a diagnostic tool as proposed in this thesis, both the microbial biomarker and the classifiers should be evaluated in additional cohorts, with other CD subtypes, disease localizations, newly diagnosed patients or different geography to test the accuracy of the diagnostic tool. Moreover, to try to overcome the geographical microbial differences, a classifier could be developed considering several cohorts from different precedencies to maximize differences caused by the disease.

Finally, together with fecal microbiome composition as a potential as diagnostic tool, as microbiota changes seem to play an important role in IBD, researchers may center they attention on finding a way to restore microbial composition on IBD patients. Fecal microbial transplantation has been studied in IBD patients, showing very modest response for both IBD subtypes, though may be more efficient in UC than in CD. This difference could be explained by the higher dysbiosis of CD patients in comparison with UC and the fact that UC patients share a high proportion of microbial composition with healthy individuals. Future investigations should focus on in improving FMT techniques for IBD patients and find a way to restore *Akkermansia muciniphila* proportion in UC patients or *Faecalibacterium prausnitzii* in CD patients. However, restoring only gut microbiota composition in IBD patients to ameliorate patient's symptoms may still be a great challenge since both the microbiome composition and the host immune system are pointed out to be the cause of the disease. Therefore, future study designs should include the evaluation of the effect of FMT combined with for instance anti-inflammatory drugs.

BIBLIOGRAPHY

Abraham, Bincy P., Tasneem Ahmed, and Tauseef Ali. 2017. “Inflammatory Bowel Disease : Pathophysiology and Current Therapeutic Approaches.”

Abubucker, Sahar, Nicola Segata, Johannes Goll, Alyxandria M. Schubert, Jacques Izard, Brandi L. Cantarel, Beltran Rodriguez-Mueller, Jeremy Zucker, Mathangi Thiagarajan, Bernard Henrissat, Owen White, Scott T. Kelley, Barbara Methé, Patrick D. Schloss, Dirk Gevers, Makedonka Mitreva, and Curtis Huttenhower. 2012. “Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome.” *PLoS Computational Biology* 8(6).

Allali, Imane, Jason W. Arnold, Jeffrey Roach, Maria Belen Cadenas, Natasha Butz, Hosni M. Hassan, Matthew Koci, Anne Ballou, Mary Mendoza, Rizwana Ali, and M. Andrea Azcarate-Peril. 2017. “A Comparison of Sequencing Platforms and Bioinformatics Pipelines for Compositional Analysis of the Gut Microbiome.” *BMC Microbiology* 17(1):194.

Ananthakrishnan, Ashwin N. 2015. “Epidemiology and Risk Factors for IBD.” *Nature Reviews Gastroenterology and Hepatology* 12(4):205–17.

Andoh, Akira, Hirotsugu Imaeda, Tomoki Aomatsu, Osamu Inatomi, Shigeki Bamba, Masaya Sasaki, Yasuharu Saito, Tomoyuki Tsujikawa, and Yoshihide Fujiyama. 2011. “Comparison of the Fecal Microbiota Profiles between Ulcerative Colitis and Crohn’s Disease Using Terminal Restriction Fragment Length Polymorphism Analysis.” *Journal of Gastroenterology* 46(4):479–86.

Aniwan, Satimai and Sang Hyoung Park. 2017. “Epidemiology, Natural History, and Risk Stratification of Crohn’s Disease.” *Gastroenterology Clinics of North America* 46(3):463–80.

Aniwan, Satimai, Sang Hyoung Park, and Edward V. Loftus. 2017. “Epidemiology, Natural History, and Risk Stratification of Crohn’s Disease.” *Gastroenterology Clinics of North America* 46(3):463–80.

Bäckhed, Fredrik, Ruth E. Ley, Justin L. Sonnenburg, Daniel A. Peterson, and Jeffrey I. Gordon. 2005. “Host-Bacterial Mutualism in the Human Intestine.” *Science* 307(5717):1915–20.

Bamias, Giorgos, Kazuhiko Sugawara, Cristiano Pagnini, and Fabio Cominelli. 2003. “The Th1 Immune Pathway as a Therapeutic Target in Crohn’s Disease.” *Current Opinion in Investigational Drugs (London, England : 2000)* 4(11):1279–86.

Barko, P. C., M. A. McMichael, K. S. Swanson, and D. A. Williams. 2018. "The Gastrointestinal Microbiome: A Review." *Journal of Veterinary Internal Medicine* 32(1):9–25.

Bartosch, Sabine, Alemu Fite, George T. Macfarlane, and Marion E. T. McMurdo. 2004. "Characterization of Bacterial Communities in Feces from Healthy Elderly Volunteers and Hospitalized Elderly Patients by Using Real-Time PCR and Effects of Antibiotic Treatment on the Fecal Microbiota." *Applied and Environmental Microbiology* 70(6):3575–81.

Baumgart, Martin, Belgin Dogan, Mark Rishniw, Gil Weitzman, Brian Bosworth, Rhonda Yantiss, Renato H. Orsi, Martin Wiedmann, Patrick McDonough, Sung Guk Kim, Douglas Berg, Ynte Schukken, Ellen Scherl, and Kenneth W. Simpson. 2007. "Culture Independent Analysis of Ileal Mucosa Reveals a Selective Increase in Invasive Escherichia Coli of Novel Phylogeny Relative to Depletion of Clostridiales in Crohn's Disease Involving the Ileum." *ISME Journal* 1(5):403–18.

Booijink, Carien C. G. M., Sahar El-Aidy, Mirjana Rajilić-Stojanović, Hans G. H. J. Heilig, Freddy J. Troost, Hauke Smidt, Michiel Kleerebezem, Willem M. De Vos, and Erwin G. Zoetendal. 2010. "High Temporal and Inter-Individual Variation Detected in the Human Ileal Microbiota." *Environmental Microbiology* 12(12):3213–27.

Braga, Raíssa Mesquita, Manuella Nóbrega Dourado, and Welington Luiz Araújo. 2016. "Microbial Interactions: Ecology in a Molecular Perspective." *Brazilian Journal of Microbiology : [Publication of the Brazilian Society for Microbiology]* 47 Suppl 1(Suppl 1):86–98.

Breiman, Leo. 2001. "Random Forests LEO." *Machine Learning*.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." *Nature Methods* 13(7):581–83.

Caporaso, J Gregory, Christian L. Lauber, William A. Walters, Donna Berg-Lyons, James Huntley, Noah Fierer, Sarah M. Owens, Jason Betley, Louise Fraser, Markus Bauer, Niall Gormley, Jack A. Gilbert, Geoff Smith, and Rob Knight. 2012. "Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms." *The ISME Journal* 6(8):1621–24.

Caporaso, J. Gregory, Christian L. Lauber, William A. Walters, Donna Berg-Lyons, James

Huntley, Noah Fierer, Sarah M. Owens, Jason Betley, Louise Fraser, Markus Bauer, Niall Gormley, Jack A. Gilbert, Geoff Smith, and Rob Knight. 2012. “Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms.” *ISME Journal* 6(8):1621–24.

Carabotti, Marilia, Annunziata Scirocco, Maria Antonietta Maselli, and Carola Severi. 2015. “The Gut-Brain Axis: Interactions between Enteric Microbiota, Central and Enteric Nervous Systems.” *Annals of Gastroenterology* 28(2):203–9.

Carding, S. R., N. Davis, and L. Hoyles. 2017. “Review Article: The Human Intestinal Virome in Health and Disease.” *Alimentary Pharmacology & Therapeutics* 46(9):800–815.

Carey, Christine M., Jennifer L. Kirk, Shivani Ojha, and Magdalena Kostrzynska. 2007. “Current and Future Uses of Real-Time Polymerase Chain Reaction and Microarrays in the Study of Intestinal Microbiota, and Probiotic Use and Effectiveness.” *Canadian Journal of Microbiology* 53(5):537–50.

Carstens, Adam, Annika Roos, Anna Andreasson, Anders Magnuson, Lars Agréus, Jonas Halfvarson, and Lars Engstrand. 2018. “Differential Clustering of Fecal and Mucosa-Associated Microbiota in ‘healthy’ Individuals.” *Journal of Digestive Diseases* 19(12):745–52.

Chao, Anne, Robin L. Chazdon, Robert K. Colwell, and Tsung-Jen Shen. 2006. “Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples.” *Biometrics* 62(2):361–71.

Le Chatelier, Emmanuelle, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, Manimozhiyan Arumugam, Jean Michel Batto, Sean Kennedy, Pierre Leonard, Junhua Li, Kristoffer Burgdorf, Niels Grarup, Torben Jørgensen, Ivan Brandslund, Henrik Bjørn Nielsen, Agnieszka S. Juncker, Marcelo Bertalan, Florence Levenez, Nicolas Pons, Simon Rasmussen, Shinichi Sunagawa, Julien Tap, Sebastian Tims, Erwin G. Zoetendal, Søren Brunak, Karine Clément, Joël Doré, Michiel Kleerebezem, Karsten Kristiansen, Pierre Renault, Thomas Sicheritz-Ponten, Willem M. De Vos, Jean Daniel Zucker, Jeroen Raes, Torben Hansen, Peer Bork, Jun Wang, S. Dusko Ehrlich, Oluf Pedersen, Eric Guedon, Christine Delorme, Séverine Layec, Ghalia Khaci, Maarten Van De Guchte, Gaetana Vandemeulebrouck, Alexandre Jamet, Rozenn Dervyn, Nicolas Sanchez,

Emmanuelle Maguin, Florence Haimet, Yohanan Winogradski, Antonella Cultrone, Marion Leclerc, Catherine Juste, Hervé Blottière, Eric Pelletier, Denis Lepaslier, François Artiguenave, Thomas Bruls, Jean Weissenbach, Keith Turner, Julian Parkhill, Maria Antolin, Chaysavanh Manichanh, Francesc Casellas, Natalia Boruel, Encarna Varela, Antonio Torrejon, Francisco Guarner, Gérard Denariaz, Muriel Derrien, Johan E. T. Van Hylckama Vlieg, Patrick Veiga, Raish Oozeer, Jan Knol, Maria Rescigno, Christian Brechot, Christine M'Rini, Alexandre Mérieux, and Takuji Yamada. 2013. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." *Nature* 500(7464):541–46.

Chen, Tianqi and Carlos Guestrin. 2016. "XGBoost."

Chiba, Mitsuro, Kunio Nakane, and Masafumi Komatsu. 2019. "Westernized Diet Is the Most Ubiquitous Environmental Factor in Inflammatory Bowel Disease." *The Permanente Journal* 23.

Cho, Judy H. 2008. "The Genetics and Immunopathogenesis of Inflammatory Bowel Disease." *Nature Reviews Immunology* 8(6):458–66.

Clarke, Siobhan F., Eileen F. Murphy, Kanishka Nilaweera, Paul R. Ross, Fergus Shanahan, Paul W. O'Toole, and Paul D. Cotter. 2012. "The Gut Microbiota and Its Relationship to Diet and Obesity: New Insights." *Gut Microbes* 3(3):186–202.

Clooney, Adam G., Fiona Fouhy, Roy D. Sleator, Aisling O'Driscoll, Catherine Stanton, Paul D. Cotter, and Marcus J. Claesson. 2016. "Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis." *PLoS ONE* 11(2):1–16.

Cohen, Louis J., Judy H. Cho, Dirk Gevers, and Hiutung Chu. 2019. "Genetic Factors and the Intestinal Microbiome Guide Development of Microbe-Based Therapies for Inflammatory Bowel Diseases." *Gastroenterology*.

Cosnes, Jacques and Antoine Cortot. 2011. "Epidemiology and Natural History of Inflammatory Bowel Diseases." *YGAST* 140(6):1785-1794.e4.

Cryan, John F., Timothy G. Dinan, Claude Bernard, Ivan Pavlov, William Beaumont, William James, Carl Lange, and Even Charles. 2012. "Mind-Altering Microorganisms : The Impact of the Gut Microbiota on Brain and Behaviour of the Nineteenth Century through the Pioneering Work." 13.

D'hoe, Kevin, Stefan Vet, Karoline Faust, Frédéric Moens, Gwen Falony, Didier Gonze,

Verónica Lloréns-Rico, Lendert Gelens, Jan Danckaert, Luc De Vuyst, and Jeroen Raes. 2018. “Integrated Culturing, Modeling and Transcriptomics Uncovers Complex Interactions and Emergent Behavior in a Three-Species Synthetic Gut Community.” *ELife* 7.

Dalal, Sushila R. and Eugene B. Chang. 2014. “The Microbial Basis of Inflammatory Bowel Diseases.” *Journal of Clinical Investigation* 124(10):4190–96.

Darfeuille-Michaud, Arlette, Jérôme Boudeau, Philippe Bulois, Christel Neut, Anne-Lise Glasser, Nicolas Barnich, Marie-Agnès Bringer, Alexander Swidsinski, Laurent Beaugerie, and Jean-Frédéric Colombel. 2004. “High Prevalence of Adherent-Invasive Escherichia Coli Associated with Ileal Mucosa in Crohn’s Disease.” *Gastroenterology* 127(2):412–21.

Derrien, Muriel, Elaine E. Vaughan, Caroline M. Plugge, and Willem M. de Vos. 2004. “Akkermansia Muciniphila Gen. Nov., Sp. Nov., a Human Intestinal Mucin-Degrading Bacterium.” *International Journal of Systematic and Evolutionary Microbiology* 54(Pt 5):1469–76.

Dong, Xiaoli, Manuel Kleiner, Christine E. Sharp, Erin Thorson, Carmen Li, Dan Liu, and Marc Strous. 2017. “Fast and Simple Analysis of MiSeq Amplicon Sequencing Data with MetaAmp.” *Frontiers in Microbiology* 8:1461.

Douglas, Gavin M., Richard Hansen, Casey M. A. Jones, Katherine A. Dunn, André M. Comeau, Joseph P. Bielawski, Rachel Tayler, Emad M. El-Omar, Richard K. Russell, Georgina L. Hold, Morgan G. I. Langille, and Johan Van Limbergen. 2018. “Multi-Omics Differentially Classify Disease State and Treatment Outcome in Pediatric Crohn’s Disease.” *Microbiome* 6(1):13.

Durbán, Ana, Juan J. Abellán, Nuria Jiménez-Hernández, Marta Ponce, Julio Ponce, Teresa Sala, Giuseppe D’Auria, Amparo Latorre, and Andrés Moya. 2011. “Assessing Gut Microbial Diversity from Feces and Rectal Mucosa.” *Microbial Ecology* 61(1):123–33.

Eckburg, Paul B., Elisabeth M. Bik, Charles N. Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R. Gill, Karen E. Nelson, and David A. Relman. 2005. “Diversity of the Human Intestinal Microbial Flora.” *Science (New York, N.Y.)*.

Edgar, Robert C. 2010. “Search and Clustering Orders of Magnitude Faster than BLAST.” *Bioinformatics*.

Edgar, Robert C. 2013. “UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads.” *Nature Methods* 10(10):996–98.

Edgar, Robert C., Brian J. Haas, Jose C. Clemente, Christopher Quince, and Rob Knight. 2011. “UCHIME Improves Sensitivity and Speed of Chimera Detection.” *Bioinformatics*.

Eeckhaut, Venessa, Kathleen Machiels, Clémentine Perrier, Carlos Romero, Sofie Maes, Bram Flahou, Marjan Steppe, Freddy Haesebrouck, Benedikt Sas, Richard Ducatelle, Severine Vermeire, and Filip Van Immerseel. 2013. “Butyricicoccus Pullicaecorum in Inflammatory Bowel Disease.” *Gut* 62(12):1745–52.

Fernandes, Melissa A., Sofia G. Verstraete, Tung G. Phan, Xutao Deng, Emily Stekol, Brandon LaMere, Susan V Lynch, Melvin B. Heyman, and Eric Delwart. 2019. “Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease.” *Journal of Pediatric Gastroenterology and Nutrition* 68(1):30–36.

Fite, Alemu, Sandra Macfarlane, Elizabeth Furrie, Bahram Bahrami, John H. Cummings, Douglas T. Steinke, and George T. MacFarlane. 2013. “Longitudinal Analyses of Gut Mucosal Microbiotas in Ulcerative Colitis in Relation to Patient Age and Disease Severity and Duration.” *Journal of Clinical Microbiology*.

Forbes, Jessica D., Chih Yu Chen, Natalie C. Knox, Ruth Ann Marrie, Hani El-Gabalawy, Teresa De Kievit, Michelle Alfa, Charles N. Bernstein, and Gary Van Domselaar. 2018. “A Comparative Study of the Gut Microbiota in Immune-Mediated Inflammatory Diseases - Does a Common Dysbiosis Exist?” *Microbiome*.

Forbes, Jessica D., Gary Van Domselaar, and Charles N. Bernstein. 2016. “Microbiome Survey of the Inflamed and Noninflamed Gut at Different Compartments Within the Gastrointestinal Tract of Inflammatory Bowel Disease Patients.” *Inflammatory Bowel Diseases* 22(4):817–25.

Fox, George E., Kenneth R. Pechman, and Carl R. Woese. 2018. *Comparative Cataloging of 16s Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics*. Vol. 14.

Franzosa, Eric A., Lauren J. Mciver, Gholamali Rahnavard, Luke R. Thompson, Melanie Schirmer, George Weingart, Karen Schwarzberg Lipson, Rob Knight, Gregory Caporaso, Nicola Segata, and Curtis Huttenhower. 2018. “Functionally Profiling Metagenomes and Metatranscriptomes at Species-level Resolution.” *Nature*

Methods accepted(November).

Franzosa, Eric A., Alexandra Sirota-Madi, Julian Avila-Pacheco, Nadine Fornelos, Henry J. Haiser, Stefan Reinker, Tommi Vatanen, A. Brantley Hall, Himel Mallick, Lauren J. McIver, Jenny S. Sauk, Robin G. Wilson, Betsy W. Stevens, Justin M. Scott, Kerry Pierce, Amy A. Deik, Kevin Bullock, Floris Imhann, Jeffrey A. Porter, Alexandra Zhernakova, Jingyuan Fu, Rinse K. Weersma, Cisca Wijmenga, Clary B. Clish, Hera Vlamakis, Curtis Huttenhower, and Ramnik J. Xavier. 2019. "Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease." *Nature Microbiology* 4(2):293–305.

Friedman, Milton. 1937. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Journal of the American Statistical Association* 32(200):675.

Frolkis, Alexandra D., Isabelle A. Vallerand, Abdel-Aziz Shaheen, Mark W. Lowerison, Mark G. Swain, Cheryl Barnabe, Scott B. Patten, and Gilaad G. Kaplan. 2018. "Depression Increases the Risk of Inflammatory Bowel Disease, Which May Be Mitigated by the Use of Antidepressants in the Treatment of Depression." *Gut* gutjnl-2018-317182.

Fujino, S., A. Andoh, S. Bamba, A. Ogawa, K. Hata, Y. Araki, T. Bamba, and Y. Fujiyama. 2003. "Increased Expression of Interleukin 17 in Inflammatory Bowel Disease." *Gut* 52(1):65–70.

Furrie, E., S. Macfarlane, J. H. Cummings, and G. T. Macfarlane. 2004. "Systemic Antibodies towards Mucosal Bacteria in Ulcerative Colitis and Crohn's Disease Differentially Activate the Innate Immune Response." *Gut*.

Gaci, Nadia, Guillaume Borrel, William Tottey, Paul William O'Toole, and Jean-François Brugère. 2014. "Archaea and the Human Gut: New Beginning of an Old Story." *World Journal of Gastroenterology* 20(43):16062–78.

Gasche, C., M. C. E. Lomer, I. Cavill, and G. Weiss. 2004. "Iron, Anaemia, and Inflammatory Bowel Diseases." *Gut* 53(8):1190–97.

Gevers, Dirk. 2015. "The Treatment-Naïve Microbiome in New-Onset Crohn ' s Disease." 15(3):382–92.

Ghosh, Subrata, Alan Shand, and Anne Ferguson. 2000. "Clinical Review Regular Review Ulcerative Colitis." 320(April):1119–23.

Gonzalez, E., F. E. Pitre, and N. J. B. Brereton. 2019. "ANCHOR: A 16S RRNA Gene Amplicon Pipeline for Microbial Analysis of Multiple Environmental Samples." *Environmental Microbiology* 1462-2920.14632.

Goodrich, Julia K., Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T. Bell, Timothy D. Spector, Andrew G. Clark, and Ruth E. Ley. 2014. "Human Genetics Shape the Gut Microbiome." *Cell* 159(4):789–99.

Gophna, Uri, Katrin Sommerfeld, Sharon Gophna, W. Ford Doolittle, and Sander J. O. Veldhuyzen van Zanten. 2006. "Differences between Tissue-Associated Intestinal Microfloras of Patients with Crohn's Disease and Ulcerative Colitis." *Journal of Clinical Microbiology* 44(11):4136–41.

Guo, Songhe, Yongfan Lu, Banglao Xu, Wan Wang, Jianhua Xu, and Ge Zhang. 2019. "A Simple Fecal Bacterial Marker Panel for the Diagnosis of Crohn's Disease." *Frontiers in Microbiology* 10:1306.

Hall, Andrew Brantley, Moran Yassour, Jenny Sauk, Ashley Garner, Xiaofang Jiang, Timothy Arthur, Georgia K. Lagoudas, Tommi Vatanen, Nadine Fornelos, Robin Wilson, Madeline Bertha, Melissa Cohen, John Garber, Hamed Khalili, Dirk Gevers, Ashwin N. Ananthakrishnan, Subra Kugathasan, Eric S. Lander, Paul Blainey, Hera Vlamakis, Ramnik J. Xavier, and Curtis Huttenhower. 2017. "A Novel Ruminococcus Gnavus Clade Enriched in Inflammatory Bowel Disease Patients." *Genome Medicine* 9(1):1–12.

Hammer, Heinz F. 2011. "Gut Microbiota and Inflammatory Bowel Disease." 550–53.

Hansen, Richard, Richard K. Russell, Caroline Reiff, Petra Louis, Freda McIntosh, Susan H. Berry, Indrani MGBRhopadhya, W. Michael Bisset, Andy R. Barclay, Jon Bishop, Diana M. Flynn, Paraic McGrogan, Sabarinathan Loganathan, Gamal Mahdi, Harry J. Flint, Emad M. El-Omar, and Georgina L. Hold. 2012. "Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's but Not in Ulcerative Colitis." *American Journal of Gastroenterology*.

Hansen, Richard, Richard K Russell, Caroline Reiff, Petra Louis, Freda McIntosh, Susan H. Berry, Indrani Mukhopadhya, W. Michael Bisset, Andy R. Barclay, Jon Bishop, Diana M. Flynn, Paraic McGrogan, Sabarinathan Loganathan, Gamal Mahdi, Harry J. Flint,

Emad M. El-Omar, and Georgina L. Hold. 2012. “Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn’s but Not in Ulcerative Colitis.” *The American Journal of Gastroenterology* 107(12):1913–22.

Harries, A. D., Baird, A. & Rhodes, J. 1982. “Non-Smoking: A Feature of Ulcerative Colitis.” *British Medical Journal (Clinical Research Ed.)* 285(6339):440.

Hayashi, H., M. Sakamoto, and Y. Benno. 2002. “Phylogenetic Analysis of the Human Gut Microbiota Using 16SrDNA Clone Libraries and Strictly Anaerobic Culture-Based Methods.” *Microbiology Immunol.* 46(8):535–48.

Hill, Kalisha A., Kim L. Wang, Steven J. Stryker, Rohit Gupta, David M. Weinrach, and M. S. Rao. 2004. “Comparative Analysis of Cell Adhesion Molecules, Cell Cycle Regulatory Proteins, Mismatch Repair Genes, Cyclooxygenase-2, and DPC4 in Carcinomas Arising in Inflammatory Bowel Disease and Sporadic Colon Cancer.” *Oncology Reports* 11(5):951–56.

Hillman, Ethan T., Hang Lu, Tianming Yao, and Cindy H. Nakatsu. 2017. “Microbial Ecology along the Gastrointestinal Tract.” *Microbes and Environments* 32(4):300–313.

Hillmann, Benjamin, Gabriel A. Al-Ghalith, Robin R. Shields-Cutler, Qiyun Zhu, Daryl M. Gohl, Kenneth B. Beckman, Rob Knight, and Dan Knights. 2018. “Evaluating the Information Content of Shallow Shotgun Metagenomics.” *MSystems* 3(6).

Hooper, Lora V., Tore Midtvedt, and Jeffrey I. Gordon. 2002. “How Host-Microbial Interactions Shape the Nutrient Environment of the Mammalian Intestine.” *Annual Review of Nutrition* 22(1):283–307.

Huang, Katherine, Arthur Brady, Anup Mahurkar, Owen White, Dirk Gevers, Curtis Huttenhower, and Nicola Segata. 2014. “MetaRef: A Pan-Genomic Database for Comparative and Community Microbial Genomics.” *Nucleic Acids Research* 42(Database issue):D617-24.

Huerta-Cepas, Jaime, Damian Szklarczyk, Kristoffer Forslund, Helen Cook, Davide Heller, Mathias C. Walter, Thomas Rattei, Daniel R. Mende, Shinichi Sunagawa, Michael Kuhn, Lars Juhl Jensen, Christian Von Mering, and Peer Bork. 2016. “EGGNOG 4.5: A Hierarchical Orthology Framework with Improved Functional Annotations for Eukaryotic, Prokaryotic and Viral Sequences.” *Nucleic Acids Research*.

Huggan, Paul J. and David R. Murdoch. 2008. “Fusobacterial Infections: Clinical Spectrum and Incidence of Invasive Disease.” *Journal of Infection* 57(4):283–89.

Hughes, Jennifer B., Jessica J. Hellmann, Taylor H. Ricketts, and Brendan J. M. Bohannan. 2001. “Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity.” *Appl. Environ. Microbiol.* 67(10):4399–4406.

Huson, Daniel H., Alexander F. Auch, Ji Qi, and Stephan C. Schuster. 2007. “MEGAN Analysis of Metagenomic Data.” *Genome Research* 17(3):377–86.

Jang, Hyo-jeong, Ben Kang, and Byung-ho Choe. 2019. “The Difference in Extraintestinal Manifestations of Inflammatory Bowel Disease for Children and Adults.” 8(13):4–15.

Jenkinson, Howard F. and Richard J. Lamont. 2005. “Oral Microbial Communities in Sickness and in Health.” *Trends in Microbiology* 13(12):589–95.

Jones, Deborah T., Mark T. Osterman, Meenakshi Bewtra, and James D. Lewis. 2008. “Passive Smoking and Inflammatory Bowel Disease: A Meta-Analysis.” *The American Journal of Gastroenterology* 103(9):2382–93.

Jordan, Cheryl, Bu’Hussain Hayee, and Trudie Chalder. 2018. “Cognitive Behaviour Therapy for Distress in People with Inflammatory Bowel Disease: A Benchmarking Study.” *Clinical Psychology & Psychotherapy* (February):1–10.

Jovel, Juan, Jordan Patterson, Weiwei Wang, Naomi Hotte, Sandra O’Keefe, Troy Mitchel, Troy Perry, Dina Kao, Andrew L. Mason, Karen L. Madsen, and Gane K. S. Wong. 2016. “Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics.” *Frontiers in Microbiology* 7(APR):1–17.

Kaakoush, Nadeem O., Andrew S. Day, Karina D. Huinao, Steven T. Leach, Daniel A. Lemberg, Scot E. Dowd, and Hazel M. Mitchell. 2012a. “Microbial Dysbiosis in Pediatric Patients with Crohn’s Disease.” *Journal of Clinical Microbiology* 50(10):3258–66.

Kaakoush, Nadeem O., Andrew S. Day, Karina D. Huinao, Steven T. Leach, Daniel A. Lemberg, Scot E. Dowd, and Hazel M. Mitchell. 2012b. “Microbial Dysbiosis in Pediatric Patients with Crohn’s Disease.” *Journal of Clinical Microbiology*.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. “KEGG as a Reference Resource for Gene and Protein Annotation.” *Nucleic Acids Research*.

Kaplan, Gilaad G. 2015. “The Global Burden of IBD: From 2015 to 2025.” *Nature Reviews Gastroenterology and Hepatology* 12(12):720–27.

Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer. 2016. “MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function.” Pp. 207–33 in *Methods in molecular biology (Clifton, N.J.)*. Vol. 1399.

Kégl, Balázs. 2013. “The Return of AdaBoost.MH: Multi-Class Hamming Trees.”

Khan, M. Tanweer, Sylvia H. Duncan, Alfons J. M. Stams, Jan Maarten van Dijl, Harry J. Flint, and Hermie J. M. Harmsen. 2012. “The Gut Anaerobe Faecalibacterium Prausnitzii Uses an Extracellular Electron Shuttle to Grow at Oxic-Anoxic Interphases.” *The ISME Journal* 6(8):1578–85.

Kim, S., H. Chinen, H. S. Said, S. Nakagome, M. Hattori, H. Oota, H. Morita, R. Kimura, K. Oshima, J. Fujita, A. Iraha, W. Suda, S. Mano, H. Ishida, and T. Dohi. 2013. “Dysbiosis of Salivary Microbiota in Inflammatory Bowel Disease and Its Association With Oral Immunological Biomarkers.” *DNA Research* 21(1):15–25.

Knoll, Rebecca L., Kristoffer Forslund, Jens Roat Kultima, Claudius U. Meyer, Ulrike Kullmer, Shinichi Sunagawa, Peer Bork, and Stephan Gehring. 2016. “Gut Microbiota Differs between Children with Inflammatory Bowel Disease and Healthy Siblings in Taxonomic and Functional Composition: A Metagenomic Analysis.” *American Journal of Physiology-Gastrointestinal and Liver Physiology*.

Kostic, Aleksandar D., Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M. Earl, Akinyemi I. Ojesina, Joonil Jung, Adam J. Bass, Josep Tabernero, José Baselga, Chen Liu, Ramesh A. Shivdasani, Shuji Ogino, Bruce W. Birren, Curtis Huttenhower, Wendy S. Garrett, and Matthew Meyerson. 2012. “Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma.” *Genome Research* 22(2):292–98.

Kruskal, William H. and W. Allen Wallis. 1952. “Use of Ranks in One-Criterion Variance Analysis.” *Journal of the American Statistical Association* 47(260):583.

Kultima, Jens Roat, Luis Pedro Coelho, Kristoffer Forslund, Jaime Huerta-Cepas, Simone S. Li, Marja Driessen, Anita Yvonne Voigt, Georg Zeller, Shinichi Sunagawa, and Peer Bork. 2016. “MOCAT2: A Metagenomic Assembly, Annotation and Profiling Framework.” *Bioinformatics* 32(16):2520–23.

Kultima, Jens Roat, Shinichi Sunagawa, Junhua Li, Weineng Chen, Hua Chen, Daniel R.

Mende, Manimozhiyan Arumugam, Qi Pan, Binghang Liu, Junjie Qin, Jun Wang, and Peer Bork. 2012. "MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit." *PLoS ONE* 7(10):1–6.

Kumari, Reena, Vineet Ahuja, and Jaishree Paul. 2013. "Fluctuations in Butyrate-Producing Bacteria in Ulcerative Colitis Patients of North India." *World Journal of Gastroenterology* 19(22):3404–14.

Laass, Martin W., Dirk Roggenbuck, and Karsten Conrad. 2014. "Autoimmunity Reviews Diagnosis and Classi Fi Cation of Crohn ' s Disease." 13:467–71.

Lam, Hugo Y. K., Michael J. Clark, Rui Chen, Rong Chen, Georges Natsoulis, Maeve O'Huallachain, Frederick E. Dewey, Lukas Habegger, Euan A. Ashley, Mark B. Gerstein, Atul J. Butte, Hanlee P. Ji, and Michael Snyder. 2011. "Performance Comparison of Whole-Genome Sequencing Platforms." *Nature Biotechnology* 30(1):78–82.

Langendijk, P. S., F. Schut, G. J. Jansen, G. C. Raangs, G. R. Kamphuis, M. H. Wilkinson, and G. W. Welling. 1995. "Quantitative Fluorescence in Situ Hybridization of Bifidobacterium Spp. with Genus-Specific 16S RRNA-Targeted Probes and Its Application in Fecal Samples." *Applied and Environmental Microbiology* 61(8):3069–75.

Legaki, Evangelia and Maria Gazouli. 2016. "Influence of Environmental Factors in the Development of Inflammatory Bowel Diseases." *World Journal of Gastrointestinal Pharmacology and Therapeutics* 7(1):112–25.

Leung, Andrea, Ho Tsoi, and Jun Yu. 2015. "*Fusobacterium* and *Escherichia* : Models of Colorectal Cancer Driven by Microbiota and the Utility of Microbiota in Colorectal Cancer Screening." *Expert Review of Gastroenterology & Hepatology* 9(5):651–57.

Lewis, James D. and Maria T. Abreu. 2017. "Diet as a Trigger or Therapy for Inflammatory Bowel Diseases." *Gastroenterology* 152(2):398-414.e6.

Ley, Ruth E., Micah Hamady, Catherine Lozupone, Peter Turnbaugh, Rob Roy, J. Stephen Bircher, Michael L. Schlegel, Tammy A. Tucker, D. Mark, Rob Knight, and Jeffrey I. Gordon. 2009. "Evolution of Mammals and Their Gut Microbes." 320(5883):1647–51.

Ley, Ruth E., Daniel A. Peterson, and Jeffrey I. Gordon. 2006. "Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine." *Cell*

124(4):837–48.

Li, Junhua, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, Manimozhiyan Arumugam, Jens Roat Kultima, Edi Prifti, Trine Nielsen, Agnieszka Sierakowska Juncker, Chaysavanh Manichanh, Bing Chen, Wenwei Zhang, Florence Levenez, Juan Wang, Xun Xu, Liang Xiao, Suisha Liang, Dongya Zhang, Zhaoxi Zhang, Weineng Chen, Hailong Zhao, Jumana Yousuf Al-Aama, Sherif Edris, Huanming Yang, Jian Wang, Torben Hansen, Henrik Bjørn Nielsen, Søren Brunak, Karsten Kristiansen, Francisco Guarner, Oluf Pedersen, Joel Doré, S. Dusko Ehrlich, Nicolas Pons, Emmanuelle Le Chatelier, Jean-Michel Batto, Sean Kennedy, Florence Haimet, Yohanan Winogradski, Eric Pelletier, Denis LePaslier, François Artiguenave, Thomas Bruls, Jean Weissenbach, Keith Turner, Julian Parkhill, Maria Antolin, Francesc Casellas, Natalia Borruel, Encarna Varela, Antonio Torrejon, Gérard Denariaz, Muriel Derrien, Johan E. T. van Hylckama Vlieg, Patrick Viega, Raish Oozeer, Jan Knoll, Maria Rescigno, Christian Brechot, Christine M'Rini, Alexandre Mérieux, Takuji Yamada, Sebastian Tims, Erwin G. Zoetendal, Michiel Kleerebezem, Willem M. de Vos, Antonella Cultrone, Marion Leclerc, Catherine Juste, Eric Guedon, Christine Delorme, Séverine Layec, Ghalia Khaci, Maarten van de Guchte, Gaetana Vandemeulebrouck, Alexandre Jamet, Rozenn Dervyn, Nicolas Sanchez, Hervé Blottière, Emmanuelle Maguin, Pierre Renault, Julien Tap, Daniel R. Mende, Peer Bork, and Jun Wang. 2014. "An Integrated Catalog of Reference Genes in the Human Gut Microbiome." *Nature Biotechnology* 32(8):834–41.

Lim, Efrem S., Yanjiao Zhou, Guoyan Zhao, Irma K. Bauer, Lindsay Droit, I. Malick Ndao, Barbara B. Warner, Phillip I. Tarr, David Wang, and Lori R. Holtz. 2015. "Early Life Dynamics of the Human Gut Virome and Bacterial Microbiome in Infants." *Nature Medicine* 21(10):1228–34.

Liu, Jimmy Z., Suzanne Van Sommeren, Hailiang Huang, Siew C. Ng, and Rudi Alberts. 2015. "Europe PMC Funders Group Association Analyses Identify 38 Susceptibility Loci for Inflammatory Bowel Disease and Highlight Shared Genetic Risk across Populations." *Nat Genet* 47(9):979–86.

Loddo, Italia and Claudio Romano. 2015. "Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis." *Frontiers in Immunology* 6(NOV):6–11.

Madsen, K. L., J. S. Doyle, L. D. Jewell, M. M. Tavernini, and R. N. Fedorak. 1999.

"Lactobacillus Species Prevents Colitis in Interleukin 10 Gene-Deficient Mice." *Gastroenterology* 116(5):1107–14.

Mahid, Suhal S., Kyle S. Minor, Roberto E. Soto, Carlton A. Hornung, and Susan Galandiuk. 2006. "Smoking and Inflammatory Bowel Disease: A Meta-Analysis." *Mayo Clinic Proceedings* 81(11):1462–71.

Malham, Mikkel, Berit Lilje, Gunnar Houen, Katrine Winther, Paal S. Andersen, and Christian Jakobsen. 2019. "The Microbiome Reflects Diagnosis and Predicts Disease Severity in Paediatric Onset Inflammatory Bowel Disease." *Scandinavian Journal of Gastroenterology*.

Mangin, IrÃ¨ne, RÃ©gis Bonnet, Philippe Seksik, Lionel Rigottier-Gois, MalÃ¨ne Sutren, Yoram Bouhnik, Christel Neut, Matthew D. Collins, Jean-FrÃ©dÃ©ric Colombel, Philippe Marteau, and JoÃ«l DorÃ©. 2004. "Molecular Inventory of Faecal Microflora in Patients with Crohn's Disease." *FEMS Microbiology Ecology* 50(1):25–36.

Manichanh, C., L. Rigottier-Gois, E. Bonnaud, K. Gloux, E. Pelletier, L. Frangeul, R. Nalin, C. Jarrin, P. Chardon, P. Marteau, J. Roca, and J. Dore. 2006. "Reduced Diversity of Faecal Microbiota in Crohn's Disease Revealed by a Metagenomic Approach." *Gut* 55(2):205–11.

Mann, H. B. and D. R. Whitney. 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other." *The Annals of Mathematical Statistics* 18:50–60.

Marchesi, Julian R., David H. Adams, Francesca Fava, Gerben D. A. Hermes, Gideon M. Hirschfield, Georgina Hold, Mohammed Nabil Quraishi, James Kinross, Hauke Smidt, Kieran M. Tuohy, Linda V Thomas, Erwin G. Zoetendal, and Ailsa Hart. 2016. "The Gut Microbiota and Host Health: A New Clinical Frontier." *Gut* 65(2):330–39.

Martinez-Medina, Margarita, Xavier Aldeguer, Ferran Gonzalez-Huix, Doroteo Acero, and Jesús L. Garcia-Gil. 2006. "Abnormal Microbiota Composition in the Ileocolonic Mucosa of Crohn's Disease Patients as Revealed by Polymerase Chain Reaction-Denaturing Gradient Gel Electrophoresis." *Inflammatory Bowel Diseases* 12(12):1136–45.

Maslowski, Kendle M. and Charles R. MacKay. 2011. "Diet, Gut Microbiota and Immune Responses." *Nature Immunology* 12(1):5–9.

Masoodi, Ibrahim, Ali S. Alshanqeeti, Shameem Ahmad, Essam J. Alyamani, Abed A. Allehibi, Adel N. Qutub, Khalid N. Alsayari, and Ahmed O. Alomair. 2019. "Microbial Dysbiosis in Inflammatory Bowel Diseases: Results of a Metagenomic Study in Saudi Arabia." *Minerva Gastroenterologica e Dietologica*.

Michail, Sonia, Matthew Durbin, Dan Turner, Anne M. Griffiths, David R. Mack, Jeffrey Hyams, Neal Leleiko, Harshavardhan Kenche, Adrienne Stolfi, and Eytan Wine. 2012. "Alterations in the Gut Microbiome of Children with Severe Ulcerative Colitis." *Inflammatory Bowel Diseases*.

Miele, Erasmo, Raanan Shamir, Marina Aloi, Amit Assa, Christian Braegger, Jiri Bronsky, Lissy de Ridder, Johanna C. Escher, Iva Hojsak, Sanja Kolaček, Sibylle Koletzko, Arie Levine, Paolo Lionetti, Massimo Martinelli, Frank Ruemmele, Richard K. Russell, Rotem Sigall Boneh, Johan van Limbergen, Gigi Veereman, and Annamaria Staiano. 2018. "Nutrition in Pediatric Inflammatory Bowel Disease: A Position Paper on Behalf of the Porto Inflammatory Bowel Disease Group of the European Society of Pediatric Gastroenterology, Hepatology and Nutrition." *Journal of Pediatric Gastroenterology and Nutrition* 66(4):687–708.

Milani, Christian, Sabrina Duranti, Francesca Bottacini, Eoghan Casey, Francesca Turroni, Jennifer Mahony, Clara Belzer, Susana Delgado Palacio, Silvia Arboleya Montes, Leonardo Mancabelli, Gabriele Andrea Lugli, Juan Miguel Rodriguez, Lars Bode, Willem de Vos, Miguel Gueimonde, Abelardo Margolles, Douwe van Sinderen, and Marco Ventura. 2017. "The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota." *Microbiology and Molecular Biology Reviews : MMBR* 81(4).

Mirsepasi-Lauridsen, Hengameh Chloé, Bruce Andrew Vallance, Karen Angeliki Krogfelt, and Andreas Munk Petersen. 2019. "Escherichia Coli Pathobionts Associated with Inflammatory Bowel Disease." *Clinical Microbiology Reviews* 32(2):1–16.

Misra, Madhusmita and Anne Klibanski. 2016. "Anorexia Nervosa and Its Associated Endocrinopathy in Young People." *Hormone Research in Paediatrics* 85(3):147–57.

Mitsuyamat, Keiichi, Atsushi Toyonaga, and Michio Sata. 2002. "Intestinal Microflora as a Therapeutic Target in Inflammatory." 73–77.

Mittermaier, Christian, Clemens Dejaco, Thomas Waldhoer, Anna Oefferlbauer-Ernst, Wolfgang Miehsler, Markus Beier, Wolfgang Tillinger, Alfred Gangl, and Gabriele

Moser. 2004. “Impact of Depressive Mood on Relapse in Patients with Inflammatory Bowel Disease: A Prospective 18-Month Follow-up Study.” *Psychosomatic Medicine* 66(1):79–84.

Moehle, Christoph, Nikolaus Ackermann, Thomas Langmann, Charalampos Aslanidis, Alexander Kel, Olga Kel-Margoulis, Anna Schmitz-Madry, Alexandra Zahn, Wolfgang Stremmel, and Gerd Schmitz. 2006. “Aberrant Intestinal Expression and Allelic Variants of Mucin Genes Associated with Inflammatory Bowel Disease.” *Journal of Molecular Medicine* 84(12):1055–66.

Molodecky, Natalie A., Ing Shian Soon, Doreen M. Rabi, William A. Ghali, Mollie Ferris, Greg Chernoff, Eric I. Benchimol, Remo Panaccione, Subrata Ghosh, Herman W. Barkema, and Gilaad G. Kaplan. 2012. “Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases with Time, Based on Systematic Review.” *Gastroenterology* 142(1):46-54.e42.

Momozawa, Yukihide, Julia Dmitrieva, Emilie Théâtre, Valérie Deffontaine, Souad Rahmouni, Benoît Charloteaux, François Crins, Elisa Docampo, Mahmoud Elansary, Ann Stephan Gori, Christelle Lecut, Rob Mariman, Myriam Mni, Cécile Oury, Ilya Altukhov, Dmitry Alexeev, Yuri Aulchenko, Leila Amininejad, Gerd Bouma, Frank Hoentjen, Mark Löwenberg, Bas Oldenburg, Marieke J. Pierik, Andrea E. Vander Meulen-De Jong, C. Janneke Van Der Woude, Marijn C. Visschedijk, Mark Lathrop, Jean Pierre Hugot, Rinse K. Weersma, Martine De Vos, Denis Franchimont, Severine Vermeire, Michiaki Kubo, Edouard Louis, Michel Georges, Clara Abraham, Jean Paul Achkar, Tariq Ahmad, Ashwin N. Ananthakrishnan, Vibeke Andersen, Carl A. Anderson, Jane M. Andrews, Vito Annese, Guy Aumais, Leonard Baidoo, Robert N. Baldassano, Peter A. Bampton, Murray Barclay, Jeffrey C. Barrett, Theodore M. Bayless, Johannes Bethge, Alain Bitton, Gabrielle Boucher, Stephan Brand, Berenice Brandt, Steven R. Brant, Carsten Büning, Angela Chew, Judy H. Cho, Isabelle Cleynen, Ariella Cohain, Anthony Croft, Mark J. Daly, Mauro D’Amato, Silvio Danese, Dirk De Jong, Goda Denapiene, Lee A. Denson, Kathy L. Devaney, Olivier Dewit, Renata D’Inca, Marla Dubinsky, Richard H. Duerr, Cathryn Edwards, David Ellinghaus, Jonah Essers, Lynnette R. Ferguson, Eleonora A. Festen, Philip Fleshner, Tim Florin, Andre Franke, Karin Fransen, Richard Gearry, Christian Gieger, Jürgen Glas, Philippe Goyette, Todd Green, Anne M. Griffiths, Stephen L. Guthery, Hakon

Hakonarson, Jonas Halfvarson, Katherine Hanigan, Talin Haritunians, Ailsa Hart, Chris Hawkey, Nicholas K. Hayward, Matija Hedl, Paul Henderson, Xinli Hu, Hailiang Huang, Ken Y. Hui, Marcin Imielinski, Andrew Ippoliti, Laimas Jonaitis, Luke Jostins, Tom H. Karlsen, Nicholas A. Kennedy, Mohammed Azam Khan, Gediminas Kiudelis, Krupa Krishnaprasad, Subra Kugathasan, Limas Kupcinskas, Anna Latiano, Debby Laukens, Ian C. Lawrance, James C. Lee, Charlie W. Lees, Marcis Leja, Johan Van Limbergen, Paolo Lionetti, Jimmy Z. Liu, Gillian Mahy, John Mansfield, Dunecan Massey, Christopher G. Mathew, Dermot P. B. McGovern, Raquel Milgrom, Mitja Mitrovic, Grant W. Montgomery, Craig Mowat, William Newman, Aylwin Ng, Siew C. Ng, Sok Meng Evelyn Ng, Susanna Nikolaus, Kaida Ning, Markus Nöthen, Ioannis Oikonomou, Orazio Palmieri, Miles Parkes, Anne Phillips, Cyriel Y. Ponsioen, Uros Potocnik, Natalie J. Prescott, Deborah D. Proctor, Graham Radford-Smith, Jean Francois Rahier, Soumya Raychaudhuri, Miguel Regueiro, Florian Rieder, John D. Rioux, Stephan Ripke, Rebecca Roberts, Richard K. Russell, Jeremy D. Sanderson, Miquel Sans, Jack Satsangi, Eric E. Schadt, Stefan Schreiber, Dominik Schulte, L. Philip Schumm, Regan Scott, Mark Seielstad, Yashoda Sharma, Mark S. Silverberg, Lisa A. Simms, Jurgita Skieceviciene, Sarah L. Spain, A. Hillary Steinhart, Joanne M. Stempak, Laura Stronati, Jurgita Sventoraityte, Stephan R. Targan, Kirstin M. Taylor, Anje Ter Velde, Leif Torkvist, Mark Tremelling, Suzanne Van Sommeren, Eric Vasiliauskas, Hein W. Verspaget, Thomas Walters, Kai Wang, Ming Hsi Wang, Zhi Wei, David Whiteman, Cisca Wijmenga, David C. Wilson, Juliane Winkelmann, Ramnik J. Xavier, Bin Zhang, Clarence K. Zhang, Hu Zhang, Wei Zhang, Hongyu Zhao, and Zhen Z. Zhao. 2018. “IBD Risk Loci Are Enriched in Multigenic Regulatory Modules Encompassing Putative Causative Genes.” *Nature Communications* 9(1).

Moore, W. E. and L. V Holdeman. 1974. “Human Fecal Flora: The Normal Flora of 20 Japanese-Hawaiians.” *Applied Microbiology* 27(5):961–79.

Morgan, Xochitl C., Timothy L. Tickle, Harry Sokol, Dirk Gevers, Kathryn L. Devaney, Doyle V Ward, Joshua A. Reyes, Samir A. Shah, Neal LeLeiko, Scott B. Snapper, Athos Bousvaros, Joshua Korzenik, Bruce E. Sands, Ramnik J. Xavier, and Curtis Huttenhower. 2012. “Dysfunction of the Intestinal Microbiome in Inflammatory Bowel Disease and Treatment.” *Genome Biology* 13(9):R79.

Mosca, Alexis, Marion Leclerc, and Jean P. Hugot. 2016. “Gut Microbiota Diversity and

Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem?" *Frontiers in Microbiology* 7:455.

Moustafa, Ahmed, Weizhong Li, Ericka L. Anderson, Emily H. M. Wong, Parambir S. Dulai, William J. Sandborn, William Biggs, Shibu Yooseph, Marcus B. Jones, J. Craig Venter, Karen E. Nelson, John T. Chang, Amalio Telenti, and Brigid S. Boland. 2018. "Genetic Risk, Dysbiosis, and Treatment Stratification Using Host Genome and Gut Microbiome in Inflammatory Bowel Disease." *Clinical and Translational Gastroenterology* 9(1):e132-8.

Navas-Molina, José A., Juan M. Peralta-Sánchez, Antonio González, Paul J. McMurdie, Yoshiki Vázquez-Baeza, Zhenjiang Xu, Luke K. Ursell, Christian Lauber, Hongwei Zhou, Se Jin Song, James Huntley, Gail L. Ackermann, Donna Berg-Lyons, Susan Holmes, J. Gregory Caporaso, and Rob Knight. 2013. "Advancing Our Understanding of the Human Microbiome Using QIIME." *Methods in Enzymology*.

Ni, Josephine, Gary D. Wu, Lindsey Albenberg, and Vesselin T. Tomov. 2017. "Gut Microbiota and IBD: Causation or Correlation?" *Nature Reviews Gastroenterology & Hepatology* 14(10):573–84.

Ni, Josephine, Gary D. Wu, Lindsey Albenberg, Vesselin T. Tomov, and Curie Boulevard. 2018. "Gut Microbiota and IBD : Causation or Correlation ?" 14(10):573–84.

Norman, Jason M., Scott A. Handley, Megan T. Baldridge, Lindsay Droit, Catherine Y. Liu, Brian C. Keller, Amal Kambal, Cynthia L. Monaco, Guoyan Zhao, Phillip Fleshner, Thaddeus S. Stappenbeck, Dermot P. B. McGovern, Ali Keshavarzian, Ece A. Mutlu, Jenny Sauk, Dirk Gevers, Ramnik J. Xavier, David Wang, Miles Parkes, and Herbert W. Virgin. 2015. "Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease." *Cell* 160(3):447–60.

Nowakowski, Jarosław, Adrian Chrobak, and Dominika Dudek. 2016. "Psychiatry Illness in Inflammatory Bowel Diseases - Psychiatric Comorbidity and Biological Underpinnings." *Psychiatria Polska* 50(6):1157–66.

O'Hara, Ann M. and Fergus Shanahan. 2006. "The Gut Flora as a Forgotten Organ." *EMBO Reports* 7(7):688–93.

Odamaki, Toshitaka, Kumiko Kato, Hirosuke Sugahara, Nanami Hashikura, Sachiko Takahashi, Jin-zhong Xiao, Fumiaki Abe, and Ro Osawa. 2016. "Age-Related Changes in Gut Microbiota Composition from Newborn to Centenarian : A Cross-

Sectional Study." *BMC Microbiology* 1–12.

Ott, S. J., M. Musfeldt, D. F. Wenderoth, J. Hampe, O. Brant, U. R. Fölsch, K. N. Timmis, and S. Schreiber. 2004. "Reduction in Diversity of the Colonic Mucosa Associated Bacterial Microflora in Patients with Active Inflammatory Bowel Disease." *Gut* 53(5):685–93.

Palmela, Carolina, Caroline Chevarin, Zhilu Xu, Joana Torres, Gwladys Sevrin, Robert Hirten, Nicolas Barnich, Siew C. Ng, and Jean Frederic Colombel. 2018. "Adherent-Invasive Escherichia Coli in Inflammatory Bowel Disease." *Gut* 67(3):574–87.

Palmer, Chana, Elisabeth M. Bik, Daniel B. DiGiulio, David A. Relman, and Patrick O. Brown. 2007. "Development of the Human Infant Intestinal Microbiota." *PLoS Biology* 5(7):1556–73.

Papa, Eliseo, Michael Docktor, Christopher Smillie, Sarah Weber, Sarah P. Preheim, Dirk Gevers, Georgia Giannoukos, Dawn Ciulla, Diana Tabbaa, Jay Ingram, David B. Schauer, Doyle V Ward, Joshua R. Korzenik, Ramnik J. Xavier, Athos Bousvaros, and Eric J. Alm. 2012. "Non-Invasive Mapping of the Gastrointestinal Microbiota Identifies Children with Inflammatory Bowel Disease." *PloS One* 7(6):e39242.

Paramsothy, Sudarshan, Ramesh Paramsothy, David T. Rubin, Michael A. Kamm, Nadeem O. Kaakoush, Hazel M. Mitchell, and Natalia Castaño-Rodríguez. 2017. "Faecal Microbiota Transplantation for Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis." *Journal of Crohn's and Colitis* 11(10):1180–99.

Pascal, Victoria, Marta Pozuelo, Natalia Borruel, Francesc Casellas, David Campos, Alba Santiago, Xavier Martinez, Encarna Varela, Guillaume Sarrabayrouse, Kathleen Machiels, Severine Vermeire, Harry Sokol, Francisco Guarner, and Chaysavanh Manichanh. 2017. "A Microbial Signature for Crohn's Disease." *Gut* 66(5):813–22.

Plovier, Hubert, Amandine Everard, Céline Druart, Clara Depommier, Matthias Van Hul, Lucie Geurts, Julien Chilloux, Noora Ottman, Thibaut Duparc, Laeticia Lichtenstein, Antonis Myridakis, Nathalie M. Delzenne, Judith Klievink, Arnab Bhattacharjee, Kees C. H. van der Ark, Steven Aalvink, Laurent O. Martinez, Marc-Emmanuel Dumas, Dominique Maiter, Audrey Loumaye, Michel P. Hermans, Jean-Paul Thissen, Clara Belzer, Willem M. de Vos, and Patrice D. Cani. 2017. "A Purified Membrane Protein from Akkermansia Muciniphila or the Pasteurized Bacterium Improves Metabolism in Obese and Diabetic Mice." *Nature Medicine* 23(1):107–13.

Pozuelo, Marta, Suchita Panda, Alba Santiago, Sara Mendez, Anna Accarino, Javier Santos, Francisco Guarner, Fernando Azpiroz, and Chaysavanh Manichanh. 2015. "Reduction of Butyrate- and Methane-Producing Microorganisms in Patients with Irritable Bowel Syndrome." *Scientific Reports* 5(1):12693.

Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Takuji Yamada, Daniel R. Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H. Bjørn Nielsen, Eric Pelletier, Pierre Renault, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, and Huanming Yang. 2010. "A Human Gut Microbial Gene Catalog Established by Metagenomic Sequencing." *Nature* 464(7285):59–65.

Qin, Junjie, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Torben Hansen, Gaston Sanchez, Jeroen Raes, Gwen Falony, Shujiro Okuda, and Mathieu Almeida. 2012. "A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes." *Nature* 490(7418):55–60.

Rajilić-Stojanović, Mirjana and Willem M. de Vos. 2014. "The First 1000 Cultured Species of the Human Gastrointestinal Microbiota." *FEMS Microbiology Reviews* 38(5):996–1047.

Ramos, Guilherme Piovezani and Konstantinos A. Papadakis. 2019. "Mechanisms of Disease: Inflammatory Bowel Diseases." *Mayo Clinic Proceedings* 94(1):155–65.

Rashid, Mamun-Ur, Egijia Zaura, Mark J. Buijs, Bart J. F. Keijser, Wim Crielaard, Carl Erik Nord, and Andrej Weintraub. 2015. "Determining the Long-Term Effect of Antibiotic Administration on the Human Normal Intestinal Microbiota Using Culture and Pyrosequencing Methods." *Clinical Infectious Diseases* 60(suppl_2):S77–84.

Rhee, Sang H., Charalabos Pothoulakis, and Emeran A. Mayer. 2009. "Principles and Clinical Implications of the Brain-Gut-Enteric Microbiota Axis." *Nature Reviews Gastroenterology and Hepatology*.

Richard, Mathias L., Bruno Lamas, Giuseppina Liguori, Thomas W. Hoffmann, and Harry Sokol. 2015. "Gut Fungal Microbiota: The Yin and Yang of Inflammatory Bowel Disease." *Inflammatory Bowel Diseases* 21(3):656–65.

Ridaura, Vanessa K., Jeremiah J. Faith, Federico E. Rey, Jiye Cheng, E. Alexis, Andrew L. Kau, Nicholas W. Griffin, Vincent Lombard, Bernard Henrissat, James R. Bain, Michael J. Muehlbauer, Olga Ilkayeva, Clay F. Semenkovich, Katsuhiko Funai, David K. Hayashi, Barbara J. Lyle, Margaret C. Martini, K. Luke, Jose C. Clemente, William Van Treuren, and William A. Walters. 2013. "Gut Microbiota from Twins Discordant for Obesity Modulate Adiposity and Metabolic Phenotypes in Mice." *Science* 341(6150):1–22.

La Rosa, Francesca, Mario Clerici, Daniela Ratto, Alessandra Occhinegro, Anna Licito, Marcello Romeo, Carmine Di Iorio, and Paola Rossi. 2018. "The Gut-Brain Axis in Alzheimer's Disease and Omega-3. A Critical Overview of Clinical Trials." *Nutrients* 10(9).

Santiago, Alba, Suchita Panda, Griet Mengels, Xavier Martinez, Fernando Azpiroz, Joel Dore, Francisco Guarner, and Chaysavanh Manichanh. 2014. "Processing Faecal Samples : A Step Forward for Standards in Microbial Community Analysis." 14(1):1–9.

Santiago, Alba, Marta Pozuelo, Maria Poca, Cristina Gely, Juan Camilo Nieto, Xavier Torras, Eva Román, David Campos, Guillaume Sarrabayrouse, Silvia Vidal, Edilmar Alvarado-Tapias, Francisco Guarner, German Soriano, Chaysavanh Manichanh, and Carlos Guarner. 2016. "Alteration of the Serum Microbiome Composition in Cirrhotic Patients with Ascites." *Scientific Reports* 6(1):25001.

Sartor, R. Balfour. 2004. "Therapeutic Manipulation of the Enteric Microflora in Inflammatory Bowel Diseases: Antibiotics, Probiotics, and Prebiotics." *Gastroenterology* 126(6):1620–33.

Sartor, R. Balfour. 2006. "Mechanisms of Disease: Pathogenesis of Crohn's Disease and Ulcerative Colitis." *Nature Clinical Practice Gastroenterology & Hepatology* 3(7):390–407.

Satsangi, J., M. S. Silverberg, S. Vermeire, and J. F. Colombel. 2006. "The Montreal Classification of Inflammatory Bowel Disease: Controversies, Consensus, and Implications." *Gut*.

Savage, DC Dwayne C. 1977. "Microbial Ecology of the Gastrointestinal Tract." *Annual Reviews in Microbiology* 31(1):107–33.

Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann,

Emily B. Hollister, Ryan A. Lesniewski, Brian B. Oakley, Donovan H. Parks, Courtney J. Robinson, Jason W. Sahl, Blaz Stres, Gerhard G. Thallinger, David J. Van Horn, and Carolyn F. Weber. 2009. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities." *Applied and Environmental Microbiology* 75(23):7537–41.

Scriven, Mary, Timothy G. Dinan, John F. Cryan, and Mary Wall. 2018. "Neuropsychiatric Disorders : Influence of Gut Microbe to Brain Signalling." 1–13.

Segata, Nicola, Daniela Boernigen, Timothy L. Tickle, Xochitl C. Morgan, Wendy S. Garrett, and Curtis Huttenhower. 2013. "Computational Meta'omics for Microbial Community Studies." *Molecular Systems Biology* 9(1):1–15.

Seksik, P., L. Rigottier-Gois, G. Gramet, M. Sutren, P. Pochart, P. Marteau, R. Jian, and J. Doré. 2003. "Alterations of the Dominant Faecal Bacterial Groups in Patients with Crohn's Disease of the Colon." *Gut* 52(2):237–42.

Sender, Ron, Shai Fuchs, and Ron Milo. 2016. "Revised Estimates for the Number of Human and Bacteria Cells in the Body." *PLoS Biology* 14(8):1–14.

Shah, Rajesh, Julia L. Cope, Dorottya Nagy-Szakal, Scot Dowd, James Versalovic, Emily B. Hollister, and Richard Kellermayer. 2016. "Composition and Function of the Pediatric Colonic Mucosal Microbiome in Untreated Patients with Ulcerative Colitis." *Gut Microbes*.

Shapiro, S. S. and M. B. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52(3/4):591.

Sheflin, Amy M., Christopher L. Melby, Franck Carbonero, and Tiffany L. Weir. 2017. "Linking Dietary Patterns with Gut Microbial Composition and Function." *Gut Microbes*.

Shin, Ji-Hee, Minju Sim, Joo-Young Lee, and Dong-Mi Shin. 2016. "Lifestyle and Geographic Insights into the Distinct Gut Microbiota in Elderly Women from Two Different Geographic Locations." *Journal of Physiological Anthropology* 35(1):31.

Silverberg, Mark S., Jack Satsangi, Tariq Ahmad, Ian D. R. Arnott, Charles N. Bernstein, Steven R. Brant, Renzo Caprilli, Jean-Frédéric Colombel, Christoph Gasche, Karel Geboes, Derek P. Jewell, Amir Karban, Edward V Loftus, A. Salvador Peña, Robert H. Riddell, David B. Sachar, Stefan Schreiber, A. Hillary Steinhart, Stephan R. Targan,

Severine Vermeire, and B. F. Warren. 2005. “Toward an Integrated Clinical, Molecular and Serological Classification of Inflammatory Bowel Disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology.” *Canadian Journal of Gastroenterology = Journal Canadien de Gastroenterologie* 19 Suppl A(September):5A-36A.

Singh, Sunny, Lesley A. Graff, and Charles N. Bernstein. 2009. “Do NSAIDs, Antibiotics, Infections, or Stress Trigger Flares in IBD.” *American Journal of Gastroenterology* 104(5):1298–1313.

Sivignon, Adeline, Amélie de Vallée, Nicolas Barnich, Jérémy Denizot, Claude Darcha, Georges Pignède, Pascal Vandekerckove, and Arlette Darfeuille-Michaud. 2015. “Saccharomyces Cerevisiae CNCM I-3856 Prevents Colitis Induced by AIEC Bacteria in the Transgenic Mouse Model Mimicking Crohn’s Disease.” *Inflammatory Bowel Diseases* 21(2):276–86.

Sokol, Harry, Valentin Leducq, Hugues Aschard, Hang Phuong Pham, Sarah Jegou, Cecilia Landman, David Cohen, Giuseppina Liguori, Anne Bourrier, Isabelle Nion-Larmurier, Jacques Cosnes, Philippe Seksik, Philippe Langella, David Skurnik, Mathias L. Richard, and Laurent Beaugerie. 2017. “Fungal Microbiota Dysbiosis in IBD.” *Gut* 66(6):1039–48.

Sokol, Harry, P. Seksik, J. P. Furet, O. Firmesse, I. Nion-Larmurier, L. Beaugerie, J. Cosnes, G. Corthier, P. Marteau, and J. Doraé. 2009. “Low Counts of Faecalibacterium Prausnitzii in Colitis Microbiota.” *Inflammatory Bowel Diseases* 15(8):1183–89.

de Souza, Heitor S. P. and Claudio Fiocchi. 2016. “Immunopathogenesis of IBD: Current State of the Art.” *Nature Reviews Gastroenterology & Hepatology* 13(1):13–27.

Spiceland, Clayton M. and Nilesh Lodhia. 2018. “Endoscopy in Inflammatory Bowel Disease: Role in Diagnosis, Management, and Treatment.” *World Journal of Gastroenterology* 24(35):4014–20.

Strauss, Jaclyn, Gilaad G. Kaplan, Paul L. Beck, Kevin Rioux, Remo Panaccione, Rebekah Devinney, Tarah Lynch, and Emma Allen-Vercoe. 2011. “Invasive Potential of Gut Mucosa-Derived Fusobacterium Nucleatum Positively Correlates with IBD Status of the Host.” *Inflammatory Bowel Diseases* 17(9):1971–78.

Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dore. 1999. “Direct Analysis of Genes Encoding 16S RRNA from Complex Communities Reveals

Many Novel Molecular Species within the Human Gut." *Appl Environ Microbiol* 65(11):4799–4807.

Takahashi, Kenichiro, Atsushi Nishida, Takehide Fujimoto, Makoto Fujii, Makoto Shioya, Hirotsugu Imaeda, Osamu Inatomi, Shigeki Bamba, Akira Andoh, and Mitsushige Sugimoto. 2016. "Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease." in *Digestion*.

Tamboli, C. P. 2003. "Dysbiosis in Inflammatory Bowel Disease." *Gut* 53(1):1–4.

Targan, Stephan R. and Loren C. Karp. 2005. "Defects in Mucosal Immunity Leading to Ulcerative Colitis." *Immunological Reviews* 206(1):296–305.

Thomas, G. A., J. Rhodes, J. T. Green, and C. Richardson. 2000. "Role of Smoking in Inflammatory Bowel Disease: Implications for Therapy." *Postgraduate Medical Journal* 76(895):273–79.

Toledo-Pereyra, Luis H. 2009. "The Strange Little Animals of Antony van Leeuwenhoek Surgical Revolution." *Journal of Investigative Surgery* 22(1):4–8.

Truong, Duy Tin, Eric A. Franzosa, Timothy L. Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, and Nicola Segata. 2015. "MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling." *Nature Methods* 12(10):902–3.

Turnbaugh, P. J., Ruth E. Ley, Micah Hamady, Claire Fraser-liggett, Rob Knight, and Jeffrey I. Gordon. 2007. "The Human Microbiome Project: Exploring the Microbial Part of Ourselves in a Changing World." *Nature* 449(7164):804–10.

Ukwenya, AY, A. Ahmed, VI Odigie, and A. Mohammed. 2011. "Inflammatory Bowel Disease in Nigerians: Still a Rare Diagnosis?" *Annals of African Medicine* 10(2):175.

Vandeputte, Doris, Gunter Kathagen, Kevin D'Hoe, Sara Vieira-Silva, Mireia Valles-Colomer, Joaõ Sabino, Jun Wang, Raul Y. Tito, Lindsey De Commer, Youssef Darzi, Séverine Vermeire, Gwen Falony, and Jeroen Raes. 2017. "Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load." *Nature* 551(7681):507–11.

Vermeire, S., G. Van Assche, and P. Rutgeerts. 2006. "Laboratory Markers in IBD: Useful, Magic, or Unnecessary Toys?" *Gut* 55(3):426–31.

Vermeire, S., M. Peeters, R. Vlietinck, S. Joossens, E. Den Hond, V. Bulteel, X. Bossuyt, B. Geypens, and P. Rutgeerts. 2001. "Anti-Saccharomyces Cerevisiae Antibodies

(ASCA), Phenotypes of IBD, and Intestinal Permeability: A Study in IBD Families." *Inflammatory Bowel Diseases* 7(1):8–15.

Vital, Marius, Adina Chuang Howe, and James M. Tiedje. 2014. "Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)Genomic Data." *MBio* 5(2).

de Vos, Willem M. 2017. "Microbe Profile: Akkermansia Muciniphila: A Conserved Intestinal Symbiont That Acts as the Gatekeeper of Our Mucosa." *Microbiology (United Kingdom)*.

Walker, Alan W., Jeremy D. Sanderson, Carol Churcher, Gareth C. Parkes, Barry N. Hudspith, Neil Rayment, Jonathan Brostoff, Julian Parkhill, Gordon Dougan, and Liljana Petrovska. 2011. "High-Throughput Clone Library Analysis of the Mucosa-Associated Microbiota Reveals Dysbiosis and Differences between Inflamed and Non-Inflamed Regions of the Intestine in Inflammatory Bowel Disease." *BMC Microbiology* 11.

Walker, John R., Jason P. Ediger, Lesley A. Graff, Jay M. Greenfeld, Ian Clara, Lisa Lix, Patricia Rawsthorne, Norine Miller, Linda Rogala, Cory M. McPhail, and Charles N. Bernstein. 2008. "The Manitoba IBD Cohort Study: A Population-Based Study of the Prevalence of Lifetime and 12-Month Anxiety and Mood Disorders." *The American Journal of Gastroenterology* 103(8):1989–97.

Wang, Wei, Liping Chen, Rui Zhou, Xiaobing Wang, Lu Song, Sha Huang, Ge Wang, and Bing Xia. 2014. "Increased Proportions of Bifidobacterium and the Lactobacillus Group and Loss of Butyrate-Producing Bacteria in Inflammatory Bowel Disease." *Journal of Clinical Microbiology* 52(2):398–406.

Wang, Z. K., Y. S. Yang, A. T. Stefka, G. Sun, and L. H. Peng. 2014. "Review Article: Fungal Microbiota and Digestive Diseases." *Alimentary Pharmacology & Therapeutics* 39(8):751–66.

Weingarden, Alexa R. and Byron P. Vaughn. 2017. "Intestinal Microbiota, Fecal Microbiota Transplantation, and Inflammatory Bowel Disease." *Gut Microbes* 8(3):238–52.

Wexler, Aaron G. and Andrew L. Goodman. 2017. "An Insider's Perspective: Bacteroides as a Window into the Microbiome." *Nature Microbiology* 2:17026.

Wilcoxon, Frank. 1945. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1(6):80.

Wilson, Brooke C., Tommi Vatanen, Wayne S. Cutfield, and Justin M. O. Sullivan. 2019. "The Super-Donor Phenomenon in Fecal Microbiota Transplantation." 9(January):1–11.

Woese, C. R. and G. E. Fox. 1977. "Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms." *Proceedings of the National Academy of Sciences of the United States of America* 74(11):5088–90.

Wright, Emily K., Michael A. Kamm, Shu Mei Teo, Michael Inouye, Josef Wagner, and Carl D. Kirkwood. 2015. "Recent Advances in Characterizing the Gastrointestinal Microbiome in Crohn's Disease." *Inflammatory Bowel Diseases* 21(6):1.

Yang, Ye and Christian Jobin. 2017. "Novel Insights into Microbiome in Colitis and Colorectal Cancer." *Current Opinion in Gastroenterology*.

Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N. Baldassano, Andrey P. Anokhin, Andrew C. Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J. Gregory Caporaso, Catherine A. Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I. Gordon. 2012. "Human Gut Microbiome Viewed across Age and Geography." *Nature*.

Zhang, Junyi, Xiao Ding, Rui Guan, Congmin Zhu, Chao Xu, Bingchuan Zhu, Hu Zhang, Zhipeng Xiong, Yingang Xue, Jing Tu, and Zuhong Lu. 2018. "Evaluation of Different 16S RRNA Gene V Regions for Exploring Bacterial Diversity in a Eutrophic Freshwater Lake." *Science of the Total Environment* 618:1254–67.

Zhang, Yue-Xi, Li-Ya Zhou, Zhi-Qiang Song, Jian-Zhong Zhang, Li-Hua He, and Yu Ding. 2015. "Primary Antibiotic Resistance of Helicobacter Pylori Strains Isolated from Patients with Dyspeptic Symptoms in Beijing: A Prospective Serial Study." *World Journal of Gastroenterology* 21(9):2786–92.

Zhao, Lele, Gang Wang, Paul Siegel, Chuan He, Hezhong Wang, Wenjing Zhao, Zhengxiao Zhai, Fengwei Tian, Jianxin Zhao, Hao Zhang, Zikui Sun, Wei Chen, Yan Zhang, and He Meng. 2013. "Quantitative Genetic Background of the Host Influences Gut Microbiomes in Chickens." *Scientific Reports* 3:1163.

Zhu, Xiqun, Yong Han, Jing Du, Renzhong Liu, Ketao Jin, and Wei Yi. 2017. "Microbiota-Gut-Brain Axis and the Central Nervous System." *Oncotarget* 8(32):53829–38.

Zou, Yuanqiang, Wenbin Xue, Guangwen Luo, Ziqing Deng, Panpan Qin, Ruijin Guo,

Haipeng Sun, Yan Xia, Suisha Liang, Ying Dai, Daiwei Wan, Rongrong Jiang, Lili Su, Qiang Feng, Zhuye Jie, Tongkun Guo, Zhongkui Xia, Chuan Liu, Jinghong Yu, Yuxiang Lin, Shanmei Tang, Guicheng Huo, Xun Xu, Yong Hou, Xin Liu, Jian Wang, Huanming Yang, Karsten Kristiansen, Junhua Li, Huijue Jia, and Liang Xiao. 2019. “1,520 Reference Genomes from Cultivated Human Gut Bacteria Enable Functional Microbiome Analyses.” *Nature Biotechnology* 37(2):179–85.

ANNEXES

ANNEX I: Grants

Amics del VHIR grant to develop a PhD thesis

ANNEX 2: A microbial signature for Crohn's disease.

Paper published with the data of this thesis.

Pascal V*, **Pozuelo M***, Borruel N, Casellas F, Campos D, Santiago A, Martinez X, Varela E, Sarrabayrouse G, Machiels K, Vermeire S, Sokol H, Guarner F, Manichanh C.

Gut. 2017 May;66(5):813-822. doi: 10.1136/gutjnl-2016-313235. Epub 2017 Feb 7.

OPEN ACCESS

ORIGINAL ARTICLE

A microbial signature for Crohn's disease

Victoria Pascal,[1] Marta Pozuelo,[1] Natalia Borruel,[1,2] Francesc Casellas,[1,2] David Campos,[1] Alba Santiago,[1] Xavier Martinez,[1] Encarna Varela,[1] Guillaume Sarrabayrouse,[1] Kathleen Machiels,[3] Severine Vermeire,[3] Harry Sokol,[4] Francisco Guarner,[1,2] Chaysavanh Manichanh[1,2]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2016-313235).

[1]Department of Gastroenterology, Vall d'Hebron Research Institute, Barcelona, Spain
[2]CIBERehd, Instituto de Salud Carlos III, Madrid, Spain
[3]Department of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium
[4]Department of Gastroenterology, AP-HP, Hôpital Saint-Antoine, Paris, France

Correspondence to
Dr Chaysavanh Manichanh, Department of Gastroenterology, Vall d'Hebron Research Institute, Pg Vall d'Hebron, Barcelona 119-129, Spain; cmanicha@gmail.com

VP and MP share co-first authorship

Received 14 October 2016
Revised 22 December 2016
Accepted 28 December 2016
Published Online First 7 February 2017

ABSTRACT

Objective A decade of microbiome studies has linked IBD to an alteration in the gut microbial community of genetically predisposed subjects. However, existing profiles of gut microbiome dysbiosis in adult IBD patients are inconsistent among published studies, and did not allow the identification of microbial signatures for CD and UC. Here, we aimed to compare the faecal microbiome of CD with patients having UC and with non-IBD subjects in a longitudinal study.

Design We analysed a cohort of 2045 non-IBD and IBD faecal samples from four countries (Spain, Belgium, the UK and Germany), applied a 16S rRNA sequencing approach and analysed a total dataset of 115 million sequences.

Results In the Spanish cohort, dysbiosis was found significantly greater in patients with CD than with UC, as shown by a more reduced diversity, a less stable microbial community and eight microbial groups were proposed as a specific microbial signature for CD. Tested against the whole cohort, the signature achieved an overall sensitivity of 80% and a specificity of 94%, 94%, 89% and 91% for the detection of CD versus healthy controls, patients with anorexia, IBS and UC, respectively.

Conclusions Although UC and CD share many epidemiologic, immunologic, therapeutic and clinical features, our results showed that they are two distinct subtypes of IBD at the microbiome level. For the first time, we are proposing microbiomarkers to discriminate between CD and non-CD independently of geographical regions.

Significance of this study

What is already known on this subject?

- Microbiome in Crohn's disease (CD) is associated with a reduction of faecal microbial diversity and plays a role in its pathogenesis.
- *Faecalibacterium prausnitzii* and *Escherichia coli*, in particular, were found decreased and increased, respectively, in CD.
- No clear comparison between dysbiosis in CD and in UC has been performed.
- Longitudinal study of the intestinal microbiome in adult patients with IBD has also been poorly investigated in large cohorts.

What are the new findings?

- Dysbiosis is greater in CD than in UC, with a lower microbial diversity, a more altered microbiome composition and a more unstable microbial community.
- Different microbial groups are associated with smoking habit and localisation of the disease in CD and UC.
- Eight groups of microorganisms including *Faecalibacterium*, an unknown Peptostreptococcaceae, *Anaerostipes*, *Methanobrevibacter*, an unknown Christensenellaceae, *Collinsella* and *Fusobacterium*, *Escherichia* could be used to discriminate CD from non-CD; the six first groups being in lower relative abundance and the last two groups in higher relative abundance in CD.

How might it impact on clinical practice in the foreseeable future?

- Considering CD and UC as two distinct subtypes of IBD at the microbiome level could help designing specific therapeutic targets.
- The microbial signature specific to CD combined with either imaging techniques or calprotectin data could help decision-making when the diagnosis is initially uncertain among CD, UC and IBS.

INTRODUCTION

CD and UC, the two main forms of IBD with a similar annual incidence (10–30 per 100 000 in Europe and North America), have both overlapping and distinct clinical pathological features.[1] Given that these conditions do not have a clear aetiology, diagnosis continues to be a challenge for physicians. Standard clinical testing to diagnose CD and UC includes blood tests and stool examination for biomarker quantification, endoscopy and biopsy. The diagnosis of IBD, particularly CD, can be missed or delayed due to the non-specific nature of both intestinal and extra-intestinal symptoms at presentation. In this regard, non-invasive, cost-effective, rapid and reproducible biomarkers would be helpful for patients and clinicians alike.

Dysbiosis, which is an alteration of the gut microbial composition, has been reported in IBD over the last 10 years.[2–5] Patients with IBD, in particular patients with CD, are associated with a

To cite: Pascal V, Pozuelo M, Borruel N, *et al*. *Gut* 2017;**66**:813–822.

lower microbial α-diversity and are enriched in several groups of bacteria compared with healthy controls (HC). Using faecal samples and culture-independent techniques, including qPCR, T-RFLP, cloning/Sanger, pyrosequencing or Illumina sequencing, several studies have reported that CD is associated with a decrease in Clostridiales such as *Faecalibacterium prausnitzii* and an increase in Enterobacteriales such as *Escherichia coli*.[6–8] Patients with UC are associated to some extent with a decrease in microbial diversity; however, no strong dysbiosis has been reported compared with healthy controls or patients with CD.[5] Although many studies have revealed a clear association between an altered microbiome and IBD, they have not addressed the differences between CD and UC at the microbiome level nor have proposed a set of biomarkers that is useful for diagnosis based on stool samples.[9]

To deeply characterise the microbiome of UC and CD, we combined 669 newly collected samples with 1376 previously sequenced ones, thus building one of the largest cohorts covering sequence data generated from four countries (Spain, Belgium, the UK and Germany). Our findings reveal that CD and UC are two distinct intestinal disorders at the microbiome level. We also developed and validated a microbial signature for the detection of CD.

METHODS

Study design

We performed a cohort study (Spanish IBD cohort) to identify microbial biomarkers for CD and validated the outcome with several other published and unpublished studies: a Belgian CD cohort, a Spanish IBS cohort, a UK healthy twin cohort and a German anorexic cohort. The Belgian CD cohort was part of an unpublished study, whereas the other cohorts were from published research. For the Spanish IBD and Belgian CD cohorts, the protocols were submitted and approved by the local Ethical Committee of the University Hospital Vall d'Hebron (Barcelona, Spain) and of the University Hospital Gasthuisberg in Leuven (Belgium), respectively. All volunteers received information concerning their participation in the study and gave written informed consent.

Study population

To study differences in the microbiome composition between IBD and healthy subjects and between inactive and active disease (remission vs recurrence), 34 patients with CD and 33 patients with UC were enrolled for a follow-up study in the Spanish cohort. Inclusion criteria were a diagnosis of UC and CD confirmed by endoscopy and histology in the past, clinical remission for at least 3 months—defined by the validated colitis activity index (CAI) for UC and the CD activity index (CDAI) for CD,[10] stable maintenance therapy (either amino-salicylates, azathioprine or no drug) and previous history of at least three clinical recurrences in the past 5 years. HC were without previous history of chronic disease. At inclusion and during the follow-up (every 3 months), we collected diagnostic criteria, location and behaviour of CD, extension of UC, and clinical data including tobacco use and medical treatment. Clinical recurrence was defined by a value of 4 or higher for CAI and higher than 150 for CDAI. Blood samples were collected to assess ESR, the blood cell count and CRP. Exclusion criteria included pregnancy or breast-feeding, severe concomitant disease involving the liver, heart, lungs or kidneys, and treatment with antibiotics during the previous 4 weeks. A total of 415 faecal samples for microbiome analysis were collected from 178 participants (111 HC and 67 patients with IBD) at various time points (table 1). Patients with CD and UC who showed recurrence during the study also provided a stool sample at the time of recurrence.

In the Belgian prospective cohort, 54 patients with CD undergoing curative ileocecal resection of the diseased bowel were included at the University Hospital Leuven. Originally, patients with CD were enrolled before ileocecal resection in order to study early triggers of inflammation and to unravel the sequence of events before and during the development of early inflammatory lesions. A total of 187 faecal samples were collected at four time points before and during the postoperative follow-up period (baseline, 1, 3 and 6 months after surgery) for microbiome analysis. Baseline characteristics are shown in table 1.

Faecal microbiome analysis

Sample collection and genomic DNA extraction

Faecal samples collected in Spain and Belgium were immediately frozen by the participants in their home freezer at −20°C for the Spanish cohort and cooled (maximum 24 hours) for the Belgian cohort and later brought to the laboratory in a freezer pack, where they were stored at −80°C. Genomic DNA was extracted following the recommendations of the International Human Microbiome Standards (IHMS; http://www.microbiome-standards.org).[11] A frozen aliquot (250 mg) of each sample was suspended in 250 µL of guanidine thiocyanate, 40 µL of 10% *N*-lauroyl sarcosine, and 500 µL of 5% *N*-lauroyl sarcosine. DNA was extracted by mechanical disruption of the microbial cells with beads, and nucleic acids were recovered from clear lysates by alcohol precipitation. An equivalent of 1 mg of each sample was used for DNA quantification using a NanoDrop ND-1000 Spectrophotometer (Nucliber). DNA integrity was examined by micro-capillary electrophoresis using an Agilent 2100 Bioanalyzer with the DNA 12 000 kit, which resolves the distribution of double-stranded DNA fragments up to 17 000 bp in length.

High-throughput DNA sequencing

For profiling microbiome composition, the hyper-variable region (V4) of the bacterial and archaeal 16S rRNA gene was amplified by PCR. On the basis of our analysis done using Primer Prospector software,[12] the V4 primer pairs used in this study were expected to amplify almost 100% of the bacterial and archaeal domains. The 5′ ends of the forward (V4F_515_19: 5′-GTGCCAGCAMGCCGCGGTAA -3′) and reverse (V4R_806_20: 5′- GGACTACCAGGGTATCTAAT -3′) primers targeting the 16S gene were tagged with specific sequences as follows: 5′-{AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGT}[12] {GTGCCAGCMGCCGCGGTAA}-3′ and 5′-{CAAGCAGAAGACGGCATACGAGAT} {Golay barcode} {AGTCAGTCAGCC} {GGACTACHVGGGTWTCTAAT}-3′. Multiplex identifiers, known as Golay codes, had 12 bases and were specified downstream of the reverse primer sequence (V4R_806_20).[13 14]

Standard PCR (0.15 units of Taq polymerase (Roche) and 20 pmol/µL of the forward and reverse primers) was run in a Mastercycler gradient (Eppendorf) at 94°C for 3 min, followed by 35 cycles of 94°C for 45 s, 56°C for 60 s, 72°C for 90 s and a final cycle of 72°C for 10 min. Amplicons were first purified using the QIAquick PCR Purification Kit (Qiagen, Barcelona, Spain), quantified using a NanoDrop ND-1000 Spectrophotometer (Nucliber) and then pooled in equal concentration. The pooled amplicons (2 nM) were then subjected to sequencing using Illumina MiSeq technology at the technical support unit of the Autonomous University of Barcelona (UAB, Spain), following standard Illumina platform protocols.

Table 1 Baseline clinical characteristics of the patients with CD and UC

Baseline clinical characteristics	CD Spanish cohort (n=34)	CD Belgian cohort (n=53)	Comparison between cohorts (p value)
Male/female (%)	13/21 (38.2/61.7)	28/25 (52.8/47.2)	0.201
Median (IQR) age at surgery (years) or at sample collection	34 (18 –58)	41.3 (26.5–52.9)	0.141
Median duration of disease (IQR) at surgery (years) or at sampling	6.5 (0–28)	15.7 (4.1–27.1)	0.0002
Maximum disease location (Montreal classification)			0.682
L1 ileal (%)	12 (35)	18 (34)	
L2 colonic (%)	0 (0)	0 (0)	
L3 ileocolonic (%)	22 (64.7)	35 (66)	
L4 isolated upper disease (%)	2 (5.8)	2 (3.8)	
Disease behaviour at surgery (Montreal classification)			0.009
B1 non-stricturing, non-penetrating (%)	3 (8.8)	2 (3.8)	
B2 stricturing (%)	22 (64.7)	21 (39.6)	
B3 penetrating (%)	5 (14.7)	30 (56.6)	
p perianal disease (%)	3 (8.8)	15 (28.3)	
Active smoking at surgery (%)	10 (29.4)	16 (30.2)	0.012
Medication at surgery or at sampling			
Mesalamine–sulfasalazine (%)	4 (11.8)	4 (7.5)	0.012
Corticosteroids (%)	2 (2.9)	10 (18.9)	0.183
Immunosuppressants (%)	14 (41.1)	12 (22.6)	0.087
Anti-TNF (%)	12 (23.5)	7 (13.2)	0.023
Antibiotics (%)	0 (0)	9 (16.9)	0.033
Methotrexate	1 (2.9)		
Other	10 (29.4)		
None	1 (2.9)		
	UC Spanish cohort 1 (n=33)	**UC Spanish cohort 2 (n=41)**	
Male/female (%)	9/24 (27.2/72.7)	17/24 (41.4/58.5)	0.595
Median (IQR) age at sample collection	43 (24–62)	43 (24–68)	0.500
Median duration of disease (IQR) at sampling	9 (1–23)	10 (1–34)	0.392
Disease behaviour at sampling			0.208
E1 proctitis	9 (27.3)	18 (43.9)	
E2 left sided colitis	11 (33.3)	10 (24.4)	
E3 pancolitis	13 (39.4)	13 (31.7)	
Medication at sampling			
Mesalamine (%)	11 (24)	26 (63.4)	0.021
Corticosteroids (%)	2 (6)	0	0.617
Immunosuppressants (%)	8 (24)	0	0.026
Other	2 (6)	3 (7.3)	0.708
None		2 (4.8)	

Comparison between cohorts have been performed; the χ^2 test was applied to categorical variables, and the t-test was applied to continuous variables; when $p<0.05$ differences were considered significant.
CD, Crohn's disease; TNF, tumour necrosis factor.

Sequence data analysis

For microbiome analysis, we first loaded the raw sequences into the QIIME 1.9.1 pipeline, as described by Navas-Molina *et al*.[14] The first step was to filter out low quality sequence reads by applying default settings and a minimum acceptable Phred score of 20. Correct primer and proper barcode sequences were also checked. After filtering, from a total of 2206 faecal samples, we obtained a total of 115.5 millions of high-quality sequences with a number of reads ranging from 1 to 223 896 per sample. We used the USEARCH[15] algorithm to cluster similar filtered sequences into Operational Taxonomic Units (OTUs) based on a 97% similarity threshold. We then identified and removed chimeric sequences using UCHIME.[16] Since each OTU can comprise many related sequences, we picked a representative sequence from each one. Representative sequences were aligned using PyNAST against Greengenes template alignment (gg_13_8 release), and a taxonomical assignment step was performed using the basic local alignment search tool to map each representative sequence against a combined database encompassing the Greengenes and PATRIC databases. The script make_phylogeny.py was used to create phylogenetic trees using the FastTree programme.[17] To correctly define species richness for the analysis of between-sample diversity, known as β diversity, the OTU table was rarefied at 6760 sequences per sample and kept for further analysis a total of 2045 samples and 115.5 millions of reads. Rarefaction is used to overcome cases in which read counts are not similar in numbers between samples. The summarise taxa table was used to classify taxa from the Domain to the Species level. To provide community α diversity estimates, we calculated the Chao1 and Shannon diversity indexes.[18 19] To

calculate between-sample diversity, weighted and unweighted UniFrac metrics were applied to build phylogenetic distance matrices, which were then used to construct hierarchical cluster trees using Unweighted Pair Group Method with Arithmetic mean and Principal Coordinate Analysis (PcoA) representations.

Statistical analyses

Statistical analyses were carried out in QIIME and in R. To work with normalised data, we analysed an equal number of sequences from all groups. The Shapiro-Wilk test[20] was used to check the normality of data distribution. Parametric normally distributed data were compared by Student's t-test for paired or unpaired data; otherwise, the Wilcoxon signed rank test was used for paired data and the Mann-Whitney U test for unpaired data. The Kruskal-Wallis one-way test of variance[21] was used to compare the mean number of sequences of the groups, that is, that of different groups of patients based on distinct parameters with that of HC, at various taxonomic levels. The Friedman test was used for one-way repeated measures of analysis of variance. We used the mixed-analysis of variance (ANOVA), a mixed-design ANOVA model, to take into account that repeated measurements are collected in a longitudinal study in which change over time is assessed. We performed analyses with the non-parametric multivariate ANOVA (NPMANOVA) called the adonis test, a non-parametric analysis of variance, to test for differences in microbial community composition. We applied Multivariate Association with Linear Models to find associations between clinical metadata (age, body mass index (BMI), gender, smoking habits, medication intake and site of disease) and microbial community abundance. When possible, the analysis provided false discovery rate (FDR)-corrected p values. FDR<0.05 considered significant for all tests.

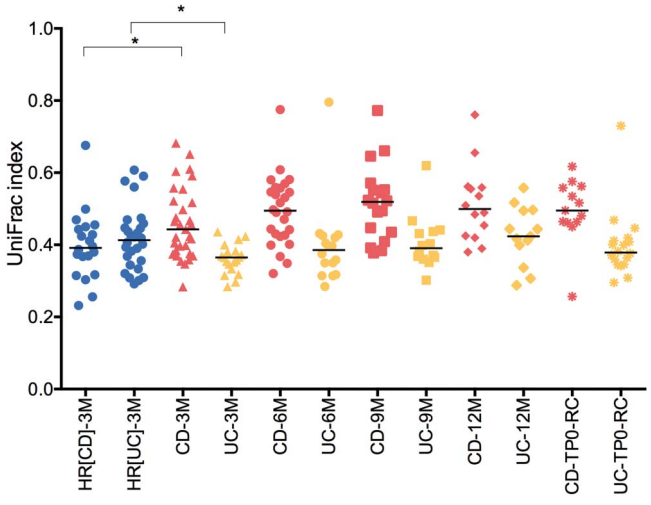


Figure 1 Microbiome stability. Unweighted UniFrac distances were calculated between different time periods for healthy relatives HR(CD) (relatives of patients with CD), HR(UC) (relatives of patients with UC), and patients with CD and UC (3M, 3 months; 6M, 6 months; 9M, 9 months; 12M, 12 months). CD-RC and UC-RC refer to samples collected during recurrence onset. At 3-month interval, patients with CD and UC presented significant differences in their UniFrac indexes compared with their HR (Mann-Whitney U test, *p=0.01). We compared the UniFrac indexes obtained between samples collected at baseline and the rest of the time points using the mixed-design ANOVA model and found that the microbiome of patients with CD was significantly more unstable than that of patients with UC (mixed-ANOVA, p<0.001). CD, Crohn's disease.

Faecal calprotectin assay

Faecal calprotectin (FC) was measured as a marker of intestinal inflammation in a subset of the Spanish participants using a commercial ELISA (Calprest; Eurospital SpA, Trieste, Italy), following the manufacturer's instructions. Optical densities were read at 405 nm with a microplate ELISA reader (Multiskan EX; Thermo Electron Corporation, Helsinki, Finland). Samples were tested in duplicate, and results were calculated from a standard curve and expressed as µg/g stool.

Validation of the microbiomarkers

Investigators interested in testing our algorithm on their own patient cohort and unable to apply by themselves the described method are invited to contact us using our dedicated email (cdmicrobiomarkers@gmail.com) to have their data processed.

RESULTS

CD more dysbiotic than UC

To characterise the microbial community of IBD we enrolled 178 participants (40 HC non-related to the patients, and 34 patients with CD and 33 patients with UC, and 36 and 35 healthy relatives (HR) of the patients with CD and UC, respectively) in a longitudinal study (discovery cohort). HR were patients' first-degree relatives. However, information on whether they were living in the same house as the patients at the time of sampling was not available. Non-related HC provided a faecal sample at a single time point, whereas HR provided two samples within a 3-month interval. Patients with UC and CD in remission provided samples at 3-month intervals over a 1-year follow-up. When the patients with IBD developed recurrence, they provided a faecal sample at the onset. During the 1-year follow-up, 13 patients with CD (38%) and 18 patients with UC (54%) developed recurrence. A total of 415 samples were collected for microbiome analysis.

Using the weighted UniFrac distance, a metric used for comparing microbial community composition between samples, we evaluated the stability of the microbiome of patients with UC and CD over time, comparing samples at baseline with the following time points: 3, 6, 9 and 12 months. Over a 3-month interval, patients with CD, but not patients with UC, showed higher UniFrac distances compared with Healthy relatives (HR) (Mann-Whitney test, p=0.01), thereby indicating a higher instability of the CD microbiome compared with controls (figure 1). Conversely, patients with UC presented a more stable microbiome than their relatives (Mann-Whitney test, p=0.015). Furthermore, over 1-year follow-up, we compared the UniFrac distances obtained between samples collected at baseline and the rest of the time points using the mixed-design ANOVA model, a repeated measures analysis of variance. The results showed that the microbiome of patients with CD was significantly more unstable than that of patients with UC (mixed-ANOVA, p<0.001).

We performed a multivariate analysis of variance on distance matrices (weighted and unweighted UniFrac) using the NPMANOVA test. The microbial community of the two groups of controls (relatives (HR) and non-relatives (HC)) were not significantly different from each other (p=0.126 for weighted and unweighted UniFrac distances), except for one genus. *Collinsella* was more abundant (Kruskal-Wallis test, 52×10^{-5} vs 1.7×10^{-5}; FDR=1.6×10^{-5}) in HR compared with HC. Conversely, the microbiome of patients with CD and UC was significantly different from that of controls (relatives and non-relatives (All-HC)) (NPMANOVA test; p=0.001 for weighted and unweighted

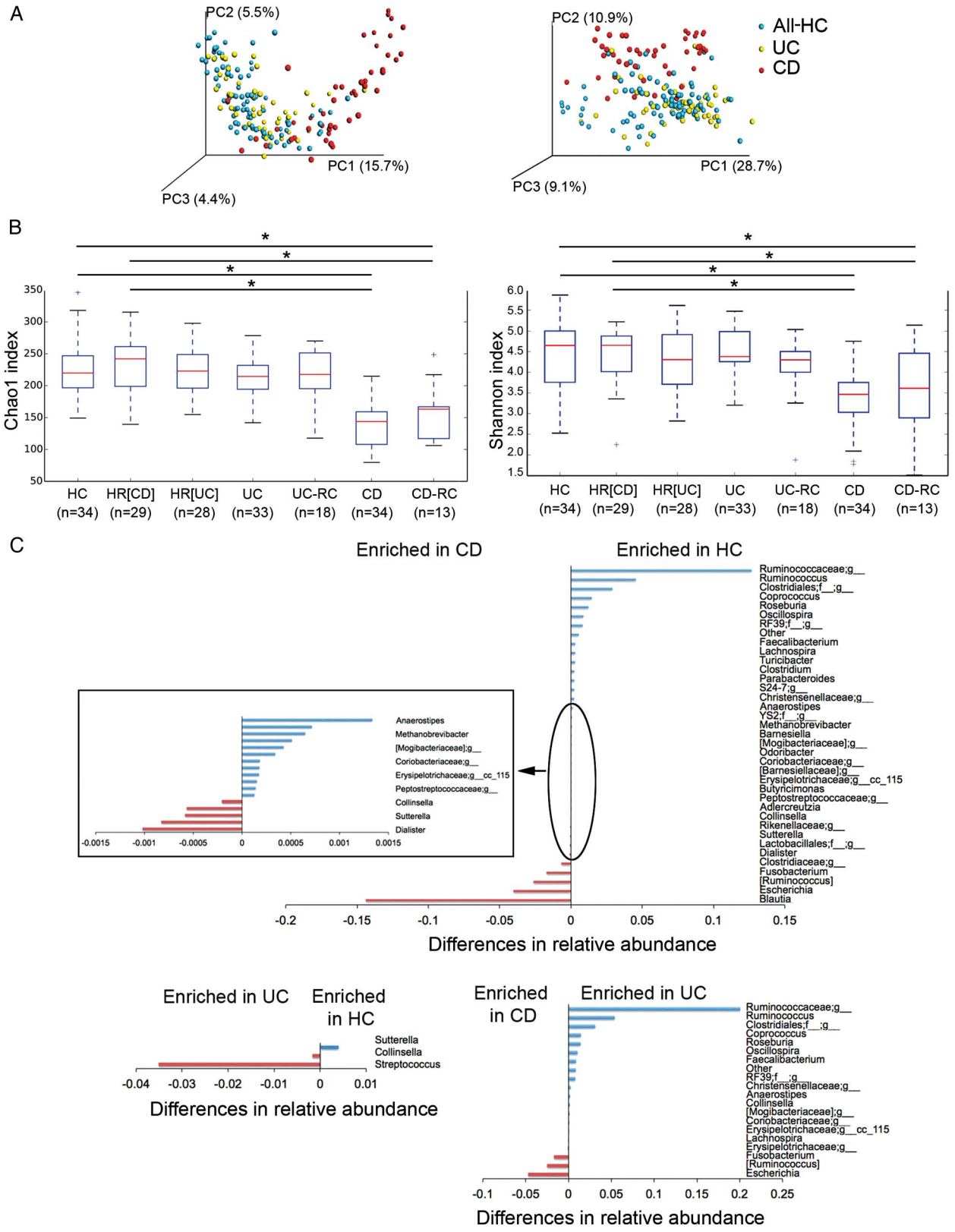


Figure 2 Dysbiosis in patients with IBD. (A) Microbiome clustering based on unweighted (left) and weighted (right) Principal Coordinate Analysis-UniFrac metrics. Significant differences were observed between all controls (All-HC, combining HC, healthy relatives HR(CD) and HR(UC)) and patients with CD (NPMANOVA test; p=0.001 for weighted and unweighted UniFrac indexes) and between all controls and patients with UC (NPMANOVA test, p=0.001 for unweighted and p=0.004 for weighted UniFrac). Microbial richness was calculated based on the Chao1 index (B, left) and microbial richness and evenness on the Shannon index (B, right). Using the Student's t-test, the microbiome of patients with CD presented significantly lower richness and evenness than healthy controls (HC, HR(CD), and HR(UC)) and patients with UC, but patients in remission and in recurrence (CD-RC and UC-RC) did not present significant differences. *p<0.05. (C) Taxonomic differences were detected between HC and UC and between HC and CD using Kruskal-Wallis test (corrected p values; false discovery rate <0.01). CD, Crohn's disease; NPMANOVA, non-parametric multivariate analysis of variance.

UniFrac distances for CD; p=0.001 for unweighted and p=0.004 for weighted UniFrac distances for UC) (figure 2A). Patients with CD and UC also showed a significant difference in their microbiome (NPMANOVA test, p=0.001 for weighted and unweighted UniFrac distances). Patients with CD but not patients with UC showed a lower microbial α diversity compared with the two groups of controls (p<0.05), as reflected by the Chao1 and Shannon indexes (figure 2B).

At baseline, six genera were enriched in patients with CD compared with 12 in HC (FDR<0.003). While only two genera were enriched in patients with UC compared with one in HC (FDR<0.03), thereby suggesting that dysbiosis is also greater in CD than in patients with UC at the taxonomic level, with a significant overall alteration in 18 genera versus 3, respectively (figure 2C). In order to uncover microbial signatures of recurrence, we used the Kruskal-Wallis test to compare the faecal samples of patients with UC and CD at the time of recurrence with those of patients who remained in remission after 1 year of follow-up. We did not find significant differences. Furthermore, in order to discover the predictive value of recurrence in patients with CD and UC, using the same test, we compared the baseline faecal samples of those who developed recurrence later on (n=13 for CD and n=18 for UC) with those who remained in remission after 1 year of follow-up (n=21 for CD and n=15 for UC). The results did not reveal any biomarker predictive of recurrence either for CD or UC.

Our results indicate that a loss of beneficial microorganisms is more associated with patients with CD than a gain of more pathogenic ones. The beneficial microorganisms include those involved in butyrate production such as *Faecalibacterium*,[22] Christensenellaceae, *Methanobrevibacter* and *Oscillospira*. Our findings confirm the results of many other studies reporting the lower relative abundance of *Faecalibacterium* in patients with CD and also show that this genus is not missing in patients with UC, thus making it a useful marker to discriminate patients with CD from patients with UC. Christensenellaceae, *Methanobrevibacter* and *Oscillospira* have been correlated with subjects with a low BMI (<25),[23–25] and they may interact with the gut immune system to maintain homeostasis. Potential pathogenic microorganisms, termed pathobionts, include *Fusobacterium* and *Escherichia*. The former is associated with infections[26] and colorectal cancer[27 28] and the latter with IBD.[8 29]

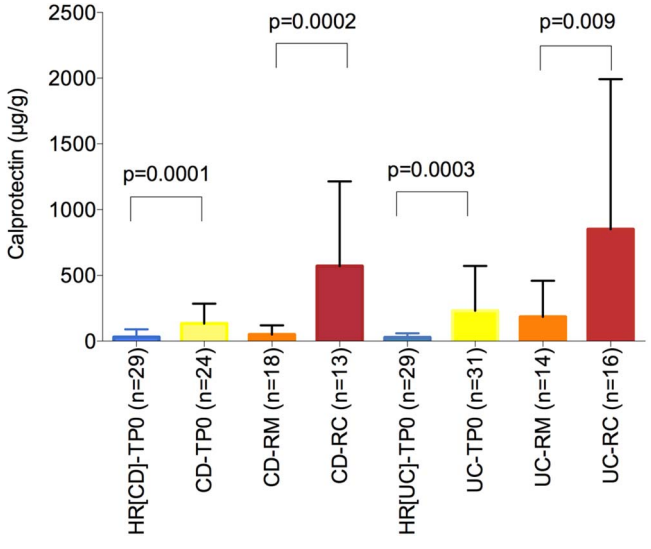


Figure 3 Calprotectin: biomarker of inflammation. Calprotectin was measured in the stool of healthy relatives of CD (HR(CD)) and UC (HR (UC)) patients, and in the stool of patients with CD and UC at baseline (TP0) and after 1-year in remission (RM) and at recurrence (RC). The Mann-Whitney test was used to compare differences between groups. CD, Crohn's disease.

Relation between microbiome, smoking habit and clinical data

Previous works have shown that smoking habit is associated with IBD.[30] Therefore, we tested the link between smoking and disease severity (remission and recurrence) using the χ^2 test. We found no link between being a smoker or ex-smoker and disease severity. We then studied the association between relative abundance of groups of bacteria and smoking habit using the Kruskal-Wallis test. In patients with CD, a genus belonging to Peptostreptococcaceae was present in a higher proportion in smokers (FDR=0.006), while *Eggerthella lenta* was found in a higher proportion in non-smokers (see online supplementary material 1). In patients with UC, we observed that smokers presented a greater abundance of *Butyricimonas*, *Prevotella* and Veillonellaceae (FDR<0.04), while non-smokers had a higher proportion of Clostridiaceae and *Bifidobacterium adolescentis* (FDR<0.03). We also examined the link between the relative abundance of groups of bacteria and disease localisation for CD and extension for UC (obtained by the Montreal classification).[31] In patients with CD, the disease was localised mostly in the ileum (L1, 35%) and in the ileocolon (L3, 64.7%). The Mann-Whitney test revealed that *Enterococcus faecalis* and an unknown species belonging to Erysipelotrichaceae were more abundant in stool when the disease was localised in the ileum than in the ileocolon. In patients with UC, the distribution of disease behaviour at sampling was as follows: proctitis (E1, 27.3%), left-sided colitis (E2, 33.3%) and pancolitis (39.4%). Using the Kruskal-Wallis test, we correlated disease behaviour and microbial community composition and found that proctitis was associated with a greater relative abundance of an unknown Clostridiales, *Clostridium*, an unknown Peptostreptococcaceae and Mogibacteriaceae (FDR<0.05) in stool. Finally, we did not find any relation between the medication use (table 1) and microbiome composition.

Microbial marker discovery

The effectiveness of FC to measure IBD activity was assessed on a subset of faecal samples (from the discovery cohort) provided by 122 participants (figure 3). For patients with CD and UC, FC was measured at baseline and either after 1-year in remission or at recurrence. During remission, FC was significantly higher in patients with CD and UC than in their HR and significantly higher during recurrence than during remission (figure 3). However, FC concentration did not differ between patients with CD and UC, either during remission or at recurrence, making them useless to discriminate the two disorders.

Groups of microbes that presented most significant differences between CD and UC and between CD and HC using the Kruskal-Wallis (FDR<0.05) test were selected to develop an algorithm with the potential to discriminate CD and non-CD (figure 4A). This algorithm retains samples that: "do not contain *Faecalibacterium*, or Peptostreptococcaceae;g, *Anaerostipes* and Christensenellaceae;g or contain *Fusobacterium* and *Escherichia* but not *Collinsella* and *Methanobrevibacter*". *Faecalibacterium*, an unknown genus of Peptostreptococcaceae, *Anaerostipes*, *Methanobrevibacter* and an unknown genus of Christensenellaceae were abundant in HC and UC and absent or almost absent in CD ones, while *Fusobacterium* and *Escherichia* were abundant in

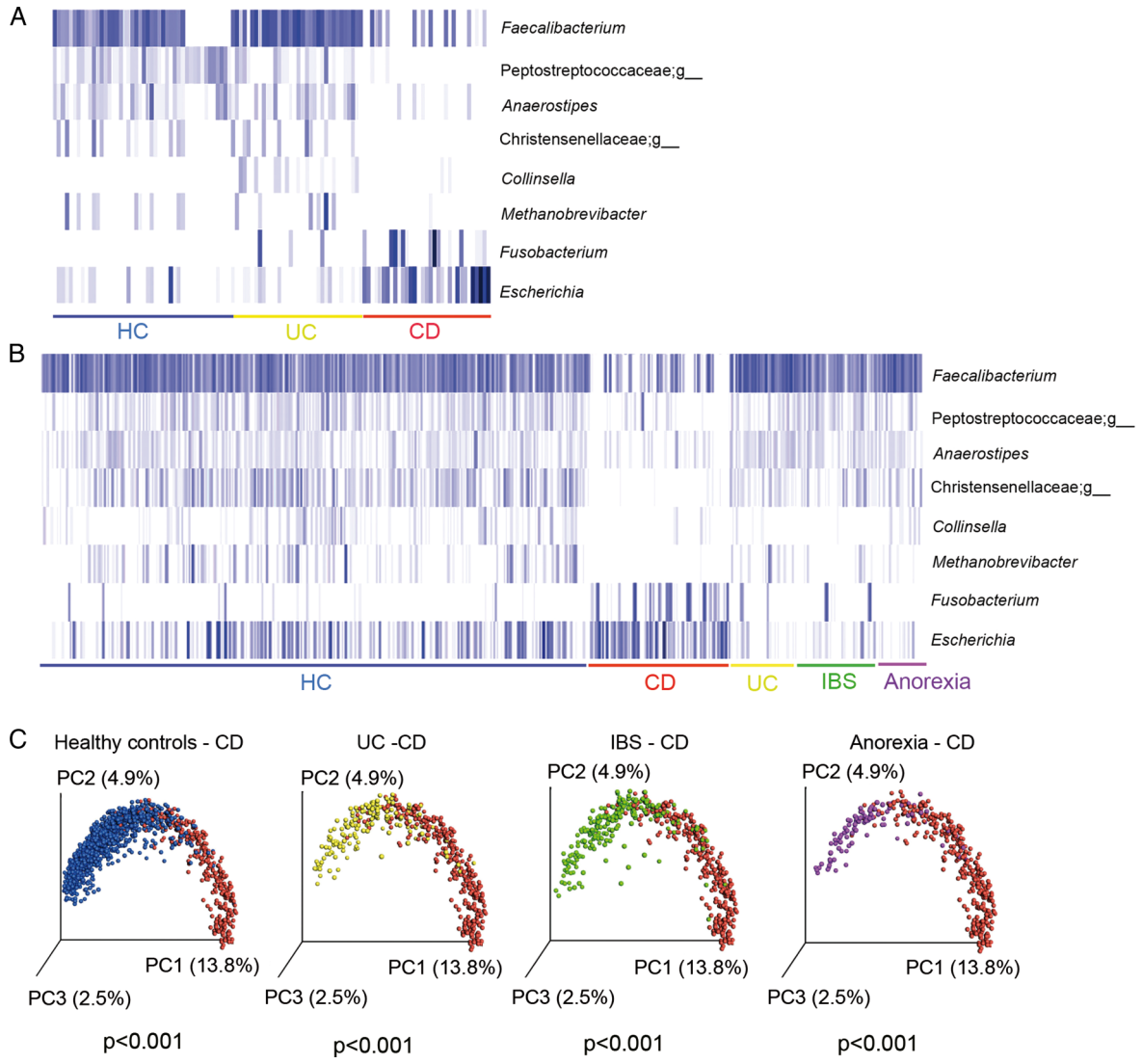


Figure 4 Microbial marker discovery and validation. Eight bacterial genera showed potential to discriminate between HC (unrelated HC) and patients with CD and UC in the discovery cohort: 34 HC, and 33 patients with UC and 34 patients with CD (A) and in the validation cohort of 2045 faecal samples from HC (n=1247), CD (n=339), UC (n=158), IBS (n=202) and anorexia (n=99) (B). Each blue bar represents the presence of each microbial group for each subject. Participants in each group are underlined with a specific colour code (blue=all HC; red=CD; yellow=UC; green=IBS and purple=anorexia). The plot was performed using an R script on relative abundance of the eight bacterial genera. The gradient of colours for the bars corresponds to white=absent, clear blue=low abundance and dark=high abundance. (C) Unweighted UniFrac Principal Coordinate Analysis representation of the various groups of subjects: HC=unrelated healthy controls, CD, Crohn's disease, Significant differences were found between CD and HC, UC, IBS and anorexia (NPMANOVA test, p<0.001). NPMANOVA, non-parametric multivariate analysis of variance.

patients with CD and almost absent in HC and UC. *Collinsella*, which was found mostly in UC cases, allowed us to discriminate between UC and CD. With these eight genera, we implemented the algorithm to identify patients with CD.

Using this algorithm, we first tested its performance on the rest of our sample set collected 3 months after baseline from relatives of HC (167 samples), and 3, 6, 9 and 12 months after baseline for patients with IBD (135 samples for CD and 135 for UC). We obtained an average of 77.7% of true positives for CD detection and an average of 7.3% and 12.8% of false positives for the detection of HC and UC, respectively (table 2). Therefore, the diagnostic accuracy for distinguishing patients with CD from HC and from patients with UC was 85.1% and 82.4%, respectively. Of the 34 patients with CD, the median duration of the disease at sampling was 6.5 years. For four patients, the diagnosis of the disease was made the same year as the sampling, and the algorithm was able to detect three of them (75%).

We validated our method with several unpublished and published data. To evaluate the sensitivity of the markers, we analysed a cohort of 54 patients with CD recruited at the University Hospital Leuven (Belgian CD cohort). Microbial DNA extraction, 16S rRNA gene amplification and sequencing and data analysis were performed in our laboratory in Spain. We generated about 5.2 million high-quality sequence reads for the 187 samples. We applied our algorithm to the whole cohort and identified an overall sensitivity of 81.8% of the samples as being CD (true positive) (table 2). Furthermore, to evaluate the predictive value of recurrence, we performed a Kruskal-Wallis analysis of the faecal samples collected before surgery, comparing patients on the basis of their Rutgeerts scores obtained 6 months after surgery. The results showed that patients who

Table 2 Detection of CD markers in HC, CD, UC, IBS, subjects with anorexia

Cohort	Number of samples	Detected	% detected	95% CI*
Discovery cohort: IBD Spain				
HC*	40	2	5	0.6 to 16.9
HC-CD_Baseline*	36	3	8.3	1.8 to 22.5
HC-CD_3M*	27	0	0	0 to 12.8
HC-UC_Baseline*	35	5	14.3	4.8 to 30.3
HC-UC_3M*	29	1	3.4	0.1 to 17.8
CD-Baseline†	34	27	79.4	62.1 to 91.3
CD-3M†	32	24	75	56.6 to 88.5
CD-6M†	27	22	81.5	61.9 to 93.7
CD-9M†	21	15	71.5	47.8 to 88.7
CD-12M†	21	17	81	58.1 to 94.6
UC-Baseline*	33	4	12.1	3.4 to 28.2
UC-3M*	26	2	7.7	0.9 to 25.1
UC-6M*	20	3	15.0	3.2 to 37.9
UC-9M*	17	2	11.8	1.5 to 36.4
UC-12M*	17	3	17.6	3.8 to 43.4
Validation cohort				
CD Belgium				
CD-Baseline†	54	39	72.2	58.4 to 83.5
CD-1M-AS†	44	37	84.1	69.9 to 93.4
CD-3M-AS†	42	35	83.3	68.6 to 93.0
CD-6M-AS†	47	42	89.4	76.9 to 96.5
UC Spain				
UC*	41	2	4.9	0.6 to 16.5
IBS Spain				
IBS-Baseline*	125	7	5.6	2.3 to 11.2
IBS-3M*	77	12	15.6	8.3 to 25.6
IBD France‡				
HC*	38	2	5.3	0.6 to 17.7
CD†	146	88	60.3	51.9 to 68.3
UC	86	28	32.6	22.8 to 43.5
Healthy UK				
HC	1017	75	7.4	5.8 to 9.2
Patients with anorexia				
AN	158	9	5.6%	2.6 to 10.5

*False positive (1-specificity).
†Sensitivity (true positive).
‡The authors of this previous work used a different region of the 16S rRNA gene (V3–V5 instead of V4; the other cohorts were analysed using V4) and a different sequencing platform (Ion Torrents).
12M, 12 months; 1M-AS, 1 month after surgery; 3M, 3 months; 3M-AS, 3 months after surgery; 6M, 6 months; 6M-AS, 6 months after surgery; 9M, 9 months; CD, Crohn's disease; HC, healthy controls; HC-CD, relatives of CD; HC-UC, relatives of UC.

developed postoperative recurrence (with a Rutgeerts score of i3 and i4, n=28) harboured a higher relative abundance of *Streptococcus* (p=0.002; FDR=0.17) than those who remained in remission (with a Rutgeerts score of i0 and i1, n=26). This result suggests that the presence of *Streptococcus* in stool samples before surgery is a predictive marker of future recurrence.

To evaluate the specificity of the markers to detect CD versus UC, we analysed a cohort of 41 patients with UC enrolled at the University Hospital Vall d'Hebron (Spanish UC cohort). The study was part of a European project (MetaHIT; http://www.metahit.eu) and included patients with UC in long-term remission. Clinical information is shown in table 1. We extracted and sequenced the faecal microbiome at baseline (ie, collected before any intervention), generating 1.5 million sequence reads and tested our algorithm on this dataset. We obtained a specificity of 95.1% for the detection of CD versus UC (table 2). We also tested the specificity of our algorithm on several non-IBD published datasets, namely on IBS, subjects with anorexia and healthy subjects. IBS and CD may present common symptoms, including abdominal pain, cramps, constipation and diarrhoea, and a simple method that distinguishes CD from IBS could also help reducing unnecessary endoscopies. Therefore, we applied our algorithm to the faecal samples of 125 subjects previously diagnosed with IBS. The sequence data were obtained from a recently published study.[32] Of the 125 patients with IBS, the algorithm identified seven as being CD, thus showing only 5.6% of false positives and a specificity of 94.4% (table 2).

The algorithm was then tested against a set of 1016 faecal samples collected at King's College (London) from a cohort of 977 healthy twin individuals[23] and against 158 faecal samples obtained from HC and patients diagnosed with anorexia.[33] Comprising healthy female adult twin pairs from the UK, the former study was originally designed to evaluate how host genetic variation shapes the gut microbiome. Our algorithm detected 75 out of 1016 samples (7.3% of false positive) as

being CD, thus showing a specificity of 92.7%. The second study was designed to address dysbiosis in patients with anorexia compared with HC and to evaluate the shift in the microbial community after weight gain in patients with anorexia.[34] As shown in this study, anorexia is associated with an alteration of gut microbiome composition. In order to evaluate whether changes occur in the gut community as a result of a condition other than IBD, we tested the algorithm on this anorexic cohort. Our tool detected 9 false positives out of 158 samples, thus showing a specificity of 94.3%.

Figure 4B illustrates the profile of the 8 microbial markers in the whole dataset of 2045 faecal samples from the various conditions: HC, CD, UC, IBS and anorexia. The results clearly confirmed that CD is characterised by a different abundance profile of the eight markers compared with the other groups, as also shown by a separate clustering based on the unweighted UniFrac PcoA representation (figure 4C).

To test the accuracy of the method, we also applied it to a set of recently published data recovered from a French cohort of IBD subjects[5] although those authors used a different method to analyse the microbial community compared with our approach. In that case, they addressed a different variable region of the 16S rRNA gene (V3–V5 instead of V4) and a different sequencing platform (Ion Torrent sequencing instead of Illumina Miseq). In that study, Sokol *et al* characterised the microbiome of 235 well-phenotyped patients with IBD and 38 HC. In spite of the technical differences, we re-ran the analysis using their raw sequence data and our sequence analysis protocol (see the Methods section). Using our quality control criteria, we recovered 8.5 million high-quality sequences for 232 patients with IBD (146 CD and 86 UC) and the 38 HC. Our method showed an accuracy of 64% for the prediction of CD versus UC (60% sensitivity and 68% specificity) and of 77% for the prediction of CD versus HC (60% sensitivity and 94.8% specificity), respectively. Moreover, we noticed that this dataset does not carry any sequences belonging to the genus *Collinsella* and a very low abundance of *Methanobrevibacter*, which in our algorithm allow the differentiation between UC and CD.

CONCLUSION

Although UC and CD share many epidemiologic, immunologic, therapeutic and clinical features, our results from the microbial community analysis confirmed that they are two distinct subtypes of IBD at the microbiome level. Based on the comparison of the microbial community between HC and CD and between HC and UC, we determined, for the first time, a non-invasive test and evaluated its potential clinical utility as a screening marker for CD in adults. We first tested its performance on the Spanish IBD cohort used as the discovery cohort and validated its sensitivity on a newly enrolled Belgian CD cohort. The overall IBD cohort comprised new-onset patients with CD and IBD in remission or with active disease. We evaluated its specificity on a healthy UK twin cohort and on several cohorts of patients with non-IBD. The test showed a sensitivity of about 80% for CD, using the Spanish and Belgian cohorts, and a specificity of 94.3%, 94.4%%, 89.4% and 90.9% of CD detection versus HC, and patients with anorexia, IBS and UC, respectively. Furthermore, all the samples from the Belgian patients with CD who took antibiotics were detected by the algorithm, thereby suggesting that antibiotics intake prior to sampling did not affect detection by the algorithm. Nevertheless, the overall sensitivity of 80% obtained with the Spanish and Belgian cohorts could have been inflated as a result of the fact that we applied the algorithm to the samples independently over time. Another limitation of our analysis is that the higher accuracy of 85.4%, to detect CD versus UC, obtained using the Spanish cohort compared with the 60% with the French cohort could be explained by a difference in the methodological approach. The low accuracy obtained with the French data may point to a limitation of this method as a diagnostic tool, as the laboratories analysing the patient's microbiome should apply the method used in this study. This finding also demonstrates the importance of the development and use of standardised methods to analyse the microbiome. Further experimental designs could be proposed to evaluate the extent to which the method used here could be implemented in a laboratory.

The rapid gathering of information on the human gut microbiome, which is the collective genomes of the gut microbiota, has been possible thanks to the following: advances in culture techniques, thus allowing a full picture of the microbial diversity present in a biological sample; the development of new sequencing technologies, which led to an exponential decrease in sequencing costs and the emergence of powerful bioinformatics tools to analyse sequence data. Together, these developments have allowed us to perform the microbiome analysis of a faecal sample for less than 150 euros on a small scale and in 1 day. On a larger scale the cost could be significantly reduced.

The non-invasive diagnostic tool described herein may be valuable when assessing patients with non-specific signs and symptoms suggestive of IBD, thereby facilitating clinical decision-making when the diagnosis of CD is initially uncertain. Indeed, this tool could be combined with either imaging techniques or calprotectin data to confirm diagnosis.

Acknowledgements We thank Julia Goodrich for sharing information on her published twin study from UK, Santiago Perez-Hoyos for his advice on statistical analysis and Andreu Schoenenberger for his statistical analyses.

Contributors CM: Study concept and design; FC, NB, FG and SV: acquisition of samples; AS, DC, GS, KM, EV and HS: acquisition of data; VP, MP and XM: analysis of data; CM: interpretation of data; CM: drafting of the manuscript; FG, HS, SV and CM: critical revision of the manuscript for important intellectual content; VP, MP: statistical analysis; CM: obtained funding). All the authors contributed to manuscript revision.

Funding This study was supported by two grants from the Instituto de Salud Carlos III/FEDER (CP13/00181, PI14/00764). KM is a postdoctoral fellow and SV a senior clinical investigator of the Fund for Scientific Research Flanders, Belgium (FWO-Vlaanderen).

Competing interests None declared.

Patient consent Obtained.

Ethics approval Local Ethical Committee of the University Hospital Vall d'Hebron in Barcelona and the University Hospital Gasthuisberg in Leuven.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol* 2015;12:720–7.
2 Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382–92.
3 Manichanh C, Rigottier-Gois L, Bonnaud E, *et al*. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;55:205–11.
4 Ott SJ, Musfeldt M, Wenderoth DF, *et al*. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 2004;53:685–93.
5 Sokol H, Leducq V, Aschard H, *et al*. Fungal microbiota dysbiosis in IBD. *Gut* 2016. doi: 10.1136/gutjnl-2015-310746. [Epub ahead of print 3 Feb 2016]

6 Joossens M, Huys G, Cnockaert M, *et al*. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;60:631–7.
7 Varela E, Manichanh C, Gallart M, *et al*. Colonisation by Faecalibacterium prausnitzii and maintenance of clinical remission in patients with ulcerative colitis. *Aliment Pharmacol Ther* 2013;38:151–61.
8 Wright EK, Kamm MA, Teo SM, *et al*. Recent advances in characterizing the gastrointestinal microbiome in Crohn's disease: a systematic review. *Inflamm Bowel Dis* 2015;21:1219–28.
9 Dubinsky M, Braun J. Diagnostic and prognostic microbial biomarkers in inflammatory bowel diseases. *Gastroenterology* 2015;149:1265–74. e3.
10 Best WR, Becktel JM, Singleton JW, *et al*. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology* 1976;70:439–44.
11 Santiago A, Panda S, Mengels G, *et al*. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol* 2014;14:112.
12 Walters WA, Caporaso JG, Lauber CL, *et al*. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* 2011;27:1159–61.
13 Caporaso JG, Lauber CL, Walters WA, *et al*. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;6:1621–4.
14 Navas-Molina JA, Peralta-Sánchez JM, González A, *et al*. Advancing our understanding of the human microbiome using QIIME. *Meth Enzymol* 2013;531:371–444.
15 Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–1.
16 Edgar RC, Haas BJ, Clemente JC, *et al*. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;27:2194–200.
17 Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–50.
18 Chao A, Chazdon RL, Colwell RK, *et al*. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 2006;62:361–71.
19 Hughes JB, Hellmann JJ, Ricketts TH, *et al*. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 2001;67:4399–406.
20 Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611.
21 Kruskal W, Wallis W. Use of ranks in one-criterion variance analysis. *J. Am. Statist Assoc.* 1952;47:583–621.
22 Khan MT, Duncan SH, Stams AJ, *et al*. The gut anaerobe Faecalibacterium prausnitzii uses an extracellular electron shuttle to grow at oxic-anoxic interphases. *ISME J* 2012;6:1578–85.
23 Goodrich JK, Waters JL, Poole AC, *et al*. Human genetics shape the gut microbiome. *Cell* 2014;159:789–99.
24 Million M, Angelakis E, Maraninchi M, *et al*. Correlation between body mass index and gut concentrations of Lactobacillus reuteri, Bifidobacterium animalis, Methanobrevibacter smithii and Escherichia coli. *Int J Obes (Lond)* 2013;37:1460–6.
25 Tims S, Derom C, Jonkers DM, *et al*. Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J* 2013;7:707–17.
26 Huggan PJ, Murdoch DR. Fusobacterial infections: clinical spectrum and incidence of invasive disease. *J Infect* 2008;57:283–9.
27 Kostic AD, Gevers D, Pedamallu CS, *et al*. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* 2012;22:292–8.
28 Leung A, Tsoi H, Yu J. Fusobacterium and Escherichia: models of colorectal cancer driven by microbiota and the utility of microbiota in colorectal cancer screening. *Expert Rev Gastroenterol Hepatol* 2015;9:651–7.
29 Darfeuille-Michaud A, Boudeau J, Bulois P, *et al*. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology* 2004;127:412–21.
30 Thomas GA, Rhodes J, Green JT, *et al*. Role of smoking in inflammatory bowel disease: implications for therapy. *Postgrad Med J* 2000;76:273–9.
31 Vermeire S, Van Assche G, Rutgeerts P. Laboratory markers in IBD: useful, magic, or unnecessary toys? *Gut* 2006;55:426–31.
32 Pozuelo M, Panda S, Santiago A, *et al*. Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome. *Sci Rep* 2015;5:12693.
33 Misra M, Klibanski A. Anorexia Nervosa and Its Associated Endocrinopathy in Young People. *Horm Res Paediatr* 2016;85:147–57.
34 Mack I, Cuntz U, Grämer C, *et al*. Weight gain in anorexia nervosa does not ameliorate the faecal microbiota, branched chain fatty acid profiles, and gastrointestinal complaints. *Sci Rep* 2016;6:26752.

ANNEX 3: Other papers

In this thesis, I developed a pipeline for the analyses of 16S rDNA data that we used and validated in the following papers.

Santiago A, Sanchez E, Clark A, **Pozuelo M**, Calvo M, Yañez F, Sarrabayrouse G, Perea L, Vidal S, Gallardo A, Guarner C, Soriano G, Manichanh C. Sequential changes in the mesenteric lymph node microbiome and immune response during cirrhosis induction in rats. (mSystems. 2019 Feb 19;4(1) pii: e00278-18. doi: 10.1128/mSystems.00278-18. IF: 5.750

Mego M, Manichanh C, Accarino A, Campos D, **Pozuelo M**, Varela E, Vulevic J, Tzortzis G, Gibson G, Guarner F, Azpiroz F. Metabolic adaptation of colonic microbiota to galactooligosaccharides: a proof-of-concept-study. Aliment Pharmacol Ther. 2017 Mar;45(5):670-680. doi: 10.1111/apt.13931. IF: 7.357

Santiago A, **Pozuelo M**, Poca M, Gely C, Nieto JC, Torras X, Román E, Campos D, Sarrabayrouse G, Vidal S, Alvarado-Tapias E, Guarner F, Soriano G, Manichanh C, Guarner C. Alteration of the serum microbiome composition in cirrhotic patients with ascites. Sci Rep. 2016 Apr 26;6:25001. doi: 10.1038/srep25001. IF: 4.122

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Host-Microbe Biology

Sequential Changes in the Mesenteric Lymph Node Microbiome and Immune Response during Cirrhosis Induction in Rats

Alba Santiago,[a] Elisabet Sanchez,[b,d] Allison Clark,[a] Marta Pozuelo,[a] Miguel Calvo,[a] Francisca Yañez,[a] Guillaume Sarrabayrouse,[a] Lidia Perea,[c] Silvia Vidal,[c] Alberto Gallardo,[e] Carlos Guarner,[b,d] German Soriano,[b,d] Chaysavanh Manichanh[a,d]

[a]Department of Gastroenterology, Vall d'Hebron Research Institute, Barcelona, Spain
[b]Department of Gastroenterology, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain
[c]Department of Immunology, IIB-Sant Pau Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain
[d]CIBERehd, Instituto de Salud Carlos III, Madrid, Spain
[e]Department of Pathology, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain

ABSTRACT Whether the interaction between the gut microbiota and the immune response influences the evolution of cirrhosis is poorly understood. We aimed to investigate modifications of the microbiome and the immune response during the progression of cirrhosis. Rats were treated with carbon tetrachloride (CCl_4) to induce cirrhosis. We then assessed microbiome load and composition in stool, ileocecal contents (ICCs), mesenteric lymph nodes (MLNs), blood, and ascitic fluids (AFs) at 6, 8, and 10 weeks or ascites production and measured cytokine production in MLNs and blood. The microbiome of MLN, blood, and AF showed a distinct composition compared to that of stool and ICCs. *Betaproteobacteria* (*Sutterella*) were found associated with the appearance of a decompensated state of cirrhosis. Microbial load increased and showed a positive correlation with the relative abundance of pathobionts in the MLN of decompensated rats. Among several genera, *Escherichia* and "*Candidatus* Arthromitus" positively correlated with elevated levels of systemic proinflammatory cytokines. "*Candidatus* Arthromitus," a segmented filamentous bacteria, was detected in ICC, MLN, and AF samples, suggesting a possible translocation from the gut to the AF through the lymphatic system, whereas *Escherichia* was detected in ICC, MLN, AF, and blood, suggesting a possible translocation from the gut to the AF through the bloodstream. In the present study, we demonstrate that microbiome changes in distinct intestinal sites are associated with microbial shifts in the MLNs as well as an increase in cytokine production, providing further evidence of the role the gut-liver-immunity axis plays in the progression of cirrhosis.

IMPORTANCE Cirrhosis severity in patients was previously shown to be associated with progressive changes in the fecal microbiome in a longitudinal setting. Recent evidence shows that bacterial translocation from the gut to the extraintestinal sites could play a major role in poor disease outcome and patient survival. However, the underlying mechanisms involving the microbiota in the disease progression are not well understood. Here, using an animal model of cirrhosis in a longitudinal and multibody sites setting, we showed the presence of a distinct composition of the microbiome in mesenteric lymph nodes, blood, and ascitic fluid compared to that in feces and ileocecal content, suggesting compartmentalization of the gut microbiome. We also demonstrate that microbiome changes in intestinal sites are associated with shifts in specific microbial groups in the mesenteric lymph nodes as well as an increase in systemic cytokine production, linking inflammation to decompensated cirrhosis in the gut-liver-immunity axis.

KEYWORDS cirrhosis complication, bacterial translocation, decompensated cirrhosis, proinflammatory response

Citation Santiago A, Sanchez E, Clark A, Pozuelo M, Calvo M, Yañez F, Sarrabayrouse G, Perea L, Vidal S, Gallardo A, Guarner C, Soriano G, Manichanh C. 2019. Sequential changes in the mesenteric lymph node microbiome and immune response during cirrhosis induction in rats. mSystems 4:e00278-18. https://doi.org/10.1128/mSystems.00278-18.

Editor Catherine Lozupone, University of Colorado Denver

Address correspondence to German Soriano, gsoriano@santpau.cat, or Chaysavanh Manichanh, cmanicha@gmail.com.

A.S. and E.S. contributed equally to this work.

Microbiome changes in GI tract are associated with shifts in specific microbial groups in the mesenteric lymph nodes as well as an increase in systemic cytokine production, linking inflammation to decompensated cirrhosis in the gut-liver-immunity axis.

Received 2 November 2018
Accepted 28 January 2019
Published 19 February 2019

Cirrhosis is defined as the presence of fibrosis and regenerating nodules in the liver due to various causes such as alcohol, hepatitis viruses, metabolic syndrome, or immune dysfunction. Cirrhosis can lead to portal hypertension and liver insufficiency and their related complications, such as ascites, infections, hepatic encephalopathy, hepatorenal syndrome, variceal bleeding, and acute-on-chronic liver failure (ACLF) (1). Before the development of complications, patients are considered to have "compensated cirrhosis," and when complications develop, they are considered having "decompensated cirrhosis," which has a poorer prognosis than the compensated stage (2).

Patients with cirrhosis present alterations in their fecal microbiome composition compared to that of healthy individuals, which could be of oral origin of some potentially pathogenic species (3, 4). The cross talk between the gut microbiome and the immune system may contribute to the development of health complications, which therefore may lead to the evolution from compensated to decompensated stages (5, 6).

The intestinal microbiota, which harbors bacteria, archaea, and eukarya, is known to play a pivotal role in the development of the host immune system and in the maintenance of host intestinal homeostasis by modulating immune responses to pathogens and by maintaining the integrity of intestinal barrier functions (7). Commensal bacteria are transported from the intestines by dendritic cells (DCs) through the lymphatic system to the mesenteric lymph nodes (MLNs), which form part of the gut-associated lymphatic tissue (GALT) and act as the first line of immune defense against pathogens from the intestines (8–10). In the MLNs, bacteria are maintained at low levels by the host mucosal immune system (11). The translocation of bacteria into MLNs has been investigated in other disorders, including Crohn's disease and ulcerative colitis (12, 13), where a distinct microbial community composition was observed between the two inflammatory disorders.

Patients with cirrhosis present alterations in the gut microbiota, intestinal permeability, and immune response, leading to bacterial translocation, which is the paracellular passage of bacteria from the intestinal lumen through the intestinal wall to the MLNs or other sites (10, 11, 14–16). Bacterial translocation then activates the gut-liver-immune axis, which stimulates the induction of proinflammatory cytokines, further perpetuating increased intestinal permeability and thus bacterial translocation (17). Pathogen-associated molecular patterns (PAMPs), such as the endotoxin lipopolysaccharide (LPS) found on the cell membranes of Gram-negative bacteria, bind to pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), causing an induction of proinflammatory cytokines such as tumor necrosis factor alpha (TNF-α) and interleukin 6 (IL-6), which tend to be elevated in patients with cirrhosis (18). MLNs also produce TNF-α in response to bacterial translocation, especially in patients with ascites (9, 18, 19). Ascites is a common complication in advanced cirrhosis that is associated with a high mortality rate and is caused by portal hypertension, leading to fluid accumulation in the abdomen. It has been hypothesized that elevated TNF-α production causes hemodynamic disturbances, leading to splanchnic vasodilatation through nitric oxide synthesis stimulation. This could contribute to altered intestinal barrier function, resulting in bacterial translocation (9, 20), which has been frequently observed in cirrhotic patients with ascites. Additionally, ascites has been shown to increase the susceptibility of host bacterial infection (17), likely due to the fact that TNF-α has been shown to loosen tight junction proteins of intestinal epithelial cells, perpetuating bacterial translocation and subsequently an inflammatory response (10).

Bajaj et al. showed that cirrhosis severity in patients was associated with progressive changes in the gut microbiome in a longitudinal study (5). In fact, recent evidence showed that bacterial translocation from the intestines could play a major role in poor disease outcome and patient survival (21, 22). However, the underlying mechanisms that involve the gut microbiota in the disease progression are not well understood.

Therefore, the aims of this study were to (i) investigate the spatial and temporal changes of the composition of the microbiome in a cirrhosis rat model, (ii) evaluate changes of the microbiome related to the progression of cirrhosis, and (iii) assess the immune modulation by the microbiome detected in extraintestinal sites.

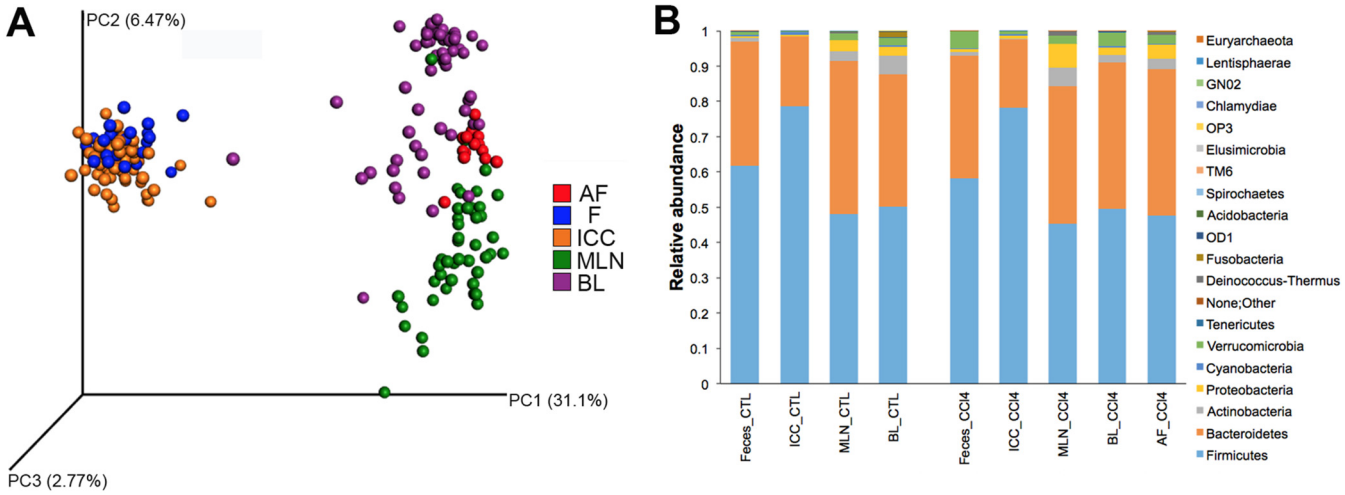


FIG 1 Spatial microbial community compositions in control and CCl_4-treated rats. (A) Weighted principal coordinate analysis (PCoA) UniFrac metrics (taxonomic clustering). (B) Relative abundance at the phylum level; $n = 23$ for stool, $n = 59$ for ICC, $n = 46$ for MLN, $n = 57$ for blood, and $n = 15$ for AF.

RESULTS

We studied 15 control rats at weeks 6 ($n = 5$), 8 ($n = 5$), and 10 ($n = 5$), 25 CCl_4-treated rats sacrificed at weeks 6 ($n = 9$), 8 ($n = 8$), and 10 ($n = 8$) of CCl_4 cirrhosis induction that were considered rats under induction of cirrhosis or compensated CCl_4-treated rats, and 19 CCl_4-treated rats when they developed ascites that were considered rats with decompensated cirrhosis. The scores of liver damage in the different groups were the following. All control rats showed score 0; in week-6 rats, six rats had scores of 2 and three had scores of 1. In week-8 rats, six rats had scores of 2 and two had scores of 1. In week-10 rats, seven rats had scores of 2 and one had a score of 1. In ascitic rats, 14 rats had scores of 2 and five had scores of 3 ($P = 0.048$ for week-10 rats, $P = 0.019$ for week-8 rats, and $P = 0.007$ for week-6 rats). Bacterial cultures were negative in all control rats, week-6 rats, and week-8 rats and were positive in two week-10 rats (both in MLNs) and in five ascitic rats ($P = 0.05$ between ascitic rats and control rats). These five ascitic rats presented a total of nine positive cultures: in MLNs in four rats, in livers and spleens in three rats, and in ascitic fluid in two rats. Isolated bacteria were *Escherichia coli* in all rats with positive cultures, except one ascitic rat in which *Enterococcus* spp. grew.

Compartmentalization of microbial communities. To assess the existence of a compartmentalized microbiome (meaning a specific microbial community at different body sites), we compared the microbiome compositions of various body sites from all rats of the study: intestinal sites such as feces (F) and ileocecal content (ICC) and extraintestinal sites such as MLN, blood, and ascitic fluid (AF). Feces and ICC specimens from control rats presented similar microbial communities ($P = 0.44$; permutational multivariate analysis of variance [PERMANOVA] test) (Fig. 1AB) dominated by *Firmicutes* (means of 61% and 78%, respectively), *Bacteroidetes* (means of 35% and 20%, respectively), *Actinobacteria* (means of 1% and 0.04%, respectively), and *Proteobacteria* (means of 0.3% and 0.4%, respectively). The microbiome of MLN, blood, and AF showed a distinctive composition compared to that of feces and ICC ($P < 0.0001$; PERMANOVA test) (Fig. 1A). MLNs and blood displayed similar microbial communities ($P = 0.616$; PERMANOVA test) dominated by *Firmicutes* (48% and 50%, respectively), *Bacteroidetes* (43% and 37%, respectively), *Proteobacteria* (3% and 2%, respectively), and *Actinobacteria* (3% and 5%, respectively) (Fig. 1B). A comparison between control and CCl_4-treated rats in the different body sites showed significant differences.

Longitudinal study: evolution of cirrhosis and microbiome modification. To evaluate the evolution of the microbiome in parallel with the progression of cirrhosis, rats were sacrificed to collect ICC, MLNs, and blood at different time points (6, 8, and

10 weeks after the initiation of CCl_4 treatment and at ascites development for the CCl_4-treated group and at matched time points for the control group). For the collection of samples, groups of animals were sacrificed at different time points. We found significant changes in the microbiome compositions in ICC and MLN samples that were associated with the progression of cirrhosis.

In the ICC samples, CCl_4-treated rats showed an increase in *Betaproteobacteria* ($P = 0.01$, $q = 0.098$) and *Erysipelotrichia* ($P = 0.001$, $q = 0.012$; Kruskal-Wallis test) at weeks 6 and 8 of CCl_4 treatment in comparison to control rats (Fig. 2A). At week 10, both groups of bacteria significantly decreased compared to that at weeks 6 and 8. When rats developed ascites, *Erysipelotrichia* almost disappeared, whereas *Betaproteobacteria* increased again. Both groups of bacteria were absent in control rats. At the genus level, *Sutterella* and *Coprococcus* increased ($P < 0.001$, $q < 0.02$; Kruskal-Wallis test) in decompensated rats compared to that in controls (Fig. 2B). At 6, 8, and 10 weeks after the initiation of CCl_4 treatment, rats without ascites presented an intermediate relative abundance of all these genera. The *Allobaculum* genus belonging to the *Erysipelotrichia* family is known as a potentially beneficial bacterial group (23) and showed an increase in the compensated CCl_4-treated rats but disappeared when they presented ascites ($P = 0.0002$, $q = 0.018$; Kruskal-Wallis test). These results suggest that groups of potentially beneficial bacteria, such as *Allobaculum*, attempted to outcompete pathogenic ones in the ileocecum such as *Sutterella*, which belongs to *Betaproteobacteria*. However, *Allobaculum* seemed to be unable to outcompete *Sutterella* in the ICC when rats presented ascites. Additionally, "*Candidatus* Arthromitus," a genus from the *Firmicutes* phylum, showed high relative abundance only in decompensated rats compared to that in compensated CCl_4-treated rats and control rats ($P = 0.023$, $q = 0.15$; Kruskal-Wallis test) (Fig. 2C).

In MLN samples from decompensated rats (with ascites), only one bacterial genus, "*Candidatus* Arthromitus," showed a significantly high relative abundance compared to that from control and compensated CCl_4-treated rats ($P = 0.0002$, $q = 0.019$; Kruskal-Wallis test) (Fig. 3A and B). "*Candidatus* Arthromitus" was identified in 26% (4 of 15) of the AF samples, in only 5% of blood samples (4 of 73), and was not found in stool samples (Fig. 3C).

In feces, *Coprococcus, Sutterella*, and *Allobaculum* showed some differences in their relative abundances between the three groups of rats (controls, compensated, and decompensated), but none of the differences were significant ($q > 0.2$; Kruskal-Wallis test). In blood samples, the *Spirochaetes* phylum was found in a high proportion in the decompensated group compared to that in the two other groups ($P = 0.001$, $q = 0.035$; Kruskal-Wallis test).

Using the weighted UniFrac distance, a metric used to compare microbial community compositions between samples, we compared the stability of the microbiome in ICC, MLN, and blood samples from rats under induction of cirrhosis without ascites (compensated CCl_4-treated rats) with that from rats with ascites (rats with decompensated cirrhosis). We observed a decreased stability of the microbiome composition of rats with ascites in blood samples ($P = 0.04$; Mann-Whitney test) (see Fig. S1 in the supplemental material), but the difference was not significant in ICC and MLN samples.

Microbial load. To complement our findings on relative abundance of the sequenced 16S rRNA genes, we evaluated the microbial load using real-time quantitative PCR of the 16S rRNA genes. Microbial load, as measured by 16S rRNA gene quantitative PCR (qPCR), was significantly higher in the MLNs from cirrhotic decompensated rats than from control rats ($P = 0.008$; Mann-Whitney test), positively correlated with the relative sequence abundance of *Proteobacteria* ($\rho = 0.673$, $P = 0.002$; Spearman test), and negatively correlated with the relative abundance of *Bacteroidetes* ($\rho = -0.637$, $P = 0.004$; Spearman test) (Fig. 4).

Correlation between microbiome and cytokine levels. The progression toward decompensated cirrhosis is associated with a high production of proinflammatory cytokines, such as TNF-α, IL-17, and IL-6, as well as anti-inflammatory cytokines such as

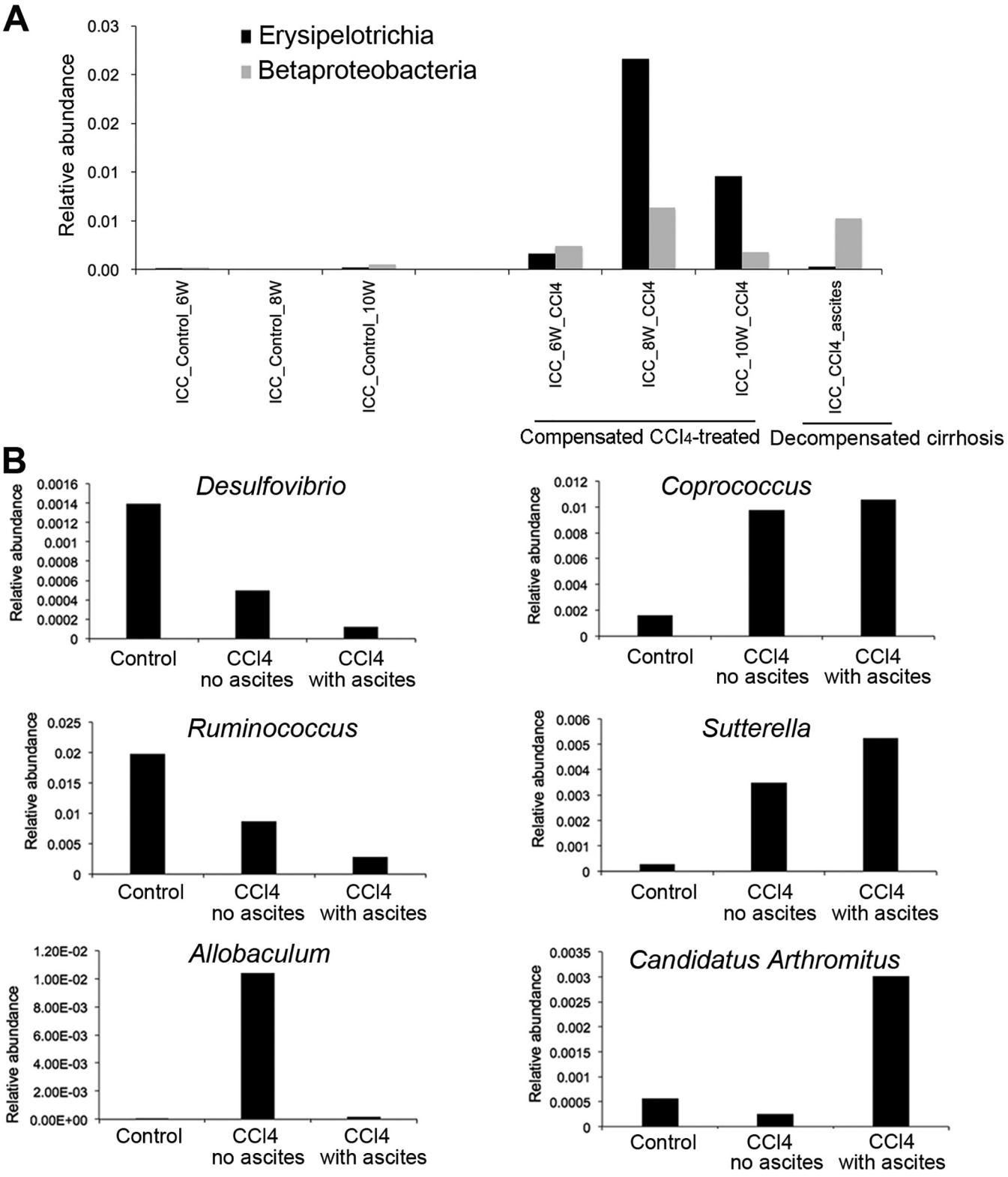


FIG 2 Microbial groups in ileocecal contents (ICC) samples involved in the severity of cirrhosis. (A) Temporal taxonomic differences between controls and CCl_4-treated rats. Two classes of bacteria, *Erysipelotrichia* ($P = 0.001$, $q = 0.012$) and *Betaproteobacteria* ($P = 0.01$, $q = 0.098$), presented significantly different relative abundances over time between the control and the CCl_4-treated groups. Statistics were performed using the Kruskal-Wallis test. (B) Taxonomic differences between controls and CCl_4-treated rats with ascites and CCl_4-treated rats before development of ascites. Two bacterial genera, *Coprococcus* ($P = 0.0001$, $q = 0.011$) and *Sutterella* ($P = 0.0005$, $q = 0.014$), were found in higher relative abundances in CCl_4-treated rats than in control rats. Two bacterial genera, *Desulfovibrio* ($P = 0.002$, $q = 0.025$) and *Ruminococcus* ($P = 0.0007$, $q = 0.013$), were found in higher relative abundances in control rats than in CCl_4-treated rats. *Allobaculum* was found in higher relative abundance ($P = 0.0004$, $q = 0.013$) only in compensated CCl_4-treated rats, and "*Candidatus* Arthromitus" ($P = 0.023$, $q = 0.15$) was in higher relative abundance in decompensated cirrhotic rats. Statistics were performed using the Kruskal-Wallis test.

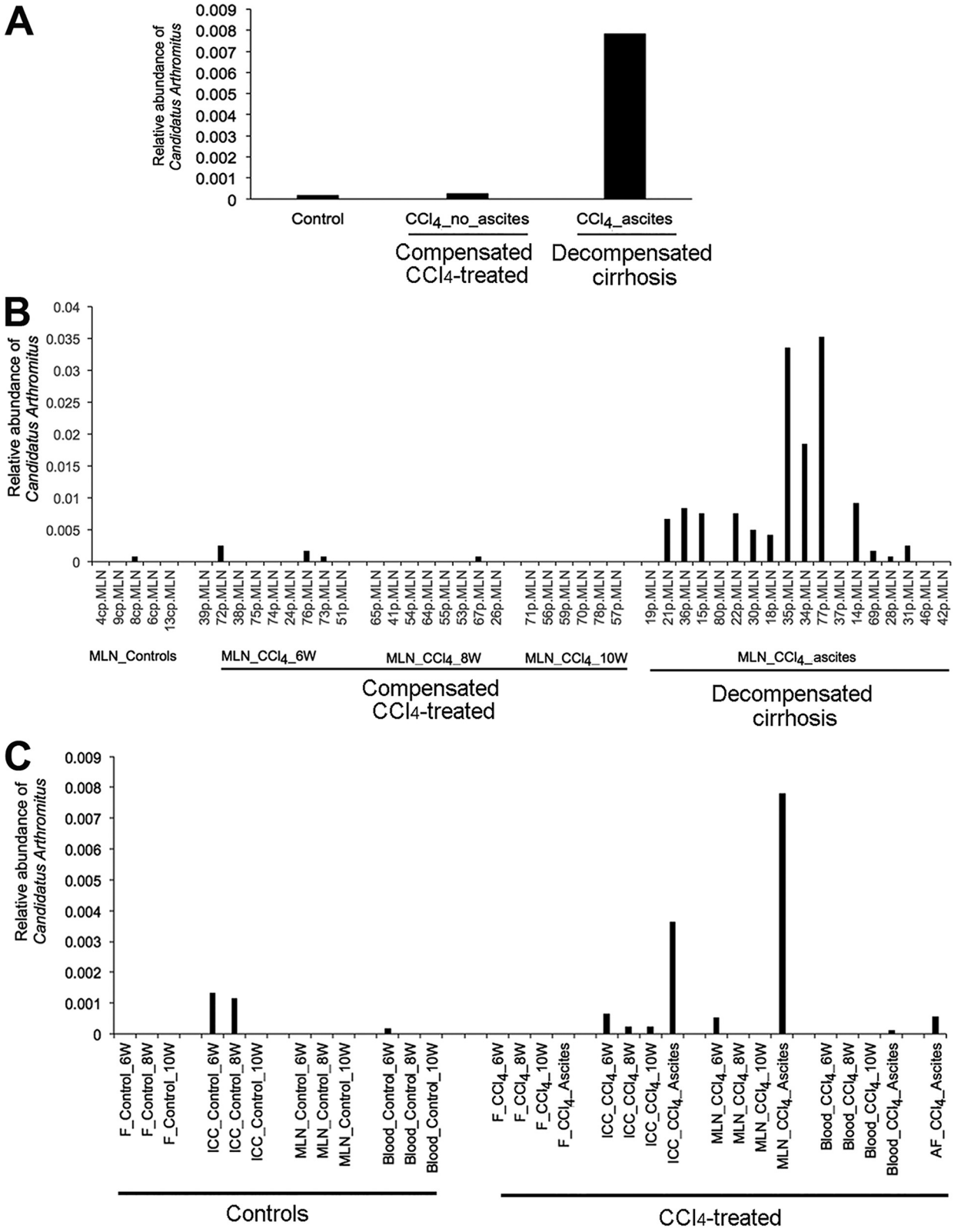


FIG 3 Detection of "*Candidatus* Arthromitus" in spatial and temporal settings. (A) "*Candidatus* Arthromitus" was found in higher relative abundance in mesenteric lymph nodes of decompensated rats ($P = 0.0002$, $q = 0.018$). (B) "*Candidatus* Arthromitus" was found in 72% (13 of 18) of mesenteric lymph node samples of decompensated rats. (C) "*Candidatus* Arthromitus" was detected in ileocecal content (ICC), mesenteric lymph node (MLN), blood, and ascitic fluid (AF) samples but not in feces (F). 6W, 6 weeks; 8W, 8 weeks; 10W, 10 weeks.

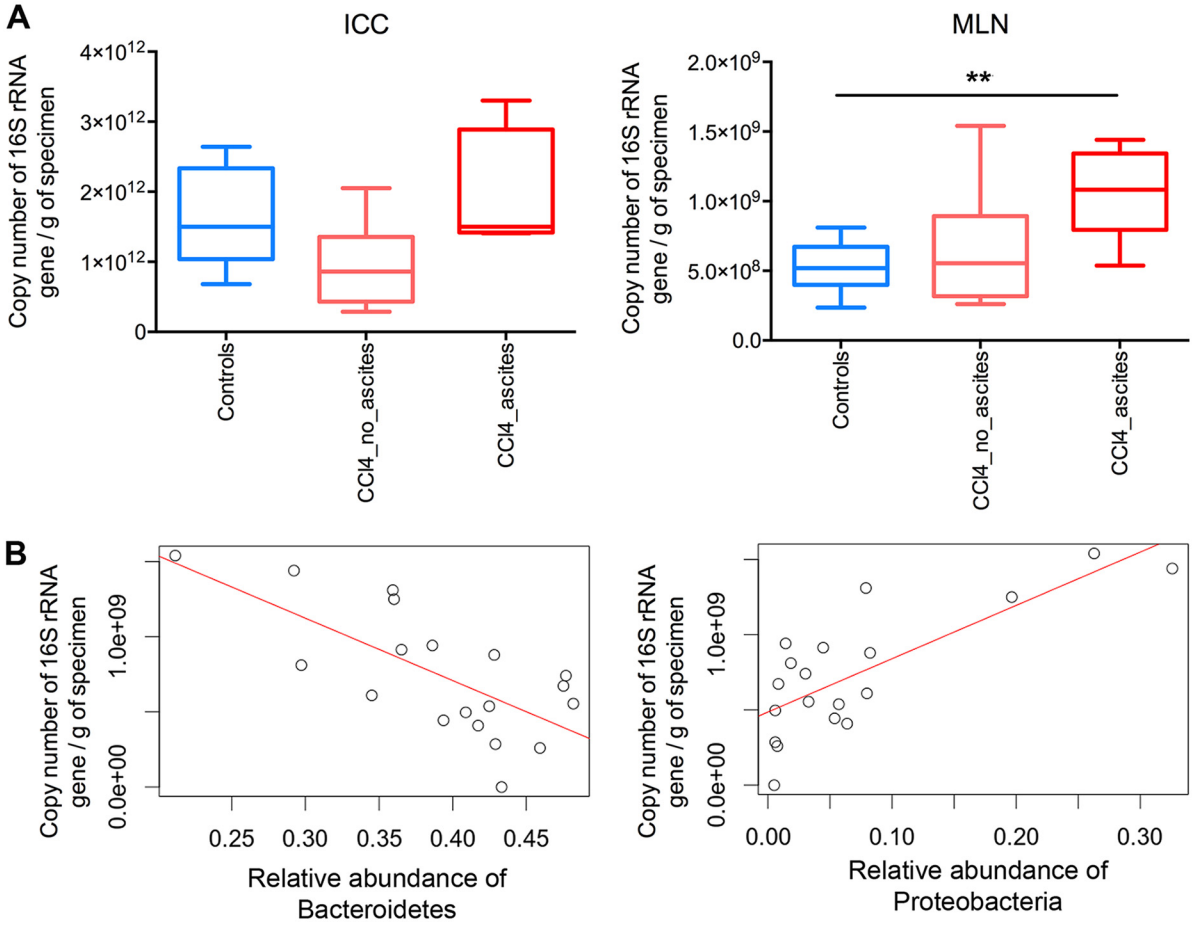


FIG 4 Quantification of the microbiota by real-time PCR on the 16S rRNA gene and correlation with microbiome composition. (A) Microbial load in ileocecal content (ICC) samples was higher in decompensated cirrhotic rats than in control and compensated CCl_4-treated rats and was significantly higher in the mesenteric lymph nodes (MLNs) of decompensated cirrhotic rats than in control rats ($P = 0.008$; Mann-Whitney test). In both ICCs and MLNs: $n = 8$ for controls (CTL), $n = 8$ for compensated CCl_4-treated rats, $n = 7$ for decompensated cirrhotic rats. (B) Spearman correlation between microbial load and relative abundance of *Bacteroidetes* ($\rho = -0.637$, $P = 0.004$) and between microbial load and relative abundance of *Proteobacteria* ($\rho = 0.673$, $P = 0.002$) in the MLNs.

IL-10 (24, 25). Using enzyme-linked immunosorbent assays (ELISAs), we measured serum and MLN levels of these cytokines. To evaluate a possible correlation between the inflammatory status and the microbial community composition, we used Spearman correlation tests to associate levels of the proinflammatory cytokine IL-17 in MLNs and the ratios of systemic IL-6/IL-10 and systemic TNF-α/IL-10 with the relative abundance of microbial groups in MLNs. Ratios between pro- and anti-inflammatory cytokines have been extensively used as biomarkers to associate an immune response with the characteristics of multiple pathologies, including liver disease. The use of ratios reduces significantly the individual variability of single cytokine production. We found a positive correlation between the IL-6/IL-10 ratio and the relative abundance of *Escherichia* ($\rho = 0.79$, $P = 9.2e{-}5$; Spearman test) and positive correlations between the TNF-α/IL-10 ratio and the relative abundances of *Escherichia* ($\rho = 0.57$, $P = 0.01$; Spearman test) and "*Candidatus* Arthromitus" ($\rho = 0.72$, $P = 0.001$; Spearman test) (Fig. 5A). Since each microorganism provides a particular array of PAMPs to signal the immune system, the final result is not necessarily a generalized increase of pro- or anti-inflammatory cytokines. In the particular case of "*Candidatus* Arthromitus," the association of one ratio but not the other (both having in common IL-10) implies that this microorganism is more involved in the production/regulation of TNF than of IL-6. We also found positive correlations between levels of IL-17 and several different microbial genera, such as those belonging to *Proteobacteria* (*Pseudomonas*, *Burkholderia* and *Sutterella*)

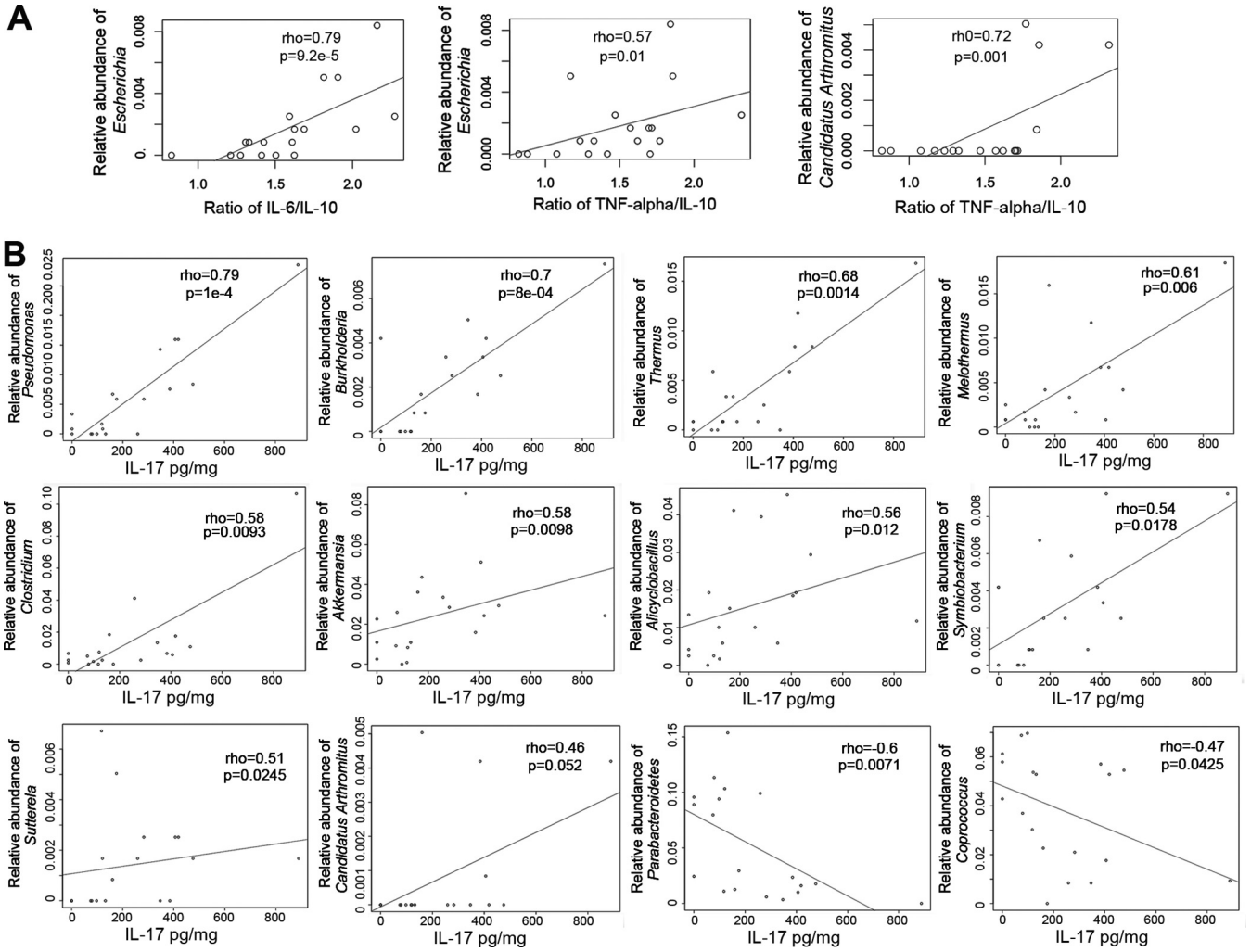


FIG 5 Correlations between proinflammatory cytokines and relative abundances of genera in MLNs. (A) Positive Spearman correlations were found between the ratio of systemic IL-6/IL-10 and *Escherichia* and between the ratio of systemic TNF-α/IL-10 and *Escherichia* and "*Candidatus* Arthromitus." (B) Spearman correlations between IL-17 levels measured in mesenteric lymph nodes (MLNs) and relative abundances of genera detected in MLNs.

and "*Candidatus* Arthromitus," and negative correlations between levels of IL-17 and *Parabacteroides* and *Coprococcus* (Fig. 5B). "*Candidatus* Arthromitus" and *Escherichia* were both detected in ICC, MLN, and AF samples, but only *Escherichia* was detected in blood samples from rats (Fig. 6).

DISCUSSION

In this study, we showed that microbiome changes in distinct intestinal sites are associated with microbial shifts in the MLNs as well as an increase in cytokine production that correlated with disease progression in rats with experimental cirrhosis. The main results of the present work are the characterization of the sequential changes of the microbiome in the progression of cirrhosis and, in particular, the cross talk between the microbiome and the host immune system using a rat model in a longitudinal and multibody sites study setting.

Loss of intestinal barrier function, dysbiosis, and systemic immune dysfunction characterize cirrhosis (17). Bacterial translocation is a result of a loss of intestinal barrier function and is considered a biomarker for cirrhosis progression and decompensation, in which intestinal bacteria travel by paracellular transport through the permeable epithelial cells to the portal vein, the liver, and systemic circulation, causing an inflammatory response (10). It has been demonstrated that microorganisms are also

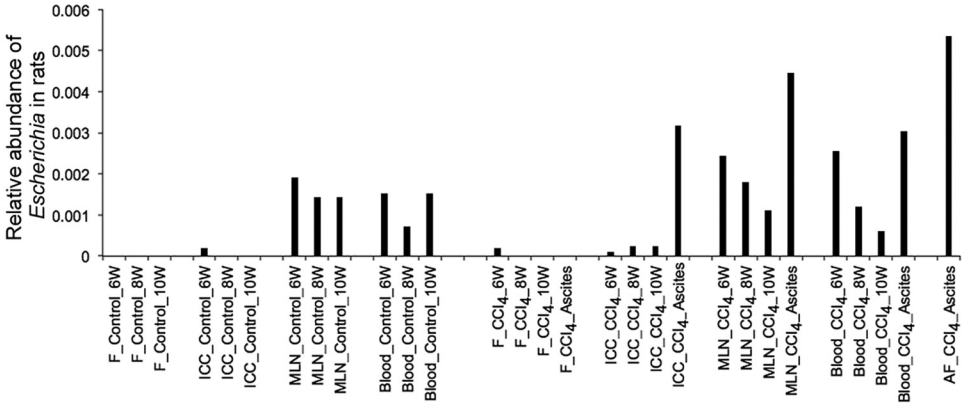


FIG 6 Detection of 16S sequences of *Escherichia* in rat samples. ICC, ileocecal content; MLN, mesenteric lymph node; AF, ascitic fluid; F, feces.

transported by dendritic cells from the intestines to the MLNs via the lymphatic system (11, 26, 27). Comparing the compositions of intestinal and extraintestinal body sites, our findings suggest a loss of barrier function in which a specific microbial community, particularly *Proteobacteria* and *Actinobacteria*, was transported from the intestine to MLNs, blood, and AF, which caused an induction of proinflammatory cytokines such as TNF-α, IL-6, and IL-17 in rats. This inflammatory response was associated with disease progression and decompensation with ascites formation.

It has been suggested that dysbiosis, which is the unfavorable shift in the microbiota community structure that compromises microbe-host homeostasis, is a major driver of cirrhosis and is also a major contributor to bacterial translocation to MLNs in cirrhosis animal models (10, 28). In the present study, we found more microbiome alterations in ICC than in fecal samples of rats with decompensated cirrhosis. This suggests that ICC samples may be more appropriate than fecal samples to detect microbiome composition alterations with smaller cohorts. This is in line with previous studies in patients suggesting that dysbiosis in advanced cirrhosis mainly occurs in upper areas of the intestine such as the ileum (10) as well as with the finding that microbiota analyses in fecal samples do not accurately represent dysbiosis in sigmoid mucosal samples obtained by biopsy during sigmoidoscopy (29). In the present study, ICC samples from CCl_4-treated rats showed an attempt of beneficial groups of bacteria, such as *Erysipelotrichia*, to outcompete potential pathobionts such as *Sutterella*, which belongs to *Betaproteobacteria*. A pathobiont is defined as a symbiont that under certain circumstances becomes a pathogen (30). This compensation failed when liver damage and inflammation increased and rats developed ascites, a condition associated with an increase of inflammatory markers, an increase in the relative abundance of *Escherichia* in all sample types except in stool, and an increase in the load of *Proteobacteria* in MLNs.

Bajaj et al. (5) characterized the composition of the fecal microbiome in patients with compensated and decompensated cirrhosis and also evaluated the stability of the microbiome composition at two time points within an interval of 6 months. Our findings confirmed their claim that the level of dysbiosis in stool samples was associated with cirrhosis progression, whereas a relatively stable microbiome composition over time was associated with stable disease. Additionally, the use of an animal model allowed us to unravel the possible cross talk between the microbiome in MLNs and the immune system that was confirmed through a correlation analysis between microbial genera identified in MLNs and levels of proinflammatory cytokines such as TNF-α, IL-6, and IL-17 in MLNs or blood. Moreover, our results are congruent with a comprehensive study that analyzed 29 different cytokines from 522 cirrhotic patients, indicating that systemic inflammation is likely the underlying cause of decompensation and acute-on-chronic liver failure in cirrhosis (8). Finally, our findings are supported by the recent

work of Muñoz et al. (6) that also demonstrated the appearance of a proinflammatory immune response driven by gut dysbiosis in decompensated cirrhosis.

We also observed a positive correlation between the abundance of *Escherichia*, which belongs to the *Proteobacteria* phylum, in MLNs and proinflammatory cytokines, confirming the results of previous studies using culture techniques and PCR of the 16S rRNA gene and making *Escherichia* a potential biomarker for cirrhosis progression (31). We have further demonstrated that an increase in overall microbial load was associated with an increase in *Proteobacteria* in the MLNs of decompensated cirrhotic rats compared to that in controls and in compensated CCl_4-treated rats, suggesting an increase in *Proteobacteria* not only in relative abundance but also in absolute amount.

Furthermore, we observed that "*Candidatus* Arthromitus," a genus from the *Firmicutes* phylum that positively correlated with IL-17 levels, was found in higher abundance in the MLNs of rats with decompensated cirrhosis than in control rats and compensated CCl_4-treated rats. This genus was detected in all sample types except feces, which again confirms that the use of stool samples would not have led to the detection of this genus in a decompensated state. "*Candidatus* Arthromitus" is a segmented filamentous bacterium (SFB) that can induce multiple adaptive immune responses, especially in Th17 cells in the small intestinal lamina propria of mice (32, 33). Th17 cells that produce the effector cytokine IL-17 are potent inducers of tissue inflammation and have been associated with the pathogenesis of many immune-mediated diseases (34). Therefore, "*Candidatus* Arthromitus" might play an important role in inducing inflammation, which may lead to a decompensated state. This genus is also well known to be refractory to *in vitro* culture techniques (35, 36), which explains why it has not been uncovered in previous studies using traditional culture techniques. This genus, which was not detected in feces and in only few blood samples, might have reached the MLNs via the lymphatic system after possible translocation in the upper regions of the gastrointestinal (GI) tract, such as the ileocecal region. *Escherichia*, differently from "*Candidatus* Arthromitus," was detected in all sample types, including in all blood samples, suggesting that its translocation to AF might be via the bloodstream.

Since this work was performed on animals, we may not be able to extrapolate all the findings to cirrhotic patients, particularly in terms of the involvement of specific genera such as "*Candidatus* Arthromitus," as the microbiome might present differences in its composition. However, our study should pave the way for the search for an equivalent genus to "*Candidatus* Arthromitus" in humans that is involved in inducing inflammation in cirrhotic patients. Also, the limited numbers of animals per group might be the reason why we did not find significant changes in fecal samples after multiple testing correction but rather a trend, compared to that of previous findings in human fecal samples (4, 37). Another limitation of our study might be the use of a unique liver disease animal model, as different etiologies might lead to different changes in the microbiome as shown by Fouts et al. (38). However, although different microbial groups might be involved in the progression of cirrhosis depending on the animal model used, our study mainly focused on the sequential changes in the microbiome in distinct intestinal sites and their association with inflammation in an advanced stage of decompensated cirrhosis. This limitation may apply to any animal models used to understand human disease.

In conclusion, our study confirmed previous studies showing that the alterations in the gut microbial community involved an increase of the ratio of pathobionts to beneficial bacteria (39) and also showed that this reflects dysbiosis present in extraintestinal sites such as MLNs, where direct cross talk between the microbiota and the immune cells takes place.

MATERIALS AND METHODS

Experimental design. (i) Animals. Male Sprague-Dawley rats weighing 35 to 49 g were purchased from Harlan Laboratories (Indianapolis, IN, USA) and provided by Research Models and Services Production (Udine, Italy). After the rats were weaned from their mothers, they were fed a rodent chow diet (2018S; Teklad, Madison, WI, USA). After 1 week of quarantine, all animals were placed in individual cages

and kept at a constant room temperature of 21°C, exposed to a 12-h light:12-h dark cycle and allowed free access to water and rodent chow (A04; SAFE, Augy, France). One week later, phenobarbital (1.5 mmol/liter) (Luminal; Kern Pharma, Barcelona, Spain) was added to the tap water given to all animals. There was no contact between rats via water, chow, or feces.

(ii) Induction of cirrhosis and study groups. Cirrhosis was induced as previously described (40). When rats reached a weight of 200 g, they were administered weekly doses of CCl_4 (Sigma-Aldrich, St. Louis, MO., USA) intragastrically using a sterile pyrogen-free syringe (ICO plus 3; Novico Médica, S.A., Barcelona, Spain) with an attached stainless-steel animal feeding tube (Popper and Sons, New Hyde Park, NY, USA) without anesthesia. The first dose of CCl_4 was 20 μl, and subsequent doses were adjusted on the basis of changes in weight 48 h after the previous dose. When rats presented ascites, the dose of CCl_4 was maintained at 40 μl.

We designed different groups of CCl_4-treated rats for which laparotomy and sample collection were performed at four different time points: after 6, 8, and 10 weeks of the first dose of CCl_4 and when ascites was suspected by the increase in abdominal girth and confirmed by paracentesis. A control group of rats not treated with CCl_4 was also included, and samples were collected at the same three time points (6, 8, and 10 weeks) as for the cirrhotic group.

Paracentesis was performed under air anesthesia with isoflurane (Forane; Abbott, Madrid, Spain) under sterile conditions, and approximately 0.1 ml of ascitic fluid was removed. One week later, a laparotomy was carried out.

(iii) Laparotomy. Laparotomy was carried out on all CCl_4-treated rats and on control rats at weeks 6, 8, and 10 or when ascites was suspected. For laparotomy, rats were anesthetized with 10 mg/kg xylazine (Rompun; Bayer, Kiel, Germany) and 50 mg/kg ketamine (Ketolar; Parke-Davis, Madrid, Spain) under sterile conditions. In brief, the abdominal fur was removed with a depilatory cream (Deliplus; Mercadona, Spain) and the skin was sterilized with iodine (Curadona; Lainco, Spain). The abdomen was then opened via a 4-cm median incision, and the remaining fluid was removed.

Biological sample collection. The sequence of sample collection at laparotomy was stool (before laparotomy), MLN, blood, liver spleen, and ileocecal content (ICC). Samples from CCl_4-treated and control rats were stored frozen at −80°C until microbiome analysis. Blood and MLN samples were also used for cytokine analysis.

Bacterial cultures. We inoculated samples of homogenized mesenteric lymph nodes, ascitic fluid, pleural fluid, spleens, and livers on Columbia blood agar, Columbia CNA agar, and the chromogenic medium CPS ID3 (bioMérieux, Marcy-l'Étoile, France). Cultures were incubated for 48 h at 37°C in an aerobic atmosphere. The isolated bacteria were presumptively identified according to their pattern of growth and morphology (41).

Liver damage. Histological liver damage was evaluated by hematoxylin-eosin and Masson's trichrome staining of 4-μm slices from paraffin blocks. A single expert pathologist blindly classified the liver samples according to a semiquantitative score: 0, normal; 1, fibrosis with porto-portal fibrous tracts; 2, regeneration nodules with thin complete fibrous tracts; and 3, regeneration nodules with thick and complete fibrous tracts (41).

Cytokine measurement. TNF-α, IL-6, and IL-10 cytokines were determined in blood samples and IL-17 in MLNs by enzyme-linked immunosorbent assays (ELISAs) according to the manufacturer's protocols (eBiosciences). Results are expressed as picograms per milliliter in blood samples and the ratio of picograms of IL-17 per milligram of total protein. Limits of detection were 30 pg/ml for TNF-α, IL-6, and IL-17 and 15 pg/ml for IL-10.

Microbiome analysis. (i) Genomic DNA extraction. All biological specimens were processed for genomic DNA extraction using protocols previously described by Santiago et al. (42) for low biomass samples such as MLN, blood, and AF and a protocol recommended by the International Human Microbiome Standard for stool samples (http://www.microbiome-standards.org/).

(ii) 16S rRNA genes gene sequencing. To prepare the DNA for sequencing, we amplified a fragment of the 16S rRNA gene by PCR using universal primers targeting the V4 hypervariable region as previously described (43). Amplicons were then purified using the QIAquick PCR purification kit (Qiagen, Barcelona, Spain), quantified using a NanoDrop ND-1000 spectrophotometer (Nucliber), and then pooled in equal concentrations. The pooled amplicons (2 nM) were then subjected to sequencing using Illumina MiSeq technology at the technical support unit of the Autonomous University of Barcelona (UAB, Spain) according to standard Illumina platform protocols.

(iii) Microbiome composition analysis. To analyze the microbiome composition, we first loaded the raw sequences into the QIIME 1.9.1 pipeline, as described by Navas-Molina et al. (44). Low-quality sequence reads were filtered out by applying default settings and a minimum Phred score of 20. From the filtering step and a total of 214 samples, we obtained a total of 2.5 million high-quality sequences with an average number of reads of 11,899. We used the USEARCH algorithm to cluster similar filtered sequences into operational taxonomic units (OTUs) based on a 97% similarity threshold. We then identified and removed chimeric sequences using UCHIME. Representative sequences were selected and aligned using PyNAST against Greengenes template alignment (gg_13_8 release), and a taxonomical assignment step was performed using the basic local alignment search tool to map each representative sequence against a combined database encompassing the Greengenes and PATRIC databases.

For β diversity analysis, we rarefied to 1,046 sequences per sample when comparing all samples simultaneously. When analyzing only low-biomass samples, we rarefied them at 1,046 sequences per sample and at 9,396 sequences per sample when analyzing stool and ICC samples. Rarefaction is used for cases in which read counts are not similar in numbers between samples. Weighted and

unweighted UniFrac metrics were applied to build phylogenetic distance matrices, which were then used to construct hierarchical cluster trees using a principal-coordinate analysis (PCoA) representation.

(iv) Microbial load assessment. To quantify microorganisms, the extracted genomic DNA was used to amplify the V4 region of the 16S rRNA gene by quantitative real-time PCR (qPCR) using the following primers: V4F_517_17 (5′-GCCAGCAGCCGCGGTAA-3′) and V4R_805_19 (5′-GACTACCAGGGTATCTAAT-3′). To calibrate the qPCR reactions, calculated amounts of a linearized plasmid containing the V4 region of the 16S rRNA gene were used. Plasmid concentration was measured using a NanoDrop ND-1000 spectrophotometer (Nucliber), and the number of plasmid copies was calculated from the plasmid's molecular weight. To extrapolate the bacterial number in each sample, serial dilutions of the plasmid were amplified. The qPCR was performed with the 7500 Fast Real-Time PCR system (Applied Biosystems) using optical-grade 96-well plates. The PCR was performed in a total volume of 25 μl using the Power SYBR green PCR master mix (Applied Biosystems) containing 100 nM each of the universal forward and reverse primers. The reaction conditions for amplification of DNA were 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. All reactions were performed in triplicates, and mean values were calculated. This experiment was also duplicated to ensure accuracy. Mean values from both experiments were taken into account. Data were analyzed using Sequence Detection Software version 1.4, supplied by Applied Biosystems.

Statistical analysis. We performed statistical analyses under QIIME and R. We used the D'Agostino-Pearson test to check for the normality of the data distribution. Parametric normally distributed data were compared by Student's *t* tests for paired or unpaired data; otherwise, the Wilcoxon signed rank test was used for paired data and the Mann-Whitney U test for unpaired data. Qualitative variables were analyzed by Fisher's exact test. The Kruskal-Wallis one-way test of variance was used to compare the mean numbers of sequences from different unpaired groups of subjects at various taxonomic levels, the Wilcoxon test was used when comparing only 2 groups. We performed analyses with the nonparametric multivariate ANOVA (PERMANOVA) called the Adonis test to test for differences in microbial communities using the UniFrac metrics. We performed Spearman tests to evaluate correlations between microbiome composition and biological parameters such as cytokine levels. When possible, the analysis provided false discovery rate (*q*)-corrected *P* values (*q* values). A *q* value of <0.05 was considered significant for all tests.

Ethics approval. The study was approved by the Animal Research Committee at the Institut de Recerca of Hospital de la Santa Creu i Sant Pau (Barcelona) and by the Department of Agriculture, Livestock and Fisheries of the Generalitat de Catalunya (Departament d'Agricultura, Ramaderia i Pesca). Animal care complied with the criteria outlined in the Guide for the Care and Use of Laboratory Animals (Committee for the Update of the Guide for the Care and Use of Laboratory Animals Institute for Laboratory Animal Research Division on Earth and Life Studies, Washington, DC, USA).

Data availability. Sequence data have been deposited in the NCBI database under accession number PRJNA448565.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00278-18.

FIG S1, TIF file, 0.4 MB.

ACKNOWLEDGMENTS

This study was supported by three grants from the Instituto de Salud Carlos III/FEDER (PI11/01026, PI14/00680, and PI17/00614).

We declare no competing interests

C.M. and G.S. designed the study. E.S. performed the experiments on rats. A.S. processed all the samples for microbiome analysis. S.V., G.S., and F.Y. performed immune response analysis. A.G. performed the histological evaluation. M.P. and M.C. performed sequence and statistical analyses. C.M. interpreted the results and wrote the manuscript. A.C., S.V., G.S., and C.G. were major contributors in revising and correcting the manuscript.

REFERENCES

1. Shah NL, Banaei YP, Hojnowski KL, Cornella SL. 2015. Management options in decompensated cirrhosis. Hepat Med 7:43–50. https://doi.org/10.2147/HMER.S62463.
2. D'Amico G, Garcia-Tsao G, Pagliaro L. 2006. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. J Hepatol 44:217–231. https://doi.org/10.1016/j.jhep.2005.10.013.
3. Oikonomou T, Papatheodoridis GV, Samarkos M, Goulis I, Cholongitas E. 2018. Clinical impact of microbiome in patients with decompensated cirrhosis. World J Gastroenterol 24:3813–3820. https://doi.org/10.3748/wjg.v24.i34.3813.
4. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, Zhou J, Ni S, Liu L, Pons N, Batto JM, Kennedy SP, Leonard P, Yuan C, Ding W, Chen Y, Hu X, Zheng B, Qian G, Xu W, Ehrlich SD, Zheng S, Li L. 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature 513:59–64. https://doi.org/10.1038/nature13568.
5. Bajaj JS, Heuman DM, Hylemon PB, Sanyal AJ, White MB, Monteith P, Noble NA, Unser AB, Daita K, Fisher AR, Sikaroodi M, Gillevet PM. 2014.

Altered profile of human gut microbiome is associated with cirrhosis and its complications. J Hepatol 60:940–947. https://doi.org/10.1016/j.jhep.2013.12.019.
6. Muñoz L, Borrero MJ, Ubeda M, Conde E, Del Campo R, Rodriguez-Serrano M, Lario M, Sanchez-Diaz AM, Pastor O, Diaz D, Garcia-Bermejo L, Monserrat J, Alvarez-Mon M, Albillos A. 10 November 2018. Intestinal immune dysregulation driven by dysbiosis promotes barrier disruption and bacterial translocation in rats with cirrhosis. Hepatology https://doi.org/10.1002/hep.30349.
7. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-bacterial mutualism in the human intestine. Science 307:1915–1920. https://doi.org/10.1126/science.1104816.
8. Claria J, Stauber RE, Coenraad MJ, Moreau R, Jalan R, Pavesi M, Amoros A, Titos E, Alcaraz-Quiles J, Oettl K, Morales-Ruiz M, Angeli P, Domenicali M, Alessandria C, Gerbes A, Wendon J, Nevens F, Trebicka J, Laleman W, Saliba F, Welzel TM, Albillos A, Gustot T, Benten D, Durand F, Gines P, Bernardi M, Arroyo V, CANONIC Study Investigators of the EASL-CLIF Consortium and the European Foundation for the Study of Chronic Liver Failure (EF-CLIF). 2016. Systemic inflammation in decompensated cirrhosis: characterization and role in acute-on-chronic liver failure. Hepatology 64:1249–1264. https://doi.org/10.1002/hep.28740.
9. Genesca J, Marti R, Rojo F, Campos F, Peribanez V, Gonzalez A, Castells L, Ruiz-Marcellan C, Margarit C, Esteban R, Guardia J, Segura R. 2003. Increased tumour necrosis factor alpha production in mesenteric lymph nodes of cirrhotic patients with ascites. Gut 52:1054–1059. https://doi.org/10.1136/gut.52.7.1054.
10. Wiest R, Lawson M, Geuking M. 2014. Pathological bacterial translocation in liver cirrhosis. J Hepatol 60:197–209. https://doi.org/10.1016/j.jhep.2013.07.044.
11. Macpherson AJ, Smith K. 2006. Mesenteric lymph nodes at the center of immune anatomy. J Exp Med 203:497–500. https://doi.org/10.1084/jem.20060227.
12. Kiernan MG, Coffey JC, McDermott K, Cotter PD, Cabrera-Rubio R, Kiely PA, Dunne CP. 2019. The human mesenteric lymph node microbiome differentiates between Crohn's disease and ulcerative colitis. J Crohns Colitis 13:58–66. https://doi.org/10.1093/ecco-jcc/jjy136.
13. O'Brien CL, Pavli P, Gordon DM, Allison GE. 2014. Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. Gut 63:1596–1606. https://doi.org/10.1136/gutjnl-2013-305320.
14. Llovet JM, Bartoli R, Planas R, Cabre E, Jimenez M, Urban A, Ojanguren I, Arnal J, Gassull MA. 1994. Bacterial translocation in cirrhotic rats. Its role in the development of spontaneous bacterial peritonitis. Gut 35:1648–1652. https://doi.org/10.1136/gut.35.11.1648.
15. Shi H, Lv L, Cao H, Lu H, Zhou N, Yang J, Jiang H, Dong H, Hu X, Yu W, Jiang X, Zheng B, Li L. 2017. Bacterial translocation aggravates CCl_4-induced liver cirrhosis by regulating $CD4^+$ T cells in rats. Sci Rep 7:40516. https://doi.org/10.1038/srep40516.
16. Gómez-Hurtado I, Such J, Francés R. 2016. Microbiome and bacterial translocation in cirrhosis. Gastroenterol Hepatol 39:687–696. https://doi.org/10.1016/j.gastrohep.2015.10.013.
17. Albillos A, Lario M, Alvarez-Mon M. 2014. Cirrhosis-associated immune dysfunction: distinctive features and clinical relevance. J Hepatol 61:1385–1396. https://doi.org/10.1016/j.jhep.2014.08.010.
18. Albillos A, Hera Ad A. d l, Reyes E, Monserrat J, Muñoz L, Nieto M, Prieto A, Sanz E, Alvarez-Mon M. 2004. Tumour necrosis factor-alpha expression by activated monocytes and altered T-cell homeostasis in ascitic alcoholic cirrhosis: amelioration with norfloxacin. J Hepatol 40:624–631. https://doi.org/10.1016/j.jhep.2003.12.010.
19. Albillos A, de la Hera A, Gonzalez M, Moya JL, Calleja JL, Monserrat J, Ruiz-del-Arbol L, Alvarez-Mon M. 2003. Increased lipopolysaccharide binding protein in cirrhotic patients with marked immune and hemodynamic derangement. Hepatology 37:208–217. https://doi.org/10.1053/jhep.2003.50038.
20. Frances R, Munoz C, Zapater P, Uceda F, Gascon I, Pascual S, Perez-Mateo M, Such J. 2004. Bacterial DNA activates cell mediated immune response and nitric oxide overproduction in peritoneal macrophages from patients with cirrhosis and ascites. Gut 53:860–864. https://doi.org/10.1136/gut.2003.027425.
21. Alexopoulou A, Agiasotelli D, Vasilieva LE, Dourakis SP. 2017. Bacterial translocation markers in liver cirrhosis. Ann Gastroenterol 30:486–497. https://doi.org/10.20524/aog.2017.0178.
22. Du Plessis J, Vanheel H, Janssen CE, Roos L, Slavik T, Stivaktas PI, Nieuwoudt M, van Wyk SG, Vieira W, Pretorius E, Beukes M, Farre R, Tack J, Laleman W, Fevery J, Nevens F, Roskams T, Van der Merwe SW. 2013. Activated intestinal macrophages in patients with cirrhosis release NO and IL-6 that may disrupt intestinal barrier function. J Hepatol 58:1125–1132. https://doi.org/10.1016/j.jhep.2013.01.038.
23. Everard A, Lazarevic V, Gaia N, Johansson M, Stahlman M, Backhed F, Delzenne NM, Schrenzel J, Francois P, Cani PD. 2014. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. ISME J 8:2116–2130. https://doi.org/10.1038/ismej.2014.45.
24. Martínez-Esparza M, Tristán-Manzano M, Ruiz-Alcaraz AJ, García-Peñarrubia P. 2015. Inflammatory status in human hepatic cirrhosis. World J Gastroenterol 21:11522–11541. https://doi.org/10.3748/wjg.v21.i41.11522.
25. Pu Y, Zhang S, Zhou R, Huang N, Li H, Wei W, Li L, Huang C, Yang J, Li Z. 2016. IL-17A up-regulates expression of endothelial tissue factor in liver cirrhosis via the ROS/p38 signal pathway. Biochem Biophys Res Commun 470:41–47. https://doi.org/10.1016/j.bbrc.2015.12.093.
26. Jang MH, Sougawa N, Tanaka T, Hirata T, Hiroi T, Tohya K, Guo Z, Umemoto E, Ebisuno Y, Yang BG, Seoh JY, Lipp M, Kiyono H, Miyasaka M. 2006. CCR7 is critically important for migration of dendritic cells in intestinal lamina propria to mesenteric lymph nodes. J Immunol 176:803–810. https://doi.org/10.4049/jimmunol.176.2.803.
27. Trivedi PJ, Adams DH. 2016. Gut-liver immunity. J Hepatol 64:1187–1189. https://doi.org/10.1016/j.jhep.2015.12.002.
28. Aguirre Valadez JM, Rivera-Espinosa L, Méndez-Guerrero O, Chávez-Pacheco JL, García Juárez I, Torre A. 2016. Intestinal permeability in a patient with liver cirrhosis. Ther Clin Risk Manag 12:1729–1748. https://doi.org/10.2147/TCRM.S115902.
29. Bajaj JS, Hylemon PB, Ridlon JM, Heuman DM, Daita K, White MB, Monteith P, Noble NA, Sikaroodi M, Gillevet PM. 2012. Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. Am J Physiol Gastrointest Liver Physiol 303:G675–G685. https://doi.org/10.1152/ajpgi.00152.2012.
30. Chow J, Tang H, Mazmanian SK. 2011. Pathobionts of the gastrointestinal microbiota and inflammatory disease. Curr Opin Immunol 23:473–480. https://doi.org/10.1016/j.coi.2011.07.010.
31. Zapater P, Francés R, González-Navajas JM, de la Hoz MA, Moreu R, Pascual S, Monfort D, Montoliu S, Vila C, Escudero A, Torras X, Cirera I, Llanos L, Guarner-Argente C, Palazón JM, Carnicer F, Bellot P, Guarner C, Planas R, Solá R, Serra MA, Muñoz C, Pérez-Mateo M, Such J. 2008. Serum and ascitic fluid bacterial DNA: a new independent prognostic factor in noninfected patients with cirrhosis. Hepatology 48:1924–1931. https://doi.org/10.1002/hep.22564.
32. Ericsson AC, Hagan CE, Davis DJ, Franklin CL. 2014. Segmented filamentous bacteria: commensal microbes with potential effects on research. Comp Med 64:90–98.
33. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. 2009. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell 139:485–498. https://doi.org/10.1016/j.cell.2009.09.033.
34. Korn T, Bettelli E, Oukka M, Kuchroo VK. 2009. IL-17 and Th17 Cells. Annu Rev Immunol 27:485–517. https://doi.org/10.1146/annurev.immunol.021908.132710.
35. Bolotin A, de Wouters T, Schnupf P, Bouchier C, Loux V, Rhimi M, Jamet A, Dervyn R, Boudebbouze S, Blottiere HM, Sorokin A, Snel J, Cerf-Bensussan N, Gaboriau-Routhiau V, van de Guchte M, Maguin E. 2014. Genome sequence of "*Candidatus* Arthromitus" sp. strain SFB-mouse-NL, a commensal bacterium with a key role in postnatal maturation of gut immune functions. Genome Announc 2:e00705-14. https://doi.org/10.1128/genomeA.00705-14.
36. Manzano M, Giusto C, Iacumin L, Patthey C, Cecchini F, Fontanillas R, Comi G. 2010. Nested PCR for the detection of *Candidatus* arthromitus in fish. FEMS Microbiol Lett 308:35–39. https://doi.org/10.1111/j.1574-6968.2010.01984.x.
37. Bajaj JS, Betrapally NS, Gillevet PM. 2015. Decompensated cirrhosis and microbiome interpretation. Nature 525:E1–E2. https://doi.org/10.1038/nature14851.
38. Fouts DE, Torralba M, Nelson KE, Brenner DA, Schnabl B. 2012. Bacterial translocation and changes in the intestinal microbiome in mouse models of liver disease. J Hepatol 56:1283–1292. https://doi.org/10.1016/j.jhep.2012.01.019.
39. Lachar J, Bajaj JS. 2016. Changes in the microbiome in cirrhosis and

relationship to complications: hepatic encephalopathy, spontaneous bacterial peritonitis, and sepsis. Semin Liver Dis 36:327–330. https://doi.org/10.1055/s-0036-1593881.

40. Runyon BA, Sugano S, Kanel G, Mellencamp MA. 1991. A rodent model of cirrhosis, ascites, and bacterial peritonitis. Gastroenterology 100:489–493. https://doi.org/10.1016/0016-5085(91)90221-6.
41. Sanchez E, Nieto JC, Vidal S, Santiago A, Martinez X, Sancho FJ, Sancho-Bru P, Mirelis B, Corominola H, Juarez C, Manichanh C, Guarner C, Soriano G. 2017. Fermented milk containing Lactobacillus paracasei subsp. *paracasei* CNCM I-1518 reduces bacterial translocation in rats treated with carbon tetrachloride. Sci Rep 7:45712. https://doi.org/10.1038/srep45712.
42. Santiago A, Pozuelo M, Poca M, Gely C, Nieto JC, Torras X, Roman E, Campos D, Sarrabayrouse G, Vidal S, Alvarado-Tapias E, Guarner F, Soriano G, Manichanh C, Guarner C. 2016. Alteration of the serum microbiome composition in cirrhotic patients with ascites. Sci Rep 6:25001. https://doi.org/10.1038/srep25001.
43. Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, Martinez X, Varela E, Sarrabayrouse G, Machiels K, Vermeire S, Sokol H, Guarner F, Manichanh C. 2017. A microbial signature for Crohn's disease. Gut 66:813–822. https://doi.org/10.1136/gutjnl-2016-313235.
44. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knight R. 2013. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol 531:371–444. https://doi.org/10.1016/B978-0-12-407863-5.00019-8.

AP&T Alimentary Pharmacology and Therapeutics

Metabolic adaptation of colonic microbiota to galactooligosaccharides: a proof-of-concept-study

M. Mego*, C. Manichanh*, A. Accarino*, D. Campos*, M. Pozuelo*, E. Varela*, J. Vulevic†, G. Tzortzis†, G. Gibson‡, F. Guarner* & F. Azpiroz*

*Digestive System Research Unit, University Hospital Vall d'Hebron, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Universitat Autònoma de Barcelona, Barcelona, Spain.
†Clasado Research Services Ltd, Science and Technology Centre, University of Reading, Reading, UK.
‡Food Microbial Sciences Unit, Department of Food and Nutritional Sciences, The University of Reading, Reading, UK.

Correspondence to:
Dr F. Azpiroz, Digestive System Research Unit, Hospital General Vall d'Hebron, 08035-Barcelona, Spain.
E-mail: azpiroz.fernando@gmail.com

Publication data
Submitted 14 June 2016
First decision 2 July 2016
Resubmitted 25 October 2016
Resubmitted 12 December 2016
Accepted 14 December 2016
EV Pub Online 12 January 2017

The Handling Editor for this article was Professor Peter Gibson, and it was accepted for publication after full peer-review.

SUMMARY

Background
Prebiotics have been shown to reduce abdominal symptoms in patients with functional gut disorders, despite that they are fermented by colonic bacteria and may induce gas-related symptoms.

Aim
To investigate changes in the metabolic activity of gut microbiota induced by a recognised prebiotic.

Methods
Healthy subjects (n = 20) were given a prebiotic (2.8 g/day HOST-G904, HOST Therabiomics, Jersey, Channel Islands) for 3 weeks. During 3-day periods immediately before, at the beginning and at the end of the administration subjects were put on a standard diet (low fibre diet supplemented with one portion of high fibre foods) and the following outcomes were measured: (i) number of daytime gas evacuations for 2 days by means of an event marker; (ii) volume of gas evacuated via a rectal tube during 4 h after a test meal; and (iii) microbiota composition by faecal Illumina MiSeq sequencing.

Results
At the beginning of administration, HOST-G904 significantly increased the number of daily anal gas evacuations (18 ± 2 vs. 12 ± 1 pre-administration; $P < 0.001$) and the volume of gas evacuated after the test meal (236 ± 23 mL vs. 160 ± 17 mL pre-administration; $P = 0.006$). However, after 3 weeks of administration, these effects diminished (11 ± 2 daily evacuations, 169 ± 23 mL gas evacuation). At day 21, relative abundance of butyrate producers (Lachnospiraceae) correlated inversely with the volume of gas evacuated ($r = -0.52$; $P = 0.02$).

Conclusion
The availability of substrates induces an adaptation of the colonic microbiota activity in bacterial metabolism, which produces less gas and associated issues. Clinical trials.gov NCT02618239.

Aliment Pharmacol Ther 2017; **45**: 670–680

doi:10.1111/apt.13931

INTRODUCTION

Prebiotics, by definition, induce beneficial effects by selectively influencing colonic microbiota.[1] They serve as selective substrates for microbiota gut metabolism.[2]

The effect of fibre on functional intestinal symptoms is controversial.[3] We recently showed that a diet rich in non-absorbable, fermentable residues in the short term increased intestinal gas production and induced digestive symptoms, such as flatulence, abdominal bloating and distension,[4] whereas a low-residue diet improved symptoms in patients with abdominal bloating and flatulence.[5] Good evidence of the clinical benefits of reducing fermentable foodstuffs has been provided by a series of studies using diets low in fermentable oligosaccharides, monosaccharides, disaccharides and polyols (FODMAP).[6–8]

In contrast to the potential effect of low-residue diets on symptoms, some specific prebiotics, despite being fermented by microbiota, have been shown to improve these types of symptoms. Specifically, a clinical trial with the prebiotic HOST-G904 demonstrated a clinical benefit in patients with irritable bowel syndrome.[9] A very elegant controlled trial in healthy university students showed that around the time of final exams, stress was associated with diarrhoea, indigestion and abdominal pain, and galactooligosaccharide supplementation reduced this stress-induced gastrointestinal dysfunction.[10.]

We hypothesised that prebiotic administration initially activates the fermentative metabolism of colonic microbiota, increasing gas production, and that this early effect is later followed by an adaptation of the microbiota with a reduction in net gas production. Our aim was to assess the effect of HOST-G904,[9] on microbiota gas production at initial exposure and then following continuous administration for a period of potential adaptation. To this end, we designed a proof-of-concept study in healthy subjects.

MATERIALS AND METHODS

Participants

Twenty-six healthy subjects without gastrointestinal symptoms or history of gastrointestinal disorders participated in the study: 20 subjects participated in the main study, and six subjects, as a control group, in an ancillary study (Table 1). All participants were instructed to fill out a clinical questionnaire based on Rome III criteria to rule out functional gastrointestinal disorders (no symptom ≥2 on a 0–10 scale) and to confirm normal bowel habits. This questionnaire has been previously shown to discriminate patients from healthy subjects.[4, 5, 11–14] Antibiotic, but not pre- or probiotic consumption during the previous 2 month was an exclusion criterium. Subjects gave written informed consent to participate in the study. The protocol, including the external control study, was approved by the Institutional Review Board of University Hospital Vall d'Hebron and was registered with ClinicalTrials.Gov [NCT02618239].

Table 1 | Demographic data

	Main study	Ancillary study
Age range, year	18–54	25–35
Women/men, *n*	12/8	5/1
BMI range, kg/m^2	19–26	22–27
Bowel movements/week, mean ± S.E.	8 ± 1	8.5 ± 1.2
Bristol score, mean ± S.E.	3.6 ± 0.2	3.8 ± 0.3

Experimental design

Participants consumed a prebiotic (2.8 g/day HOST-G904, HOST Therabiomics, Jersey, Channel Islands) for 3 week. For three 3-day periods, pre-administration (day −2 to 0), at the early administration period (day 1–3) and the late administration period (day 19–21), different outcomes (see below) were measured (evaluation periods).

Diet

During the study participants consumed their habitual diet except during the evaluation periods when the diet was standardised, as follows. During the evaluation periods, subjects were instructed to consume a low fibre diet[5] restricted to the following foodstuffs: (i) meat, fish, fowl and eggs; (ii) salad; (iii) rice, pasta and bread; (iv) dairy products and (v) strained orange juice, tangerine, pears, apples and berries. This low-residue diet was complemented with one portion per day of the following: whole crackers, lentils, chickpeas, beans, peas, artichoke, Brussels' sprouts, banana, peach or prunes; the portion size of each specific foodstuff was adjusted to contain 12 g fibre. For the rest of the administration period, the participants consumed their usual diet. For the duration of the study, fermented dairy products and any tablets, pills or food supplements containing pre-or probiotics were not allowed (Figure 1). During the 3-day evaluation periods, participants were instructed to fill out a diary specifying the foods they consumed, to assess compliance with the diet and to calculate dietary intake.[15–17]

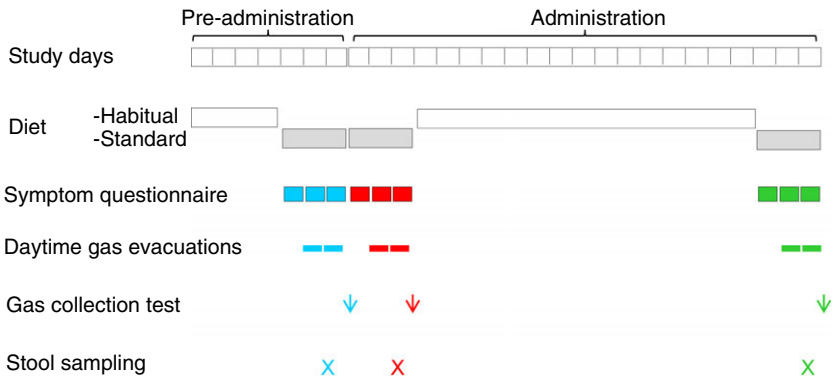


Figure 1 | Experimental design. All participants completed the main study (n = 20) and were included for analysis. Note the colour code: pre-administration in blue, early administration red, and late administration period green.

Outcomes

The following data were collected during the 3-day evaluation periods before, at the beginning and at the end of administration.

Daily symptom questionnaire. During the 3 day of each evaluation period, the participants were instructed to fill out daily questionnaires that included the following parameters: (i) subjective sensations of flatulence (defined as anal gas evacuation), abdominal bloating (pressure/fullness), abdominal distension (sensation of girth increment), borborygmi and abdominal discomfort/pain using 0–10 analogue scales; (ii) digestive well-being using a 10-point scale graded from +5 (extremely pleasant sensation/satisfaction) to −5 (extremely unpleasant sensation/dissatisfaction) and mood on similr scale graded from +5 (very positive) to −5 (very negative). This questionnaire has been previously used and was shown sensitive to detect effect of dietary interventions in different populations.[4, 5, 14, 18]

Number of anal gas evacuations. The number of anal gas evacuations was measured during the day using an event marker (Hand Tally Counter No 101, Digi Sport Instruments, Shanggiu, China). Participants were instructed to carry the event marker during the day and register each passage of anal gas. To facilitate compliance, anal gas evacuations were only registered during the last 2 day of each evaluation period. This method has been previously used with reproducible and consistent results[4, 5]; furthermore, studies measuring the number of gas evacuations by an event marker and continuously recording anal gas evacuations have shown a very good correlation ($r > 0.95$; $P < 0.05$).[19–22]

Microbiota composition. Faecal samples were collected on the last day of each evaluation period, i.e. immediately before and on the third and 21st day of administration (60 samples in total). After collection and homogenisation, the samples were immediately frozen by the participants in their home freezers at −20 °C and later brought to the laboratory in a freezer pack, where they were stored at −80 °C.

Genomic DNA extraction: A frozen aliquot (250 mg) of each sample was suspended in 250 μL of guanidine thiocyanate, 0.1 M Tris (pH 7.5), 40 μL of 10% N-lauroyl sarcosine and 500 μL 5% N-lauroyl sarcosine. DNA was extracted by mechanical disruption of microbial cells with beads, and recovery of nucleic acids from clear lysates was achieved by alcohol precipitation, as previously described.[23] An equivalent of 1 mg of each sample was used for DNA quantification using a NanoDrop ND-1000 Spectrophotometer (Nucliber, Madrid, Spain).

High-throughput DNA sequencing: For profiling the microbiome composition, the hyper-variable region (V4) of the bacterial and archaeal 16S rRNA gene was amplified by PCR. On the basis of our analysis performed using PrimerProspector software,[24] the V4 primer pairs used in this study were expected to amplify almost 100% of the bacterial and archaeal domains. The 5′ ends of the forward (V4F_515_19: 5′-GTGCCAGCAMGCCGCG GTAA-3′) and reverse (V4R_806_20: 5′-GGACTACCAG GGTATCTAAT-3′) primers targeting the 16S gene were tagged with specific sequences as follows: 5′-{AATGAT ACGGCGACCACCGAGATCTACACTATGGTAATTGT} {GTGCCAGCMGCCGCGGTAA}-3′ and 5′-{CAAGCAG AAGACGGCATACGAGAT} {Golay barcode} {AGTCA GTCAGCC} {GGACTACHVGGGTWTCTAAT}-3′. Multiplex identifiers, known as Golay codes, had 12 bases and were specified downstream of the reverse primer sequence (V4R_806_20).[25] Standard PCR (0.75 units of Taq polymerase (Roche, Barcelona, Spain) and 20 pmol/μL of the forward and reverse primers) was run in a Mastercycler gradient (Eppendorf, Madrid, Spain) at 94 °C for 3 min,

Aliment Pharmacol Ther 2017; 45: 670–680

followed by 35 cycles of 94 °C for 45 s, 56 °C for 60 s, 72 °C for 90 s, and a final cycle of 72 °C for 10 min. Amplicons were purified using a QIAquick PCR Purification Kit (Qiagen, Barcelona, Spain), quantified using a NanoDrop ND-1000 Spectrophotometer (Nucliber, Madrid, Spain), and then pooled in equal concentrations. Pooled amplicons (2 nM) were then subjected to sequencing using Illumina MiSeq technology in the technical support unit of the Autonomous University of Barcelona (UAB, Spain) following standard Illumina platform protocols.

Sequence analysis: Sequences obtained from the 60 faecal samples after the sequencing step were analysed with QIIME (Quantitative Insights Into Microbial Ecology) 1.9.1[26] using an in-house script that performs upstream and downstream analyses. Low-quality raw sequences with a Phred score of less than 20 were removed from the analysis. Each read was assigned back to its corresponding sample during a demultiplexing step and barcodes were removed from the sequences. After filtering, we obtained a total of 2 460 589 high-quality sequences. The USEARCH (Ultra-fast Sequence Analysis)[27] tool was used to cluster similar sequences into Operational Taxonomic Units (OTUs) or taxa based on a 97% similarity and to remove chimeric sequences with the UCHIME (Ultra-fast Chimeric search) algorithm. From each of these OTUs, one representative sequence was selected and then aligned using PyNAST (Python Nearest Alignment Space Termination tool) against a Greengenes template alignment from the most recent version of the database (gg_13_8). Then, a taxonomical assignment step was performed using the basic local alignment search tool (BLAST) to map each representative sequence against a combined database encompassing the Greengenes and PATRIC (Pathosystems Ressource Integration Center) databases. A phylogenetic tree using the FastTree programme and an OTU table were built. To avoid false positive OTUs, we eliminated those that did not represent at least 0.2% of the sequences in at least two samples. The final OTU table was rarefied at 15396 sequence reads per sample. Rarefaction is used to overcome cases in which read counts were not similar between samples.

Quantification of Bifidobacterium*:* To assess *Bifidobacterium* genus quantification, the extracted genomic DNA was used to amplify the 16S rRNA gene by quantitative real-time PCR (qPCR) using the following specific primers Bifgenus_F: 5′-TGG CTC AGG ATG AAC GCT G-3′ and Bifgenus_R: 5′-TGA TAG GAC GCG ACC CCA T-3′ and TaqMan MGB probe (FAM™ dye-labeled): 5′-CAT CCG GCA TTA CCA-3′. To calibrate the qPCR reactions, we used calculated amounts of extracted DNA from three isolated *Bifidobacterium* species (*B. breve, B. longum and B. infantis*). Serial dilutions of the pooled DNA were amplified (copy number ranging from 25 to 2.5 × 106) to extrapolate the bifidobacterial number in each sample. The qPCR was performed with the 7500 Fast Real-Time PCR System (Applied Biosystems, Barcelona, Spain) using optical-grade 96-well plates. The PCR reaction was performed in a total volume of 25 μL using the TaqMan Universal PCR Master Mix (Applied Biosystems), containing 300 nM of each primer and 100 nM of MGB probe. The reaction conditions for amplification of DNA were 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. All reactions were performed in triplicate and mean values were calculated Data were analysed using Sequence Detection Software version 1.4, supplied by Applied Biosystems.

Response to a probe meal. The day following each evaluation period, participants reported to the laboratory after an overnight fast and the response to a probe meal was evaluated. The *probe meal* consisted of a ham omelet (100 g), 46 g of white bread, 10 g of butter, 25 g of jam and 200 mL of fruit juice (a 400 Kcal caloric content, 350 mL of total volume, 1.5 g of fibre). The first dose of HOST-G904 was administered after the first collection test; in the other two sets (early and late administration period) HOST-G904 was administered with the probe meal.

Anal gas evacuation: The volume of gas evacuated per anus was measured for 4 h after the probe meal, as previously described.[4, 28, 29] In brief, gas was collected using a rectal balloon catheter (20 F Foley catheter, Bard, Barcelona, Spain) connected via a gas-tight line to a barostat, and the volume was continuously recorded. The intrarectal balloon was inflated with 5 mL of water to prevent anal gas leaks.

Abdominal symptoms: Perception of abdominal sensations was measured every 30 min during the 4-h gas collection period using the same scales as described above: 0–10 scales for scoring abdominal bloating (pressure/fullness), abdominal distension (sensation of

girth increment), borborygmi and abdominal discomfort/pain; −5 to + 5 scales for scoring digestive well-being and mood.

Girth measurement: The method has been previously described.[30] Briefly, a non-stretch belt (48-mm wide) was placed over the umbilicus. The belt had a metric tape with marks at 1 mm intervals fixed over it. The overlapping ends of the belt were adjusted carefully by two elastic bands to ensure that the belt was in constant contact with the abdominal wall. Girth measurements down to the one-millimetre level were taken without manipulation by the investigator at 30-min intervals during the study. Previous studies validated the reproducibility of the measurements and the sensitivity of this method to consistently detect small variations in girth induced by various experimental conditions.[12, 30–33] Changes in girth during the infusion period were compared to the measurements during the basal period.

Ancillary study: external control group

In the control group (see Participants) HOST-G904 was administered at the same dose (2.8 g/day) for 2 weeks while participants consuming their usual diet (see Experimental design). The number of anal gas evacuations (see above) was measured during daytime before administration (day 0), and on day 2, 3, 5, 7, 12 and 15 during administration.

Statistical analysis

Microbiota analysis. The Shapiro–Wilk test was used to check the normality of the data, and pairwise comparisons were made between the study groups with the non-parametric Kruskal–Wallis one-way analysis of variance test, which compares means between groups. A false discovery rate (FDR) of corrected *P*-values was taken into account to consider the significance of the results.

Overall comparisons. The means (±S.E.) of the variables measured were calculated. The Kolmogorov–Smirnov test was used to check the normality of the data distribution. Parametric normally distributed data were compared by Student's *t*-test for paired or unpaired data; otherwise, the Wilcoxon signed-rank test was used for paired data and the Mann–Whitney *U* test for unpaired data. The association of parameters was analysed using linear regression analysis.

RESULTS

Study flow and dietary intake

All participants included in the study ($n = 26$) completed the protocols (main or ancillary study) and were included for analysis. Participants reported adherence to study instructions. Based on the diaries, dietary intake during each 3-day evaluation period was calculated, and mean daily intake during the three evaluation periods was similar (Table 2).

Symptoms and gas volume

Pre-administration. Before administration, participants tolerated the standardised diet (Table 2) without a significant perception of abdominal symptoms, except for a mild-sensation of flatulence. Interestingly, participants scored a positive sensation of digestive well-being and positive mood (Figure 2). Using the event marker, a mean of 12 ± 1 daytime anal gas evacuations was recorded (Figure 3). During the 4-h gas collection period after the probe meal, subjects evacuated 160 ± 17 mL of gas (Figure 4) without reporting a significant perception or abdominal distension (Figure 5); the scores of abdominal sensation, well-being and mood were similar to those recorded in diaries on previous day.

Early administration period. At the beginning of the HOST-G904 administration, no changes in abdominal sensations, well-being or mood were detected (Figure 2), but a clear effect on colonic gas production was observed. Indeed, on the standard diet (Table 2) the number of daytime gas evacuations increased by 39 ± 9%; up to 18 ± 2 daytime evacuations ($P < 0.001$ vs. pre-administration), and the effect was already present on the 2nd day of administration (Figure 3).

Table 2 | Daily dietary intake during evaluation periods

	Study periods: HOST-G904 administration		
	Before	Early phase	Late phase
GOS*, g	0.17 ± 0.04	0.19 ± 0.03	0.19 ± 0.03
Fructans, g	2.8 ± 0.3	2.9 ± 0.2	2.6 ± 0.2
Fibre, g	18.4 ± 0.4	18.4 ± 0.4	18.8 ± 0.4
Carbohydrates, g	292 ± 9	292 ± 9	286 ± 7
Lipids, g	47 ± 1	46 ± 1	48 ± 1
Proteins, g	86 ± 2	83 ± 2	86 ± 2

Data are means of 3 day in each evaluation period.

* Galacto-oligosaccharides.

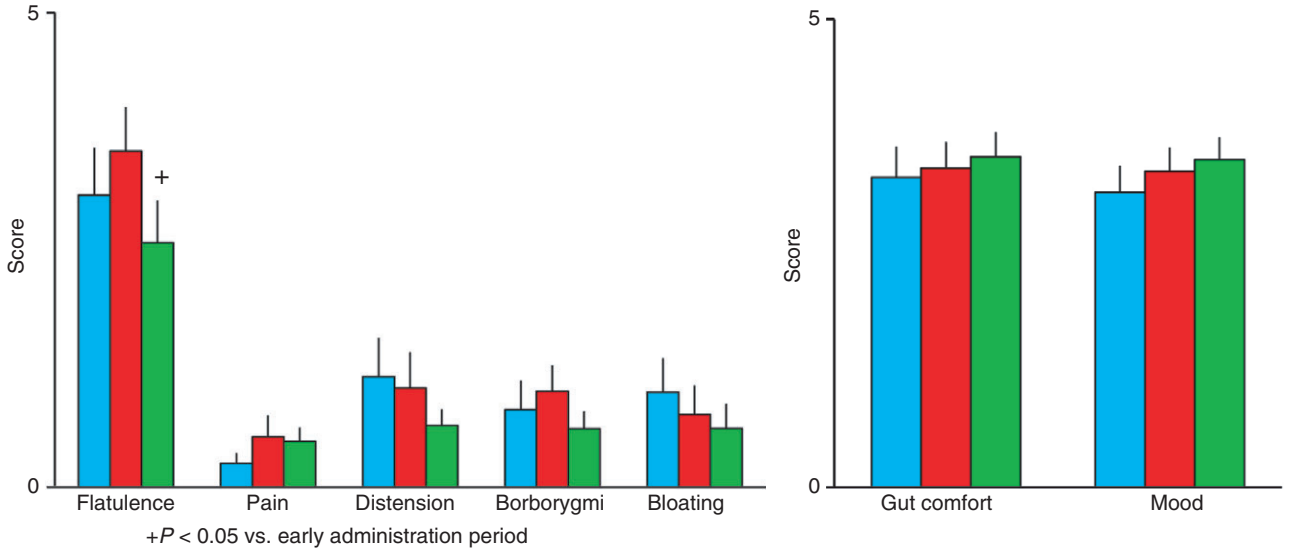


Figure 2 | Symptoms measured by daily questionnaires pre-administration (day −2 to 0, blue), in the early administration period (day 1–3, red), and in the late administration period (day 19–21, green) (n = 20). Data are average over each 3-day periods.

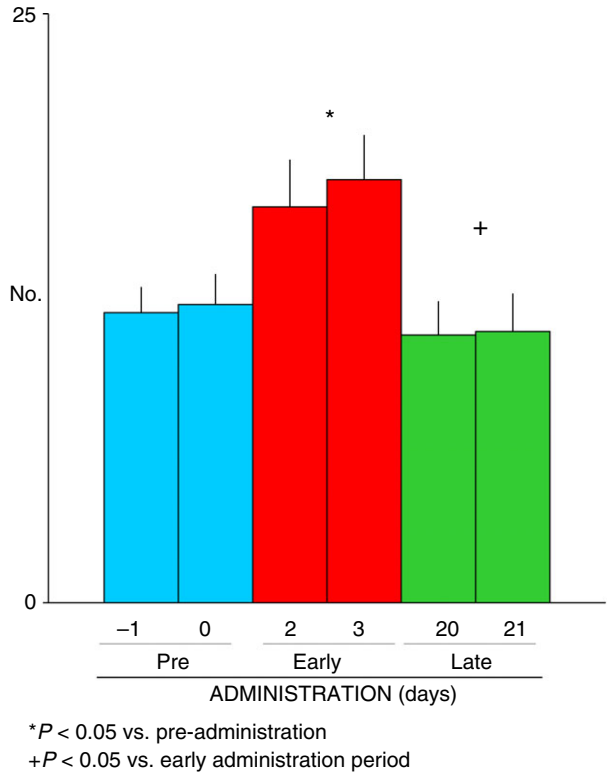


Figure 3 | Number of daytime anal gas evacuations during the last 2 day of the evaluation periods: pre-administration, in the early treatment period and the late administration period (n = 20).

Likewise, the volume of gas collected after the probe meal increased by 64 ± 16%; up to 236 ± 23 mL (P = 0.006 vs. pre-administration) (Figure 4), but this did not affect sensation scores measured during the gas-collection period after the probe meal (Figure 5).

Late administration period. After 3 weeks of HOST-G904 administration, on the standard diet the abdominal sensation were not significantly different than before administration (Figure 2). After 3 weeks of administration, the number of anal gas evacuations on the standard diet (Table 2) significantly decreased as compared to the early administration period (P = 0.001) and returned to the pre-administration level (11 ± 2 daytime evacuations; P = 0.351 vs. pre-administration) (Figure 3). The same adaptive effect was observed on the volume of gas evacuated after the probe meal (169 ± 23 mL; P = 0.002 vs. early administration; P = 0.733 vs. pre-administration) (Figure 4).

Microbial changes during the intervention

Cluster analysis of the microbial profiles in the 60 faecal samples (three time points per subject) was performed using the unweighted UniFrac principal coordinates analysis (PcoA) (Figure 6). Samples of the three time points clustered together in most subjects, indicating that intra-individual fluctuations of the microbiota during

the intervention were less distinctive than the inter-individual differences. Changes in composition during HOST-G904 administration followed different patterns in different individuals and no significant statistical differences in specific taxa were found when comparing day 0 vs. day 3 or day 21 samples in the overall group of 20 subjects. However, the relative abundance of bifidobacteria increased in 13 subjects with low levels at baseline, as defined by relative abundance below 0.5% of total bacteria. In these subjects (n = 13) abundance of bifidobacteria at day 3 (0.49 ± 0.23%) and at day 21 (0.28 ± 0.10%) was significantly higher than at day 0 (0.14 ± 0.06%; P = 0.042 and P = 0.031, respectively); values at day 3 and day 21 were not significantly different (P = 0.200). The remainder seven individuals with abundance above 0.5% at baseline showed no significant changes in bifidobacteria. There were no differences in gas volumes between subjects who increased bifidobacteria numbers during HOST-G904 consumption (from day 0 to day 21) and those with stable abundance.

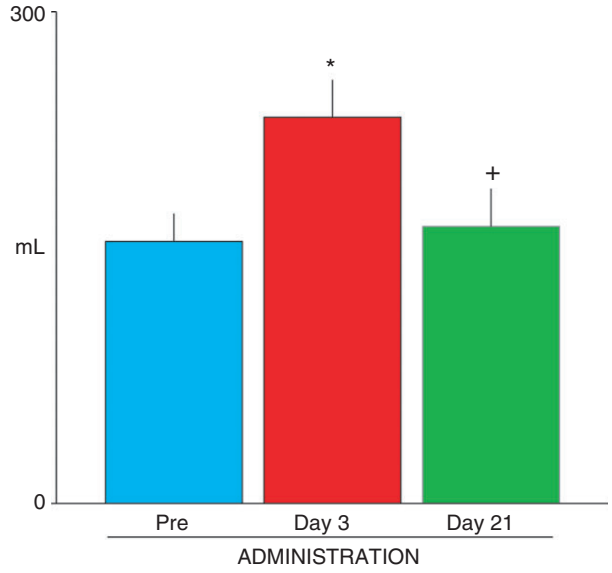


Figure 4 | Volume of gas evacuated in response to probe meal pre- administration, in the early treatment period (day 3), in the late administration period (day 21) (n = 20).

Volumes of gas recorded at day 21 correlated inversely with abundances of Lachnospiraceae (r = −0.52, P = 0.02), Clostridiaceae (r = −0.41, P = 0.07) and an unknown clostridiales species (r = −0.45, P = 0.04) in faecal samples at day 21. In addition, seven subjects harbouring methanogens (Methanobrevibacter) also produced low volumes of gas at day 21, although the correlation in the whole group (n = 20) was not significant due to the fact that in 13 subjects methanogens were negligible. Figure 7 shows the 3D display of abundances of Methanobrevibacter, Lachnospiraceae and Clostridiaceae, where the black solid dots represent the 5

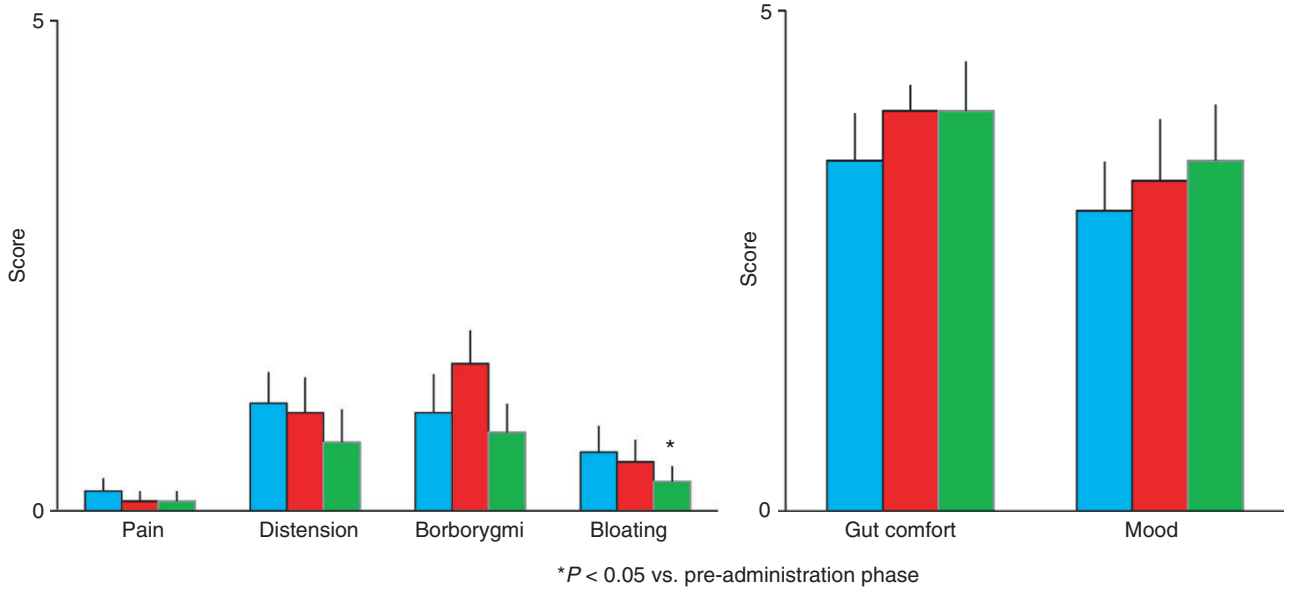


Figure 5 | Symptoms in response to probe meal pre- administration (blue), in the early administration period day 3 (red), in the late administration period day 21 (green) (n = 20).

Aliment Pharmacol Ther 2017; 45: 670–680

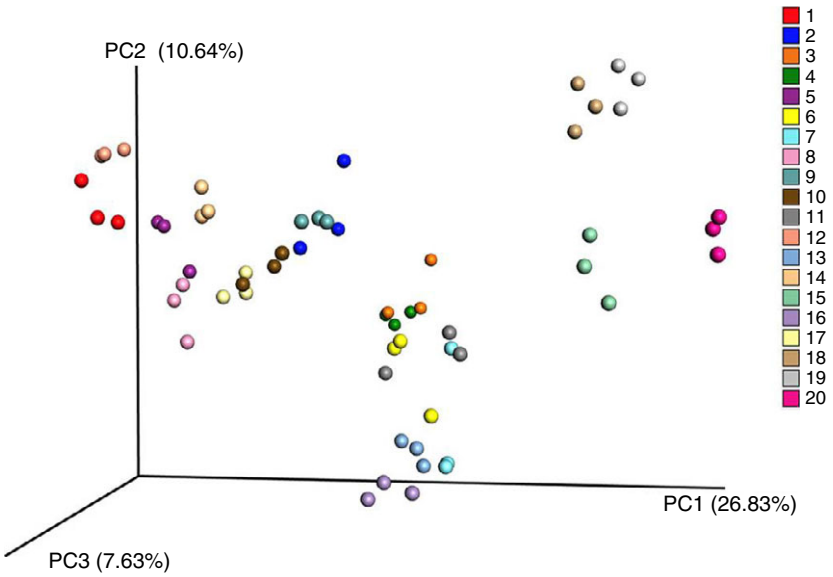


Figure 6 | Principal coordinates analysis of the microbial profiles in faecal samples from 20 subjects at 3 time points (day 1, 3 and 21). Samples from the same subject (same colour) clustered together in most cases, indicating that intra-individual fluctuations of the gut microbiota during the intervention were less distinctive than the inter-individual differences.

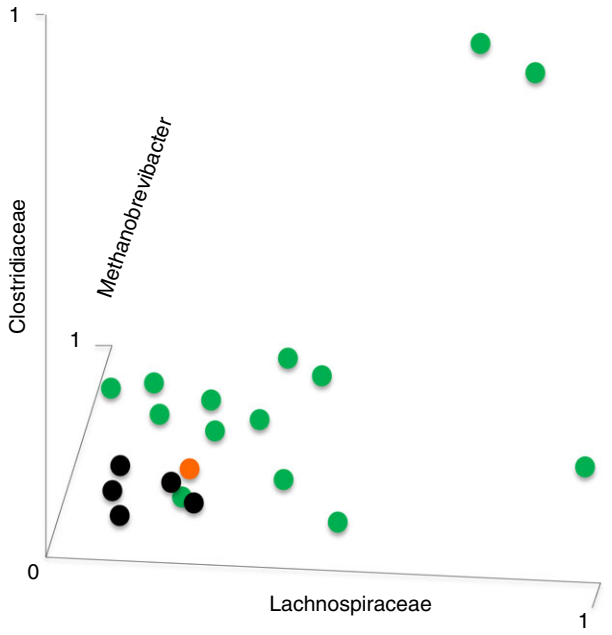


Figure 7 | A three-axis plot shows relative abundance of *Methanobrevibacter, Lachnospiraceae* and *Clostridiaceae* in faecal samples at day 21. The black solid dots are the samples from five subjects who produced more than 200 mL of gas at day 21. The remainder 15 subjects produced less than 200 mL of gas, and the red dot is a sample with high abundance of an unknown clostridiales species.

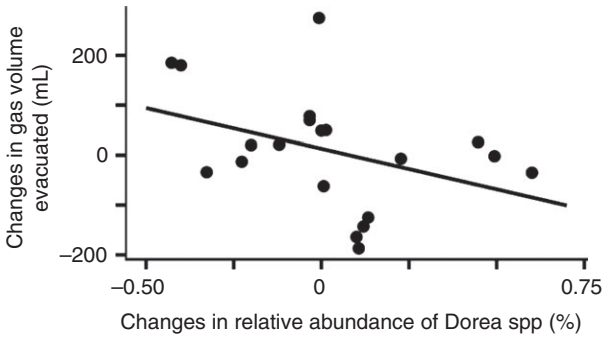
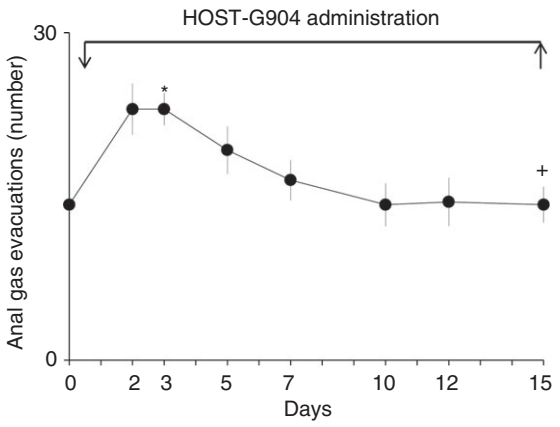


Figure 8 | Inverse correlation ($r = -0.48$; $P = 0.03$) between changes in gas volume from day 0 to 21, and changes in the abundance of *Dorea* spp. from day 0 to 21. Increases in *Dorea* spp. were associated to reduced gas production by the end of the prebiotic administration period.

subjects who produced more than 200 mL of gas at day 21. Thus, the five individuals with high gas production after 21 day prebiotic consumption had low abundance of the above-mentioned taxa; in contrast, in the remainder 15 subjects with low gas production at day 21 (<200 mL), gas production had decreased during the administration period, and 14 of them showed higher abundance of methanogens, Lachnospiraceae, Clostridiaceae or an unknown clostridiales species (Figure 6). Finally, Figure 8 shows changes relative to baseline in gas production and in abundance of the genus *Dorea* in faeces. An inverse correlation ($r = -0.48$, $P = 0.03$) suggests that increases in *Dorea* spp. were associated to reduced gas production by the end of the prebiotic administration period. Lachnospiraceae, Clostridiaceae and *Dorea* species ferment sugars and produce organic acids.



*P = 0.015 vs. day 0; + P = 1 vs. day 0 and P = 0.007 vs. day 3

Figure 9 | Control study. Time effect of HOST-G904 administration on number of daytime anal gas evacuations measured in the ancillary study (n = 6).

Ancillary study: external control group

With participants on their habitual diet, the number of anal gas evacuations significantly increased at the beginning of HOST-G904 administration and gradually declined back to the baseline level by 10 day administration (Figure 9).

DISCUSSION

Our data demonstrate the adaptation of gut microbiota to the intraluminal environment: increased availability of HOST-G904 led to a change in microbiota that had more efficient metabolic routes. High volumes of gas correlated with low methanogenic populations in some participants. This may be explained by increases in hydrogen, which is the usual substrate for gut-derived methane.[34, 35] Conversely, lower gas production was concomitant with high methanogenic populations, presumably as hydrogen was converted to methane.

The effect of HOST-G904 administration was tested with participants on a standardised diet, that is, a diet low in fermentable residues, as previously tested in our laboratory,[5] but complemented with fixed portions of a choice of foodstuffs containing equivalent amounts of fibre that were also previously used in a high-flatulogenic diet.[4] Under these conditions, the amount of daytime anal gas evacuations was in the expected range based on previous data from healthy subjects under various dietary regimes.[4, 5, 36] Similarly, the volume of gas evacuated per anus measured for 4 h after a low-residue probe meal fit the anticipated values derived from previous observations under various experimental conditions.[4, 5, 36] Interestingly, a relatively small supplement with a non-absorbable product (2.8 g/day HOST-G904) in addition to the dietary fibre intake, initially produced a marked effect on the microbiota metabolic activity, as reflected by gas production and as measured by the number of daytime anal gas evacuations and volume of gas evacuated after the probe meal. The effect of HOST-G904 on the microbiota metabolic activity was already present 24 h after the first administration without major changes observed 24 h later. The colonic transit time of inert residues averages 35 h in healthy subjects[37, 38]; conceivably, HOST-G904 was consumed at an earlier stage before reaching the distal colon.

The volume of intestinal gas produced depends in part on the amount of fermentable residues reaching the colon and the metabolic pathways used by the microbiota to consume them.[39] Hence, on the same diet, the amount of gas produced differs among individuals as a function of their microbiota profile. A portion of the gas produced by fermentation is consumed by other pools of microorganisms in the formation of less oxidised products, a portion is absorbed into the blood and cleared by breathing, and the rest is evacuated per anus.[39] At the first exposure to HOST-G904, the increase in gas production was conceivably related to the availability of substrates. A decrease in anal gas evacuation after adaptation, that is, a decrease in net gas production, could be related to the proliferation of microorganisms using non-fermentative pathways to metabolise the substrates with less gas production and/or to the up-regulation of the gas-consuming activity. Indeed, reduced gas production was related to the proliferation of methanogens that use H_2 to reduce CO_2 to CH_4, reducing the volume of gas by 1:5.[39] Conversely, individuals with low counts of methanogens and of some specific organic acid producers exhibited a poor adaptation at the end of the administration period.

To ensure similar testing conditions within and between individuals, participants were put on standardised diet during the 3-day evaluation periods before, at the beginning and at the end of prebiotic administration. We wish to acknowledge that diet standardisation may have different effects depending on individuals′ habitual diet with potential increase or decrease in fibre intake, and this might interfere with the effect of the prebiotic. To account for this potential limitation, the prebiotic was also tested in an external control group of subjects on their habitual diet and the same response, in terms or number of anal gas evacuations, was observed. We acknowledge the inherent limitations and potential bias

of external controlled trials, and that in the control group the volume of gas evacuated and microbiota were not measured.

It is interesting that a relatively small amount of HOST-G904 relative to the daily dietary fibre intake had a remarkable effect, initially on microbiota metabolic activity and subsequently inducing adaptation. Conceivably, not all fermentable residues have the same capability; the power to induce adaptation might be a crude indicator of prebiotic activity. We based the test dose on previous studies[9]; a higher dose might compromise selectivity of fermentation, which is a requirement for prebiotic effects. The dose of HOST-G904 that was used activated microbiota metabolism and increased the number of anal gas evacuation without inducing abdominal symptoms in healthy subjects. We cannot ascertain how this dose would be tolerated by patients with functional gut disorders, but conceivably, potential symptoms at first exposure would subside with adaptation. The ancillary study showed that the number of daily anal gas evacuations started to decrease after 5 day administration and by 10 day returned to pre-administration level; it remains to be determined whether symptom adaptation in patients follows the same time pattern.

HOST-G904 has been shown to improve symptoms in patients,[9] which was attributed to beneficial changes in gut microbiota. Our study in healthy subjects has potential clinical implications. Indeed, our data would support advising patients to allow for a period of adaptation before prebiotic effects become apparent.

AUTHORSHIP

Guarantor of the article: Fernando Azpiroz.

Author contributions: MM: study management, conduction of experiments, and data analysis; CM. microbiota analysis, manuscript revision; AA: supervision of studies; DC: microbiota analysis; MP: microbiota analysis; EV: microbiota analysis; JV: study design, interpretation of results, manuscript revision; GT: study design, interpretation of results, manuscript revision; GG: study design, interpretation of results, manuscript revision; FG: study design, interpretation of results, manuscript revision; FA: study design, data interpretation, and manuscript preparation.

All authors approved the final draft of the manuscript.

ACKNOWLEDGEMENTS

The authors thank Anna Anna Aparici, Sara Mendez for technical support; Gloria Santaliestra for secretarial assistance.

Declaration of personal interests: Jelena Vulevic, employee Clasado, George Tzortzis, employee Clasado, Glen Gibson, Francisco Guarner and Fernando Azpiroz served as advisory board members for Clasado. Rest of authors, no competing interests declared. Writing Assistance: American Journal Experts for English editing of the manuscript (Certificate Verification Key: BBC7-69E1-C87F-F6A6-3087) funded by SAF 2013-43677-R.

Declaration of funding interests: The present study was supported by the Spanish Ministry of Economy and Competitiveness (Dirección General de Investigación Científica y Técnica, SAF 2013-43677-R; SAF 2016-76648-R) and Instituto de Salud Carlos III/FEDERI (PI14/00764); Ciberehd was funded by the Instituto de Salud Carlos III.

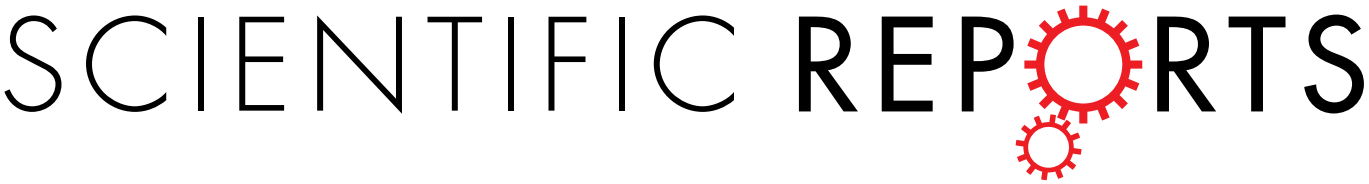
LINKED CONTENT

This article is linked to Staudacher paper. To view this article visit https://doi.org/10.1111/apt.13976.

REFERENCES

1. Roberfroid M, Gibson GR, Hoyles L, *et al.* Prebiotic effects: Metabolic and health benefits. *Br J Nutr* 2010; **104** (Suppl. 2): S1–63.
2. Macfarlane S, Macfarlane GT, Cummings JH. Review article: prebiotics in the gastrointestinal tract. *Aliment Pharmacol Ther* 2006; **24**: 701–14.
3. Francis CY, Whorwell PJ. Bran and irritable bowel syndrome: time for reappraisal. *Lancet* 1994; **344**: 39–40.
4. Manichanh C, Eck A, Varela E, *et al.* Anal gas evacuation and colonic microbiota in patients with flatulence: effect of diet. *Gut* 2014; **63**: 401–8.
5. Azpiroz F, Hernandez C, Guyonnet D, *et al.* Effect of a low-flatulogenic diet in patients with flatulence and functional digestive symptoms. *Neurogastroenterol Motil* 2014; **26**: 779–85.
6. Peters SL, Yao CK, Philpott H, Yelland GW, Muir JG, Gibson PR. Randomised clinical trial: the efficacy of gut-directed hypnotherapy is similar to that of the low FODMAP diet for the treatment of irritable bowel syndrome. *Aliment Pharmacol Ther* 2016; **44**: 447–59.
7. Rao SS, Yu S, Fedewa A. Systematic review: dietary fibre and FODMAP-restricted diet in the management of constipation and irritable bowel syndrome. *Aliment Pharmacol Ther* 2015; **41**: 1256–70.
8. Halmos EP, Power VA, Shepherd SJ, Gibson PR, Muir JG. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. *Gastroenterology* 2014; **146**: 67–75.
9. Silk DB, Davis A, Vulevic J, Tzortzis G, Gibson GR. Clinical trial: the effects of a trans-galactooligosaccharide prebiotic on faecal microbiota and symptoms in irritable bowel syndrome. *Aliment Pharmacol Ther* 2009; **29**: 508–18.
10. Hughes C, Davoodi-Semiromi Y, Colee JC, *et al.* Galactooligosaccharide supplementation reduces stress-induced gastrointestinal dysfunction and days of cold or flu: a randomized, double-blind, controlled trial in healthy university students. *Am J Clin Nutr* 2011; **93**: 1305–11.
11. Barba E, Burri E, Accarino A, *et al.* Abdomino-thoracic mechanisms of functional abdominal distension and correction by biofeedback. *Gastroenterology* 2015; **148**: 732–8.
12. Burri E, Barba E, Huaman JW, *et al.* Mechanisms of postprandial abdominal bloating and distension in functional dyspepsia. *Gut* 2014; **63**: 395–400.
13. Malagelada C, Drozdzal M, Segui S, *et al.* Classification of functional bowel disorders by objective physiological criteria based on endoluminal image analysis. *Am J Physiol Gastrointest Liver Physiol* 2015; **309**: G413–9.
14. Malagelada C, Barba I, Accarino A, *et al.* Cognitive and hedonic responses

to meal ingestion correlate with changes in circulating metabolites. *Neurogastroenterol Motil* 2016; **28**: 1806–14.
15. Muir JG, Rose R, Rosella O, *et al.* Measurement of short-chain carbohydrates in common Australian vegetables and fruits by high-performance liquid chromatography (HPLC). *J Agric Food Chem* 2009; **57**: 554–65.
16. Biesiekierski JR, Rosella O, Rose R, *et al.* Quantification of fructans, galacto-oligosacharides and other short-chain carbohydrates in processed grains and cereals. *J Hum Nutr Diet* 2011; **24**: 154–76.
17. Odimet, Organizador dietético Metabólico. Hospital Clínico Universitario de Santiago de Compostela, Spain 29 October 014. Available at: http://www.guiametabolica.org/recurso/odimet-organizador-dietetico-metabolico
18. Malagelada C, Accarino A, Molne L, *et al.* Digestive, cognitive and hedonic responses to a meal. *Neurogastroenterol Motil* 2015; **27**: 389–96.
19. Serra J, Azpiroz F, Malagelada J-R. Gastric distension and duodenal lipid infusion modulate intestinal gas transit and tolerance in humans. *Am J Gastroenterol* 2002; **97**: 2225–30.
20. Serra J, Azpiroz F, Malagelada J-R. Mechanisms of intestinal gas retention in humans:impaired propulsion versus obstructed evacuation. *Am J Physiol* 2001; **281**: G138–43.
21. Serra J, Azpiroz F, Malagelada J-R. Intestinal gas dynamics and tolerance in humans. *Gastroenterology* 1998; **115**: 542–50.
22. Serra J, Azpiroz F, Malagelada J-R. Impaired transit and tolerance of intestinal gas in the irritable bowel syndrome. *Gut* 2001; **48**: 14–9.
23. Cardona S, Eck A, Cassellas M, *et al.* Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol* 2012; **12**: 158.
24. Walters WA, Caporaso JG, Lauber CL. Primer-Prospector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* 2011; **27**: 1159–61.
25. Caporaso JG, Lauber CL, Walters WA, *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012; **6**: 1621–4.
26. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, *et al.* Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol* 2013; **531**: 371–444.
27. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010; **26**: 2460–1.
28. Serra J, Salvioli B, Azpiroz F, Malagelada JR. Lipid-induced intestinal gas retention in the irritable bowel syndrome. *Gastroenterology* 2002; **123**: 700–6.
29. Hernando-Harder AC, Serra J, Azpiroz F, *et al.* Colonic responses to gas loads in subgroups of patients with abdominal bloating. *Am J Gastroenterol* 2010; **105**: 876–82.
30. Tremolaterra F, Villoria A, Azpiroz F, Serra J, Aguade S, Malagelada J-R. Impaired viscerosomatic reflexes and abdominal wall dystony associated with bloating. *Gastroenterology* 2006; **130**: 1062–8.
31. Passos MC, Tremolaterra F, Serra J, Azpiroz F, Malagelada J-R. Impaired reflex control of intestinal gas transit in patients with abdominal bloating. *Gut* 2005; **54**: 344–8.
32. Salvioli B, Serra J, Azpiroz F, *et al.* Origin of gas retention and symptoms in patients with bloating. *Gastroenterology* 2005; **128**: 574–9.
33. Caldarella MP, Serra J, Azpiroz F, Malagelada JR. Prokinetic effects in patients with intestinal gas retention. *Gastroenterology* 2002; **122**: 1748–55.
34. Strocchi A, Furne J, Ellis C, Levitt MD. Methanogens outcompete sulphate reducing bacteria for H2 in the human colon. *Gut* 1994; **35**: 1098–101.
35. Suarez F, Furne J, Springfield J, Levitt M. Insights into human colonic physiology obtained from the study of flatus composition. *Am J Physiol* 1997; **272**: G1028–33.
36. Mego M, Accarino A, Malagelada JR, Guarner F, Azpiroz F. Accumulative effect of food residues on intestinal gas production. *Neurogastroenterol Motil* 2015; **27**: 1621–8.
37. Metcalf AM, Phillips SF, Zinsmeister AR, MacCarty RL, Beart RW, Wolff BG. Simplified assessment of segmental colonic transit. *Gastroenterology* 1987; **92**: 40–7.
38. Fort JM, Azpiroz F, Casellas F, Andreu J, Malagelada J-R. Bowel habit after cholecystectomy: physiological changes and clinical implications. *Gastroenterology* 1996; **111**: 617–22.
39. Azpiroz F. Intestinal gas. In: Feldman M, Friedman LS, Brand LJ, eds. *Pathophysiology, Diagnosis, Management*. 10th ed. Philadelphia, USA: Elsevier, 2015; 242–50.



OPEN

Alteration of the serum microbiome composition in cirrhotic patients with ascites

Received: 25 February 2016
Accepted: 08 April 2016
Published: 26 April 2016

Alba Santiago[1], Marta Pozuelo[1], Maria Poca[2,3], Cristina Gely[2], Juan Camilo Nieto[4], Xavier Torras[2,3], Eva Román[2,3,5], David Campos[1], Guillaume Sarrabayrouse[1], Silvia Vidal[4], Edilmar Alvarado-Tapias[2], Francisco Guarner[1,3], German Soriano[2,3], Chaysavanh Manichanh[1,3] & Carlos Guarner[2,3]

The progression of cirrhosis is associated with alterations in the composition of the gut microbiome. To assess microbial translocation, we compared the serum microbial composition of patients with and without ascites and characterized the ascitic fluid microbiome using 16S rDNA high-throughput sequencing data. A complex and specific microbial community was detected in the serum and ascitic fluid of patients with cirrhosis but barely detectable in the serum of healthy controls. The serum microbiome of patients with ascites presented higher levels of lipopolysaccharide binding protein, a marker of microbial translocation, associated with higher diversity and relative abundance of Clostridiales and an unknown genus belonging to the Cyanobacteria phylum compared to patients without ascites. The composition of the fecal microbiome was also more altered in patients with than without ascites, confirming previous studies on fecal microbiome. We propose that alteration of the serum and fecal microbiome composition be considered indicators of cirrhosis progression.

Liver cirrhosis is a major cause of global health loss. In this regard, its incidence increased from 676,000 patients in 2008 to over 1 million in 2010[1]. It is the final phase of chronic liver disease, in which inflammation is associated with dying hepatic cells and fibrosis, leading to poor liver function and portal hypertension. Alterations in the gut microbiota, which represents the collective microbial cells present in the digestive tract, or its products, are linked to the progression of liver disease and the complications of cirrhosis[2]. Over the last decade, advances in molecular techniques and bioinformatics, as well as the exponential decrease in the cost of sequencing, have allowed comprehensive characterization of the composition and function of the gut microbial community. Using these techniques, recent studies on the gut microbiome have demonstrated an alteration of the composition of the stool microbial community in cirrhotic patients compared to healthy controls[3,4]. Furthermore, this level of alteration appears to be positively correlated with the severity of the disease[5].

More specifically, bacterial translocation has been suspected to play an important role in the pathogenesis and complications of cirrhosis. By administering green fluorescent protein (GFP)-labeled *Escherichia coli* orally to cirrhotic rats, Teltschik *et al.*[6] revealed the presence of bacteria not only in the intestinal lumen but also in mesenteric lymph nodes (MLNs) and ascites. We also recently described that rat MLNs harbor a high microbial diversity[7]. However, very little is known about the microbiome of extra-intestinal sites such as the systemic circulation and ascitic fluid in patients with cirrhosis.

This study sought to: (a) characterize the microbiome of serum and fecal samples of patients with cirrhosis and compare them with those of healthy controls; (b) define the serum microbiome associated with severity of liver disease; and (c) identify the microbiome of ascitic fluid.

[1]Department of Gastroenterology, Vall d'Hebron Research Institute, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. [2]Department of Gastroenterology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [3]Centro de Investigaciòn Biomédica en Red en el Área temática de Enfermedades Hepáticas y Digestivas, CIBERehd, Instituto de Salud Carlos III, Madrid, Spain. [4]Department of Immunology, Institut de Recerca-IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [5]Escola Universitària d'Infermeria EUI-Sant Pau, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. Correspondence and requests for materials should be addressed to C.M. (email: cmanicha@gmail.com) or G.S. (email: gsoriano@santpau.cat)

Results

Enrollment process. A total of 60 outpatients with cirrhosis were evaluated. Thirty-three were excluded due to treatment with non-absorbable disaccharides and/or antibiotics ($n = 11$), current alcohol intake ($n = 7$), hepatocellular carcinoma ($n = 5$), spontaneous bacterial peritonitis ($n = 1$), other infections or suspicion of infection ($n = 3$), severe comorbidities ($n = 4$), or because they were unwilling to participate in the study ($n = 2$). Therefore, a total of 27 patients were included—13 with ascites and 14 without ascites. Seventeen healthy controls were included for stool ($n = 17$) and serum ($n = 7$) microbiome analysis.

Patient characteristics. The characteristics of patients are shown in Supplementary Table 1. The main differences between the two groups of patients consisted, as expected, of a more advanced liver insufficiency as determined by the Child-Pugh score and a higher incidence of previous ascites in patients with than in those without ascites. When analyzing other factors that could influence the microbiome composition, we did not find statistical differences between the two groups regarding age, body mass index or etiology of cirrhosis. Patients without ascites showed a trend towards a lower prevalence of diabetes than those with ascites and they were more frequently receiving treatment with beta-blockers or proton pump inhibitors. These differences, however, did not reach statistical significance.

No patient in either group presented symptoms, signs at physical examination or analytical data suggesting infection. Microbial cultures were negative, and neutrophil count was $<250/mm^3$ in all ascitic fluid samples. Therefore, all patients with ascites were considered to have a non-infected ascitic fluid.

Microbiome in stool. The stool microbiome of 27 patients with cirrhosis was compared to that of 17 healthy controls. Alpha-diversity analysis showed that the fecal microbial community of healthy controls presented a higher diversity than that of patients with cirrhosis (Fig. 1a). However, the diversity was similar in patients with or without ascites (Fig. 1b). Together, these results suggest that a loss of microbial diversity in fecal samples is associated with cirrhosis without ascites, but the progression to ascites is not associated with a further loss of diversity. Clustering analysis using PCoA and UPGMA methods based on UniFrac metrics showed that the stool microbiome of cirrhotic patients clustered separately from that of healthy controls (Fig. 1c,d). At the taxonomic level, patients with cirrhosis were depleted of six species (FDR < 0.05; Kruskal-Wallis test): unknown Clostridiales, *Roseburia faecis*, *Alistipes putredinis*, unknown *Oscillospira*, unknown Mogibacteriaceae, and unknown *Dehalobacterium*, but were enriched in an unknown Peptostreptococcaceae compared to healthy controls (FDR < 0.05; Kruskal-Wallis test; Fig. 1e). Proteobacteria, at the phylum level, were more abundant in cirrhotic patients than in healthy controls but the difference did not reach significance (FDR $= 0.42$; Kruskal-Wallis test). All together, these results confirm previous findings that the microbiome composition of cirrhotic patients is altered[4].

Cirrhosis can progress to ascites, which is defined as the accumulation of fluid in the peritoneal cavity. Interestingly, when we analyzed the stool microbiome of patients with ascites and those without ascites separately, only the former displayed a significant dysbiosis at the species level, with depletion of unknown Ruminococcaceae, Clostridiales and Peptostroptococcaceae, *Roseburia faecis* and *Alistipes putredinis* and with an enrichment of *Veillonella dispar* compared to healthy controls (FDR < 0.05; Kruskal-Wallis test; Fig. 1f). For several of these species, such as *Roseburia faecis*, *Alistipes putredinis* and *Veillonella dispar,* our findings are in line with those of Qin *et al.*[4] and further support the notion that the progression of the disease is associated with a greater dysbiosis, as reported by Bajaj *et al.*[5]. Patients without ascites presented only a trend towards lower relative abundance of unknown Mogibacteriaceae and *Alistipes* (FDR $= 0.053$; Kruskal-Wallis test) compared to healthy controls.

Microbiome in fluids. Standard diagnostic microbiological analysis revealed that the serum and ascitic fluid samples were negative for bacterial growth. We analyzed the microbiome serum from 7 healthy controls and from the 27 patients and ascitic fluid from 11 patients. Analysis of the 16S rRNA gene of such low-biomass samples may generate contamination at various steps of the process. Therefore, we applied strict protocols for sample collection, DNA extraction, and PCR amplification. For sample collection, we used gloves and proceeded in sterile conditions. For DNA extraction, we used chemicals such as DNA terminator (Biotools, B & M Labs, Spain) to degrade any trace of contaminant DNA in laboratory equipment, and we added negative controls (blanks) during extraction. During PCR amplification, we used UV to clean consumables and H2O and also added PCR blanks.

The amplicons were analyzed in an electrophoretic gel and their presence was indicated by a DNA band at about 400 bp (Supplementary Fig. 1). No DNA band was observed for four control serum samples out of seven, one serum sample from patient with ascites and one ascitic fluid sample, or for the negative controls added during the extraction (NEG1 and NEG2) and PCR (NEG3) procedures. The PCR amplifications of serum and ascitic fluid samples provided a gradient of intensity in the DNA bands, as analyzed in the electrophoretic gel (Supplementary Fig. 1), in the following order: healthy control serum < cirrhotic patients without ascites < cirrhotic patients with ascites < ascitic fluid, thereby also suggesting a gradient in the microbial load. To remove potential false positive OTUs during sequence analysis, we subtracted sequences with abundant taxa generated in the blanks from the serum and ascitic fluid samples and applied a more restricted filter to the data obtained from samples in order to remove taxa with a low abundance, as specified in the method section. The contamination present in the negative controls was identified as being mostly Proteobacteria (69%) at the phylum level and Pseudomonas (30%), Halomonas (18%) and unknown (12%) at the genus level. After this filtering step and at a rarefaction of 1000 sequences per sample, we obtained sequence data for 24 out of 27 serum samples from patients and for eight out of 11 ascitic fluid samples and no sequence data were recovered from healthy controls. Supplementary Fig. 2 shows the taxonomic profiling of the three sample types at the phylum level before and after the sequence-filtering step.

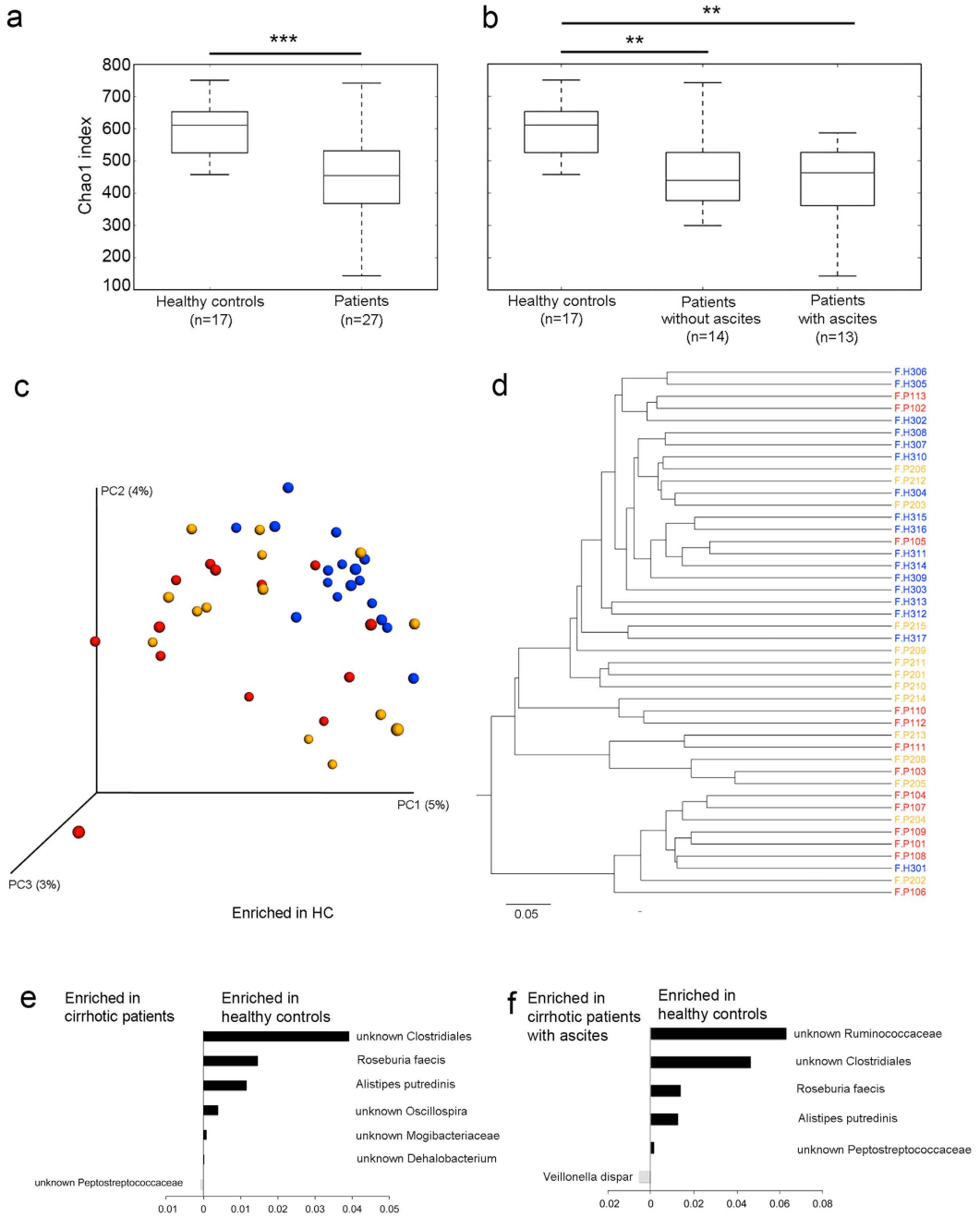


Figure 1. Fecal microbiome of cirrhotic patients and healthy controls. (**a,b**) Healthy controls presented higher microbial diversity compared to all cirrhotic patients (**a**) and to patients with and patients without ascitic fluid (**b**) as assessed by the Chao1 index. The two groups of patients with and without ascites were not significantly different. (**c,d**) Unweighted UniFrac PcoA (**c**) and weighted UniFrac UPGMA (**d**) clustering analysis. Blue: healthy controls; orange: patients without ascites; and red: patients with ascites. (**e,f**) Relative abundance of microbes differentially present at the species level between healthy controls and all cirrhotic patients (**e**) and between healthy controls and cirrhotic patients with ascites (**f**) (Kruskal-Wallis; FDR < 0.05). Analyses were performed on 16 S rRNA V4 region data, obtained from stool samples, rarefied to a depth of 19,930 reads per sample. Healthy controls ($n = 17$); patients ($n = 27$); patients with ascites ($n = 13$); patients without ascites ($n = 14$); $^{***}P = 0.001$; $^{**}P = 0.003$.

Beta-diversity analysis, which studies the variation in composition between samples, showed a similar microbial composition between serum and ascitic fluid samples. However, the microbial community differed greatly between these two sample types and the stools (Fig. 2a), although 89% and 86% of the serum and ascitic fluid microbiome was shared with the stool microbiome at the genus level (Supplementary Fig. 3). Euryarchaeota (phylum level) was detected only in stool samples and Thermi and Deinococcus-Thermus were detected only in ascitic fluid (Fig. 2b). Firmicutes and Bacteroidetes were the two most dominant phyla in the three sample types.

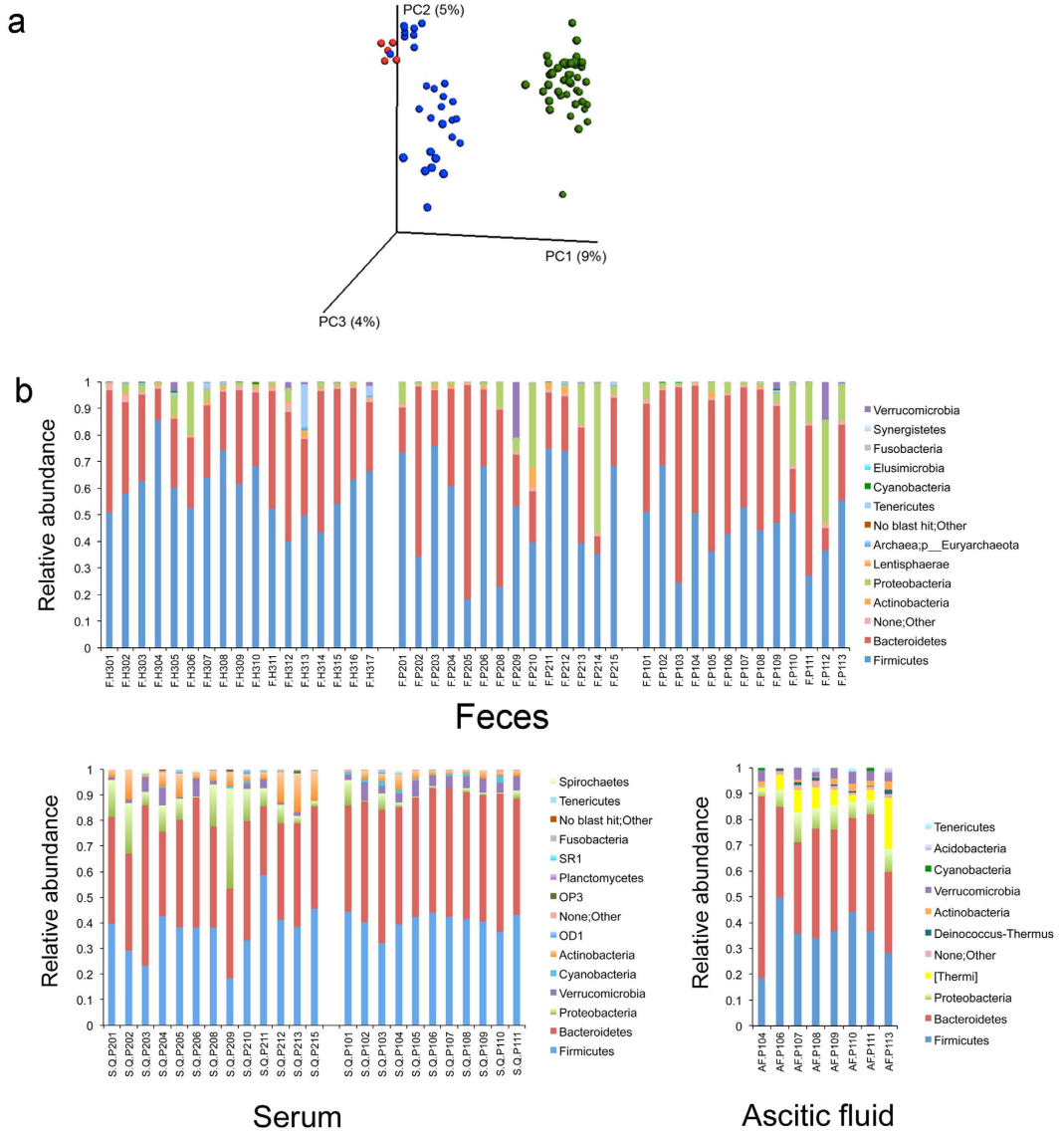


Figure 2. Fecal, serum and ascitic fluid microbiome. (**a**) Clustering of samples using unweighted UniFrac PcoA representation. (**b**) Taxonomic composition at the phylum level of the three sample types: Feces, serum, and ascitic fluid. Analyses were performed on 16 S rRNA V4 region data, rarefied to a depth of 19,930 reads for stool and 1,000 reads for serum and ascitic fluid samples. Green: stool; blue: serum; red: ascitic fluid. F.H = Feces of healthy controls; F.P = Feces of patients with cirrhosis; S.Q.P = Serum of patients; AF.P = ascitic fluid of patients. 201 to 215 = patients without ascites; 101 to 113 = patients with ascites.

From serum and ascitic fluid, we detected six and eight groups of microbes at the phylum level, 26 and 28 groups at the family level, and 36 and 38 groups at the genus level, respectively. At the phylum level, Firmicutes (41%), Bacteroidetes (37%) and Proteobacteria (14%) accounted for 92% of the sequence data of the serum microbiome, whereas in ascitic fluid Firmicutes (46%), Bacteroidetes (27%), Thermi (10%) and Proteobacteria (8%) accounted for 92%. Serum and ascitic fluid were similar in terms of diversity and richness, as assessed by an abundance-based richness estimator (Chao1) (Fig. 3a). However, serum specimens of patients with ascitic fluid presented a more diverse microbiome ($P = 0.008$) than those of patients without (Fig. 3b), and a significantly higher concentration of lipopolysaccharide binding protein (LBP) ($P = 0.02$, Mann Whitney test), a marker of microbial translocation (Fig. 3c). This observation could be explained by patients with ascites, who are expected to have a greater deterioration of the intestinal barrier integrity, also having a higher degree of microbial translocation than those without ascites, thus leading to a higher microbial diversity in serum.

Furthermore, using an UPGMA clustering method of the serum microbiome based on an unweighted UniFrac metric, the microbiome of patients with and without ascites clustered separately (Fig. 4a). This result suggests that a specific serum microbiome is linked to the presence of ascites.

Taxonomic comparison showed that an unknown group of microbes at the family level, belonging to the Clostridiales order, displayed a higher relative abundance in serum of patients with ascites (FDR = 0.03;

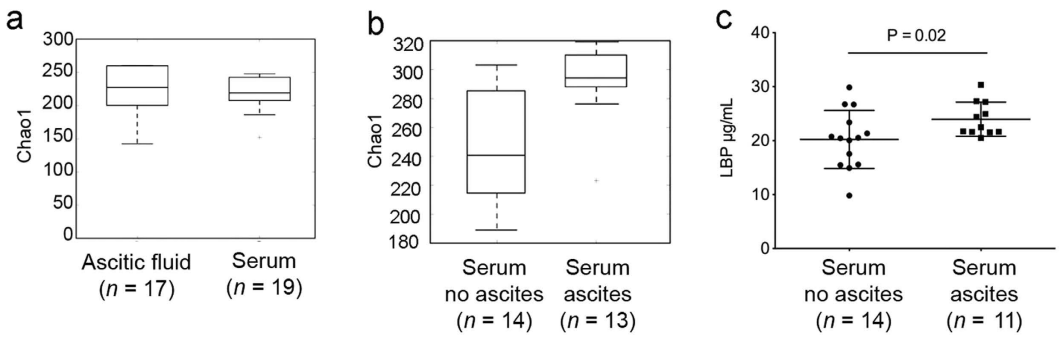


Figure 3. Microbial of extra-intestinal sites and marker of translocation. (**a**) Alpha-diversity of the microbial fluid samples as assessed by Chao1 index of diversity. Ascitic fluid ($n = 11$); Serum of patients with cirrhosis ($n = 19$; instead of 27 due to rarefaction depth with ascitic fluid samples). (**b**) Higher alpha-diversity of serum microbiome of cirrhotic patients with ascites compared to that of patients without ($P < 0.05$). Analyses were performed on 16 S rRNA V4 region data, rarefied to a depth of 1,000 reads per sample. (**c**) Lipopolysaccharide binding protein (LBP) levels as assessed by specific ELISA; serum of patients with ascites ($n = 11$ available samples).

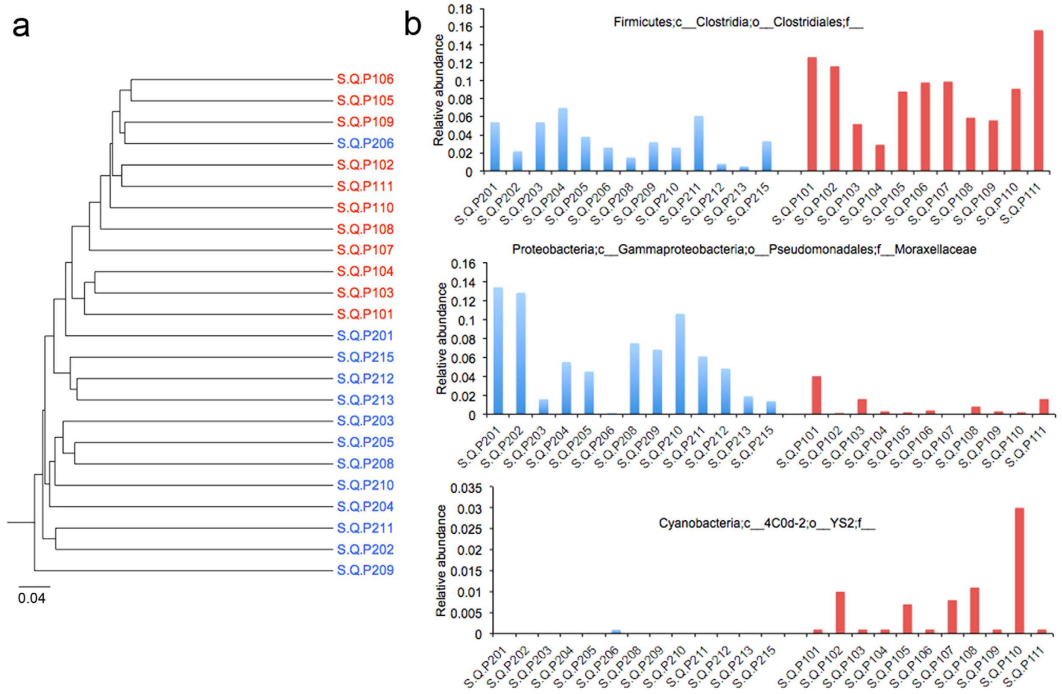


Figure 4. Serum microbiome of patients with and without ascites. (**a**) UPGMA clustering based on unweighted UniFrac metric of serum samples of cirrhotic patients with and without ascites. (**b**) Relative abundance of microbes or groups of microbes significantly different between serum microbiome of cirrhotic patients with and without ascites. Analyses were performed on 16S rRNA V4 region data, rarefied to a depth of 1,000 reads per sample.

Kruskal-Wallis test) and another group, Moraxellaceae, showed a lower relative abundance in patients with ascites compared to those without (Fig. 4b). Interestingly, this group of bacteria was also found in ascitic fluid samples (Supplementary Fig. 4), thereby supporting the notion of translocation from serum to ascitic fluid. At the genus and species level, an unknown genus related to Cyanobacteria (FDR = 0.002) was found in higher relative abundance in patients with ascites compared to those without.

Microbial translocation. In order to study whether the presence of bacterial DNA in ascitic fluid and blood derived from the gastrointestinal tract, we counted the taxa common to stool and serum, stool and ascitic fluid, and serum and ascitic fluid. For this purpose, we first counted the number of taxa in each sample type, finding an average of 397 (SD = 94), 283 (SD = 76) and 97 (SD = 25) taxa in stool, serum and ascitic fluid, respectively. By comparing the taxa between samples, we detected on average 37 taxa common to both stool and serum, 20 to

serum and ascitic fluid, and three to ascitic fluid and stool (Supplementary Fig. 5). These results indicate that the three sites share few common microbial taxa and therefore suggest that the microbial taxa present in the serum, but not detected in stool, could either take root in extra-intestinal sites such as the lung or the vagina for women or were in too low abundance in the stool to be detectable but when they reached the serum, a more appropriate environment for their growth, they became detectable.

Discussion

This is the first study to validate the presence of polymicrobial DNA in both the serum and ascitic fluid of patients with cirrhosis using high-throughput sequencing techniques. Our findings showed that the microbial community in serum and ascitic fluid, although showing more than 80% similarity with that of the stool microbiome at the genus level, is specific and complex at the taxa level. Previous studies using a variety of techniques, mainly conventional PCR, reported the presence of bacterial DNA in ascitic fluid and/or blood only in up to 30–60% of these patients[8–11]. Moreover, most of the DNA detected in these studies was monomicrobial, identified as being *Escherichia coli* or *Staphylococcus aureus*[8–11]. A recent study has reported the characterization of the microbial composition of the ascitic fluid of cirrhotic patients[12]. However, the authors amplified the 16 S gene from only one individual out of seven and this individual was positive for *Escherichia coli* in culture. Using shotgun-sequencing technique on two pools of ascitic fluid obtained from three patients, they were able to identify only 0.1% of bacterial DNA, for which the majority was identified as being *Escherichia*. However, according to our findings, *Escherichia* belonging to the Proteobacteria phylum could also be found in the extraction and PCR blanks. We therefore recommend that future studies on samples with a low biomass include several blanks and minimize the amount of Taq polymerase used during the PCR amplification, since it may contain contaminant DNA. The detection of polymicrobial DNA in the serum and ascitic fluid observed in the present study is in line with our previous findings in rats, showing a high microbial diversity in MLNs of a model of CCl4-induced liver injury, as well as in those of control rats[7].

We were unable to analyze the serum microbiome of the seven healthy controls at a sufficient rarefaction depth compared to all other samples. Indeed, the presence of DNA bands in the electrophoretic gel after serum amplification could be due to the presence of human DNA combined with contaminant DNA during extraction and amplification, thus impeding analysis of the microbiome of these samples after filtering out the contaminant sequences. As the same method of sample collection and processing was used for patients with cirrhosis, this finding supports that the detection of bacterial DNA in patients with cirrhosis was not caused by contamination. This observation suggests that healthy individuals harbor a very low or undetectable microbial load in blood, which is in agreement with a recent study demonstrating the presence of a gut-vascular barrier that controls the systemic dissemination of bacteria in healthy individuals but not in patients with celiac disease and liver damage[13]. In cirrhotic patients, the similarity of the microbiome composition between serum and ascitic fluid compared to stool samples could be due, in part, to the body site selecting only microorganisms capable of growing in a liquid and relatively aerobic environment. The differences found in diversity (Chao1 index) and in composition and structure of the serum microbiome between patients with and without ascites are alterations that are associated with cirrhosis progression, thereby validating the assumption of previous studies[14].

The decrease in stool microbial diversity and the depletion of several commensal groups of bacteria (unknown Ruminococcaceae, Clostridiales and Peptostroptococcaceae, *Roseburia faecis* and *Alistipes putredinis*) in patients with cirrhosis is also in agreement with the findings of previous studies[3,4]. However, in contrast to other authors[5], we did not observe a significant increase in potential pathogenic bacteria such as Enterobacteria, but only a trend towards an increase in Proteobacteria or Streptococcaceae. This observation could be explained by a smaller sample size and the fact that the patients in our study presented a relatively preserved liver function, as reflected by the low Child-Pugh and MELD scores, in comparison with other studies that included groups with more advanced liver failure.

Our study presents several limitations such as a small sample size, DNA contamination that may remain after sequence curating (despite the multiple precautions to avoid this as mentioned above), and confounding factors. To reduce possible confounding factors, we excluded patients with recent alcohol intake and those treated with antibiotics or non-absorbable disaccharides. We did not find statistically significant differences between patients with and without ascites in other possible confounding factors, such as diabetes and the use of beta-blockers or proton-pump inhibitors. However, we cannot exclude that the non-significant differences observed in these parameters could have influenced the results reported here.

Despite these limitations, we conclude that serum and ascitic fluid of patients with cirrhosis contain a complex and specific microbial community and that our method of low-biomass analysis could be applied to other conditions of gut-vascular barrier failure[13]. We propose that alteration of the serum and fecal microbiome composition be considered indicators of cirrhosis progression.

Methods

Ethical statement. The study included consecutive outpatients with cirrhosis treated at the *Hospital de la Santa Creu i Sant Pau*, a tertiary care hospital in Barcelona, Spain. The methods conformed to the Declaration of Helsinki and Guidelines for Good Clinical Practice in Clinical Trials and were carried out in accordance with the Clinical Research Ethics Committee of the *Hospital de la Santa Creu i Sant Pau*. All experimental protocols were approved by the same Ethics Committee. All patients received information concerning their participation in the study and gave written informed consent.

Patient information. Cirrhosis was diagnosed by clinical, analytical, and ultrasonographic findings or by liver biopsy. Exclusion criteria were the following: hospitalization in the previous month due to decompensation of cirrhosis; hepatocellular carcinoma or other neoplasia; alcohol intake in the previous 3 months;

current infection or overt hepatic encephalopathy; marked symptomatic comorbidities (cardiac, pulmonary, renal, untreated active depression); treatment with antibiotics or non-absorbable disaccharides in the previous 3 months; and life expectancy of less than 6 months. Patients were carefully evaluated to exclude active infection when joining the study. Patients were classified into two groups, namely those with ascites and those without. The former group consisted of stable patients with refractory ascites attending the day hospital for regular therapeutic paracentesis. A group of age- and gender-matched healthy controls was included to compare their stool and blood microbiome composition with that of patients with cirrhosis.

Sample collection. Fecal samples were collected by the patients or controls as previously described[15]. Blood and ascitic fluid samples were collected in sterile conditions by peripheral vein puncture and during therapeutic paracentesis, respectively. For patients with cirrhosis, we performed routine blood analysis to assess the degree of liver failure, renal function, blood white cell count, and ascitic fluid neutrophil count to rule out ascitic fluid infection (spontaneous bacterial peritonitis). Samples of blood and ascitic fluid were cultured in blood culture bottles (BactAlert®) to assess for microbial growth. Additional samples of blood and ascitic fluid were collected in in SST™ Tubes (BD Vacutainer®) tubes and 15 ml centrifuge tubes respectively, and frozen at −80 °C until DNA analysis.

Lipopolysaccharide binding protein levels. Serum was tested for lipopolysaccharide binding protein (LBP) concentration to assess exposure to bacteria and their endotoxins as an index of bacterial translocation[16,17], using specific ELISA (Biometec GmbH, Greifswald, Germany) according to the manufacturer's instructions. LBP was quantified with standard curves provided by the corresponding ELISA kit. The detection limit was 5 ng/mL.

DNA extraction, PCR amplification, and sequencing. We analyzed the microbiome of samples from healthy controls (stool, $n = 17$; serum, $n = 7$) and cirrhotic patients (stool, $n = 27$; serum, $n = 27$; ascitic fluid, $n = 11$). In order to identify possible contamination in low-biomass samples and subtract the sequences of the potentially contaminated DNA generated during the extraction and PCR amplification, we introduced negative controls (blanks) during these two technical steps.

A frozen aliquot of fecal sample (250 mg) from each individual ($n = 44$) was subjected to genomic DNA extraction using a previously described method, referred to here as the "Godon" method[15,18]. Each sample was suspended in 250 μl of guanidine thiocyanate, 0.1 M Tris (pH 7.5), 40 μl of 10% N-lauroyl sarcosine, and 500 μl 5% N-lauroyl sarcosine. DNA was extracted by mechanical disruption of the microbial cells with beads. RNA was removed by the addition of 2 μl of a 10-mg/ml solution of RNAase, and nucleic acids were recovered from clear lysates by alcohol precipitation. Twenty-seven and seven serum samples were collected from patients and healthy controls, respectively, and subjected to genomic DNA extraction using beads to disrupt the microbial cells followed by the QIAamp® DNA Blood Midi Kit (Qiagen, Madrid, Spain), following the manufacturer's protocol. We obtained 11 ascitic fluid samples (4 ml) from 13 patients. Microbial DNA was extracted using a modified "Godon" protocol. In this regard, after a 10-min centrifuge at 14000 rpm, the pellet was subjected to the same procedure as the fecal samples. However, the final resuspension of the nucleic acids was carried out with 30 μl of a Tris-EDTA buffer solution.

An equivalent of 1 mg of each sample was used for DNA quantification using a NanoDrop ND-1000 Spectrophotometer (Nucliber). DNA integrity was examined by micro-capillary electrophoresis using an Agilent 2100 Bioanalyzer with the DNA 12,000 kit, which resolves the distribution of double-stranded DNA fragments up to 17,000 bp in length.

For profiling microbiome composition, the hyper-variable region (V4) of the bacterial and archaeal 16 S rRNA gene was amplified by PCR. On the basis of our analysis done with Primer Prospector software, the V4 primer pairs used in this study were expected to amplify almost 100% of the bacterial and archaeal domains. The 5′ ends of the forward (V4F_515_19: 5′-GTGCCAGCMGCCGCGGTAA-3′) and reverse (V4R_806_20: 5′-GGACTACHVGGGTWTCTAAT-3′) primers targeting the 16S gene were tagged with specific sequences as follows: 5′-{AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGT}[3,15,18] {GTGCCAGCMGCCGCGGTAA}-3′ and 5′-{CAAGCAGAAGACGGCATACGAGAT} {Golay barcode} {AGTCAGTCAGCC} {GGACTACHVGGGTWTCTAAT}-3′. Multiplex identifiers, known as Golay codes, had 12 bases and were specified downstream of the reverse primer sequence (V4R_806_20)[19,20].

Standard PCR using 0.75 units of Taq polymerase (Roche) and 20 pmol/μL of the forward and reverse primers was run in a Mastercycler gradient (Eppendorf) at 94 °C for 3 min, followed by 35 cycles of 94 °C for 45 sec, 56 °C for 60 sec, 72 °C for 90 sec, and a final cycle of 72 °C for 10 min. Amplicons were first purified using the QIAquick PCR Purification Kit (Qiagen, Barcelona, Spain), quantified using a NanoDrop ND-1000 Spectrophotometer (Nucliber) and using an Agilent 2100 Bioanalyzer with the DNA 1000 kit, and then pooled in equal concentration. The pooled amplicons (2 nM) were then subjected to sequencing using Illumina MiSeq technology at the technical support unit of the Autonomous University of Barcelona (UAB, Spain), following standard Illumina platform protocols.

Sequence analysis. Sequences obtained from stool, ascitic fluid, and serum, together with negative controls from the extraction and PCR methods, were analyzed with QIIME 1.8.0[21] using an in-house script. Raw sequences of low quality were filtered out with a minimum acceptable Phred score of 20. A demultiplexing step was performed to assign back each read to its corresponding sample and to remove barcodes. A total of 3,393,253 high quality sequences were finally recovered (2,910,686 for feces and 482,567 for serum and asctic fluid samples). UCLUST algorithm based on 97% of similarity was used to cluster similar sequences into Operational Taxonomic Units (OTUs) or taxa. Representative sequences of each OTU were aligned using PyNAST against Greengenes

template alignment (gg_13_8). Chimeric sequences were then identified and removed with ChimeraSlayer. Finally, a taxonomical assignment for each OTU was performed with the basic local alignment search tool (BLAST) and the combination of two microbial databases (Greengenes and PATRIC). A phylogenetic tree was obtained with the FastTree program. The general OTU table was split into various tables in order to individually analyze feces, serum, and ascitic fluid samples.

In order to avoid false positive OTUs in stool samples, we eliminated those that did not represent at least 0.2% of the sequences. For samples with a low biomass, such as serum and ascitic fluid, we removed the OTUs that did not account for at least 0.2% of the sequences in at least 3 samples. Moreover, OTUs detected in negative controls were also removed for downstream analyses. Unknown bacteria assigned by BLAST against Greengenes and PATRIC databases were additionally checked against the NCBI database, and OTUs identified as from human origin were removed from the dataset. The final total, mean, minimum and maximum number of sequences per sample type were computed, and OTU tables were rarefied at several rarefaction depths (Supplementary Table 2).

Statistical analyses. The characteristics of patients with and without ascites were compared using Fisher's exact test for categorical variables and Mann-Whitney test for quantitative variables. For sequence analysis, pair-wise comparisons were performed using OTU tables generated from each sample type. Samples that contained fewer reads than the rarefaction depth were removed for the alpha and beta diversity analyses. The Shapiro-Wilk test was used to check normality of the data, and pairwise comparisons were made between the study groups with the non-parametric test Kruskal-Wallis one-way analysis of variance, which compares means between groups. False discovery rate (FDR) corrected p-values were taken into account to consider significant results. Richness provided by alpha diversity was computed with Chao1 index. Sample clustering was performed using UPGMA and PCoA methods based on UniFrac metrics.

References

1. Mokdad, A. A. *et al.* Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. *BMC Med* **12,** 145 (2014).
2. Macnaughtan, J. & Jalan, R. Clinical and Pathophysiological Consequences of Alterations in the Microbiome in Cirrhosis. *Am J Gastroenterol* **110,** 1399–1410 (2015).
3. Bajaj, J. S. *et al.* Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J Hepatol* **60,** 940–947 (2014).
4. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513,** 59–64 (2014).
5. Bajaj, J. S., Betrapally, N. S. & Gillevet, P. M. Decompensated cirrhosis and microbiome interpretation. *Nature* **525,** E1–2 (2015).
6. Teltschik, Z. *et al.* Intestinal bacterial translocation in rats with cirrhosis is related to compromised Paneth cell antimicrobial host defense. *Hepatology* **55,** 1154–1163 (2012).
7. Cuenca, S. *et al.* Microbiome composition by pyrosequencing in mesenteric lymph nodes of rats with CCl4-induced cirrhosis. *J Innate Immun* **6,** 263–271 (2014).
8. Bruns, T. *et al.* Identification of bacterial DNA in neutrocytic and non-neutrocytic cirrhotic ascites by means of a multiplex polymerase chain reaction. *Liver Int* **29,** 1206–1214 (2009).
9. Serste, T. *et al.* Detection of bacterial DNA in serum and ascitic fluid of asymptomatic outpatients with cirrhosis and non-neutrocytic ascites. *Liver Int* **31,** 494–498 (2011).
10. Soriano, G. *et al.* Bacterial DNA in the diagnosis of spontaneous bacterial peritonitis. *Aliment Pharmacol Ther* **33,** 275–284 (2011).
11. Such, J. *et al.* Detection and identification of bacterial DNA in patients with cirrhosis and culture-negative, nonneutrocytic ascites. *Hepatology* **36,** 135–141 (2002).
12. Feng, Y. *et al.* Application of next-generation sequencing to study ascitic microbiome in cirrhotic patients with or without spontaneous bacterial peritonitis. *J Microbiol Immunol Infect* **48,** 504–509 (2015).
13. Spadoni, I. *et al.* A gut-vascular barrier controls the systemic dissemination of bacteria. *Science* **350,** 830–834 (2015).
14. Frances, R. *et al.* A sequential study of serum bacterial DNA in patients with advanced cirrhosis and ascites. *Hepatology* **39,** 484–491 (2004).
15. Santiago, A. *et al.* Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol* **14,** 112 (2014).
16. Albillos, A. *et al.* Increased lipopolysaccharide binding protein in cirrhotic patients with marked immune and hemodynamic derangement. *Hepatology* **37,** 208–217 (2003).
17. Koutsounas, I., Kaltsa, G., Siakavellas, S. I. & Bamias, G. Markers of bacterial translocation in end-stage liver disease. *World J Hepatol* **7,** 2264–2273 (2015).
18. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol* **63,** 2802–2813 (1997).
19. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6,** 1621–1624 (2012).
20. Navas-Molina, J. A. *et al.* Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol* **531,** 371–444 (2013).
21. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7,** 335–336 (2010).

Acknowledgements

We thank the controls subjects and patients who provided samples for this study. This study was partially funded by grants from the Instituto de Salud Carlos III/FEDER (PI11/1026, CP13/00181, PI14/00764, PI14/00680). CIBERehd is funded by the Instituto de Salud Carlos III.

Author Contributions

G.S., C.G., F.G. and C.M. were involved in the study design; C.G., X.T., E.R., E.A.T. and G.S. were involved in the sample collection; A.S., D.C., J.C.N. and G.S. were involved in the experimental work; M.P. and C.M. were involved in the bioinformatics and statistical analysis, and interpretation of the data; C.M. and G.S. were involved in the drafting of the manuscript; M.P., A.S., S.V., X.T. and F.G. were involved in the critical revision of the manuscript.

Additional Information

Accession Code: The 16S rRNA gene sequences have been deposited in the NCBI-SRA database under the accession BioProject ID: SUB1268941 (Temporary Submission ID).

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Santiago, A. *et al.* Alteration of the serum microbiome composition in cirrhotic patients with ascites. *Sci. Rep.* **6**, 25001; doi: 10.1038/srep25001 (2016).