Unidad mixta de investigación en Genómica y Salud
Centro Superior de Investigación en Salud Pública -
Instituto Cavanilles de Biodiversidad y Biología Evolutiva

Facultat de Ciències Biològiques

VNIVERSITAT DE VALÈNCIA

Variability of the Human Gut Microbiota in Space, Time, and associated with the Irritable Bowel Syndrome

Memoria presentada por Ana Durbán Vicente
para optar al grado de Doctor en Biotecnología

Directores: Juan José Abellán Andrés,
Amparo Latorre Castillo, Andrés Moya Simarro

Valencia, 2012

Unidad mixta de investigación en Genómica y Salud
Centro Superior de Investigación en Salud Pública -
Instituto Cavanilles de Biodiversidad y Biología Evolutiva

Facultat de Ciències Biològiques

VNIVERSITAT DE VALÈNCIA

Variability of the human gut microbiota in space, time, and associated with the irritable bowel syndrome

Memoria presentada por Ana Durbán Vicente
para optar al grado de Doctor en Biotecnología

Directores: Juan José Abellán,
Amparo Latorre, Andrés Moya

Valencia, 2012

ANDRÉS MOYA y AMPARO LATORRE, Catedráticos del Departamento de Genética de la Universitat de València, y JUAN JOSÉ ABELLÁN, Investigador del Centro Superior de Investigación en Salud Pública (CSISP) de la Generalitat Valenciana,

CERTIFICAN que la memoria titulada "Variability of the human gut microbiota in space, time, and associated with the irritable bowel syndrome" ha sido realizada bajo su dirección en la unidad mixta de investigación en Genómica y Salud CSISP-UVEG/Instituto Cavanilles por ANA DURBÁN VICENTE.

Y para que así conste, firman el presente certificado.

Valencia, a de de 2012

Juan José Abellán Amparo Latorre Andrés Moya

Agradecimientos

Quiero agradecer a Andrés y Amparo la confianza que depositaron en mí y el afecto que siempre me han mostrado. En particular, quiero agradecer a Andrés el escogerme para este proyecto "arriesgado" y "dinámico", pero también enriquecedor, y a Amparo, sus esfuerzos para que nos entendamos. A Juanjo, su dedicación a este proyecto, el ofrecerse a ser uno de mis directores, sus cuidados, y las "patadas" que tan bien me han venido. A Nuria, su constante apoyo e implicación, su esfuerzo para poner a punto todo, y sus intentos de "cortarme las alas" cuando hacía falta.

También quiero dar las gracias a todos los compañeros de Genómica y Salud del CSISP y del Cavanilles, los que son y los que fueron, por aguantarme en el ámbito laboral y/o en los momentos de distensión, según se aplique, y por crear entre todos un ambiente tan agradable. En especial, a los de la sala de becarios I, fuente de compañerismo, mimos, y chistes malos: la "Granpositiva" Ana Elena, la dulce Leo, Pedro, que me lo soluciona todo, el casi siempre entrañable Raúl, Peris, Bea, y todos los demás. A Marc, miembro adoptivo del grupo, por su cariño y su pasión por "la ciencia". A Mariam y Llúcia, por los momentos de desconexión al humo del cigarrillo y la consecuente amistad. A la gente de Biología Computacional de Sistemas de la Universidad de Viena, por la atención que me brindaron durante mi estancia. A mi amiga Ana, por estar a mi lado desde hace tanto compartiendo penas y alegrías (como las que se dan en cualquier tesis que se precie). A mi madre y hermano, por su apoyo y por no poner en duda el valor de lo que hago.

Por último, quiero reiterar mi simpatía hacia todos los investigadores consolidados, en fase de consolidación, y, en general, cualquiera susceptible de llegar a ser algo en la vida, que nunca se sabe a quién se acaba pidiendo trabajo.

The picture of the back cover was adapted from Phillips (2009).

Caption: The gut contains thousands of microbes, including: 1) *Ruminococcus*, 2) *Bifidobacterium*, 3) *Peptostreptococcus*, 4) *Staphylococcus*, 5) *Lactobacillus*, 6) *Acidaminococcus*, 7) *Fusobacterium*, 8) *Eubacterium*, 9) *Clostridium*, 10) *Coprococcus*, 11) *Escherichia*, 12) *Butyrivibrio*, 13) *Bacteroides*, 14) *Brachyspira*.

Contents

1 Introduction

1.1 First insights into the human microbiota

Microorganisms came to light in the mid-1600s when the Dutch merchant Anton van Leeuwenhoek directly observed tiny "animalcules", later known as bacteria, using a single-lens microscope of his own design. It was also the first time that human-associated microbes were seen in tooth plaque. From there, some physicians and microbiologists were defenders of the "germ theory" of disease, a theory that proposes that microorganisms are the cause of many diseases. In the late 19th century, the German physician Robert Koch, working on cholera, anthrax, and tuberculosis, was the first scientist who devised a series of tests to assess the "germ theory" and finally proved it.

After microorganisms were shown to cause disease in humans, the popular and scientific views of the microbial world became dominated by its role in microbial disease, which led to significant progress in medicine through improved hygiene, vaccinations, and antibiotics. Nowadays, the search of isolated disease agents has been complemented with the recognition that also imbalances in the complex microbial communities naturally inhabiting the human body can be related to some diseases.

However, most interactions between humans and microorganisms are not pathogenic for the host but beneficial to both. This new view of the human-associated microbiota began at the end of the 19th century, when the German-Austrian pediatrician Theodore Escherich realised that the intestinal microorganisms should interact with each other and influence properties of their host. He carried out an extensive work on the relationship of the intestinal bacteria with the physiology of digestion in infants (Escherich, 1886). The following statement of T. Escherich would probably be subscribed by many researchers working on the gut microbiota nowadays:

"At a time when microbiologic research has gained us so many laurels by

following the research methods of Koch into the regions of the aetiology and pathology of infectious diseases, it would appear to be a pointless and doubtful exercise to examine and disentangle the apparently randomly appearing bacteria in normal faeces and the intestinal tract, a situation that seems controlled by a thousand coincidences. If I have nevertheless devoted myself ... to this special study, it was with the conviction that the accurate knowledge of these conditions is essential for the understanding of not only the physiology of digestion, ... but also the pathology and therapy of microbial intestinal diseases" (Escherich, 1885).

There was a decreasing interest in the study of the "normal" microbiota of humans in the early decades of the 20th century. Some reasons were the great effort devoted to search for aetiological agents of disease and the lack of effective methods for culturing bacteria that grow only in oxygen-free atmospheres (anaerobes). As it is now well-known, most microorganisms found in the large bowel and in faeces are strict anaerobes. The development of these methods was key to the explosion of research in the field during the 1960s, when many microbiologists all over the world made important contributions to the current knowledge of the intestinal microbiota. See Savage (2001) for an excellent review. Some key concepts regarding the intestinal microbiota derived from these studies are:

a) The human body surfaces are colonised by microbial cells that outnumber the host cells by at least one order of magnitude. The vast majority of these microbes lies in the gastrointestinal (GI) tract (more than one hundred trillion cells).

b) The communities are established in newborns in an ecological succession.

c) The microbiota of adult humans is composed of anaerobic bacteria.

d) Some species reside in the mucosa layer covering the epithelium and some in the lumen.

e) The microbiota is regulated by many factors, some exerted by the host and the intestinal environment and some exerted by the microbes themselves.

These concepts will be further explained and complemented with new ones in Section 1.2. The work of these pioneer researchers provided a new perspective of humans as composites of animal and microbial cells, with the microbiota assuming biological functions that are essential to the host. Under these assumptions, the research in the area has continued until today.

1.2 Current knowledge of the human microbiota

1.2.1 Microbial habitats in the human body

Like other mammals, humans have co-evolved with complex microbial communities that live in their bodies. Microbial symbionts of humans were once considered only commensals or opportunistic pathogens. Actually, the interaction can be recognised as mutualistic in many cases. Microorganisms benefit from protected environments and nutrients from which they extract the energy that require to survive, while make vital contributions to human development, immunity, nutrition, and physiology, and offer protection against potentially dangerous invaders (Guarner and Malagelada, 2003; Dethlefsen et al., 2007).

Microbes constitute a major component of the human body. For example, adult humans have 10^{12} microbes associated with their epidermis, and 10^{14} microbes in their GI tract. It is estimated that the total number of microorganisms in the human body is about 10 times greater than the total number of host cells (Luckey, 1972).

Microbial habitats in the human body include skin surfaces (Grice et al., 2009), penis (Price et al., 2010), vagina (Fredricks, 2011), upper respiratory tract (Charlson et al., 2010), oral cavity (Aas et al., 2005), and gut (Zoetendal et al., 2006). These habitats can in turn be subdivided into microhabitats for microbial colonisation. The most striking cases of habitat diversification are probably found within the skin and the GI tract. Human microbial communities are overwhelmingly dominated by bacterial species. Other members are several archaeal species, yeast, and other microscopic eukaryotes. More than 50 bacterial phyla have been detected, but Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria contain most of the diversity in all habitats (**Figure 1.1**) (Dethlefsen et al., 2007; Spor et al., 2011).

The composition of the human-associated microbial communities is determined primarily by body habitat. Each habitat harbours a characteristic microbiota and a relatively stable set of prevalent phyla across people and over time, although the interpersonal variability within the same body habitat can be high, especially at the species/strain level (Costello et al., 2009). For example, Bacteroidetes and Firmicutes predominate in the colon and Actinobacteria in the skin in almost all humans. By contrast, Firmicutes predominate in the vagina of most women, but Actinobacteria predominate in a small fraction of

women (Dethlefsen et al., 2007). Finally, the many local microbial communities harboured by humans can interact through direct dispersal or indirect interactions modulated by the host immune system (Gonzalez et al., 2011).

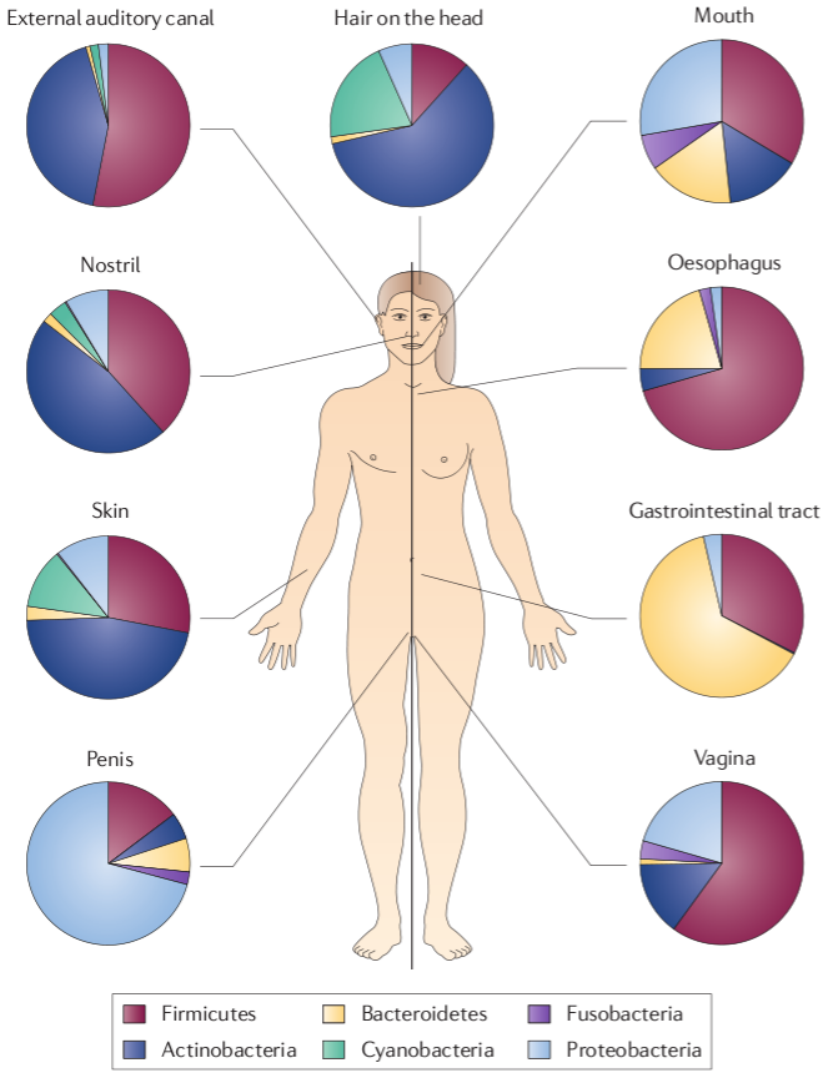


Figure 1.1: Composition of the bacterial community (relative abundances of the dominant phyla) at different body locations in a healthy human. Adapted from Spor et al. (2011).

1.2.2 The microbial communities inhabiting the GI tract

1.2.2.1 Microbial habitats in the GI tract

The human GI tract refers to all the organs from the mouth to the anus, that is, the oesophagus, the stomach, the small intestine, and the large intestine (**Figure 1.2**). The composition of the GI microbiota is not uniform. It varies longitudinally along the GI tract and transversally between the intestinal lumen and the mucosa lining the epithelium. Some of the microbial habitats within the GI tract are described bellow.

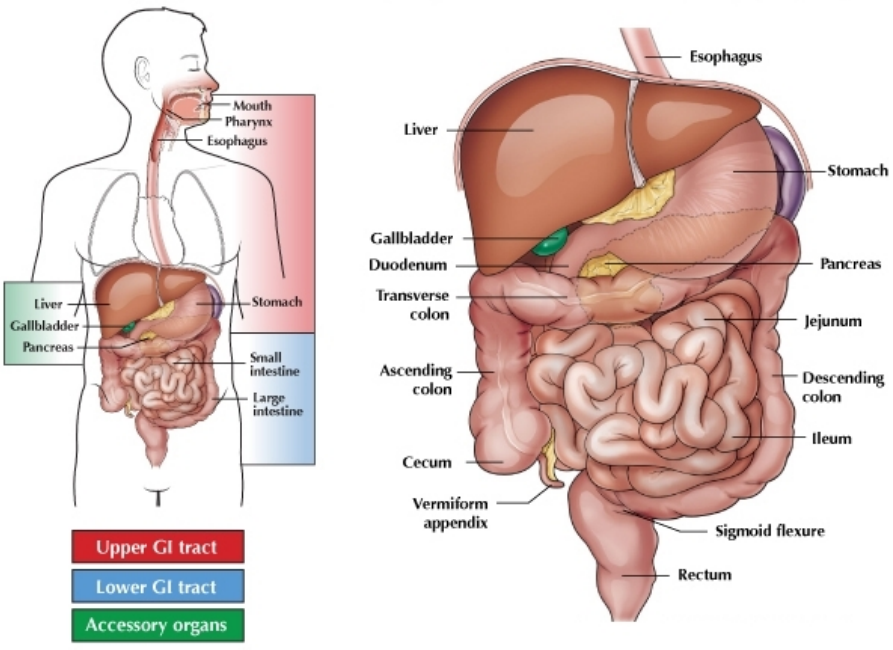


Figure 1.2: Regions of the human GI tract and accessory organs of digestion. Adapted from http://mybioversa.com/qleanza_by_bioversa.

The stomach. The stomach is a dilated part of the GI tract located between the oesophagus and the small intestine. Masticated food enters the stomach and is mixed with hydrochloric acid, proteases, and fluid through muscular contractions of the stomach wall. The acidic pH (1-2) allows the activity of the secreted proteases and kills or inhibits many ingested microbes. Some microorganisms can survive these conditions and reach low

population levels, such as the opportunistic pathogen *Helicobacter pylori*, which can be found normally in association with the gastric mucosa (Lee, 1999).

The small intestine. The small intestine is subdivided into the duodenum, the jejunum, and the ileum. It is about three meters long, but its absorptive surface is increased to about six meters by the special structure of the epithelium, with evaginations of the mucosa (villi) and the brush border of the epithelial cells (microvilli). Most of the digestion by host pancreatic enzymes and the absorption of food takes place in the small intestine. Proteins and peptides are degraded into amino acids, lipids into fatty acids and glycerol, and some carbohydrates (CHOs) are degraded into oligosaccharides and monosaccharides. Many complex polysaccharides can not be digested by host enzymes and pass into the large intestine where intestinal bacteria digest them (this is one of the main contributions of the microbiota). Food is also blended with bile salts and bicarbonate in the duodenum. Bile salts emulsify fats and help in their digestion. Bicarbonate neutralises the potentially harmful acid coming from the stomach and provides the pH needed for the digestive enzymes. The short transit time (1-4h) limits the microbial growth in the lumen of the small intestine. There is a gradient of bacterial colonisation from 10^3-10^4 cells per gram of luminal content near the stomach to 10^7-10^8 near the colon, where the transit slows down (Finegold et al., 1983).

The large intestine. The large intestine comprises the caecum, the colon (which is in turn subdivided into the ascending, transverse, descending, and sigmoid colon), and the rectum. The large intestine is about 1.5 meters long. Excess water, salts, and products of the microbial activity on nutrients that have escaped digestion in the small intestine are absorbed through the colonic mucosa. The colon is densely populated by microorganisms, with concentrations of 10^{11}-10^{12} cells per gram of luminal content, one of the highest recorded for any microbial habitat (Whitman et al., 1998). Regarding nutrition, they ferment dietary compounds that are not degraded in the upper GI tract and endogenous mucus. This and other benefits to the host are discussed in Section 1.2.2.4.

Faeces. Undigested food and waste material from the body are eliminated with faeces. Faeces are compacted in the large intestine through the absorption of excess water, and are stored into the rectum until defaecation. Bacteria are a

major component of faeces representing about 55% of their dry weight (Stephen and Cummings, 1980).

Faeces could be considered another microbial habitat. Faecal bacteria could be a mixture of luminal and shed or poorly adhered mucosal bacteria. Besides, growing conditions such as temperature, oxygen concentration, and nutrient availability change dramatically after evacuation, potentially leading to alterations in the community composition due to differential bacterial death or growth. Faeces have been extensively employed for the study of the GI microbiota. However, faecal microbial communities may not entirely represent those found in other parts of the GI tract (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005).

Radial differentiation of intestinal habitats. The intestinal lumen is the cavity where food passes through, digestion takes place, and nutrients are absorbed from. Microorganisms in the lumen can be dispersed in the liquid or bound to the solid particles found in it (Walker et al., 2008).

The intestinal epithelium is responsible for the absorption of nutrients. The epithelium also creates a barrier between external and internal environments and is the first line of pathogen recognition by the immune system. The epithelium of the GI tract is a single layer of columnar cells, with the exception of the multilayered squamous epithelium of the esophagus and anus. Mucins are high molecular weight glycoproteins produced by globet cells within the intestinal epithelium (and foveolar cells in the stomach). They are the major component of the mucus layer that overlies and protects the epithelium. Mucus forms a double protective layer: a very dense, firmly attached, and quite sterile inner mucus layer, and a less dense, loose, and more strongly colonised outer mucus layer (Johansson et al., 2011). The microbial composition of the inner mucus layer, the outer mucus layer, and the lumen differ from each other (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005; Johansson et al., 2011).

The epithelium rests on the *lamina propria*, a layer of connective tissue populated by cells of the immune system and supported by a thin layer of smooth muscle called *muscularis mucosa*. The epithelium, *lamina propria* and *muscularis mucosa* form the *mucosa*, which lays on the *submucosa*, a dense layer of connective tissue with nerves, blood vessels, and lymphatic vessels. The *submucosa* is supported by the *muscularis externa*, which usually consists of an inner circular and an outer longitudinal layers of smooth muscle. The

coordinated contractions of these layers is called peristalsis and push the food through the GI tract. In most areas, a thin layer of connective tissue and the surface mesothelial cells form the *serosa* that encases the GI tract.

1.2.2.2 Composition and location of GI microbial communities

The human GI microbiota refers to the communities of microorganisms, including bacteria, archaea, and microbial eukaryotes, that populate the human GI tract. The current knowledge of the composition of the human microbiota comes from culture-based studies, and more recently, from culture-independent molecular approaches that identify microorganisms from the small subunit ribosomal RNA (16S rRNA) genes.

Microbial eukaryotes constitute a minor fraction of the GI microbiota. The eukaryotic diversity within the GI tract appears to be low, rather stable over time, and dominated by different subtypes of *Blastocystis*. *Blastocystis* spp. are unicellular protozoa, at least some with pathogenic potential. The cultivable fungal fraction is dominated by yeast of the genus *Candida* (Cohen et al., 1969; Scanlan and Marchesi, 2008). However, culture-independent analysis has shown that other species are more frequently found in faeces than *Candida* spp., for example *Gloeotinia/Paecilomyces* and *Galactomyces* (Scanlan and Marchesi, 2008).

Archaea in the human GI tract comprise several methanogenic species. Methanogens utilise several substrates for methane generation, including hydrogen (H_2). This is, in fact, a major pathway for the removal of the H_2 produced by bacterial fermentation in the distal gut. The predominant species in the human colon is *Methanobrevibacter smithii*. Few other methanogens have been detected, such as *Methanosphaera stadtmanae* and *Methanobrevibacter oralis* (Scanlan et al., 2008).

Bacteria are by far the most abundant and widely studied group of microbes in the GI microbiota (Zoetendal et al., 2008). The diversity of major lineages (phyla) of bacteria is low. Most GI microbes belong to the Firmicutes and the Bacteroidetes phyla, which together comprise more than 90% of the bacteria in the faeces and colonic mucosa of adults. Proteobacteria and Actinobacteria are common but ususally not dominant. Some other phyla can be found at low abundance in some GI sites and individuals, for example Cyanobacteria, Deferribacteres, Deinococcus–Thermus, Fusobacteria, Lentisphearae, Spirochaetes, Verrucomicrobia, and the candidate phyla TM7 and SR1. Within these few deep lineages, the diversity of bacterial species and strains is extremely high

(**Figure 1.3**).

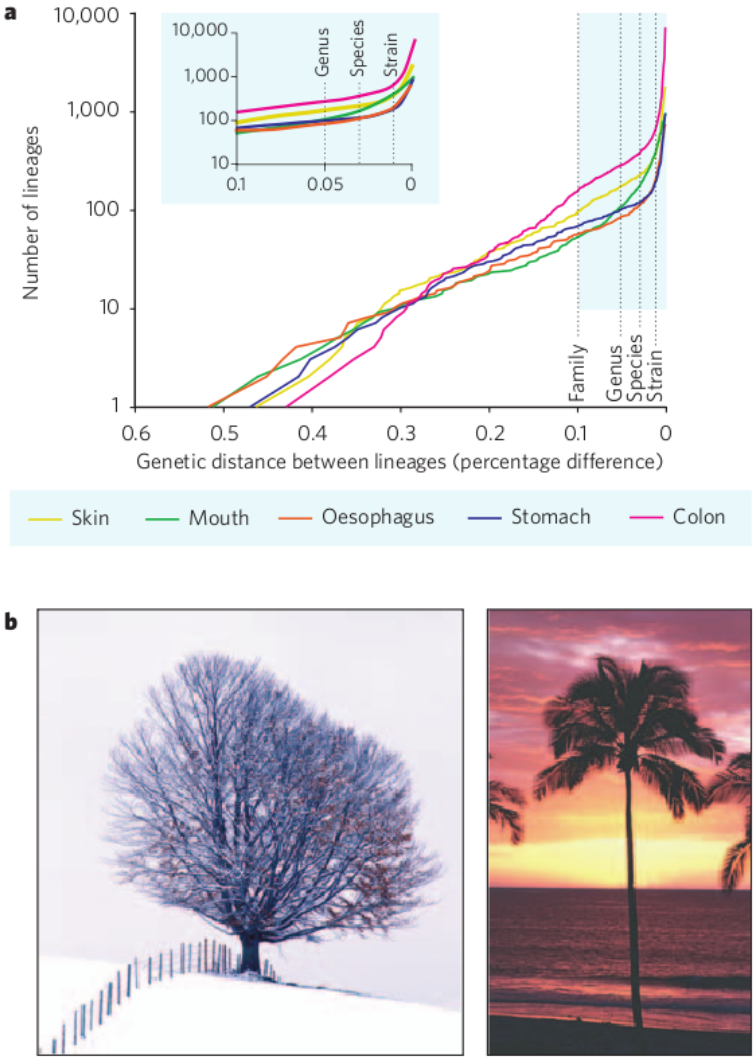


Figure 1.3: Patterns of human-associated microbial diversity. a) Richness estimates in specific human habitats. The x axis shows the percentage difference threshold for delineating separate lineages. The y axis shows the number of distinct lineages at each distance threshold. b) Patterns of microbial diversity in soil and aquatic environments generally resemble the tree shape on the left, with new branches arising at all distances from the root. In contrast, patterns of diversity in human-associated communities resemble the tree shape on the right, with few branches arising close to the root and many branches arising close to the branch tips. Adapted from Dethlefsen et al. (2007).

A variety of habitats and niches is found along the proximal-distal axis of the GI tract that result in compositionally distinct microbial communities. Some features of these GI habitats are:

a) The oral cavity contains 600-700 bacterial species, with the predominance of species of the genus *Streptococcus*, within the Firmicutes phylum. Multispecies biofilms are typical in the mouth, where surface attachment is key in order to persist. Different surfaces, such as the supragingival and subgingival dental plaque, the cheek, and the tongue, contain distinct bacterial communities (Aas et al., 2005).

b) Few species have been detected in the oesophagic mucosa, most of which are similar or identical to residents of the oral cavity (Pei et al., 2004).

c) The acidic condition in the stomach kills many ingested microbes. Contrary to what occurs in other GI sites, Proteobacteria and Actinobacteria constitute a major fraction of the microbiota in the stomach. More than one hundred species have been detected, but *Helicobacter pylori*, associated with the gastric mucosa, is the only undisputed resident (Bik et al., 2006).

d) Microbial growth is limited in the small intestine by the short retention time and the host secretions. The abundance and diversity of bacteria increase from the proximal to the distal small intestine. Facultative anaerobes predominate in the jejunum and the ileum, but the proportion of strict anaerobes increases towards the distal gut (Hayashi et al., 2005; Wang et al., 2005).

e) The largest bacterial populations of the GI tract are found in the large intestine because of the long retention time and the high amount of substrates available for bacterial fermentation. Relatively small differences in the microbial community composition have been found between mucosal sites along the large intestine (Zoetendal et al., 2002; Eckburg et al., 2005; Lepage et al., 2005). Richness estimates suggest around 1000 bacterial species, with a dominance of obligate anaerobes of the Bacteroidales order and the *Clostridium* clusters IV and XIVa (Zoetendal et al., 2008). At the species level, the composition of the colonic and faecal microbiota in healthy adults is highly host-specific (Eckburg et al., 2005; Tap et al., 2009).

Different habitats are found across the radial axis of the GI tract: the epithelium, the mucus layer covering the epithelium, and the intestinal lumen. Distinct selective pressures acting in these habitats presumably influence the composition and activity of the local microbial communities. It has been hypothesised that microorganisms targeted by host defence molecules in

the mucus layer are restricted to the lumen, while those non-targeted form the mucosa-associated microbial community. Mucosal colonisers could also be those able to adhere to mucus, digest it, and/or deal with the oxygen gradient along the mucus layer (Roos and Jonsson, 2002; Derrien et al., 2004; Ruas-Madiedo et al., 2008; Kankainen et al., 2009). Additionally, some microorganisms might be able to interact directly with the epithelial cells. With their closer interaction with the intestinal epithelium, mucosal microorganisms would be more relevant for the shaping of the immune system and the stimulation of the epithelial cell proliferation and differentiation (Van den Abbeele et al., 2011). In contrast, luminal microorganisms would be important for nutrient digestion. The microbial communities associated with the intestinal mucosa and those of faecal samples from the same individual differ from each other (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005). Also, there is stratification within the mucus layer, the inner mucus layer usually being devoid of bacteria (Johansson et al., 2011). Moreover, liquid and particulate faecal fractions have distinct bacterial composition (Walker et al., 2008).

The microorganisms found in any GI habitat may not be characteristic of it and be just passing through. Such microbes may come from ingested food or water, or from other habitats in the human body where they could be normal colonisers, such as skin, upper respiratory tract, or GI habitats above the one in which they are found. In practice, however, it is challenging to distinguish between transient and resident microbes. Resident microbiota fills the available niches in its habitat and avoids the colonisation by non-indigenous microbes unless the habitat is perturbed. Thus, some criteria for recognising indigenous microbes in a GI habitat include stable population levels in normal adults, intimate association with the mucosal epithelium (and thus not being eliminated by the flowing stream), and/or metabolic activity in that habitat. Recent longitudinal studies have shown that only a small subset of the bacterial species in faecal communities is permanently detected and maintains a relatively uniform abundance, while the stability of more inclusive taxa is high (Caporaso et al., 2011; Dethlefsen and Relman, 2011). Molecular approaches targeting RNA help deciphering which faecal microorganisms are metabolically active and reveal that the activity level of some microbial taxa can not be predicted from their relative abundance (Turnbaugh et al., 2010; Peris-Bondia et al., 2011).

Finally, it should be noted that, contrary to the high inter-subject variability in the species composition, the gene content of the faecal microbiota seems to be largely shared among individuals, defining a common core at the functional level (Turnbaugh et al., 2009a; Qin et al., 2010). The fact that different species assemblages converge towards conserved functions implies a high degree of functional redundancy between microbes and microbial consortia within the gut microbiota. It also suggests that the intestinal environment selects for microbial traits rather than for specific microbes.

1.2.2.3 Factors influencing the GI microbiota

Host genotype

Evidence of the effect of host genotype on the GI microbiota comes from the correlation between genetic relatedness and community composition similarity, and from the influence of gene polymorphisms on the configuration of the microbiota, both in humans and in murine models of disease. A higher similarity with higher degree of relatedness has been observed in a study of the dominant bacterial populations of faecal samples of monozygotic twins, dizygotic twins, and unrelated children (Stewart et al., 2005). Most of the genes shown to have an impact on the composition of the gut microbiota are components of the immune system, and a few others have roles in metabolism (see Spor et al. (2011) for a review of host genes with known effects on the microbiota). For example, mutations of the gene MEFV that lead to an autoinflammatory disorder in humans, the familial Mediterranean fever (FMF), have been associated to a depletion in the total number of bacteria, loss of diversity, and major shifts in several bacterial populations (Khachatryan et al., 2008). Mice deficient in Toll-like receptor 5 develop insulin resistance and greater adiposity, and their microbiota is somehow altered because, when it is transferred to wild-type mice, many features of the donors' phenotype are reproduced (Vijay-Kumar et al., 2010). Similarly, obese mice due to a mutation in the gene encoding leptin, a proteic hormone with a central role in energy intake and control of appetite, have lower Bacteroidetes/Firmicutes ratio than lean mice (Ley et al., 2005). Complex host genetic control on the composition of the microbiota has been demonstrated in mice, as several quantitative trait loci control the relative abundance of specific microbes (Benson et al., 2010).

Host age

Humans undergo certain life stages, during which the structure and stability of the gut microbiota change. Various host and environmental factors are thought to influence the microbiota throughout the lifetime (**Figure 1.4**).

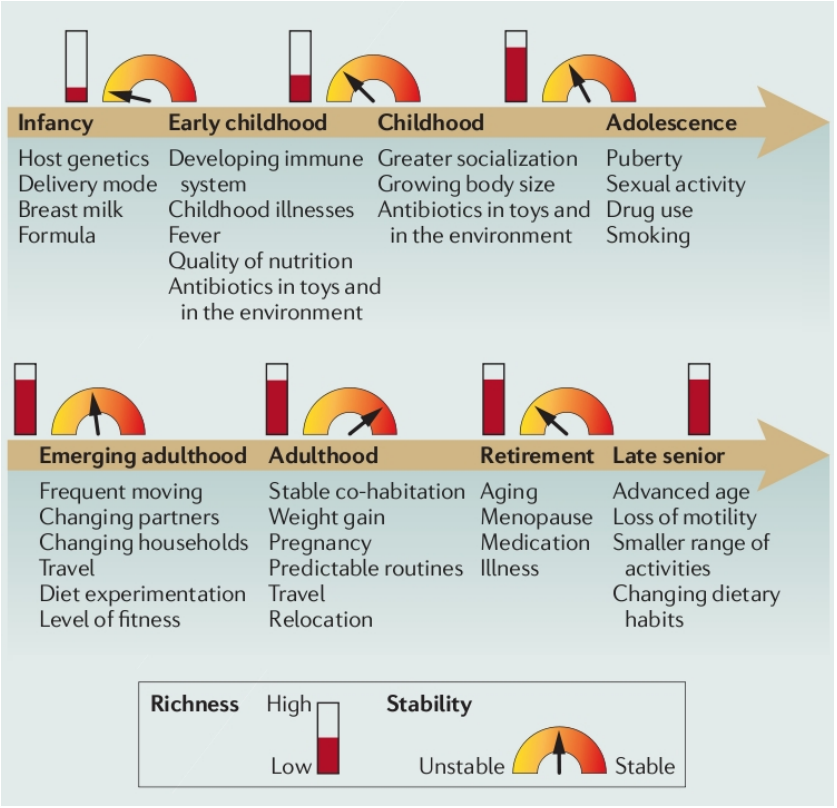


Figure 1.4: Overview of host and environmental factors potentially affecting the gut microbiota, and evolution of microbial diversity and stability throughout life. Adapted from Spor et al. (2011).

The GI tract has been traditionally considered sterile at birth. However, recent studies have shown that the meconium, the earliest stools of a newborn, composed of materials ingested during the stay in the uterus, is non-sterile (Jiménez et al., 2008; Gosalbes et al., 2012). Anyway, the GI tract of newborns is rapidly colonised by bacteria of maternal origin from the vagina, skin, milk, and any other environment to which they are exposed. Initially, the type of delivery (natural birth versus caesarean) and the type of diet (breast milk

versus formula milk) affect the colonisation pattern (Yoshioka et al., 1983; Grönlund et al., 1999; Harmsen et al., 2000; Penders et al., 2006; Dominguez-Bello et al., 2010). In the first years of life, the GI tract progresses from near sterility to dense colonisation. The postnatal development of the gut microbial communities is remarkably variable between individuals. To a large extent, it is determined by the specific bacteria to which the baby is exposed. However, the gut microbiota of infants becomes progressively similar to one another, converging into a complex assembly similar in composition and function to that found in adults (Palmer et al., 2007; Vaishampayan et al., 2010; Koenig et al., 2011). The predictable end-point of the succession could respond to pressures selecting for bacteria that typically dominate the adult colon and that overcome the initial advantage of early colonisers less adapted to the adult gut environment. Examples of these pressures could be the introduction of solid food and changes in the growing conditions due to the development of the intestinal mucosa and the action of the microbiota.

It is claimed that the gut microbiota is stable throughout adult life in the absence of perturbations. The temporal stability of the faecal microbiota has been evaluated by repeated sampling of microbial communities over time, which allows the identification of persistent and transient components of the communities (Franks et al., 1998; Zoetendal et al., 1998; Vanhoutte et al., 2004; Caporaso et al., 2011), and the measurement of the robustness and resilience to perturbations (De La Cochetière et al., 2005; Manichanh et al., 2010; Dethlefsen and Relman, 2011). Gut communities have been considered stable because the temporal variation within individuals is smaller than the variation between individuals. Stability has typically been examined with samples collected at intervals of weeks or months, finding that the microbiota of healthy individuals remains fairly constant over these long-term intervals at family or genus level. However, variation is found in the presence and abundance of some taxa, while others remain more constant. Recent studies examining the daily variation in the faecal microbiota confirm these results, but also reveal pronounced fluctuations in the abundance of bacterial taxa and that few species are permanent members of the communities (Caporaso et al., 2011; Dethlefsen and Relman, 2011). Moreover, the gut microbiota of adults shows considerable resilience, that is, ability to return to its original state after being perturbed (De La Cochetière et al., 2005; Dethlefsen and Relman, 2011).

Modifications of the gut microbiota have been reported in elderly people, including reduction in bacterial diversity, increased numbers of facultative

anaerobes, and decrease in bifidobacteria. These shifts are linked to changes in host physiology associated with ageing, such as loss of appetite, decreased intestinal motility, and persistent activation of the innate immunity (Woodmansey, 2007; Claesson et al., 2011). Together, these alterations may result in increased putrefaction in the colon and a greater susceptibility to GI infections in the elderly.

Forces exerted by the host

Humans select GI microorganisms based on multiple physical and defence mechanisms. Some forces exerted by the host influence the microbial composition equally in all regions of the GI tract. For example, microorganisms with an optimum growth temperature of about 37 °C are favoured. However, most factors vary along the tract, thus setting up particular microbial communities in different parts of it.

Oxygen concentration decreases progressively along the GI tract and it is very low, if any, in the large intestine (Levitt, 1971). The majority of GI microbes in an adult human are therefore anaerobes, reflecting the anoxic lumen, and the proportion of facultative anaerobes decreases while the proportion of strict anaerobes increases as we advance towards the end of the intestine. Also, mucosal microbes are exposed to an oxygen gradient because oxygen is released from the blood vessels that irrigate the epithelium towards the mucus layer, so aerotolerant microbes have a niche there.

Luminal pH strongly affects the microbiota (Fallingborg, 1999; Duncan et al., 2009). The very acidic condition of the stomach (pH 1-2) kills or inhibits the growth of many microbes. Bicarbonate released in the small intestine neutralises pH. Then, microorganisms themselves are the ones that modify the luminal pH. In the proximal colon, fermentation processes produce high levels of short-chain fatty acids (SCFAs) that low pH to 5-6, limiting the growth of acid-sensitive microbes. In contrast, the substrate availability is reduced in the distal colon and pH is close to neutral.

The short retention time in the stomach (1-4 h) and the small intestine (2-6 h) prevents microbial colonisation in the lumen. The ability to attach to the epithelium is key in the upper GI tract to remain long enough to grow, while most luminal microbes are probably ingested and transient ones. In contrast, the colon is characterised by a long retention time (10 h to several days) that, along with the more favourable growing conditions, provides an ideal environment for a heavy microbial colonisation.

Some host-derived substances have an impact on the gut microbiota. In the small bowel, microbes are exposed to the action of pancreatic juices and bile salts. It has been suggested that bile salts repress bacterial growth through direct antimicrobial effects and/or regulation of the host mucosal defences, and bile acids have been shown to regulate the composition of the gut microbiota in rats (Inagaki et al., 2006; Islam et al., 2011). The constant shedding of epithelial cells and the attached mucus requires the replacement of mucosa-associated microbes, but, on the other hand, it provides nutrients for luminal microbes. The mucus layer favours the colonisation by microorganisms that have developed specific mechanisms to adhere to it or to use mucins as energy and carbon source (Roos and Jonsson, 2002; Derrien et al., 2004; Ruas-Madiedo et al., 2008; Kankainen et al., 2009).

The intestinal epithelium forms a physical barrier against the invasion by pathogens, and, importantly, by the vast amount of indigenous non-pathogenic microbiota. Tight junctions hold the intestinal epithelial cells (IECs) together, so they form a virtually impermeable barrier. Specific nutrients or microbial antigens are absorbed through controlled mechanisms. The physical barrier is strengthened by the mucus layer.

The barrier function of the intestinal mucosa also includes the innate and the adaptive immunity (**Figure 1.5**). The innate mucosal immunity involves the recognition of conserved microorganism-associated molecular patterns (MAMPs) through the pathogen recognition receptors (PRRs) of IECs and dendritic cells (DCs), which can migrate between the IECs and through the *lamina propria*. This leads to the secretion of innate immune effectors with broad specificity, such as antimicrobial peptides (AMPs) (Müller et al., 2005; Meyer-Hoffert et al., 2008). The immune system distinguishes pathogens thanks to the strategic location of PRRs: those basolateral in the cellular membrane or intracytosolic can sense pathogens invading the epithelium (Cario and Podolsky, 2000; Gewirtz et al., 2001; Girardin et al., 2001). Besides, microorganisms and microbial antigens from the gut lumen are continually sampled by IECs, DCs, and M cells of the epithelium overlying the gut-associated lymphoid tissue (GALT) of the Peyer's patches. Antigens are engulfed by antigen presenting cells (APCs), that is, DCs, B lymphocytes, and macrophages. DCs activate cells of the adaptive immune system, leading to the differentiation of the T lymphocyte cell subsets Th1 (resulting in cell-mediated immunity and inflammation), Th2 (humoral immunity), Th17 (cell-mediated immunity and inflammation), and Treg (immunosuppression).

Tolerance to the indigenous microbiota results from the induction of Treg cells and the secretion of anti-inflammatory cytokines such as interleukin-10 (IL-10) and transforming growth factor beta (TGF-β) (Neutra et al., 2001; Mazmanian et al., 2005). Th2 cells stimulate the development of B lymphocytes into plasma cells that produce specific immunoglobulin (Ig) A antibodies (Macpherson and Uhr, 2004). IgA-coated bacteria can not adhere to epithelial cells. The mucus layer traps AMPs and IgA, thus preventing microbes from breaching the epithelium. These mechanisms restrict the inflammatory response to the indigenous microbiota to the level seen under normal conditions (physiologic inflammation). It has been hypothesised that microorganisms targeted by host defence molecules in the mucus layer are restricted to the lumen, while microorganisms non-targeted by or resistant to defence molecules can colonise the mucosa (Van den Abbeele et al., 2011).

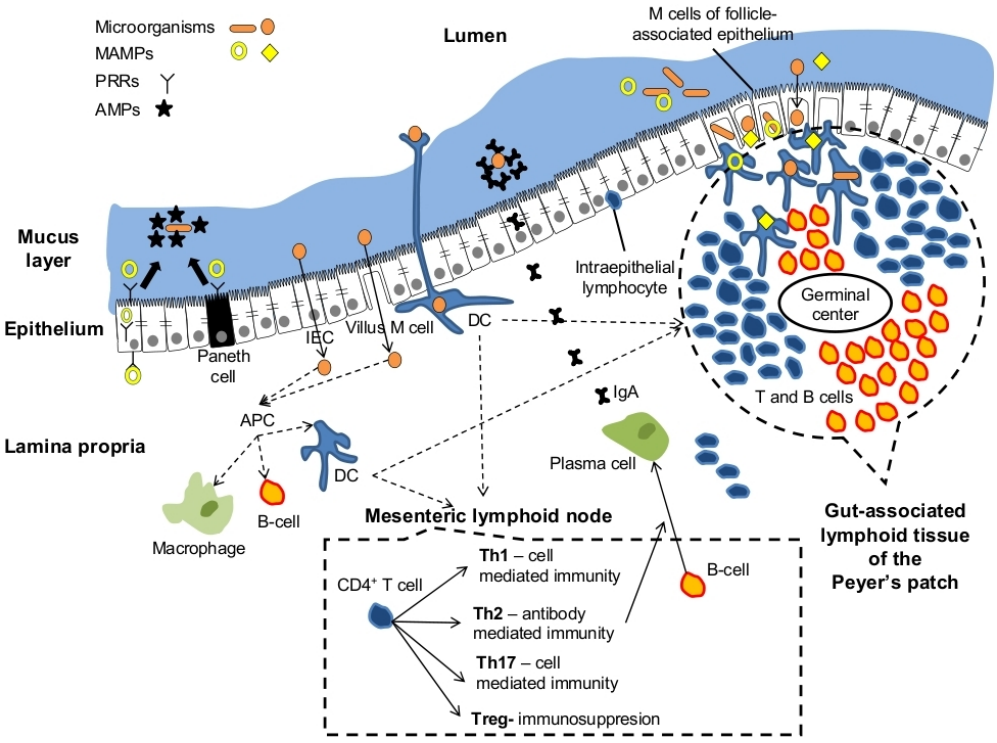


Figure 1.5: The mammalian immune system and the intestinal microbiota. Adapted from Van den Abbeele et al. (2011).

Diet

Diet is an obvious factor that potentially affects the GI microbiota. There are some reports in support of this idea, although the exact mechanisms whereby diet influences gut microbes remain to be characterised.

The faecal microbiota of human infants differs between breast-fed and formula-fed newborns, and it is strongly affected by the introduction of solid food (Yoshioka et al., 1983; Harmsen et al., 2000; Penders et al., 2006; Koenig et al., 2011). Populations of *Bifidobacterium* spp. dominate in breast-fed babies, whereas babies drinking formulas develop a more diverse microbiota that contains, in addition to bifidobacteria, also Bacteroides, enterobacteria, enterococci, streptococci, and Clostridia. Once solid food is introduced, the level of strict anaerobes increase progressively with the concomitant decrease in the level of facultatives.

It also appears that changes in diet affect the faecal microbial composition in adults. A research group on obesity found that the relative proportion of Bacteroidetes increased in humans following two types of low-calorie diet (fat-restricted and CHO-restricted) (Ley et al., 2006), and that the change in "humanised" gnotobiotic mice (raised in germ-free conditions and then colonised with human faecal microbial communities) from a diet low in fat and rich in fibre to a diet high in fat and low in fibre altered the composition and gene expression of the microbiota (Turnbaugh et al., 2009b). Wu et al. (2011) linked the composition of the faecal microbiota with long-term dietary patterns: communities with relatively high levels of *Bacteroides* were associated with high intake of animal protein, amino acids, and saturated fats, whereas communities with relatively high levels of *Prevotella* were associated with low values for these nutrients, but high values for CHOs and simple sugars. Other group found striking differences between the faecal microbial composition of European children, eating a modern Western diet, and that of children living in a rural African village, where the diet is very low in fat and proteins and rich in fibre (De Filippo et al., 2010). Even though the two populations differed in many environmental factors, the authors hypothesised that diet had a major impact on microbial composition. They correlated the presence of some bacterial genera in the African cohort with a higher capacity of energy extraction from ingested fibre, which was observed in turn by determining the concentration of SCFAs produced by bacterial fermentation in the faecal samples. In addition, studies of specific dietary compounds have shown their effects on specific taxa. For example, dietary sulfate favours sulfate-reducing bacteria over

methanogenic archaea (Gibson et al., 1993), and inulin increases the abundance of *Bifidobacterium* (Kolida et al., 2007).

The acquisition of a new diet is a fundamental driver for the evolution of new species. The comparison of the faecal microbiota of extant mammals revealed that both host diet and host phylogeny influence gut bacterial diversity, which increases from carnivory to omnivory to herbivory, and that the gut microbiota of humans living a modern lifestyle is typical of omnivorous primates (Ley et al., 2008). Ancestral mammals were carnivores. The adaptation to a plant-based diet was an evolutionary breakthrough that resulted in massive radiations. Gut microbes were pivotal for the emergence of herbivory as they provide most of the enzymatic repertoire needed for extracting energy from complex plant CHOs such as celluloses and resistant starches. Herbivores enlarged the GI tract in several ways to prolong the gut retention time and thus expose fibre to the bacterial activity (Stevens and Hume, 1998, 2004). Therefore, the nutrient sources in which different animals were specialised, together with the subsequent structural adaptation of the GI tract, influenced the composition of their gut microbiota and the animal nutritional needs that are covered through the activity of the microbiota.

Clinical interventions

Antibiotic use is often associated with adverse GI symptoms and can induce disturbances in the intestinal microbiota. However, modulation of the gut microbiota by antibiotics, probiotics, prebiotics, and synbiotics are attractive approaches to improve host health. In addition, transplantation of whole faecal microbial communities from healthy donors is sometimes used as a therapy for some intestinal disorders.

Antibiotics. Antibiotics are widely used in clinics as antimicrobial agents to prevent and treat infections caused by pathogenic microorganisms. However, antibiotics not only affect the target pathogens, but also other members of the indigenous microbiota (Sjölund et al., 2003; De La Cochetière et al., 2005). They inhibit the growth of susceptible organisms and select for resistant ones. This can result in dysbiosis of the intestinal microbiota, subsequently leading to intestinal problems such as antibiotic-associated diarrhoea (Högenauer et al., 1998; Young and Schmidt, 2004). The antibiotic-induced alterations in the composition of the microbiota can be temporary, but medium- and long-term disturbances have been also reported (Jernberg et al., 2007; Dethlefsen and

Relman, 2011).

Besides, several studies have demonstrated that broad-spectrum antibiotics eradicate small intestinal bacterial overgrowth and improve GI symptoms, including studies in a subset of IBS patients (Baidoo et al., 2006; Basseri et al., 2011).

Probiotics and prebiotics. Minimising the side-effects of antibiotic therapy has been one reason for the growing interest in probiotic research in the last decade (Katz, 2006; Engelbrektson et al., 2009). According to the definition of the World Health Organisation, probiotics are "live microorganisms which, when administered in adequate amounts, confer a health benefit to the host". Prebiotics are "non-digestible food ingredients that, when consumed in sufficient amounts, selectively stimulate the growth and/or activity of a limited number of microorganisms in the gut microbiota that confer health benefits to the host" (Meyer and Stasse-Wolthuis, 2009). Mixtures of probiotics and prebiotics are referred to as synbiotics. Probiotic microorganisms are mainly lactic acid bacteria and bifidobacteria. They are often taken in food such as yogurts and cheese, food supplements, or as drugs. Prebiotics include inulin, fructo-oligosaccharides, and galacto-oligosaccharides.

Probiotics have been shown to modulate systemic and mucosal immune function, improve intestinal barrier function, protect against physiological stress, inhibit the colonisation by pathogens, and promote the stability of the microbiota (Fooks and Gibson, 2002; Ng et al., 2009; Wallace et al., 2011).

The therapeutic or preventive effects of certain probiotics have been documented in chronic intestinal inflammation, irritable bowel syndrome, constipation, diarrhoea (including antibiotic-associated diarrhoea), rotavirus enteritis, obesity, and metabolic disorders. The effectiveness of probiotics has also been studied in a number of autoimmune and neoplastic diseases (Gareau et al., 2010; Iannitti and Palmieri, 2010; Weichselbaum, 2010).

Faecal microbiota transplantation. Faecal microbiota transplantation (FMT), also known as faecal bacteriotherapy, involves the repeated infusion of faecal suspensions (or cellular suspensions obtained from faeces) from a healthy individual into the intestine of an ill individual with a disease believed to result, at least partially, from microbiota-related dysfunctions. FMT has been used sporadically for over 50 years and has been proved as a safe and highly efficient treatment for recurrent *Clostridium difficile* infection (CDI)

refractory to standard antibiotic therapies (Gough et al., 2011). FMT may be working through the restoration of a "healthy" microbiota. Indeed, a recent study on FMT for the treatment of CDI demonstrated the long-term persistence of the donor's microbiota in the recipient patients (Grehan et al., 2010). The increasing evidence that many diseases can be related to perturbations in the intestinal microbiota points to FMT as a promising alternative treatment, but there is a need for controlled trials (Borody and Khoruts, 2011).

Stochastic events

Unpredictable events also affect the microbiota of every individual. These include the colonisation history, bacterial infections that produce a response in the host that subsequently disturbs the already established community, and bacterial predation by phages.

Early colonisation. In the first days to months of life, the composition and temporal progression of the faecal microbiota vary widely from baby to baby, although eventually the microbiota acquires an adult profile. The earliest colonisation events are determined to a large extent by accidental exposures to bacteria from the infant environment, such as maternal vaginal, faecal, or skin microbiota, and breast milk (Palmer et al., 2007; Dominguez-Bello et al., 2010). The early life environment could affect the composition of the adult intestinal microbiota, as it has been shown in animal models (Deloris Alexander et al., 2006; Mulder et al., 2009; Ubeda et al., 2012). Due to functional redundancy, several microorganisms can potentially fill the same niche within a GI habitat. The first ones to arrive can settle and then could select for microbes with whom to establish cooperative interactions, or for microbes with non-overlapping niches, as well as promote the exclusion of competitors. For example, indigenous bacteria can regulate the availability of host-derived nutrients for their own benefit, strengthen the mucosal barrier, and overall prevent the growth of other bacteria introduced later in the ecosystem (Bernet et al., 1994; Hooper et al., 1999, 2001).

Enteric infections. The invasion of enteric pathogens impacts the already established microbiota. For example, host-mediated inflammation in response to infection by the mouse pathogen *Citrobacter rodentium* produces a reduction in the total numbers of colonic bacteria, preferentially of the Cytophaga-Flavobacter-Bacteroides class, and the overgrowth of either resident or invader

aerotolerant bacteria, particularly of the Enterobacteriaceae family (Lupp et al., 2007). Similar events occur after infection by *Salmonella* (Barman et al., 2008). In these cases, the original microbial community is largely restored after clearance of the pathogen. Long-term effects of bacterial gastroenteritis over the GI microbiota are suspected in humans with post-infectious irritable bowel syndrome, in which persistent symptoms begin after an enteric infection (Spiller and Garsed, 2009).

Phage attacks. The human GI tract harbours abundant and diverse viral communities (Breitbart et al., 2003; Reyes et al., 2010). Phages affect the abundance and diversity of bacteria in a community, as it has been shown in aquatic ecosystems (Suttle, 1994; Sime-Ngando and Colombet, 2009). However, Reyes et al. (2010) did not detect such viral-microbial dynamics in human faeces. Anyway, it is known that phage attacks can cause rapid community changes by decimating established bacterial species or strains. Phages can also serve as a reservoir of genes that can be horizontally transferred to their microbial hosts, thus expanding their functional diversity (Dinsdale et al., 2008).

Microbial interactions

Interactions established within the microbial community contribute to the diversity and individuality of the gut microbiota. Gut microorganisms can influence each other through direct interactions or indirectly by modifying intestinal habitats. Examples of microbial interactions are explained below.

Cooperative interactions. Microbial transformations often involve several microorganisms organised in syntrophic associations in which the metabolic end-products of one species are used by others. The microbial partners may become very complementary and dependent, resulting in resilient associations within the microbial community. An example of such consortia is that involved in the degradation of polysaccharides in the colon (**Figure 1.6**) (Flint, 2004; Derrien et al., 2010; Koropatkin et al., 2012). Numerous species are needed for the initial stages of food digestion. Some microorganisms have a versatile repertoire of enzymes for the binding and hydrolysis of the complex polymers, mostly belonging to the Bacteroidetes phylum, such as *Bacteroides thetaiotaomicron* (Salyers et al., 1977; Martens et al., 2009), but also to Actinobacteria, Proteobacteria, and Firmicutes. At the same time, the degradation of specific

polymers is often species- and even strain-dependent. The fermentation of the resultant sugars involves multiple species that use different pathways and depend on each other to produce certain SCFAs (Walker et al., 2005; Belenguer et al., 2006; Falony et al., 2006). Bifidobacteria and lactobacilli are dominant acetate and/or lactate producers. Other microorganisms convert acetate, lactate, and/or partially degraded CHOs to butyrate (such as members of the *Clostridium* clusters IV and XIVa) and propionate (such as members of the *Clostridium* cluster IX and *Bacteroides* spp.). The H_2 generated during fermentation is removed by methanogenic archaea, sulphate-reducing bacteria, and acetogenic bacteria (Nakamura et al., 2010). Otherwise, the accumulation of H_2 would inhibit the fermentation processes (Macfarlane and Macfarlane, 2003).

Moreover, cooperative interactions between microorganisms are not restricted to sequential metabolic reactions. For example, some microorganisms consuming products of the degradation of complex CHOs in the colon do not contribute with digestive enzymes, but stimulate strains that do so through the release of growth factors (Flint, 2004).

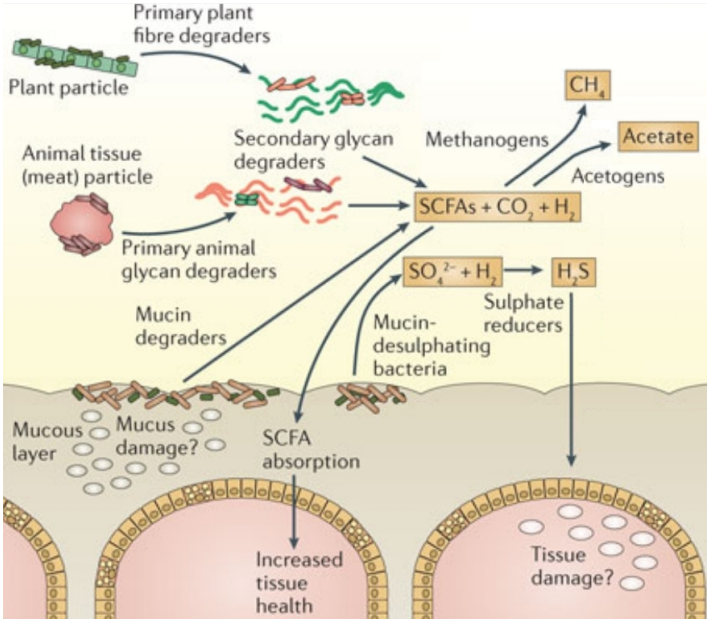


Figure 1.6: Schematic illustration of the ways in which different gut microorganisms are thought to interact in the processing of dietary and endogenous glycans. Adapted from Koropatkin et al. (2012).

Competitive interactions. The members of the microbiota compete for the nutrients available in the GI tract. Besides, symbionts can indirectly prevent the growth of competitors by regulating the production of host-derived resources (Hooper et al., 1999). Microbes also compete for the binding sites on the surface of the epithelial cells and the mucus layer (Bernet et al., 1994). Resource competition provides resistance to pathogen colonisation, one of the main roles of the gut microbiota.

In addition, one strain can be directly inhibited by another without competition for resources, which is called interference. Some compounds are specifically produced to inhibit competitors, such as bacteriocins. In other cases, metabolic by-products of one species inhibit the growth of another (Brook, 1999). Bacteriocins are peptides released by a microbe that kill other microbes. Bacteriocins are active against specific strains, usually closely related to the producer one, but can also be effective against more phylogenetically distant microorganisms (Servin, 2004; Hütt et al., 2006). Symbionts can indirectly inhibit the growth of pathogens by inducing the intestinal epithelium to overproduce antimicrobial peptides (Vaishnava et al., 2008). Lactate and SCFAs produced by the bacterial fermentation in the colon are bactericidal and bacteriostatic for some species through direct mechanisms or by lowering the luminal pH, which affects acid-sensitive microbes (Alakomi et al., 2000; Shin et al., 2002). Sulfate-reducing bacteria produce hydrogen sulphide (H_2S), which can be toxic for the host and for other microbes (Nakamura et al., 2010). Some lactobacilli produce hydrogen peroxide (H_2O_2), which may inhibit or kill organisms that produce little or no H_2O_2-scavenging enzymes (Wheater et al., 1952).

1.2.2.4 Benefits of the GI microbiota to the host

In the symbiotic relationship between humans and the GI microbiota, the microbial partners make specific contributions of great impact on the host health. Microbial functions include salvage of energy and nutrients, trophic effects on the intestinal epithelium and on the immune system, and protection against invasion by exogenous microbes (**Figure 1.7**).

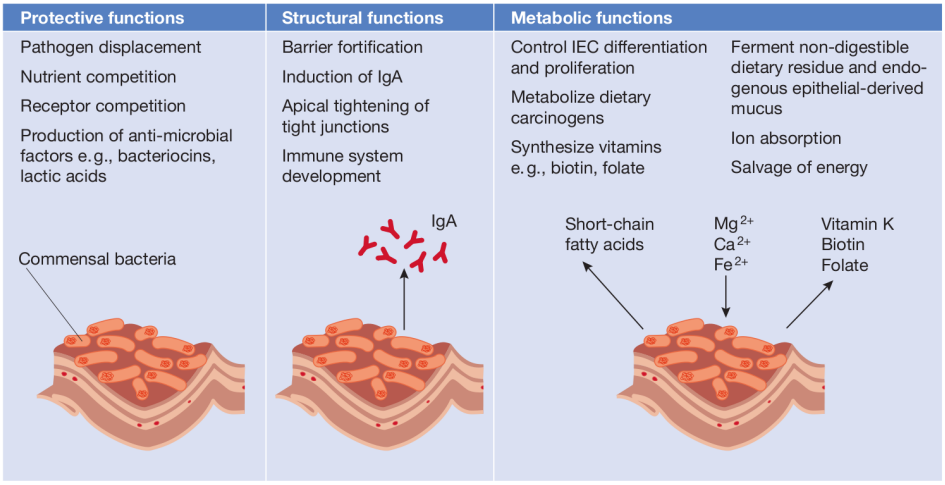


Figure 1.7: Beneficial functions of the intestinal microbiota. Adapted from O'Hara and Shanahan (2006).

Fermentation of non-digestible dietary substrates and endogenous mucus

The colonic microbiota degrades CHOs that can not be metabolised by humans and endogenous mucin-linked glycans, which are a major source of energy for the host and for the microbiota itself. Non-digestible CHOs include large polysaccharides (resistant starches, fructans, cellulose, hemicellulose, pectins, inulins, and gums), some oligosaccharides that escape digestion, and unabsorbed sugars and alcohols. Specific members of the gut microbiota have the enzymatic machinery needed to break and digest these compounds (Flint, 2004; Derrien et al., 2010; Koropatkin et al., 2012). The major end-products of fermentation are SCFAs, primarily acetate, propionate, and butyrate (Cummings et al., 1987). These SCFAs have important and specific functions in host physiology. Most of the SCFAs are absorbed by the host and provide 6-9% of his total energy requirement (McNeil, 1984). Butyrate is the preferred source of energy for colonic epithelial cells and is almost completely consumed by them (Roediger, 1982; Cummings et al., 1987). Acetate and propionate pass into the bloodstream and are metabolised by the liver (mainly propionate) and peripheral tissues, particularly muscle (acetate) (Cummings et al., 1987). Also, SCFAs lower the luminal pH, potentially inhibiting the growth of acid-sensitive

pathogens (Alakomi et al., 2000; Shin et al., 2002), and improve the absorption of ions (Roediger and Moore, 1981).

The colonic microbiota anaerobically metabolise proteic dietary residues and proteins of endogenous origin (pancreatic enzymes, mucus, and shed epithelial cells) (Macfarlane et al., 1986). These processes also produce SCFAs. Once CHO sources are exhausted in the proximal colon, protein fermentation becomes more prevalent (Macfarlane et al., 1992). Catabolism of proteins by the colonic microbiota provides less energy than that of CHOs, but its importance lies mainly in the effects that its end-products have on host intermediary metabolism and as potential toxins.

Other metabolic functions

Other important metabolic functions of the gut microbiota are the synthesis of essential amino acids and vitamins (such as vitamin K, folate, and biotin) (Conly et al., 1994; Hill, 1997; Torrallardona et al., 2003), the improvement of ion absorption by the gut epithelium (including magnesium, calcium, sodium, and iron) (Roediger and Moore, 1981; Younes et al., 2001), the deconjugation of bile salts and the subsequent bile acid modification, which influence lipid metabolism and bile-controlled endocrine functions (Houten et al., 2006; Ridlon et al., 2006), and the detoxification of xenobiotics (Gill et al., 2006). At the organism level, the gut microbiota impacts global metabolism by modulating energy harvest from the diet and energy storage in the host (Bäckhed et al., 2004; Martin et al., 2007).

Control of epithelial cell growth, differentiation, and function

Studies in germ-free (GF) animals have proved that the gut microbiota stimulates the intestinal angiogenesis (Stappenbeck et al., 2002), the renewal of epithelial cells (Savage et al., 1981), the differentiation of secretory globet and enteroendocrine cells (Bates et al., 2006), and the secretion of mucus (Szentkuti et al., 1990), while also influences the composition of mucins (Meslin et al., 1999).

The three major SCFAs stimulate the proliferation and differentiation of gut epithelial cells (Frankel et al., 1994). Furthermore, butyrate may have a role in the prevention of colon carcinogenesis, as it inhibits the proliferation and promotes the differentiation of neoplastic epithelial cells, and promotes the reversion of neoplastic phenotypes (Gibson et al., 1992; Siavoshian et al., 2000).

Regulation of the development and homeostasis of the immune system

The interaction between the host and the gut microbiota is crucial for the development of both the intestinal and the systemic immune systems. GF animals have low densities of lymphoid cells in the gut mucosa, inappropriate balance of Th cell subsets, and low blood levels of immunoglobulins (Ig) (Butler et al., 2000; Mazmanian et al., 2005). The microbial colonisation of the GI tract contributes to the development of the gut-associated lymphoid tissue (GALT) (Cebra et al., 1998). Gut symbionts stimulate the maturation of the adaptive immune system. It has been shown that a single normal symbiont, *Bacteroides fragilis*, induces the maturation of T cells to Th1 in GF animals (while the default pathway is skewed towards Th2), and stimulates the development of immunosuppressive T cells (Tregs) in an animal model of colitis, which results in reduced inflammation (Mazmanian et al., 2005, 2008).

The constant monitoring of the gut microbiota induces the local production of specific antibodies (IgA) that limit the colonisation of the mucosa and select the mucosa-associated microbiota (Macpherson and Uhr, 2004). The microbiota also induces the secretion of a variety of broad-spectrum antimicrobial molecules that strengthen the mucosal barrier (Müller et al., 2005; Meyer-Hoffert et al., 2008).

A role for SCFAs in the tolerance of mutualistic microorganisms is suggested by their immunomodulatory effects: SCFAs repress the secretion of pro-inflammatory cytokines in cultured epithelial cells and leukocytes, and attenuate inflammation in animal models (Vinolo et al., 2011).

Protection against pathogens

The observation that GF animals had increased susceptibility to infection by enteric pathogens led to the concept of colonisation resistance, that is, the normal resident gut microbiota limits the establishment and/or growth of potential pathogens. The barrier effect implies several mechanisms. First, the resident microbiota competes for the attachment sites on the epithelial surface, thus preventing the colonisation and invasion by pathogens (Bernet et al., 1994). Second, the normal microbiota competes for available nutrients. Moreover, some symbionts may regulate the amount of resources provided by the host and thus prevent overproduction that would favour the growth of exogenous microbes (Hooper et al., 1999). Finally, some symbionts secrete antimicrobial compounds called bacteriocins that kill or inhibit the growth of their com-

petitors, including host pathogens (Servin, 2004; Hütt et al., 2006). SCFAs produced during bacterial fermentation can also be bacteriostatic for some bacterial species, through a direct effect or by lowering pH (Alakomi et al., 2000; Shin et al., 2002). In addition, symbionts can regulate host immune response against pathogens. They can alert the host epithelium that pathogens are present in the mucosa, resulting in enhanced secretion of antimicrobial peptides to avoid the infection (Vaishnava et al., 2008).

1.2.2.5 The GI microbiota and disease

The GI microbiota has been implicated in the aetiology of a number of complex multifactorial pathologies as diverse as obesity, cardiovascular diseases, chronic inflammatory and autoimmune diseases, neurological and psychiatric diseases, and colorectal cancer (Gerritsen et al., 2011; Tlaskalová-Hogenová et al., 2011). An improved understanding of the interactions between the indigenous microbiota and the host organism might bring new insights into the mechanisms of such diseases. As a consequence, the gut microbiota could be the target of new tools for disease prevention, diagnosis, and treatment. The following sections summarise the current understanding of several disorders in which the influence of the gut microbiota has been widely explored.

Obesity

Obesity is a complex disease characterised by excess fat accumulation in the body leading to adverse effects on health. The modulation of the host metabolism by the gut microbiota suggests a role for the microbiota in promoting obesity and related diseases, such as insulin resistance, type 2 diabetes, and atherosclerosis.

A number of studies in animal models have provided insights into potential mechanisms underlying this relationship. Gordon and co-workers pioneered the investigation of the gut microbiota as a factor that regulates fat storage. They found that colonisation of young GF mice with the gut microbiota of conventionally reared animals increases the energy harvest from diet, induces hepatic lipogenesis, and promotes triglyceride accumulation in adypocites, thereby contributing to the increase in body weight (Bäckhed et al., 2004). Furthermore, GF mice were found to be resistant to diet-induced obesity (Bäckhed et al., 2007). Key modulators are the fasting-induced adipose factor (FIAF) and the AMP-activated protein kinase (AMPK). The former is a secreted lipoprotein lipase (LPL) inhibitor whose expression is normally supressed by the micro-

biota in the gut epithelium, thus enhancing LPL activity in adipocytes. The activity of the latter and its downstream targets involved in fatty acid oxidation is increased in the liver and skeletal muscle of GF animals.

Obesity has been directly associated with phylum-level changes in the composition and function of the gut microbiota. Gordon and co-workers found that genetically obese mice had a reduction in the relative abundance of Bacteroidetes and a proportional increase in Firmicutes when compared with lean mice (Ley et al., 2005; Turnbaugh et al., 2006). It was corroborated in humans, together with the increase in the Bacteroidetes/Firmicutes ratio accompanying weight loss in obese subjects following low-calorie diets (Ley et al., 2006). Obese mice also harboured more methanogenic Archaea, which may increase the efficiency of bacterial fermentation. These same researchers demonstrated that microbiota transplantation from the caecum of obese mice results in a greater increase in body fat in recipient mice than colonisation with microbiota from lean donors (Turnbaugh et al., 2006). However, these specific changes remain controversial. Several studies have found no evidence that the relative proportions of Bacteroidetes and Firmicutes differ between obese and non-obese subjects (Duncan et al., 2008; Zhang et al., 2009), while another study reported a Bacteroidetes/Firmicutes ratio in favour of Bacteroidetes in the obese group, totally contradicting the findings of Gordon *et al.* (Schwiertz et al., 2010). Altogether, these data suggest that instead of phylum-level changes, more subtle changes in the composition of the gut microbiota could be associated with the development of obesity.

Irrespective of the correlation between obesity and the relative abundance of the various bacterial groups, the amount of SCFAs produced by bacterial fermentation in the colon is higher in obese mice and humans (Turnbaugh et al., 2006; Schwiertz et al., 2010). SCFAs can contribute directly to weight gain by acting as energy source for the host *de novo* lipogenesis, and indirectly through the modulation of intestinal motility and nutrient absorption (Halldén and Aponte, 1997; Xiong et al., 2004; Samuel et al., 2008).

Inflammatory bowel disease

The major forms of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC), are chronic inflammatory disorders of the GI tract. Substantial advances in the understanding of their pathogenesis have revealed an inadequate mucosal immune response to the GI microbiota in genetically predisposed hosts.

Genome-wide searches have been successful in identifying susceptibility loci for IBD, especially in CD. Many of the mutations are found in genes involved in the recognition, processing, and killing of microorganisms, and the regulation of immune processes (Gaya et al., 2006; Xavier and Podolsky, 2007).

Mouse models of colitis, which exhibit many of the features of UC, provide the best evidence that the presence of the gut microbiota is necessary for triggering colitis since GF animals do not develop the disease (Nell et al., 2010). Moreover, the microbiota of the TRUC immunodeficient mouse model of colitis is able to induce colitis after being transmitted to wild-type mice (Garrett et al., 2007).

Recently, the "hygiene hypothesis" has been implicated in the aetiology of IBD. This hypothesis, first proposed by D. P. Strachan in 1989, suggests that the increasing incidence of inflammatory, autoimmune and allergic diseases in developed countries is related to the reduction in the exposure to microbes in early life as a result of the improvement in hygienic measures, which results in an impaired development of the immune system (Strachan, 1989; Guarner et al., 2006).

Several studies have demonstrated a deficit of mucosal defensin in CD. Defensins are antibacterial peptides produced by the Paneth cells of the small intestine and play a major role in the innate immunity (Wehkamp et al., 2003, 2004). In addition, several studies have reported high concentrations of antibodies targeting the antigens of non-pathogenic indigenous bacteria in the GI tract of IBD patients (Macpherson et al., 1996).

The above evidence supports a key role of the gut microbiota in IBD, as it carries the antigens that fuel the chronic inflammation. Dysbiosis in the gut microbiota is highly suspected in the onset, perpetuation, and/or aggravation of IBD. Its understanding has increased in recent years thanks to the application of molecular methods for assessing microbial community composition. See Seksik (2010) for a review of the alterations detected in the microbiota of IBD patients by separating the diseases (CD and UC) and the explored compartments (faeces and gut mucosa). Among the structural alterations confirmed by several studies, there are an abnormally elevated bacterial load in the intestinal mucosa, which could be related to the deficiency of the mucosal barrier, an overall reduction in microbial diversity, a decrease in bacteria within the Firmicutes phylum, particularly in the *Clostridium leptum* group in CD patients, which includes the butyrate-producing bacteria *Faecalibacterium prausnitzii*, and in the *Clostridium coccoides* group in UC

patients, an increase in facultative anaerobes such as Enterobacteriaceae, and the presence of unusual bacteria. Several studies searching for a dysbiosis restricted to mucosal lesions did not demonstrate a significant difference between the microbiota of normal and injured areas. In addition, IBD is amongst the first diseases that have been the subject of a metagenomic study. Qin et al. (2010) found that IBD patients can be clearly separated from healthy individuals, and CD from UC patients, based on the global structure of the faecal microbiota (in that case, the relative abundance of 155 species). Besides, the microbiota of IBD patients had a lower gene count than that of healthy individuals, which suggests that the functional diversity is also reduced in IBD patients.

Irritable bowel syndrome

The role of the gut microbiota in the irritable bowel syndrome (IBS) is a central topic of this thesis and is introduced in detail in the next section.

1.3 The irritable bowel syndrome

1.3.1 Definition, clinical presentation, and diagnosis

IBS is a functional bowel disorder that produces chronic symptoms (Longstreth et al., 2006). The main ones are abdominal pain, bloating, discomfort, and alteration of bowel habits. No organic or structural cause explains these symptoms. Patients may experience diarrhoea, constipation, or an alternation of both, and the severity of symptoms can be mild, moderate, or severe. Symptoms are typically worse at times of increased stress.

The many presentations of IBS and its functional nature make the diagnosis challenging. The Rome criteria are a system developed to classify the functional GI disorders and to improve their diagnosis (Thompson, 2006). They were created in 1989 and have become since then the most widely accepted diagnostic criteria for IBS. According to the Rome criteria, IBS is characterised by continuous or recurrent abdominal pain or discomfort that is:

a) relieved by defaecation and/or

b) associated with change in the frequency of stool and/or

c) associated with change in the form of stool.

The versions of the Rome criteria produced over time differ in how strict they are about the period of time since the symptoms started, the frequency

of presentation of symptoms, the number of main criteria required for being diagnosed, and whether additional supportive criteria are required for the diagnosis. For example, the Rome II criteria, established in 1999, base the diagnosis on the presence of abdominal pain or discomfort in at least twelve weeks in the preceding twelve months plus at least two of the three main criteria (Drossman, 1999b).

The diagnosis of IBS has been called one of exclusion, but actually, there is no need of a differential diagnosis to discard other conditions. Some conditions share symptoms with IBS, such as several IBDs, lactose intolerance, coeliac disease, colon cancer, functional chronic constipation, and functional chronic diarrhoea. Misdiagnosis of other disease could occur, but only the presence of some "alarm" symptoms should alert the physician to discard another diagnosis. They include anaemia, GI bleeding, weight loss, fever, palpable abdominal or rectal mass, familial history of colon cancer or IBD, and new or recent onset in patients older than 50 years (Longstreth et al., 2006).

Some subtyping systems for IBS have been suggested according to the predominant stool pattern. For example, the criteria proposed by Longstreth et al. (2006) are:

a) IBS with constipation (IBS-C): hard or lumpy stools $\geq$25% and mushy or watery stools <25% of bowel movements.

b) IBS with diarrhoea (IBS-D): mushy or watery stools $\geq$25% and hard or lumpy stools <25% of bowel movements.

c) Alternating type IBS (IBS-A): hard or lumpy stools $\geq$25% and mushy or watery stools $\geq$25% of bowel movements.

d) Unsubtyped IBS: insufficient abnormality of stool consistency to meet criteria for IBS-C, IBS-D, or IBS-A.

The Bristol stool scale for the form of faeces can be used to identify constipation as types 1 and 2, and diarrhoea as types 6 and 7 (Lewis and Heaton, 1997).

1.3.2 Epidemiology

IBS is the most common functional disorder of the GI tract, affecting approximately 10–15% of the Western population (Quigley et al., 2006). It has been reported a wide variation in the prevalence between countries. One should be cautious when comparing prevalence rates between studies because these estimates depend on the employed diagnostic criteria and are affected by the

tendency of sufferers to seek for health care and that of physicians to diagnose IBS, which may differ between countries. Two population surveys in which the same diagnostic criteria were applied revealed a prevalence of 14.1% in the United States (Hungin et al., 2005), and a lower prevalence in Europe (an overall prevalence of 9.6%, ranging from 6.2 to 12% across countries) (Hungin et al., 2003). Although IBS is so common in the general population, it is estimated that up to 70% of sufferers may not have been formally diagnosed (Hungin et al., 2003, 2005).

Women are affected more often than men, with female:male ratios ranging from 3:2 to 3:1 (Chang and Heitkemper, 2002; Quigley et al., 2006). IBS affects people of all ages, even children, but most patients are aged 20-40. Symptoms begin before 35 years old in 50% of patients, and almost all patients report onset before 50 years old (Hungin et al., 2003). Onset in older ages is rare, and it may indicate organic pathology. The first onset of symptoms can follow an episode of infectious gastroenteritis. This is called post-infectious IBS, and affects 6-17% of the IBS patients (Spiller and Garsed, 2009). Also, the prevalence of IBS is higher among people who experienced some physical or psychological abuse in the past, and among people with psychiatric problems (Drossman, 1999a; Lea and Whorwell, 2003; Hood et al., 2008).

In spite of being a benign condition, IBS can have a profound negative impact on patients' quality of life, comparable to or greater than that of other chronic conditions such as IBDs (Pace et al., 2003; Tang et al., 2008), rheumatoid arthritis, asthma, and migraine (Frank et al., 2002). The impact of IBS ranges from mild inconvenience to severe debilitation. It is associated with psychological disturbances (anxiety, depression), sleep problems, physical difficulties in daily life, and sexual dysfunction (Drossman, 1999a; Hungin et al., 2003; Lea and Whorwell, 2003; Hood et al., 2008).

IBS has a high economic impact (Inadomi et al., 2003; Quigley et al., 2006). Direct costs include physician visits, diagnostic tests, medical treatments, and hospitalisations. Approximately 3% of primary care consultations are due to IBS; about 30% are derived to gastroenterologists and other specialists. Indirect costs include work absenteeism and reduced productivity due to illness. The absolute economic cost of IBS is unknown, partly due to the difficulty of estimating indirect costs and to the high number of non-diagnosed IBS sufferers.

1.3.3 Aetiology

The aetiology of IBS remains unclear. IBS has been attributed to psychologic factors or post-infectious alterations in the GI tract neuromuscular function, but actually it should be viewed as a multifactorial disorder. Today, the main mechanisms proposed are visceral hypersensitivity, altered gut motility and secretion, autonomic and/or central nervous system dysfunction, along with psychosocial stress, environmental factors, and gut flora alterations (Talley and Spiller, 2002; Karantanos et al., 2010; Salonen et al., 2010). The interaction of these factors could trigger and sustain IBS symptoms.

Visceral hypersensitivity

The abdominal pain feeling in IBS is mainly due to visceral hypersensitivity. The enteric nervous system (ENS) is localised in the submucosa layer and between the smooth muscle fibres of the GI wall. It transmits stimuli of abdominal pain and visceral reflexes to the central nervous system (CNS). Visceral sensitivity is regulated at the levels of the ENS, spinal chord, thalamus, and cerebral cortex. In response to injury of the enteric mucosa or visceral stimuli like bowel distension or irritative substances in the lumen, a variety of mediators are released that participate in a signalling cascade leading to the recognition of abdominal pain by the brain (Bueno and Fioramonti, 2002). Any alteration of this system could result in the characteristic abdominal pain of IBS. Different studies have shown that IBS patients have a reduced threshold for abdominal pain compared with healthy people (Serra et al., 2001; Bouin et al., 2002). Also, antagonists of receptors of those mediators, such as the neurotransmitter serotonin (5-HT), act as visceral analgesics in animal models of abdominal pain as well as in IBS patients (Banner and Sanger, 1995; Spiller, 2011).

Abnormal gut motility and secretion

The ENS also controls the neuromuscular and secretory functions of the GI tract. Stimuli are integrated in the ENS and in the spinal cord and hypothalamus. A variety of mediators and their receptors participate in this system. 5-HT stimulates intestinal secretion and peristalsis. Therapeutic agents targeting this system are employed for the treatment of IBS symptoms: antagonists of 5-HT receptors delay the GI transit in IBS patients with diarrhoea, while activation of the receptors promotes transit in IBS patients

with constipation (Talley et al., 1990; Martínez et al., 2004).

Autonomic nervous system dysfunction

The autonomic nervous system (ANS) regulates visceral sensitivity, gut motility, and gut secretion. Most of the IBS symptoms can be related to specific alterations of the ANS. Increased activity of the sympathetic nervous system and decreased activity of the parasympathetic nervous system are frequent in IBS (Adeyemi et al., 1999). Furthermore, differences in the ANS function are suspected to explain the variation between IBS subtypes and between genders (Aggarawal et al., 1994; Elsenbruch and Orr, 2001; Tillisch et al., 2005).

Low-grade mucosal inflammation

Although inflammation in the intestinal mucosa was initially discarded in IBS patients, the presence of low-level inflammation has been observed in several studies (Chadwick et al., 2002; Aerssens et al., 2008). Inflammation in IBS could arise from infection, stress, food allergy, and changes in the GI microbiota, and it is a normal symptom in post-infectious IBS. Inflammation could be seen as an overlap between IBS and IBDs, but actually it is of different nature in each condition (Spiller, 2009).

Genetic predisposition

Evidence of the contribution of genetic factors to IBS comes from familial aggregation, twin studies, and analysis of gene polymorphisms. Some studies have reported a higher prevalence of IBS among first-degree relatives of patients diagnosed with this disorder (Kalantar et al., 2003). Interestingly, one study reported an elevated IBS prevalence among first-degree relatives of patients with IBD, which may suggest the involvement of shared genetic factors (Aguas et al., 2011). However, these familial clusters could also reflect shared psychological and environmental factors. There is a higher concordance for IBS in monozygotic twins than in dizygotic twins (Levy et al., 2001; Bengston et al., 2006). Some polymorphisms in genes of the serotonergic and adrenergic systems and genes encoding immunomodulatory proteins have been related to IBS (Fukudo and Kanazawa, 2011). For example, the gene of the serotonin transporter (SERT), involved in the reuptake of serotonin from the synaptic cleft, has a long allele that has been related to constipation in homozygosis. The effect of the short allele is controversial: it has been related to diarrhoea in some studies but with constipation in others.

Psychosocial factors

IBS is not a psychiatric disorder *per se*, but psychosocial factors play an important role in the development and persistence of IBS symptoms (Drossman, 1999a; Hungin et al., 2003; Lea and Whorwell, 2003; Hood et al., 2008). Traumatic life events such as a history of physical abuse are strongly correlated with the development of IBS and the severity of its symptoms. Studies have repeatedly shown a high incidence of psychiatric comorbidity in IBS patients, such as anxiety, depression, and hypochondriasis. Anxiety is associated with rapid bowel transit and increased stool frequency, whereas depression is generally associated with delayed transit. Moreover, symptoms are typically worse at times of increased stress.

Diet

Diet is not the cause of IBS but can influence the development of symptoms. Indeed, many patients believe that food intolerance contribute to their symptoms. IBS has been related to malabsorption of certain CHOs such as lactose, sorbitol, or fructose. A hypersensitive reaction to the bowel distension caused by incomplete absorption of these CHOs is likely to occur. Careful eating may reduce IBS symptoms, but an individualised dietary control is needed to identify the specific substances causing symptoms (McKenzie et al., 2012).

Altered GI microbiota

Evidence of dysbiosis has emerged in patients with IBS. Since IBS could be considered a disorder of the gut-brain axis, some authors have proposed a model, illustrated in **Figure 1.8**, to incorporate dysbiosis into the pathogenesis of IBS (Collins and Bercik, 2009). Known risk factors for IBS, such as acute gastroenteritis, antibiotic therapy, or stress, could produce long-term changes in the indigenous microbiota of IBS patients due to an underlying instability of the gut physiology, in contrast to the transient changes seen in otherwise normal subjects. Intestinal dysbiosis could then result in inflammation that is subclinical but perturbs gut function and produces GI symptoms. In addition, intestinal dysbiosis might contribute to the psychiatric comorbidity that sometimes occurs in IBS patients.

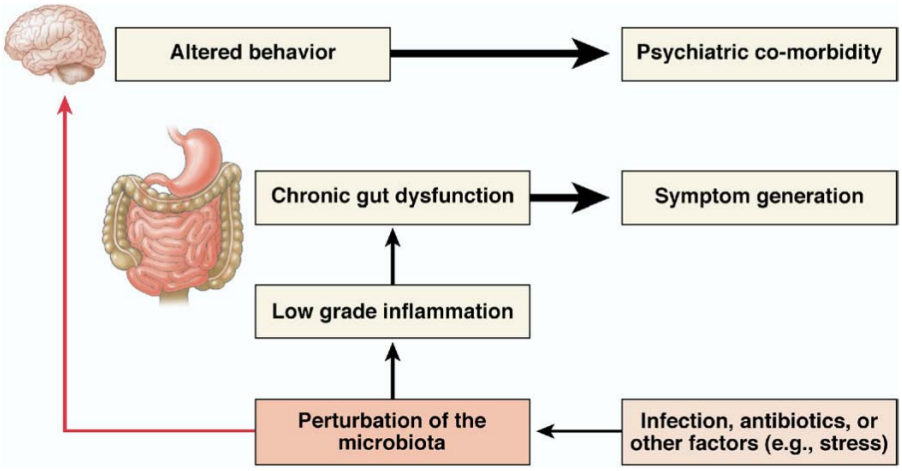


Figure 1.8: A hypothetical model describing the role of the intestinal microbiota in the pathogenesis of IBS. Adapted from Collins and Bercik (2009).

While this model requires testing in clinical studies, the involvement of the GI microbiota in IBS is supported by several observations:

a) Post-infectious IBS (PI-IBS) is a special subtype of IBS (Spiller and Garsed, 2009). About 6-17% of IBS patients believe their symptoms began after an episode of acute gastroenteritis, in most cases due to bacterial infection by *Campylobacter*, *Salmonella*, or *Shigella*. Infectious enteritis is the strongest risk factor for developing IBS. The risk increases with the duration of the initial infection and the virulence of the infecting microorganism. It is higher for females and in co-occurrence with psychological alterations (Neal et al., 2002). The mechanisms that cause PI-IBS are unknown but could include residual mucosal damage and inflammation or persistent changes in mucosal immunocytes, enterochromaffin cells (which store serotonin), mast cells (which store histamine and heparin), enteric nerves, and the GI microbiota.

b) Clinical trials targeting the microbiota seem to alleviate IBS symptoms (Parkes et al., 2008). Although antibiotic treatment can trigger IBS (Mendall and Kumar, 1998; Maxwell et al., 2002), its therapeutic use has been explored and found effective in reducing symptoms, in particular in patients diagnosed with small intestinal bacterial overgrowth (SIBO) (Basseri et al., 2011). The results of randomised controlled trials of probiotics in IBS are not consistent. Some showed reduction of symptoms, but others failed to prove benefits to the

patients (Moayyedi et al., 2010). Anyway, probiotics are viewed as a promising strategy to treat IBS.

c) An increased number of bacteria in the small intestine, a condition known as SIBO, has been observed in a subset of IBS patients (Lee and Pimentel, 2006). SIBO is an abnormal colonisation of the proximal small intestine by species commonly found in the lower GI tract. However, reports of SIBO in IBS are controversial. Firstly, IBS symptoms may be mimicked by SIBO. Secondly, the hydrogen breath test commonly used for the detection of SIBO seems to have a high false-positive rate (Simrén and Stotzer, 2006). Studies employing other methods suggest that the incidence of SIBO in IBS patients is not significantly different from that of controls (Posserud et al., 2007).

d) Some studies suggest an altered colonic microbiota in IBS patients, with specific features depending on the predominant bowel pattern (see Salonen et al. (2010) for a review of the recent insights into the intestinal microbiota of IBS patients generated in molecular studies). It has been found a higher instability in the faecal communities of IBS patients compared with those of controls in studies based on fingerprinting of 16S rRNA amplicons (Mättö et al., 2005; Maukonen et al., 2006). Other studies have looked for differences between IBS patients and healthy controls in the abundance of specific bacterial groups. The available data do not reveal pronounced IBS-related alterations and the results are often inconsistent between studies, even when the same Finnish IBS cohort is analysed under complementary methodologies (Salonen et al., 2010). Despite these limitations, some repeated findings in faecal samples of patients with IBS are a reduction in the levels of Actinobacteria (especially bifidobacteria) and *Lactobacillus*, and an increase in the levels of Enterobacteriaceae, observed both by culture and molecular methods (Balsari et al., 1982; Si et al., 2004; Malinen et al., 2005; Kassinen et al., 2007; Kerckhoffs et al., 2009; Krogius-kurikka et al., 2009; Tana et al., 2010). In addition, IBS-D and IBS-C appear to have a distinct microbial composition, and IBS-D patients differ from healthy controls in their faecal microbiota more than IBS-C patients (Kassinen et al., 2007; Rajilic-Stojanovic, 2007; Lyra et al., 2009).

There are several explanations for the lack of reproducibility of the changes in microbial composition detected in the studies carried out so far. One is that dysbiosis in IBS seems to consist in subtle alterations in the microbiota, in contrast to the global differences detected in other pathologies such as obesity or IBDs (Ley et al., 2006; Qin et al., 2010). The detection of such subtle changes is difficult due to the complexity of the GI microbiota and the high level

of between-subject variability in microbial composition due to reasons other than the pathological state. Also, cohorts are often relatively small, which reduces the statistical power to detect significant alterations. The disagreement between studies in the specific microbial groups found altered may also be partly due to the methods employed and to phenotypic variation within the patient's group. IBS is a complex disorder in its aetiology and clinical presentation. Patients with heterogeneous aetiology, type and/or severity of symptoms at the time of sampling could have been included in different studies, and even within the same study, further difficulting the detection of patterns.

Most of the studies on IBS are based on faecal samples because they are the most accessible source of GI microbiota. There is little information about the differences between IBS patients and healthy controls in mucosa-associated microbiota. Mucosal communities may be more important to the disorder because of their closer association with the host epithelium and thus their relevant role in the regulation of intestinal immunity and inflammation. The few studies targeting both mucosal and faecal communities in IBS have shown differences in the diversity and abundance of specific taxa between sampling sites, as seen in healthy controls. However, some changes were common to the mucosal and faecal samples, such as a decrease in *Bifidobacterium* and an increase in *Pseudomonas* in IBS patients (Kerckhoffs et al., 2009, 2011). Thus, each intestinal compartment seems to be differently affected by IBS, and its study could provide complementary information to address the role of the microbiota in IBS.

Currently, most of the detected alterations can not be functionally interpreted due to the lack of information about the role of many taxa. Nevertheless, some information is available on functional attributes of the gut microbiota that can contribute to symptoms, such as the fermentation end-products SCFAs and H_2. In one study, IBS patients showed a significant increase in the levels of acetic and propionic acids that correlated with worsening of symptoms and higher amounts of *Lactobacillus* and *Veillonella* (Tana et al., 2010). Noteworthy, acetate is a known irritant and at high concentrations induces mucosal lesions and abdominal cramps in laboratory animals (Burton and Gebhart, 1995). H_2 could be involved in some IBS symptoms, such as bloating, flatulence, and distension (Serra et al., 2001). H_2 is removed by methanogenic archaea, sulphate-reducing bacteria (SRB), and acetogenic bacteria (Nakamura et al., 2010). Methane produced by methanogens has been linked to constipation (Pimentel et al., 2006), and the H_2 disposal through SRB generates H_2S,

a toxic with potential detrimental impact on the host. However, the results of the current studies addressing the impact of H_2-removing microorganisms on IBS are inconclusive.

Further research is needed to assess the implication of the gut microbiota in IBS from a functional perspective, taking advantage of the recently developed high-throughput tools. Follow-up studies, in which samples are collected at moments with different symptomatology, could be very useful. They would help to mitigate the confusion caused by between-subject variability and heterogeneity within IBS that is problematic in cross-sectional studies. Also, they would reveal the structural and functional dynamics of the microbiota in IBS.

1.4 Methods for the study of the GI microbiota

The study of the microbial communities associated with humans involves the identification of the organisms present, their activity, and their interactions with each other and with the host. Progress in the study of the GI microbiota has been largely conditioned by the methodology available at the time. Some milestones in the development of methods for studying the human microbiota are discused below.

1.4.1 Anaerobic culture techniques

Anaerobic microbes, that is, microbes that grow in atmospheres without oxygen, were discovered lately in the history of microbiology (Pasteur, 1860). Effective methods for culturing anaerobes were developed in the middle decades of the 20th century (Hungate, 1950; Aranki et al., 1969) and were put in use for culturing anaerobes from the GI tract of ruminants and humans (Hungate, 1966; Moore and Holdeman, 1974). They allowed the access to the strict anaerobes of the GI tract and the recognition that they constitute the majority of the intestinal microbial communities of adult humans (Savage, 2001). Cultivation techniques are still widely used for the isolation and detailed characterisation of new representatives of the GI microbiota, which is being facilitated by the development of specific culture media (Zoetendal et al., 2003; Derrien et al., 2004) and high-throughput culture technologies (Zengler et al., 2002; Ingham et al., 2007).

1.4.2 Animal models for exploring the host-microbiota interactions

A gnotobiotic animal is one in which the status of the microbiota is known. GF animals have no microorganisms living in or on them (technically, they are also gnotobionts). Such animals are born and raised in sterile conditions. Gnotobionts are obtained after exposing GF animals to selected microbes (Faith et al., 2010). The comparison of GF and conventionally-raised mice has provided relevant information about the effects of the gut microbiota on host physiology and development (Section 1.2.2.4). Gnotobionts are used to study *in vivo* the symbiotic interaction between an animal and one or more of the microorganisms that inhabit its body (Sonnenburg et al., 2006; Mahowald et al., 2009). Besides, transplantation of gut microbial communities from conventional mice to GF recipients is used to determine how much of the donor phenotype is transferable via the microbiota and the underlying mechanisms (Turnbaugh et al., 2006; Garrett et al., 2007; Vijay-Kumar et al., 2010). Mice with genetically-determined or chemically-induced phatological phenotypes, such as models of obesity and colitis, are used to investigate the contribution of the microbiota to these diseases (Ley et al., 2005; Nell et al., 2010).

1.4.3 Microbial diversity estimates based on the SSU rRNA

The SSU rRNA gene is present in all life organisms due to its house-keeping function (it encodes an RNA component of the ribosome involved in the translation of messenger RNAs (mRNAs) into proteins). Therefore, its sequence is relatively conserved, but also includes variable regions that can be used for discriminating between species (Baker et al., 2003). The classification of organisms using the SSU rRNA gene led Carl Woese to define a new domain of life, Archaea, and to reconstruct the phylogenetic tree of life with the three domains, Archaea, Bacteria, and Eukarya (Woese, 1987; Woese et al., 1990). Although the concept of species remains unclear for prokaryotes, the analysis of 16S rRNA genes has become the standard method for their definition. Sequences are clustered into groups based on their sequence similarity. Arbitrary cut-offs are used to delimit microbial species, commonly ranging from 97 to 99% of sequence identity.

Polymerase chain reaction (PCR) amplification and sequencing of 16S

rRNA genes obtained from microbial communities has revolutionised microbial ecology by revealing that the majority of microbes in a variety of ecosystems were previously unknown, as they are yet unculturable. For example, estimates of cultivability of bacteria in the GI tract range from 10 to 50%, depending on the study (Wilson and Blitchington, 1996; Suau et al., 1999). The resistance of many microbes to be cultured could be due to their unknown growth requirements, the stress imposed by the culture procedure, and the almost impossible simulation of the interactions between microbes (and with the host) that occur in their natural environment.

A wide variety of molecular approaches based on the sequence variability of the 16S rRNA gene has been applied during the last 15 years to explore the diversity and dynamics of microbial communities (Zoetendal et al., 2004).

Sequencing of 16S rRNAs and their genes allows the identification of microorganisms, often to the species level, and the detection of even minoritary members of the community. 16S rRNA sequences can be obtained either from their genes after PCR amplification or directly from the 16S rRNAs after reverse-transcription and subsequent PCR. Genes encoding 16S rRNA reveal the composition of the community. 16S rRNAs discriminate the metabolically active microbial populations, since the number of ribosomes and the quantity of rRNA per cell depend on the growth phase and the level of activity, being higher in active than in dormant cells (Wagner, 1994). 16S rRNA sequences are assigned to taxonomic groups after comparison with sequences deposited in 16S rRNA databases. Phylogenetic analysis can be performed, and species-level phylogenetic types can be defined within the samples based on sequence similarity (**Figure 1.9**). Until recently, sequencing of PCR amplicons was very laborious and expensive because it required the generation of clone libraries before sequencing with the classical dideoxynucleotide chain termination method (Sanger sequencing). The advent of high-throughput sequencing (the Roche 454 GS-FLX technology and the Illumina Solexa technology, based respectively on pyrophosphate release and bridge amplification) has made possible the direct sequencing of amplicons, thus avoiding the cloning step (Metzker, 2010). The new sequencing platforms produce high numbers of reads in a single run, while the cost has dramatically dropped. In addition, sequencing of hypervariable regions of the 16S rRNA gene can be combined with the tagging of samples with specific barcode sequences, which enables mixing of multiple samples to be sequenced in parallel (Sogin et al., 2006; Andersson et al., 2008).

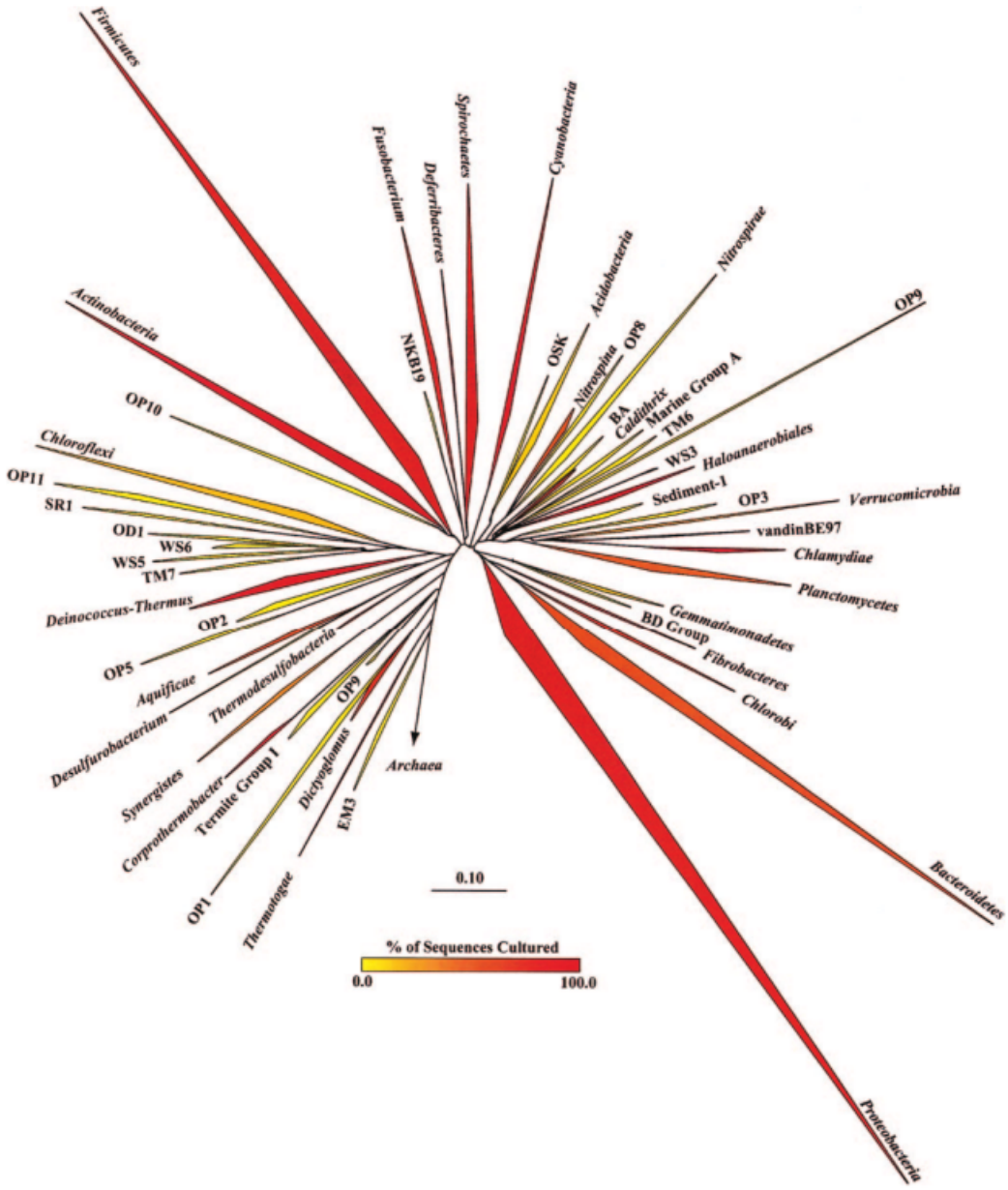


Figure 1.9: Phylogenetic tree of Bacteria showing established phyla (italicised Latin names) and candidate phyla within the November 2003 ARB database (http://arb-home.de). The vertex angle of each wedge indicates the relative abundance of sequences in each phylum; the length of each wedge indicates the range of branching depth found in each phylum; the redness of each wedge corresponds to the proportion of sequences in that phylum obtained from cultured representatives. Adapted from Handelsman (2004).

Fingerprinting techniques consist in the PCR amplification of 16S rRNA genes and the subsequent generation of profiles representing the sequence diversity within the selected samples. They include techniques such as denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE), that are based on sequence-specific melting behaviour of amplicons. They are very useful for a rapid monitoring and comparison of microbial communities. However, these diversity profiles only recover the most dominant bacteria, and sequencing is necessary for identification of the community members.

Several methods are used to quantify 16S rRNA concentrations in environmental samples. Real-time PCR uses species- or group-specific primers to quantify specific 16S rRNAs relative to the total concentration of 16S rRNA. Other approaches involve hybridisation with specific oligonucleotide probes instead of PCR amplification. Fluorescence *in situ* hybridisation (FISH) combines the hybridisation with fluorescent 16S rRNA probes with microscopy or flow cytometry for the direct enumeration of individual cells. Microscopy allows the direct visualisation of microbes and the determination of their spatial organisation in the samples (Swidsinski et al., 2005). Flow cytometry coupled with cell sorting allows the study of specific fractions of the community or even of single cells in a culture-independent manner (Peris-Bondia et al., 2011). A limitation of quantitative PCR and FISH is that only a few primers/probes can be used per analysis, so they are laborious at the species level. In contrast, high-throughput quantification of 16S rRNAs is possible with diversity microarrays (or phylogenetic microarrays). Diversity microarrays are small glass surfaces spotted with arrays of covalently linked species-specific 16S rRNA probes that are available for hybridisation. They allow the simultaneous quantification of thousands of microbes. Microarrays focused on the microbial communities of specific ecosystems have been constructed, including some that target human-associated microbiota (Rajilic-Stojanovic et al., 2009).

Studies that analyse the 16S rRNA sequence variability have rapidly expanded the knowledge about the composition and dynamics of microbial communities. However, these approaches have some well-known limitations (von Wintzingerode et al., 1997; Hongoh et al., 2003; Sipos et al., 2007; Hong et al., 2009). First, the design of primers and probes is limited by the available information of the 16S rRNA sequence diversity. Moreover, although these oligos are designed to target either the whole microbial community or particular taxa, their hybridisation efficiency may be not uniform. FISH depends on

the permeability of the target cells, while conditions of the PCR, such as the annealing temperature and the number of cycles, affect the estimated relative abundances of microbial groups in the samples. Additionally, little information can be derived about the activity of the microorganisms from these approaches, as there is a lack of information of their functional attributes in most of the cases and/or they have only moderate similarity to characterised microorganisms.

1.4.4 Genomics and functional genomics of microbial communities

Beyond the description of microbial diversity, new culture-independent approaches have been recently developed to gain insight into the activities of microbes in complex ecosystems (**Figure 1.10**) (Zoetendal et al., 2008).

Metagenomics (also called community genomics or environmental genomics) is the analysis of the collection of genomes directly isolated from an environment, that is, the metagenome. A metagenomic project can be sequence-driven or function-driven (Handelsman, 2004). Massive sequencing of the metagenome is performed to explore the genomic and genetic diversity within the ecosystem, the functional potential and the phylogenetic composition of the microbial community, and the distribution and redundancy of functions. Population properties can also be derived from metagenomic data, and comparative metagenomics can reveal the influence of environmental factors on microbial communities and their adaptations to specific habitats (Raes et al., 2007). As metagenomics tries to capture sequences from many diverse organisms simultaneously, the coverage of a particular organism in the sample at a given depth of sequencing is determined by its relative abundance and genome size. On the other hand, function-driven metagenomics involves the construction of expression libraries from metagenomic DNA that are screened to search for genes with specific functions of interest. This approach leads to the discovery of new genes for known or new functions, such as antimicrobial peptides, antibiotic resistance genes, and degradative enzymes. Metagenomics has revealed extensive information about the gene content of microbial communities in a variety of habitats, including the GI tract (Venter et al., 2004; Daniel, 2005; Delong et al., 2006; Gill et al., 2006; Kurokawa et al., 2007).

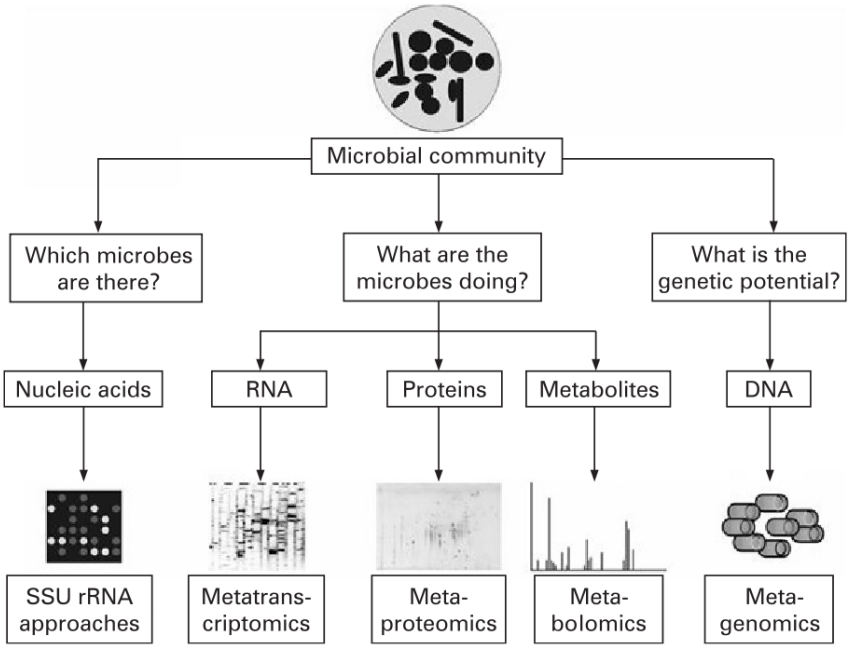


Figure 1.10: Schematic representation of metagenomics and other community-based "omic" approaches. Adapted from Zoetendal et al. (2008).

A limitation of metagenomics is that it is not known whether the predicted genes are expressed and, if so, under what conditions and to what extent. In addition, it is not possible to determine whether environmental DNA comes from active, dormant, or dead cells. Other meta-"omic" approaches that focus on activity biomarkers, such as RNA (metatranscriptomics) (Frias-Lopez et al., 2008; Urich et al., 2008; Shi et al., 2009; Turnbaugh et al., 2010; Gosalbes et al., 2011), proteins (metaproteomics) (Ram et al., 2005; Verberkmoes et al., 2009; Kolmeder et al., 2012), or metabolites (metabolomics) (Antunes et al., 2011; Ponnusamy et al., 2011), are needed to understand the *in situ* activity of microorganisms. Metatranscriptomics is the most advanced of these functional community-based approaches and has been applied mainly to aquatic and soil ecosystems, but also to the GI tract. In metatranscriptomics, complementary DNA (cDNA) is synthesised from total RNA extracted from an environmental sample and is used for massive sequencing. Total RNA is dominated by rRNAs, but also includes functionally relevant molecules such as mRNAs, which will be translated into proteins, and small non-coding RNAs (sRNAs), which act

as regulatory elements or ribozymes. Therefore, metatranscriptomics allows to assess the structure of the most active members of the community and the genes that are expressed under particular conditions. This offers the opportunity to link community structure and function and to get a clearer picture of the *in situ* activity of the microorganisms in an ecosystem.

High-throughput sequencing is generating a vast amount of complex data that are challenging to analyse. In fact, the study of microbial communities has prompted the adaptation and further development of computational and statistical methods and tools employed in genomics and microbial ecology, as well as to the continuous development of new bioinformatic analytical tools.

2 Objectives

The advent of high-throughput molecular techniques has given a boost to the research on environmental microbial communities during the last decade. Those associated with human hosts are being investigated from ecological and biomedical perspectives by many researchers worldwide. This thesis focuses on some basic issues concerning the distal gut microbiota such as its spatial distribution, its temporal dynamics, and its association with a disease in which a link with the microbiota is suspected. Usually, we were not the first to address these issues and other studies on the same topics were published during the development of our work, as one can expect in this hot research area. However, we tried to go beyond what was known at the time of the experimental design of each study included here. We also tried to take advantage of the available knowledge for comparison with our results.

The objectives we tackled are stated below and will be presented in specific sections:
Study 1. Bacterial diversity in rectal mucosa and faeces of healthy subjects.
Study 2. Structural alterations of colonic mucosal and faecal bacteria in the irritable bowel syndrome.
Study 3. Stability and host-specificity of faecal bacteria in healthy subjects.
Study 4. Follow-up of faecal microbial communities in the irritable bowel syndrome.
Study 5. Response of a *Prevotella*-dominated human faecal microbiota to a ketogenic diet.

Our main objective was to study the potential relationship of the colonic microbiota with IBS, one of the most common functional bowel disorders in Western countries. The definition and treatment of IBS is challenging due to its largely unknown aetiology and the variety of symptoms that patients can

present. Alterations in the GI microbiota have been longly suspected to be one of the numerous factors potentially involved in the onset and persistence of IBS (Section 1.3.3).

To address this, we first tried to understand the intrinsic variation in the GI microbiota within and between individuals. This led us to run a few other studies.

We first measured the within-subject variability of the bacterial composition of faeces and colonic mucosal biopsies in healthy subjects in order to evaluate the extent to which the faecal bacteria resemble the mucosa-associated ones (**Study 1**).

Then, our first approach to IBS was to compare the bacterial composition of IBS patients with different predominant bowel habits with that of healthy subjects (**Study 2**). We sampled faecal and colonic mucosa-associated microbial communities, unlike most studies on IBS, which only analyse faecal samples.

We also measured the short-term temporal variation in faecal bacterial communities in healthy subjects (**Study 3**). Our aims were to evaluate the normal dynamics of the community structure and to establish putative patterns of interaction between specific bacteria.

Our results in the cross-sectional study on IBS, as well as the conclusions about the normal temporal variation in the structure of the microbiota, led us to undertake a longitudinal study on IBS (**Study 4**). It allowed us to compare samples from single patients at moments with different type and/or severity of symptoms, thus overcoming the confusion introduced by the inter-subject variability and the heterogeneity within IBS.

Additionally, we realised that some obese subjects included in our cohorts had an unusual high prevalence of bacteria within the *Prevotella* genus. Together with some observations extracted from the literature, this led us to suggest a potential link between this genus and some type of obesity, which was first tested through the monitoring of one volunteer following a CHO-restricted diet (**Study 5**).

When volunteers were followed over time (Studies 3-5), sampling was restricted to faeces because they are easily collected in a non-invasive manner. The composition of mucosal and faecal bacterial communities was analysed through massive sequencing of 16S rRNA amplicons in Studies 1, 2, 3, and 5. We adopted metagenomic and metatranscriptomic approaches in Study 4 to assess potential changes in the genetic potential and the gene expression of the microbiota.

3 Material and Methods

3.1 General remarks

We studied the human microbiota of several colonic mucosal sites and faeces, both in healthy volunteers and IBS sufferers. The microbial communities were analysed by sequencing their nucleic acids without isolation and culturing of the community members.

The general "wet lab" methodology for the specific studies was overall similar. Protocols started with the isolation of nucleic acids. On one hand, the DNA of the whole community was used for the analysis of the taxonomic composition by PCR amplification and sequencing of the 16S rRNA genes. Sometimes, it was also used for the analysis of the gene composition by direct sequencing (metagenomics). On the other hand, the RNA of the whole community was used for the analysis of the gene expression by sequencing the cDNA obtained by retro-transcription of the total RNA (metatranscriptomics).

There were diferences between studies in the experimental procedure due to several factors. Regarding the isolation of nucleic acids, in our first studies we were only interested in the DNA content of the samples, so we employed protocols optimised for the DNA isolation from each sample type, biopsies and faeces. Later on, to analyse also the RNA content, we adopted a protocol to perform the co-extraction of both types of nucleic acid instead of the previous ones. We were also conditioned by the sequencing technologies available at each time, which have experienced a revolution in the last years with the advent of next-generation sequencing technologies. We started sequencing with the classical Sanger method, but we moved to pyrosequencing when it became available, thus avoiding the cloning step prior to sequencing. We then moved to the improved pyrosequencing chemistry to obtain longer reads. The sequencing technology employed in each study affected in turn the choice of primers and conditions for the PCR amplification.

The detailed protocols of the specific objectives (**Studies 1-5**) will be explained below. Section 3.2 describes the characteristics of the study participants and the procedure of sample collection. Section 3.3 provides details about the processing of samples, from the isolation of nucleic acids to their sequencing. Section 3.4 specifies the bioinformatic and statistical analyses applied to sequence data.

3.2 Sampling

3.2.1 Study participants

All participants gave prior informed written consent to the study protocol, which was approved by the Ethics Committee of La Fe University Hospital (Valencia, Spain). Volunteers were administered a questionnaire face to face about lifestyle and relevant clinical features.

Study 1
The study subjects comprised nine volunteers (hereinafter referred to as individuals 1-9). None had a previous history of GI disease or systemic comorbidities, recent treatment with antibiotics (except for V7, who had taken antibiotics the previous month), immunomodulating therapy, anti-diarrhoeal medication, or laxatives. Relevant volunteers' details are summarised in **Table 3.1**.

Study 2
The study subjects comprised nine healthy controls (subjects without GI disorders or systemic comorbidities) enrolled from a screening program for colon cancer prevention, and sixteen IBS patients classified into diarrhoea subtype (IBS-D, thirteen patients) and constipation subtype (IBS-C, three patients) according to the Rome II criteria (Drossman, 1999b; Longstreth et al., 2006). Hereinafter they will be referred to as controls C01-C09 and patients I01-I16. Relevant volunteers' details are summarised in **Table 3.2**.

Table 3.1: Characteristics of the volunteers of Study 1 and sample collection date. BMI: body mass index, mass (kg)/(height (m))^2

Volunteer	Age	Sex	Nationality	BMI	Antibiotics (last 3 months)	Collection date Biopsies	Faeces
1	29	F	Spanish	21.5	No	27-11-07	08/02/08
2	26	F	Spanish	20.2	No	27-11-07	11/12/07
3	36	F	Spanish	23.7	No	27-11-07	02/01/08
4	61	F	Spanish	22.6	No	27-11-07	08/02/08
5	42	F	Spanish	32	No	11/12/07	11/12/07
6	33	M	Italian (48 months in Spain)	25.4	No	02/06/08	02/06/08
7	37	M	Spanish	31.3	Ampicilin (1 month earlier)	03/06/08	03/06/08
8	40	M	Spanish	24.4	No	03/06/08	03/06/08
9	36	M	Mexican (8 months in Spain)	24.4	No	03/06/08	03/06/08

Study 3

The study subjects were three male volunteers (hereinafter referred to as individuals A, B, and C). None had a previous history of GI disease or systemic comorbidities, recent (in the last 3 months) treatment with antibiotics, immunomodulating therapy, anti-diarrhoeal medication, or laxatives. A, B, and C were 40, 39, and 29 years old, respectively. A was normal-weight; B and C were obese class I according to the classification by the body mass index of the World Health Organisation ($30 \leq \text{BMI} < 35$). They followed a Mediterranean diet that remained unchanged throughout the follow-up.

Study 4

Initially, four IBS patients with diarrhoea subtype, according to the Rome II criteria (Drossman, 1999b; Longstreth et al., 2006), were included in the study (Patients 1-4). The husband of Patient 1 was included as an age-matched control that also shared her environment (Control 1). Patients underwent medical examination every two weeks. Additionally, patients collected daily data on symptoms during the follow-up, specifically, presence of abdominal pain or discomfort, abdominal distension, and defaecatory urgency, and number and type of stools (according to the Bristol scale, **Figure 3.1**) (Lewis and Heaton, 1997). Patients 3 and 4 remained fairly asymptomatic over the follow-up, so their samples were not selected for analysis. Relevant volunteers' details are summarised in **Table 3.3**.

Table 3.2: Characteristics of the volunteers of Study 2.

(a)

Subject	Group	Age	Sex	BMI	Nationality	Educational level	Physical activity	High fibre food consumption
C01	Control	60	Female	25.1	Ecuadorian (7 years in Spain)	High school	Moderate	Daily
C02	Control	56	Male	27.5	Spanish	Primary school	Moderate	Daily
C03	Control	55	Female	27.9	Spanish	Primary school	Moderate	Daily
C04	Control	62	Female	26.4	Spanish	University/College	Moderate	Daily
C05	Control	55	Male	31.4	Spanish	University/College	Moderate	Daily
C06	Control	54	Male	NA	Spanish	Primary school	Moderate	Daily
C07	Control	50	Male	26.1	Spanish	University/College	Moderate	Daily
C08	Control	62	Female	23	Spanish	University/College	Sedentary	Daily
C09	Control	66	Male	25.5	Spanish	Primary school	Moderate	Weekly
I01	IBS-diarrhoea	49	Female	18.7	Spanish	University/College	Moderate	Daily
I02	IBS-diarrhoea	33	Male	29.3	Spanish	High school	Intense	Monthly
I03	IBS-diarrhoea	32	Female	18.9	Spanish	University/College	Sedentary	Weekly
I04	IBS-diarrhoea	62	Female	31.5	Spanish	University/College	Sedentary	Daily
I05	IBS-diarrhoea	35	Male	24.7	Spanish	University/College	Intense	Daily
I06	IBS-diarrhoea	56	Male	34	Spanish	University/College	Moderate	Weekly
I07	IBS-constipation	60	Female	NA	Spanish	University/College	Moderate	We ekly
I08	IBS-diarrhoea	38	Male	24	Spanish	High school	Moderate	Daily
I09	IBS-diarrhoea	53	Female	24	Spanish	University/College	Moderate	Daily
I10	IBS-constipation	46	Female	15.4	Spanish	High school	Intense	Daily
I11	IBS-diarrhoea	27	Male	23.2	Spanish	University/College	Sedentary	Daily
I12	IBS-constipation	46	Male	21.6	Spanish	Primary school	Moderate	Daily
I13	IBS-diarrhoea	26	Female	25.5	Spanish	High school	Sedentary	Monthly
I14	IBS-diarrhoea	42	Female	17.3	Spanish	Primary school	Moderate	Weekly
I15	IBS-diarrhoea	57	Male	23.7	Spanish	University/College	Sedentary	Daily
I16	IBS-diarrhoea	27	Male	20.8	Spanish	University/College	Sedentary	Daily

(b)

Subject	Alcohol consumption	Smoking	Antibiotics (last 3 months)	Drug consumption at sampling time
C01	Monthly	Non-smoker	No	None
C02	Daily	Smoker	No	None
C03	Never	Ex-smoker	No	Hypertension, Diabetes
C04	Weekly	Smoker	No	Gastroesophageal reflux, Bowel spasms
C05	Monthly	Ex-smoker	Yes	Gastroesophageal reflux
C06	Never	Non-smoker	No	None
C07	Daily	Ex-smoker	No	None
C08	Daily	Smoker	No	None
C09	Daily	Ex-smoker	No	Gastroesophageal reflux, Hypertension, Dietary supplements
I01	Daily	Non-smoker	No	Psychotropics, Hyperaldosteronism
I02	Never	Ex-smoker	No	None
I03	Weekly	Smoker	Yes	None
I04	Monthly	Ex-smoker	No	Gastroesophageal reflux, Hypertension
I05	Daily	Smoker	Yes	Gastroesophageal reflux
I06	Weekly	Non-smoker	No	None
I07	Never	Smoker	No	None
I08	Never	Ex-smoker	Yes	Gastroesophageal reflux
I09	Never	Ex-smoker	No	Psychotropics, Hypertension
I10	Monthly	Ex-smoker	No	Gastroesophageal reflux, Bowel spasms, Psychotropics
I11	Monthly	Non-smoker	No	None
I12	Monthly	Non-smoker	No	Psychotropics, Osteoporosis
I13	Monthly	Smoker	No	Contraceptives
I14	Daily	Smoker	No	Psychotropics, Hyperlipidemia
I15	Weekly	Ex-smoker	No	Gastroesophageal reflux, Hypertension, Hyperlipidemia
I16	Monthly	Non-smoker	No	Gastroesophageal reflux, Alopecia

Table 3.3: Characteristics of the volunteers of Study 4. C1, P1, and P2 are Control 1, Patient 1, and Patient 2, respectively.

Subject	C1	P1	P2
Group	Control	IBS-diarrhoea	IBS-diarrhoea
Age	66	62	21
Sex	Male	Female	Female
BMI	31.7	27.5	30.4
Nationality	Spanish	Spanish	Spanish
Educational level	University/College	University/College	High school
Physical activity	Moderate	Moderate	Moderate
High fibre food consumption	Daily	Weekly	Daily
Alcohol consumption	Monthly	Never	Never
Smoking	Ex-smoker	Ex-smoker	Smoker
Antibiotics (last 3 months)	No	No	No
Drug consumption	Hypertension	Hypertension	Bowel spasms
at sampling time	Hyperlipidemia	Arthrosis	Contraceptives
	Urinary-tract disorders	Osteoporosis	

Study 5

An obese volunteer (individual B in **Study 3**) subject to a ketogenic diet (high in fat and protein and virtually without CHOs) was monitored for two months. His daily food intake and weight loss were registered during the follow-up.

3.2.2 Sample collection

Biopsies were obtained from the colonic mucosa by endoscopy using a standard colonoscopy (Olympus) and single-use biopsy forceps (Radial JawTM 4, Boston Scientific), recovered in dry tubes, preserved on ice, and immediately frozen at -80 °C. Faeces were self-collected by the volunteers in tubes containing 10 mL of phosphate-buffered saline (PBS, containing, per liter, 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4, and 0.24 g of KH_2PO_4 [pH 7.2]) and stored in the volunteers' home freezers until its release. Faecal samples were kept at 4 °C for 1-2 hours before being stored at -80 °C. All samples were stored at -80 °C until further processing.

Study 1

Biopsies of the rectal mucosa and faeces were obtained from each subject. Biopsies were obtained by rectoscopy. Neither laxatives nor enema were adminis-

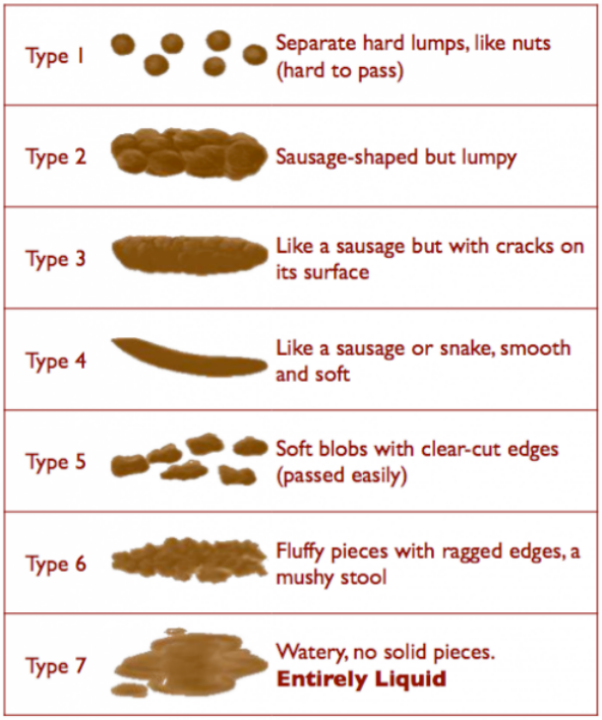


Figure 3.1: Categories of the Bristol stool scale for the form of faeces. Types 1 and 2 indicate constipation, while types 6 and 7 show diarrhoea. Adapted from the Wikipedia file "Bristol Stool Chart.png".

tered prior to endoscopy to avoid the potential disturbance of the mucosal microbiota associated with these procedures. Biopsies were taken in the absence of macroscopic faeces. Although the possibility of contamination of biopsies by faecal material could not be ruled out, it is unlikely because stools usually have hard consistency and are not attached to the mucosa in the rectum. Faeces were collected within the shortest possible time-lapse to biopsies in order to minimise potential temporal changes in community composition. Initially, all faecal samples were collected prior to rectoscopy and within the same day. However, for four of the volunteers, the faecal sample did not provide enough DNA for the study and eventually a second sample was obtained between 2 and 8 weeks after rectoscopy (see **Table 3.1** for details).

Study 2
Three types of sample were obtained per subject: biopsies of the ascending and the descending colonic mucosa, and faeces. Biopsies were obtained by endoscopy, and participants were administered laxatives prior to the endoscopic procedure. Faeces were collected prior to endoscopy and within the same day, thus minimising potential temporal changes in community composition between faecal and mucosal samples.

Study 3
The study participants collected faecal samples daily in the morning over a period of fifteen consecutive days.

Study 4
The sampling design was to start the follow-ups at a moment of acute symptoms and to collect faecal samples every two days the first week. Then, weekly samples would be collected. Additional samples would be collected in case of worsening of symptoms. Patient 1 and Control 1 provided faecal samples at days 1, 3, 5, 7, 14, 21, 28, 37, and 42 of their follow-ups. Faecal samples were collected at days 1, 3 (1st and 2nd), 4, 5, 7, 14, 21, 27, 28, 35, 42, 49, and 56 from Patient 2. A summary of patients' symptoms in the days of sampling is shown in **Table 3.4**.

Study 5
The volunteer provided faecal samples daily before, during, and after a dietary intervention of 24 days. From them, samples at days -6, -5, -2, -1 (pre-diet), 1, 2, 3, 4, 7, 8, 9, 13, 14, 15, 19, 24 (during diet), and +1, +2, +3, +4, +27, +57 (post-diet) were analysed.

3.3 Sample processing

3.3.1 Nucleic acid extraction

DNA was extracted from the biopsy samples in **Studies 1** and **2** using the QIAamp DNA Mini Kit (QIAGEN) and its protocol for DNA purification from tissues. The standard protocol was modified to maintain overnight the incubation at 56 °C with ATL buffer and proteinase K, and to extend the incubation at 70 °C with RNase A from 10 min to 30 min. DNA was extracted from the

Table 3.4: Intestinal symptons of Patient 1 (P1) and Patient 2 (P2) in the days of sampling, the previous days, and the subsequent ones. It is indicated the presence of abdominal pain or discomfort, abdominal distension, and defaecatory urgency, the number of stools, and the number of stools of each type according to the Bristol scale (T1-T7). Data for day 29 in Patient 2 were not recorded.

	Day	Sampling	Pain	Distension	Urgency	Stools	T1	T2	T3	T4	T5	T6	T7
P1	1	Yes	Yes	Yes	Yes	2	-	-	1	-	-	1	-
	2	No	Yes	Yes	Yes	2	-	-	1	-	-	1	-
	3	Yes	Yes	Yes	Yes	2	-	-	1	-	-	1	-
	4	No	Yes	Yes	Yes	2	-	-	-	1	-	1	-
	5	Yes	Yes	Yes	Yes	3	-	-	1	1	-	1	-
	6	No	Yes	Yes	Yes	4	-	1	-	-	-	3	-
	7	Yes	Yes	Yes	Yes	2	-	1	-	-	-	1	-
	8	No	Yes	Yes	Yes	2	-	-	1	1	-	-	-
	13	No	Yes	Yes	Yes	2	-	-	1	1	-	-	-
	14	Yes	Yes	Yes	Yes	4	-	-	1	-	1	2	-
	15	No	Yes	Yes	Yes	2	-	-	1	-	-	1	-
	20	No	Yes	Yes	Yes	2	-	-	1	-	1	-	-
	21	Yes	Yes	No	No	3	-	-	1	-	-	2	-
	22	No	Yes	Yes	No	1	-	1	-	-	-	-	-
	27	No	Yes	Yes	Yes	2	-	1	-	-	-	1	-
	28	Yes	Yes	Yes	Yes	3	-	1	-	-	-	2	-
	29	No	Yes	Yes	Yes	2	-	-	-	-	-	2	-
	36	No	Yes	Yes	Yes	2	-	-	-	-	-	2	-
	37	Yes	Yes	Yes	Yes	2	-	-	-	-	-	2	-
	38	No	Yes	Yes	Yes	1	-	-	-	-	-	1	-
	41	No	Yes	Yes	Yes	3	-	-	-	-	-	3	-
	42	Yes	Yes	Yes	Yes	2	-	-	-	-	-	2	-
	43	No	Yes	Yes	Yes	2	-	2	-	-	-	-	-
P2	1	Yes	No	No	Yes	3	-	-	-	-	1	1	1
	2	No	Yes	Yes	No	3	-	-	-	-	-	3	-
	3	Yes	Yes	Yes	Yes	11	-	-	-	-	-	-	11
	4	Yes	Yes	Yes	Yes	5	-	-	-	-	-	2	3
	5	Yes	Yes	Yes	Yes	4	-	-	-	-	-	3	1
	6	No	No	No	No	5	-	-	-	-	-	3	2
	7	Yes	No	No	No	4	-	-	-	-	-	3	1
	8	No	Yes	Yes	Yes	5	-	-	-	-	-	3	2
	13	No	Yes	Yes	No	5	-	-	-	-	-	5	-
	14	Yes	Yes	Yes	Yes	5	-	-	-	-	-	5	-
	15	No	No	No	No	3	-	-	-	1	-	2	-
	20	No	Yes	Yes	Yes	5	-	-	-	-	-	-	5
	21	Yes	Yes	Yes	Yes	5	-	-	-	-	-	3	2
	22	No	Yes	Yes	Yes	7	-	-	-	-	-	5	2
	26	No	Yes	Yes	Yes	6	-	-	-	-	-	3	3
	27	Yes	Yes	Yes	Yes	4	-	-	-	-	-	4	-
	28	Yes	Yes	Yes	Yes	7	-	-	-	-	-	3	4
	34	No	Yes	Yes	Yes	4	-	-	-	-	-	-	4
	35	Yes	Yes	Yes	No	2	-	-	-	-	2	-	-
	36	No	Yes	Yes	No	6	-	-	-	-	-	4	2
	41	No	Yes	Yes	Yes	3	-	-	-	3	-	-	-
	42	Yes	No	No	No	5	-	-	-	-	-	5	-
	43	No	Yes	Yes	Yes	7	-	-	-	-	-	-	7
	48	No	Yes	Yes	Yes	3	-	-	-	3	-	-	-
	49	Yes	Yes	Yes	Yes	4	-	-	-	-	-	-	4
	50	No	Yes	No	Yes	2	-	-	-	-	-	-	2
	55	No	Yes	Yes	Yes	8	-	-	-	-	-	-	8
	56	Yes	Yes	Yes	Yes	4	-	-	-	-	-	-	4
	57	No	Yes	Yes	Yes	6	-	-	-	-	-	-	6

faecal samples in **Studies 1**, **2**, and **3** using the QIAamp DNA Stool Mini Kit (QIAGEN) and its protocol for isolation of DNA for pathogen detection. Before the DNA extraction, the faecal samples were resuspended in PBS and centrifuged at 4000 rpm for 3 min to remove faecal debris as far as possible. Between 1-3 mL of the supernatants were centrifuged at 14000 rpm for 5 min and the pellets were resuspended in 2 mL of ASL buffer from the QIAGEN kit. Then, we went on to step 3 of the protocol. DNA extractions were stored at -20 °C.

DNA and RNA were simultaneously extracted from the faecal samples in **Studies 4** and **5** using the AllPrep DNA/RNA Mini Kit (QIAGEN) and its protocol for simultaneous purification of genomic DNA and total RNA from animal cells. Before the DNA/RNA extraction, the faecal samples were treated as described above but being the pellets resuspended in RLT Plus buffer. DNA and RNA extractions were stored at -20 °C and -80 °C, respectively.

3.3.2 Amplification and sequencing of 16S rRNA genes

DNA from samples in **Studies 1**, **2**, **3**, and **5** was used to amplify bacterial 16S rRNA genes by PCR using broad range bacterial primers. Then, amplicons were sequenced using the same primers (**Figure 3.2**). The PCR conditions were selected to minimise biases in the estimation of the distribution of bacterial taxa in the samples potentially introduced by the amplification process (use of degenerate primers, low primer annealing temperature, and low number of amplification cycles).

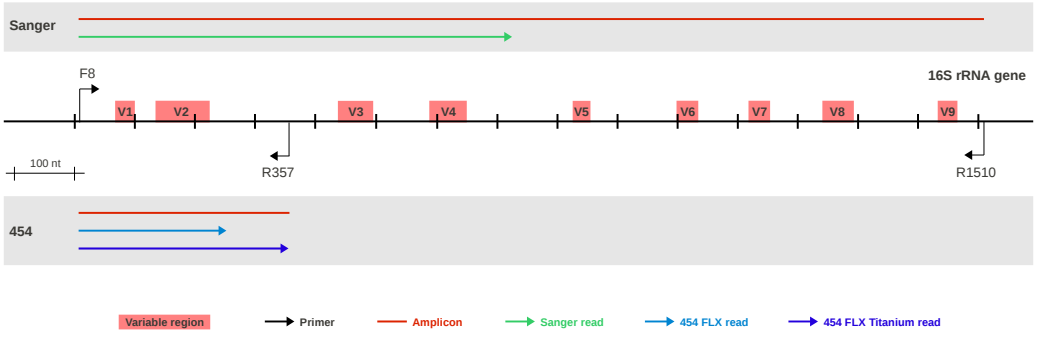


Figure 3.2: Schematic view of the amplicons and reads generated by Sanger and 454 sequencing in the studies of this thesis. Positions are mapped along the 16S rRNA gene based on the *Escherichia coli* reference sequence.

Study 1

The 16S rRNA genes of the samples were amplified using the bacterial forward primer B8F (5'-AGAGTTTGATCMTGGCTCAG-3') and the bacterial reverse primer B1510R (5'-TACGGYTACCTTGTTACGACTT-3') (Baker et al., 2003).

Each PCR mixture was composed of 25 μL of GoTag Green Master Mix (Promega), 1 μL of each primer (20 μM), and 1 μL of template DNA in a total volume of 50 μL. The PCR conditions were 5 min of initial denaturation at 95 °C followed by 25 cycles of denaturation (30 s at 95 °C), annealing (30 s at 56 °C), and elongation (90 s at 72 °C), with 8 min of final extension at 72 °C. The PCR products were purified by ethanol precipitation.

The PCR products were ligated to pCR-XL-TOPO vectors using the TOPO XL PCR Cloning kit (Invitrogen). One-Shot TOP10 electrocompetent *E. coli* cells (Invitrogen) were transformed, according to the manufacturer's instructions. Approximately 800 transformant colonies from each library were picked up randomly and plasmid extraction was performed using the Montage Plasmid MiniPrep96 Kit (Millipore) and a MultiPROBE II Robotic Liquid Handling System (Packard BioScience).

The 5' half of the cloned 16S rRNA genes was determined by Sanger sequencing using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and the B8F primer. Sequences were analysed on an ABI 3730 Sequencer (Applied Biosystems).

Study 2

The 16S rRNA genes of the samples were amplified using the composite forward primer (5'-GCCTCCCTCGCGCCATCAGNNNNNNTC*AGAGTTTGATCM-TGGCTCAG*-3') and the composite reverse primer (5'-GCCTTGCCAGC-CCGCTCAGGC*TGCTGCCTCCCGTAGGAGT*-3'), where the underlined sequences are the adaptors A and B for the pyrosequencing GS FLX chemistry (454 Life Sciences, Roche), respectively, NNNNNN designates a unique six-nucleotide barcode used to tag each PCR product, and in italics are the bacterial primers B8F and B357R, respectively (Baker et al., 2003).

Each PCR mixture was composed of 0.4 μL of FastStart Taq DNA polymerase (Roche), 5 μL of 10X PCR reaction buffer with $MgCl_2$ (Roche), 2 μL of dNTPs (10 mM), 1 μL of each composite primer (20 μM), and 1 μL of template DNA in a total volume of 50 μL. The PCR conditions were 5 min of initial denaturation at 95 °C followed by 25 cycles of denaturation (30 s at 95 °C),

annealing (30 s at 52 °C), and elongation (30 s at 72 °C), with 8 min of final extension at 72 °C. The PCR products were purified by vacuum filtration, and equal amounts of barcode-tagged PCR products from different samples were pooled.

The mixtures were sent for pyrosequencing with adaptor A on a Genome Sequencer FLX system and the GS FLX chemistry (454 Life Sciences, Roche).

Study 3

The 16S rRNA genes of the samples of individual A were amplified using the composite primers employed in **Study 2**. The 16S rRNA genes of the samples of individuals B and C were amplified using the composite forward primer (5'-CGTATCGCCTCCCTCGCGCCATCAGNNNNNNNNTC*AGAGTTTGATCMTGGCTCAG*-3') and the composite reverse primer (5'-CTATGCGCCTTGCCAGCCCGCTCAG*TGCTGCCTCCCGTAGGAGT*-3'), where the underlined sequences are the adaptors A and B for the pyrosequencing GS FLX Titanium chemistry (454 Life Sciences, Roche), respectively, NNNNNNNN designates a unique eight-nucleotide barcode used to tag each PCR product, and in italics are the bacterial primers B8F and B357R, respectively (Baker et al., 2003).

The PCR mixtures, reaction conditions, and PCR purification method were the same as for samples in **Study 2**, except for the number of amplification cycles that was 20.

The mixtures were sent for pyrosequencing with adaptor A on a Genome Sequencer FLX system and the GS FLX chemistry for samples of subject A, and the GS FLX Titanium chemistry for samples of subjects B and C (454 Life Sciences, Roche). This difference was due to the change in the sequencing chemistry while the study was ongoing.

Study 5

The 16S rRNA genes of the samples were amplified using the bacterial primers B8F and B357R, with eight-nucleotide barcodes added to the forward primer used to tag each PCR product. In this case, the adaptors A and B for the pyrosequencing GS FLX Titanium chemistry (454 Life Sciences, Roche) were added later by the sequencing service, thus minimising potential biases introduced during the PCR due to the adaptor sequences when added to the primers.

The PCR mixtures, reaction conditions, and PCR purification method were

the same as for samples in **Study 3**.

The mixtures were sent for pyrosequencing with adaptor A on a Genome Sequencer FLX system and the GS FLX Titanium chemistry (454 Life Sciences, Roche).

3.3.3 Sequencing of metagenomes

Total DNA extracted from samples in **Study 4** was sent directly for pyrosequencing on a Genome Sequencer FLX system and the GS FLX Titanium chemistry (454 Life Sciences, Roche).

3.3.4 Sequencing of metatranscriptomes

Total RNA extracted from samples in **Study 4** was incubated with DNase I (Ambion) 30 min at 37 °C. The digestion of DNA was checked with a PCR using the broad range bacterial primers B8F and B357R and the treated RNA as template. The integrity of total RNA was assessed using standard agarose gel electrophoresis.

Total RNA (DNase-treated) was amplified with enrichment in mRNA using the MessageAmp II-Bacteria Kit (Ambion) following the manufacturer's instructions. Briefly, the method is based on polyadenilation of the 3' end of RNA using the *E. coli* Poly(A) Polymerase (PAP). The poly(A)-tailed RNA is reverse-transcribed to cDNA with an oligo(dT) primer containing a recognition site for a restriction enzyme (BpmI). These steps favour the preferential amplification of bacterial mRNA compared to rRNA because the 3' end of intact rRNA is sterically blocked, which obstructs their polyadenilation (Wendisch et al., 2001). Then, cDNA is linearly amplified by an *in vitro* transcription yielding large amounts of antisense RNA (aRNA).

This aRNA was further reverse-transcribed to single-strand cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), and to double-strand cDNA using an enzymatic mix containing *E. coli* DNA Polymerase I, *E. coli* DNA Ligase, RNase H, and T4 DNA Polymerase (New England Biolabs).

Finally, the double-strand cDNA was digested with BpmI (Fermentas), purified by ethanol precipitation, and sent for pyrosequencing on a Genome Sequencer FLX system and the GS FLX Titanium chemistry (454 Life Sciences, Roche).

3.4 Sequence analyses

3.4.1 Initial processing of sequences

Amplicon reads generated by the Sanger method

In **Study 1**, base-calling of each sequence was performed using the *Pregap4* program; sequences were then manually revised using the *Trev* and *Gap4* programs, all in the *Staden* v1.12 package (Staden et al., 2000). After adjusting for quality values, the average read length was around 700 nucleotides.

Amplicon reads generated by pyrosequencing

In **Studies 2**, **3**, and **5**, processing of the pyrosequencing output was done using the *GS FLX Data Analysis Software* (454 Life Sciences, Roche), whereas sorting of reads into those from each sample based on the barcodes, trimming off the barcodes, removal of sequences with low average quality scores (<20), and removal of sequences with short read lengths (<200 nt for sequences obtained using the GS FLX chemistry, <250 nt for sequences obtained using the GS FLX Titanium chemistry) were done using the *Pyrosequencing Pipeline* of the Ribosomal Database Project (RDP) II (Cole et al., 2007, 2009). The remaining sequences were checked for potential chimeras using the *chimera.slayer* tool incorporated into the *mothur* v1.13.0 package (Schloss et al., 2009).

Metagenomic and metatranscriptomic reads generated by pyrosequencing

In **Study 4**, processing of the pyrosequencing output and filtering of reads based on signal quality were done using the *GS FLX Data Analysis Software* (454 Life Sciences, Roche) with default parameters. Sequences shorter than 50 nt were left out of the analyses.

3.4.2 Annotation of sequences

3.4.2.1 Taxonomic affiliation of 16S rRNAs

In **Study 1**, the taxonomic affiliation of sequences was performed with a *BLASTN* search (Altschul et al., 1990) against a curated 16S rRNA database made as follows: from the original dataset of the RDP-II database (Cole et al., 2007, 2009), a non-redundant dataset of sequences with known taxonomic affi-

liation was obtained after performing a clustering at 99% of sequence identity using the *cd-hit-est* tool of the *cd-hit* v4.0 program (Li and Godzik, 2006). Best-hit sequences were used to assign a confident taxonomic position to each sequence.

In **Studies 2**, **3**, **4**, and **5**, the taxonomic affiliation of sequences was determined using the *Classifier* tool of the RDP-II (Cole et al., 2007, 2009), with a bootstrap threshold of 70% for 16S rRNA gene amplicons, and a lower one (50%) for 16S rRNA genes and cDNAs identified in metagenomes and metatranscriptomes (see Section 3.4.2.3) because of their typically shorter length.

3.4.2.2 Phylotype definition for 16S rRNAs

Clustering at 97% or 98% of sequence identity was carried out using the *cd-hit-est* tool of the *cd-hit* v4.0 program (**Studies 1** and **2**) (Li and Godzik, 2006) or the *cluster* tool of the *usearch* v5.0 package (**Studies 3**, **4**, and **5**) (Edgar, 2010). Input sequences to *cluster* were previously sorted by decreasing abundance as recommended for 16S rRNA amplicons. The resulting phylotypes were used to study sample composition at the "species" level. The species of the closest isolated strain of phylotypes in **Studies 2** and **5** was determined with a *BLASTN* search (Altschul et al., 1990) against the 16S rRNA genes of bacterial isolates in the RDP-II database (Cole et al., 2007, 2009).

3.4.2.3 rRNA search in metagenomes and metatranscriptomes

We used the Small Subunit rRNA Reference Database (SSUrdb) and the Large Subunit rRNA Reference Database (LSUrdb) described in Urich et al. (2008) to search for genes and cDNAs of the 16S and 23S rRNAs, respectively. To select the correct parameters for the *BLASTN* comparisons, we used SSU rRNA, LSU rRNA, and mRNA test datasets. 1000 SSU rRNA human gut-associated sequences were collected from the environmental division of the NCBI through the EnvDB database (Pignatelli et al., 2009). The same number of LSU and mRNA sequences was collected from Genbank (ftp://ftp.ncbi.nlm.nih.gov/genbank) using regular expressions. Fragments of 100 nucleotides were obtained by randomly sampling out the collected sequences. We compared these datasets with the SSUrdb and the LSUrdb using *BLASTN* (Altschul et al., 1990) with different maximum e-values. This analysis showed that e-value thresholds of 10^{-16} for the SSUrdb and 10^{-4} for the LSUrdb gave the lowest rates of "cross-contamination". Then, metagenomic and metatranscriptomic sequences were

compared with the SSUrdb using *BLASTN* (Altschul et al., 1990). All sequences with positive matches were labelled as 16S rRNAs and used to determine the phylogenetic diversity. The remaining reads in metatranscriptomes were compared with the LSUrdb, all putative 23S rRNAs were discarded, and the remaining fraction was used to infer the functional content.

3.4.2.4 Functional annotation of protein coding genes and mRNAs

Assembly of metagenomes

The reads from each metagenome were assembled using the *runAssembly* tool of the *Newbler* v2.6 package (454 Life Sciences, Roche) with 95% of minimun overlap identity and 50 bp of minimun overlap length.

Mapping of metatranscriptomes to metagenomes

The assembled metagenomes from different samples of the same patient were concatenated. Then, metatranscriptomic reads with no significant similarity with any of the rRNA databases (see Section 3.4.2.3) were aligned to the concatenated metagenomic assembly of the same patient using the *runMapping* tool of the *Newbler* v2.6 package (454 Life Sciences, Roche) with 95% of minimun overlap identity.

Gene finding in metagenomic assemblies

Putative coding regions were identified in the metagenomic assemblies from the coordinates of best hits in a *BLASTX* search (e-value threshold of 10^{-3}) (Altschul et al., 1990) against a subset of the NCBI non-redundant (nr) protein sequence database containing bacterial sequences (ftp://ftp.ncbi.nlm.nih.gov/blast/db). Additional open reading frames were searched using the gene finder program *Glimmer* v3.02 (Salzberg et al., 1998) (http://www.cbcb.umd.edu/software/glimmer).

BLAST-based searches

Putative coding regions previously identified in the metagenomic assemblies were compared with the KEGG GENES database (Kanehisa et al., 2004) (http://www.genome.ad.jp/kegg) using *BLASTX* (Altschul et al., 1990) with an e-value threshold of 10^{-5} to assess the metabolic or regulatory pathways present in the samples. Metatranscriptomic reads that aligned to a genomic region adopted its annotation.

Profile-based searches

Putative coding regions previously identified in the metagenomic assemblies were aligned against the TIGRFAM database of global and fragment models of protein families (Selengut et al., 2007) (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi) using *HMMER* v3.0 (Eddy, 1998) (http://hmmer.janelia.org) with an e-value threshold of 0.1. Metatranscriptomic reads that aligned to a genomic region adopted its annotation.

Metatranscriptomic reads non-annotated as putative mRNAs (see Section 3.4.2.3) were aligned against the Rfam database of RNA families (Griffiths-Jones et al., 2003) (http://rfam.sanger.ac.uk) using *INFERNAL* v1.0.2 (Nawrocki et al., 2009) (http://infernal.janelia.org) with an e-value threshold of 0.1 to identify known small RNAs.

3.4.2.5 Taxonomic affiliation of protein coding genes and mRNAs

The taxonomic assignment of the putative coding regions identified in the metagenomic assemblies from the BLASTX search against the NCBI non-redundant (nr) protein sequence database (see above) was assessed using the *Blast2lca* v0.02 tool developed in our lab (http://github.com/emepyc/Blast2lca). For this assignment, *Blast2lca* calculates the lowest common ancestor (LCA) of each query sequence in a *BLAST* result from its best-hits. It works in a similar way to the well-recognised *MEGAN* tool for exploring the taxonomic content of metagenomic data (Huson et al., 2007). Metatranscriptomic reads that aligned to a genomic region adopted its taxonomic affiliation.

3.4.3 Phylogenetic analysis of 16S rRNAs

In **Study 1**, the representative sequences of phylotypes defined at 98% of sequence identity were aligned using the *align.seqs* tool of the *mothur* v1.13.0 package (Schloss et al., 2009) and the aligned sequences of the Greengenes Core Set as template alignment (DeSantis et al., 2006). The closest template of each sequence was found using 9-mer searching, and the pairwise alignment between the sequences and the templates was made using the Gotoh algorithm. A neighbor-joining tree was obtained from the alignment with the programs *DNADIST* (using the F84 model of nucleotide substitution) and *NEIGHBOR* of the *PHYLIP* v3.5 package (Felsenstein, 1989). The derived tree was used as input for *UniFrac* together with taxa abundance in the different samples

(Lozupone et al., 2006). The *UniFrac* metric measures the difference between two communities in terms of the branch length that is unique to one community or the other. We employed weighted *UniFrac*, which weights the branches based on the relative abundance of a given sequence for each particular community. To compare multiple communities, we used principal coordinates analysis (PCoA) (Quinn and Keough, 2002, p 474).

3.4.4 Statistical analyses

Most of the statistical analyses were carried out using the statistical environment R (R Development Core Team, 2010) (http://www.R-project.org) and the R *vegan* package (Oksanen et al., 2011). Other R packages were used in specific studies. These are cited in the corresponding places below.

3.4.4.1 Exploratory statistical analyses

The Shannon diversity index (Shannon, 1948) and the Chao 1 and Abundance-based Coverage (ACE) estimators of richness (Chao, 1987; Chazdon et al., 1998; Chao et al., 2000) were calculated to assess bacterial diversity and richness in the samples. Rarefaction curves were also computed. In **Study 3**, re-sampling of the samples to identical sequencing depth (the number of sequences of the smallest sample) was done with the *multiple.rarefactions* tool of the *QIIME* v1.3.0 package (Caporaso et al., 2010). It allowed to evaluate the impact of the number of sequences per sample on the estimation of richness, biodiversity, and sample composition.

The similarity between samples according to bacterial taxonomic and functional composition was assessed with correspondence analysis (CA) or detrended correspondence analysis (DCA) (Quinn and Keough, 2002, p 459), and with cluster analysis using Euclidean distances or Bray-Curtis distances depending on the type of data used as input (Quinn and Keough, 2002, p 488).

Analysis of similarities (ANOSIM) based on distance matrices (Quinn and Keough, 2002, p 484) was used to assess differences between groups and sampling sites in **Study 2**, and between individuals in **Studies 3** and **4**.

3.4.4.2 Study-specific statistical analyses

Study 1: Patterns of variation in the bacterial composition
We considered a Bayesian hierarchical model, which is set up in levels or hierarchies. In the first layer, the model assumes a multinomial distribution for the observed counts of sequences in each taxonomic group:

$$\mathbf{Y}_{ik} \sim \text{Multinomial}(n_{ik}, \boldsymbol{\pi}_{ik})$$

where $\mathbf{Y}_{ik} = (Y_{ik1}, \ldots, Y_{ikJ})'$, with Y_{ikj} being the number of sequences in taxon $j = 1, \ldots, J$ found in the sample of type $k = 1, 2$ (faeces or biopsies) from individual $i = 1, \ldots, 9$, n_ik is the total number of sequences in that sample, and $\boldsymbol{\pi}_{ik} = (\pi_{ik1}, \ldots, \pi_{ikJ})'$, where π_{ikj} is the actual (unknown) proportion of taxon j in the community corresponding to sample of type k from individual i. The variability of these proportions is in turn broken down into:

$$\log \pi_{ikj} = \alpha + \lambda_i + \theta_j + \delta_k + \varphi_{ij} + \gamma_{kj} + \varepsilon_{ikj}$$

where α is an overall intercept, λ_i, θ_j, and δ_k are individual, taxon, and sample type main effects, respectively, and φ_{ij} and γ_{kj} are order-two interactions between individuals and taxa, and between sample type and taxa, respectively.

In the Bayesian paradigm, prior information about the parameters (if available) is combined with the information provided by the data (likelihood) via the Bayes theorem. The output is the posterior probability distribution of the parameters, from which we can extract the relevant summaries (mean, median, other quantiles, standard deviation, etc.). It is therefore necessary to assign prior distributions to the parameters. We assigned non-informative prior distributions to all model parameters. In particular, we chose a flat normal prior for the intercept, and the following normal prior distributions for the main effects and the interactions:

$$\begin{aligned}
\lambda_i &\sim \text{Normal}(0, \sigma_\lambda) \\
\theta_j &\sim \text{Normal}(0, \sigma_\theta) \\
\delta_k &\sim \text{Normal}(0, \sigma_\delta) \\
\varphi_{ij} &\sim \text{Normal}(0, \sigma_\varphi) \\
\gamma_{kj} &\sim \text{Normal}(0, \sigma_\gamma) \\
\varepsilon_{ikj} &\sim \text{Normal}(0, \sigma_\varepsilon)
\end{aligned}$$

Note that the standard deviations σ represent the amount of variability of the proportions π_{ikj} (on the log-odds scale) due to the corresponding main

effects and interactions. These (hyper)parameters were in turn assigned prior distributions, namely inverse gamma priors with shape = 0.05 and rate = 0.0005.

The posterior distribution of the parameters does not have a closed analytical expression, so we resorted to Markov chain Monte Carlo (MCMC) simulation methods as implemented in the JAGS package to obtain a sample from the posterior distribution of the parameters (http://mcmc-jags.sourceforge.net). This model allows to estimate the unknown proportions π_{ikj} as well as to decompose their variation into different sources while taking into account the sampling variation due to the different number of sequences in each sample. The estimated proportions π_{ikj} (on their log-odds transformation) are then used for further statistical analyses.

Study 1: Closeness between samples

In order to assess the similarity between samples according to their bacterial composition, we computed Euclidean distances between samples based on their compositions (on the log-odds scale) estimated with the above Bayesian model.

Study 1: Prediction of sample type from composition

We applied linear discriminant analysis (LDA) (Fisher, 1938) and classification and regression trees (CART) (Breiman et al., 1984) to assess whether community composition could characterise sample type.

Study 2: Univariate analysis

To assess differences in the microbial composition between groups (IBS-D/IBS-C and controls), we used taxa abundances in the individual samples and also in the pooled samples obtained by pooling the individual ones by sample type (ascending colon, descending colon, or faeces) within each group. Analogously, for the comparison of the bacterial communities in the three sampling sites, we pooled the individual samples by sample type.

Univariate chi-square tests were applied to assess the homogeneity in the relative abundance of each bacterial taxon in the samples of IBS patients compared with those of controls differentiating by IBS subtype and separately for each sample type (ascending colon, descending colon, and faeces). For many taxa, the assumption of the asymptotic chi-square distribution did not hold, hence we computed Monte-Carlo p-values based on 10^5 replicas. Bonferroni's correction for multiple testing was applied.

Odds ratios (ORs) were calculated to measure the magnitude of the over-representation or underrepresentation of taxa between samples. We calculated ORs using (a/b)/(c/d), where a is the number of sequences assigned to taxon i in sample X, b is the number of sequences assigned to all other taxa in sample X, c is the number of sequences assigned to taxon i in sample Y, and d is the number of sequences assigned to all other taxa in sample Y. An OR greater than one indicates overrepresentation of taxon i in sample X compared with sample Y, whereas an OR lower than one suggests underrepresentation.

We proceeded similarly to compare the bacterial composition between sampling sites.

Study 2: Variation between individuals

Gini coefficients (Gini, 1912) were calculated to assess the homogeneity in the relative abundance of taxa between the individual samples that were pooled in the IBS and control pooled samples. The Gini coefficient measures the evenness of a distribution. It ranges between 0 and 1, with 0 representing complete homogeneity (i.e. all individuals have the same proportion of sequences of a given taxon) and 1 maximum concentration (one individual concentrates all the sequences of a given taxon).

The robustness of the results in the IBS versus control homogeneity tests was assessed by repeating the analysis with 999 pooled samples obtained after randomly labelling the individual samples as IBS patient or control and then aggregating them to build new pooled samples. This procedure allowed us to obtain a distribution of OR values under the null hypothesis of no difference between IBS cases and controls to assess the significance of the differences found between these cohorts. Extreme OR values of the actual pooled samples when compared with those from the replicas were indicative of true association.

Within-subject comparisons were performed to assess the consistency of the differences between sampling sites across individuals.

Study 2: Evaluation of the questionnaires

The homogeneity of the clinical and lifestyle variables in patients and controls was assessed using chi-square tests for non-numerical variables, and t-tests for numerical variables. Canonical correspondence analysis (CCA) (Quinn and Keough, 2002, p 467) was carried out to explore the patterns of variation between the microbial composition of patients and controls adjusting for questionnaire variables.

Study 3: Multivariate time series modelling

From a statistical point of view, the data on daily taxa abundances can be regarded as a multivariate time series. The potential interactions between taxa are expected to be reflected in the correlations between taxa, but also some temporal correlation is expected to be present in the data. In other words, if Y_{it} represents the matrix with the number of sequences of taxon $i = 1, \ldots, n$ found in the sample collected on day $t = 1, \ldots, T$ for a given individual, both rows and columns present correlation structures of different nature. Abundances in a given row are likely to be affected by temporal correlation, whereas values in a specific column may be subject to the correlations generated by the underlying interactions between taxa. To model both correlation structures simultaneously, we applied a Bayesian hierarchical model to the follow-up data for each individual. Our model specification is as follows. Let $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{nt})'$ be the taxonomic distribution of sequences on day t. Our model first assumes that $\mathbf{Y}_t$ follows a multinomial distribution:

$$\mathbf{Y}_t \sim \text{Multinomial}(\boldsymbol{\pi}_t, N_t)$$

where N_t is the total number of sequences on day t and $\boldsymbol{\pi}_t = (\pi_{1t}, \ldots, \pi_{nt})'$, π_{it} being the unknown proportion in which taxon i is present in the community on day t. The proportions π_{it} are in turn decomposed, on the log-odds scale, into:

$$\log\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = \alpha_i + \nu_{it} + \epsilon_{it}$$

where α_i is a taxon-specific intercept that picks up the average relative abundance of taxon i over the $T = 15$ days, and ν_{it} and ϵ_{it} are random effects intended to pick up time structured and unstructured variation, respectively. To this end, we chose a normal prior distribution for ϵ_{it}, and a multivariate random walk of order one for $\boldsymbol{\nu}_t,\ t = 1, \ldots, T$

$$\boldsymbol{\nu}_t \mid \boldsymbol{\nu}_{t-1} \sim \text{MVN}\left(\boldsymbol{\nu}_{t-1}, \Sigma\right)$$

where Σ is the $n \times n$ variance-covariance matrix between taxa abundances. For convenience, we take $\boldsymbol{\nu}_0 = \mathbf{0}_{n\times 1}$. This conditional specification is a particular case of the intrinsic multivariate conditional autoregressive (MCAR) models (Kim et al., 2001; Gelfand and Vounatsou, 2003), for which the full conditional distribution is

$$\boldsymbol{\nu}_t \mid \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{t-1}, \boldsymbol{\nu}_{t+1}, \ldots, \boldsymbol{\nu}_T \sim \begin{cases} \text{MVN}\left(\boldsymbol{\nu}_{t+1}, \Sigma\right) & t = 1 \\ \text{MVN}\left(\frac{\boldsymbol{\nu}_{t-1}+\boldsymbol{\nu}_{t+1}}{2}, \frac{\Sigma}{2}\right) & t = 2, \ldots, T-1 \\ \text{MVN}\left(\boldsymbol{\nu}_{t-1}, \Sigma\right) & t = T \end{cases}$$

that is, $\boldsymbol{\nu}_t$ follows a multivariate normal distribution centred in the average of its temporal neighbours, and a variance-covariance matrix inversely proportional to the number of neighbours. The joint distribution of $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{n1}, \nu_{12}, \ldots, \nu_{n2}, \ldots, \nu_{1T}, \ldots, \nu_{nT})'$ is a zero-mean multivariate normal distribution with precision matrix $\Omega = (D - W) \otimes \Sigma^{-1}$, where W is a $T \times T$ matrix with $W_{tt'} = 1$ if time points t and t' are adjacent and $W_{tt'} = 0$ otherwise, D is a $T \times T$ diagonal matrix with D_{tt} equal to the number of neighbours of time point t (i.e. $D_{11} = D_{TT} = 1$ and $D_{tt} = 2\ \forall t = 2, \ldots, T-1$), and $\otimes$ represents the Kronecker product for matrices. The matrix $D - W$ is singular, which makes this distribution improper. However, with our choice of W and D, Ω satisfies the so-called symmetry condition that ensures propriety of the posterior. In practice, this impropriety is overcome by using the proper full conditionals for $\boldsymbol{\nu}_t$ and imposing n sum-to-zero constraints. See e.g. Banerjee et al. (2004, pp 247-251) for further details.

We fitted our model using Markov chain Monte Carlo (MCMC) simulation techniques as implemented in the *WinBUGS* software (Lunn et al., 2000) and the *R2WinBUGS* package (Sturtz et al., 2005) for the R statistical software (R Development Core Team, 2010) (http://www.R-project.org). We ran two chains with 50000 iterations, discarded the first 10000 as burn-in and kept every 40th to reduce autocorrelation in the chains. Therefore, inference for each parameter is based on a thinned sample of size 2000 from its posterior distribution.

Study 3: Putative interactions between taxa

Graphical Gaussian networks (GGNs) (Schäfer and Strimmer, 2005a,b) have been used to recover gene regulation network structures using gene expression data as input. They aim at predicting interaction networks between genes. Here, however, we apply for the first time GGNs to explore patterns of association between taxa using the partial correlations between their abundance profiles. A strong partial correlation between two species is indicative of some form of association. The estimation of that matrix is tricky because typically it is sparse and has large dimensionality. However, GGNs allow to estimate

efficiently the partial correlation matrix from the variance-covariance matrix. We therefore applied GGNs to the covariance matrix Σ (which measures covariances between taxa abundances on the log-odds scale) obtained with the above Bayesian model to infer a network of potential associations between taxa. This was done using the R package *corpcor* (Schafer et al., 2010). The statistical significance of the estimated partial correlations was calculated using the algorithms proposed by Opgen-Rhein and Strimmer (2007). The idea is to model the partial correlations with a mixture of two components. The first component tries to capture the null partial correlations, whereas the second component intends to pick up the sizeable ones. Opgen-Rhein and Strimmer (2007) suggest a method to estimate the mixture components and also a "local false discovery rate" (LFDR) procedure to assess the statistical significance of each partial correlation. They show that their methods perform very well both in simulations as well as in application to large-scale real expression data in the context of gene association networks. We applied these methods as implemented in the R package *GeneNet* (Schaefer et al., 2009). These analyses were done for each individual separately.

The output from this analysis is a graph, with nodes corresponding to taxa and edges representing a statistically significant partial correlation between taxa (taking as such that the probability, 1-LFDR, for the partial correlation to be different from zero is above 0.95). Graphics were generated with the R package *Rgraphviz* (Gentry et al., 2008).

Study 4: Evaluation of symptom fluctuation

Principal components analysis (PCA) (Quinn and Keough, 2002, p 443) was employed to analyse the variation in symptom presence/absence and symptom intensity between patients and within patients throughout the follow-ups. It was used for the classification of sampling days based on the severity of symptoms.

Study 4: Univariate analysis

Univariate chi-square tests were applied to assess the homogeneity in the relative abundance of each bacterial taxon in samples from days with acute symptoms and days with milder symptoms within each patient. Samples were pooled for the analysis based on their similarity in symptom intensity. For many taxa, the assumption of the asymptotic chi-square distribution did not hold, hence we computed Monte-Carlo p-values based on 10^5 replicas.

Bonferroni's correction for multiple testing was applied.

Odds ratios (ORs) were calculated to measure the magnitude of the overrepresentation or underrepresentation of taxa between samples.

Study 4: LEfSe analysis

The linear discriminant analysis (LDA) effect size (LEfSe) algorithm (Segata et al., 2011) was applied to identify intermediate functional categories in the KEGG PATHWAY and the TIGRFAM hierarchies characterising the differences between patients and between days with severe and mild/moderate symptoms within each patient. For the comparisons between samples of a single patient, samples were pooled based on their similarity in symptom intensity.

4 Results and Discussion

4.1 Bacterial diversity in rectal mucosa and faeces of healthy subjects

4.1.1 Background

Our knowledge of the bacteria living in the human GI tract has mainly been obtained by studying the sequence variability of the 16S rRNA genes in representative samples of GI habitats. Such studies have been carried out on both healthy people and patients affected by different disorders (Zoetendal et al., 2008; Salonen et al., 2010; Seksik, 2010). The purposes were to get a first description of the biodiversity and the spatial and temporal variation in the composition of the GI microbiota, and to assess the implication of imbalances in the GI microbiota in the aetiology of diseases, as well as changes in composition in response to therapeutic treatments.

Many studies on the GI microbiota are based on faecal samples. This is because faeces are easily collected in a non-invasive manner. However, the faecal microbiota may not entirely represent the microbial communities living within the GI tract, which nevertheless seem to be relatively similar in the mucosal fraction along the colon (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005). This might be due to several reasons. The faecal microbiota could be a mixture of luminal microbes and shed or poorly adhered mucosal microbes. Also, inadequate storage of faecal samples can lead to alterations in the composition of the faecal microbiota (Ott et al., 2004). For instance, a delay of several hours between the collection of faecal samples and their adequate storage is quite common. This may affect faecal microbial composition due to its dynamic nature, which depends on growing conditions such as nutrient availability, oxygen concentration, and temperature. These condi-

tions change dramatically after evacuation, potentially leading to alterations in the community composition due to differential bacterial death or growth.

Unlike faeces, biopsies of the colonic mucosa are directly collected from the GI tract and thus could be a more suitable option for the study of microbiota-related gut pathologies or treatments targeting the GI microbiota. Furthermore, biopsies can be extracted under controlled conditions and preserved immediately by cryopreservation. Unfortunately, there are some important drawbacks to using biopsy samples. The main one is that they are collected by an invasive procedure (endoscopy) that can not be used routinely. Moreover, the endoscopic procedure is usually carried out after a bowel cleansing, which can have an impact on the mucosal bacterial community. Finally, biopsy samples of an individual may pose problems to some molecular techniques because there may not be enough microbial material to work with.

Several studies have dealt with differences between faecal and colonic mucosal samples so far (Zoetendal et al., 2002; Ott et al., 2004; Ouwehand et al., 2004; Eckburg et al., 2005; Lepage et al., 2005; Bibiloni et al., 2008; Momozawa et al., 2011). Most of them used fingerprinting techniques and showed differences between the two types of sample, except that by Bibiloni et al. (2008), who found a high similarity between the bacterial profiles obtained from mucosal biopsy, luminal aspirate, and faeces of the same subject. Despite providing a rapid method for the comparison and monitoring of microbial ecosystems, the diversity profiles produced by fingerprinting techniques only recover the most dominant members of the community, and sequencing is still necessary for their identification. Before us, Eckburg et al. (2005) employed massive sequencing of PCR-amplified 16S rRNA genes for the comparison of mucosal and faecal samples of three individuals. They detected differences between the two sample types, but faeces were collected one month after the intestinal biopsies, a lag that might have introduced changes in the composition of the microbiota. After the publication of our study (Durbán et al., 2011), Momozawa et al. (2011) also reported differences between the two sample types using high-throughput sequencing. A common feature of these two studies is that individuals received a bowel cleansing prior the collection of biopsies.

From a clinical point of view, it is necessary to more accurately determine to what extent faecal microbial communities actually represent the communities within the gut. There is a critical question in this respect: how can we assess whether the gut microbiota is involved in the aetiology of a particular disease when the mucosal fraction has not actually been observed? Or put it in other

terms, how reliable are the results obtained based only on faeces? A first step is to assess how well faecal microbiota represents the mucosa-associated one. To this end, it is essential that both samples are collected at the same time. Also, the bowel cleansing prior to colonoscopy may introduce perturbations in the composition of the mucosal community (related to this, Bibiloni et al. (2008) attributed the high similarity they found between the bacterial profiles of biopsies, luminal aspirates, and faeces to contamination of biopsies by the fluid that pools in the bowel after cleansing). We addressed both issues in this study (see Section 3.2.2).

Our objectives were to analyse the variability in the composition of gut bacterial communities between healthy individuals, to analyse the within-subject variability in the bacterial composition of faeces and colonic biopsies, and to measure the extent to which the bacterial composition of faeces serves as a predictor of the bacterial composition of colonic biopsies. To this end, PCR-amplified 16S rRNA gene libraries were obtained from rectal biopsies and faeces of nine healthy volunteers.

4.1.2 Results

We sequenced around 740 clones for each of the eighteen clone libraries of 16S rRNA genes. Trimmed sequences had an average length of 710 nt. Sequences were classified into phylotypes defined at 98% of sequence identity.

4.1.2.1 Dataset coverage and bacterial diversity

To determine the fraction of operational taxonomic units (OTUs, considering as such any of the extant taxonomic units under study) that were recovered in the samples, we carried out a rarefaction analysis (**Figure 4.1**). Rarefaction curves were obtained by plotting the number of observed OTUs against the number of cloned sequences. At the family level, curves reached or nearly reached a plateau for most samples, whereas for curves calculated using phylotypes, the upward phase was ongoing. This means that we observed most of the families in most samples, but also that quite a few phylotypes were missed. The rarefaction curves also showed that bacterial communities were usually less diverse in the faecal samples than in the respective biopsies.

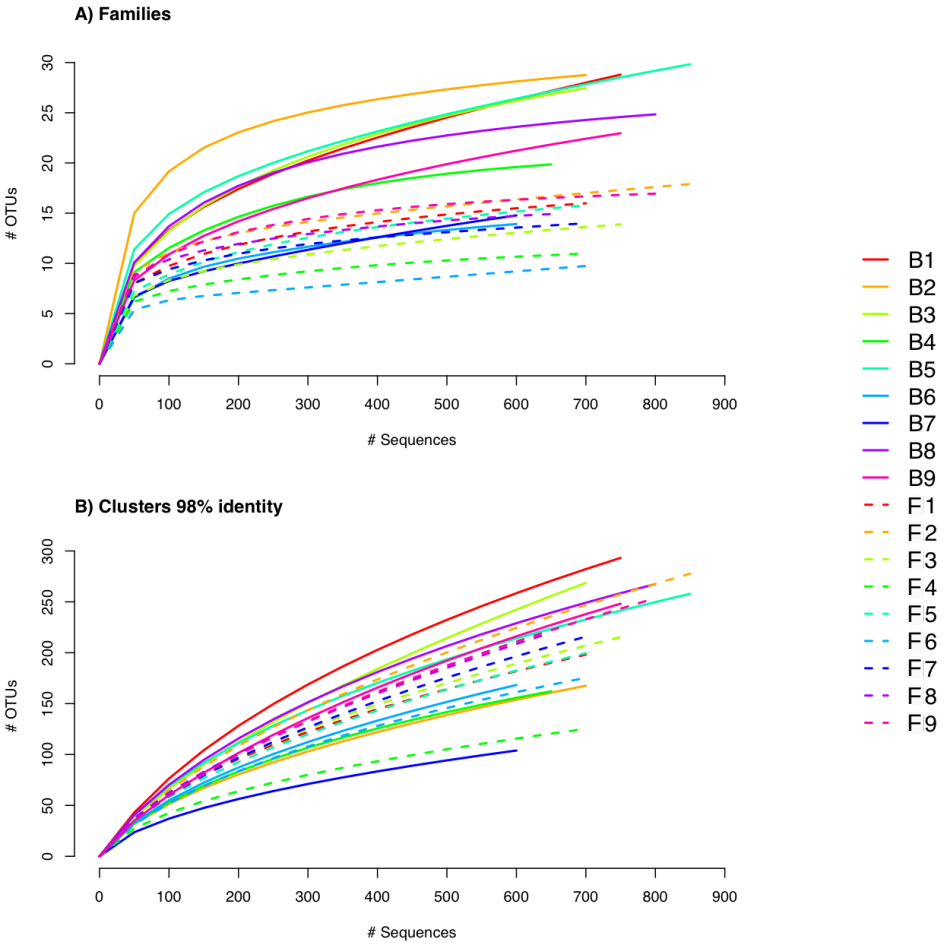


Figure 4.1: Rarefaction curves for each sample calculated at the levels of family (A) and phylotype at 98% of sequence identity (B). B: rectal biopsies; F: faeces.

We employed the Chao1 richness estimator to estimate the number of families and phylotypes in the sampled communities (**Table 4.1**). In agreement with the rarefaction curves, the comparison of the observed and the estimated number of phylotypes indicated substantial numbers of unseen phylotypes in the samples, which could be detected only after sequencing many more clones (on average, 470 phylotypes were estimated, while 215 were detected). This can be put down to the fact that many phylotypes appear at very low fre-

quencies. The Shannon biodiversity index (H), that correlates positively with species richness and evenness, was calculated at both family and phylotype levels (**Table 4.1**). Overall, Chao1 estimates and Shannon indices indicated great diversity of the intestinal bacterial communities. Furthermore, in most of the paired samples, the mucosa-associated community was more diverse than the faecal one.

Table 4.1: Observed richness, Chao1 richness estimator (and standard error), and Shannon biodiversity index (H) calculated for each sample at the levels of family and phylotype at 98% of sequence identity. B: rectal biopsies; F: faeces; nc: not computable.

Volunteer	Sample	No. reads	Families			Clusters 98% identity		
			Obs.	Chao1	Shannon H	Obs.	Chao1	Shannon H
1	B	763	29	51 (33.41)	1.91	296	545 (55.79)	5.24
	F	705	16	17 (2.29)	1.69	199	406 (59.84)	4.64
2	B	742	29	31 (3.49)	2.58	173	345 (53.44)	4.17
	F	867	18	23 (17.14)	1.92	281	680 (96.40)	4.77
3	B	751	28	33 (5.92)	1.71	281	746 (108.90)	4.87
	F	772	14	17 (11.66)	1.41	219	491 (75.53)	4.63
4	B	681	20	21 (2.29)	1.94	166	288 (38.92)	4.18
	F	708	11	12 (3.74)	1.26	126	242 (44.53)	3.74
5	B	863	30	48 (28.64)	1.99	260	519 (66.49)	4.97
	F	740	16	18 (5.29)	1.67	206	400 (52.71)	4.52
6	B	617	14	15 (2.29)	1.48	171	365 (61.23)	4.27
	F	749	10	16 (nc)	1.01	182	346 (49.22)	4.14
7	B	623	15	20 (10.17)	1.37	106	192 (34.65)	3.36
	F	701	14	17 (nc)	1.74	216	449 (60.83)	4.54
8	B	832	25	27 (3.49)	1.81	273	482 (49.90)	5.03
	F	678	15	15 (1.31)	1.86	226	679 (128.31)	4.69
9	B	754	23	30 (10.27)	1.76	249	501 (59.69)	4.58
	F	822	17	18 (3.74)	1.56	258	663 (101.23)	4.56

4.1.2.2 Bacterial taxonomic composition

The distribution of the 16S rRNA gene sequences at phylum, class, and family levels is shown in **Figure 4.2**. Most sequences were assigned to the Firmicutes and Bacteroidetes phyla, which have repeatedly been described as major and functionally significant components of the human intestinal microbiota. Proteobacteria was the third most abundant phylum and its presence was lower in faeces than in rectal biopsies. Low prevalence of other phyla was also found in the biopsy samples, such as Actinobacteria, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Tenericutes, and Verrucomicrobia, which were lower or even absent in faecal samples. The relatively low abundance of Actinobacteria could be

a result of an insufficiently rigorous cell lysis procedure (this phylum has been found as a major constituent of the GI microbiota by using other molecular approaches) (Zoetendal et al., 2008).

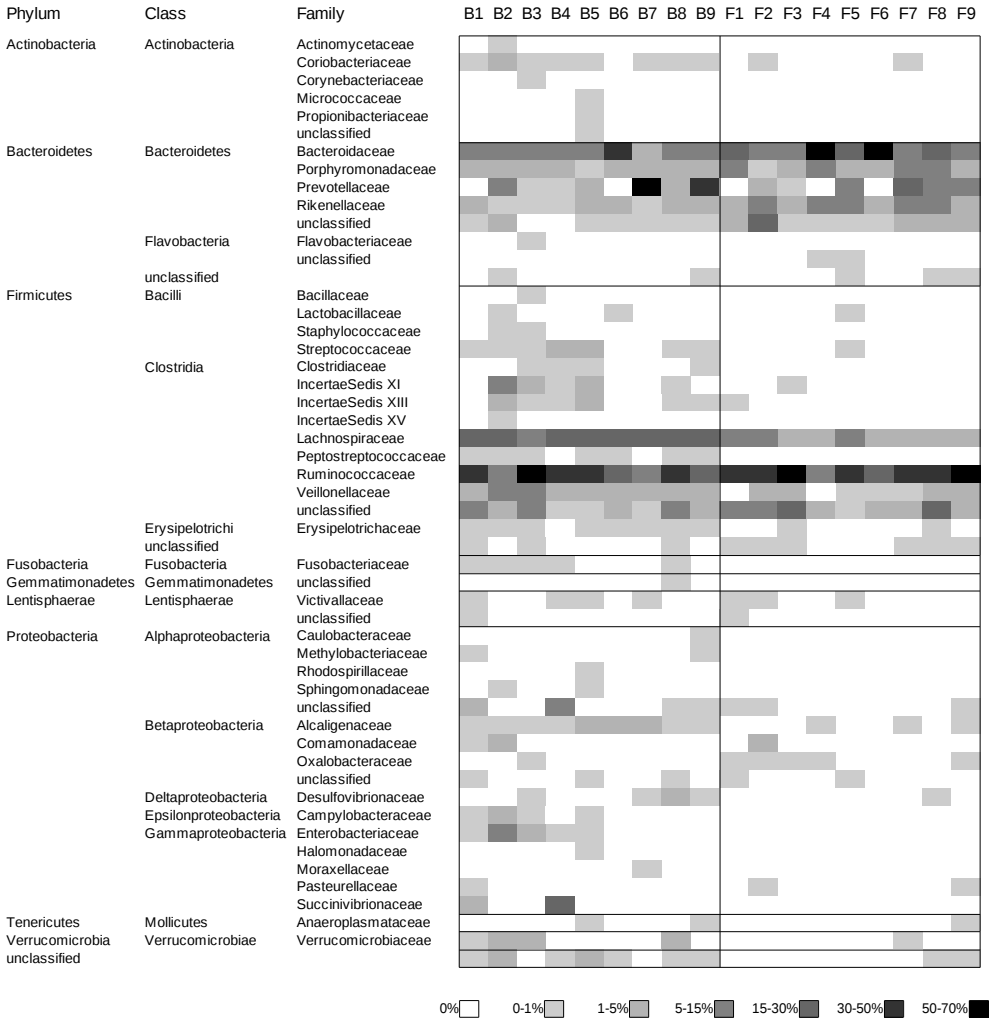


Figure 4.2: Percentage of sequences (grouped into six intervals) at phylum, class, and family levels in rectal biopsies (B) and faeces (F) of volunteers 1-9.

The Firmicutes phylum covered 59.4% of the total number of sequences, and 67.9% of the phylotypes. Most (96.9%) of the Firmicutes sequences belonged to

the Clostridiales order, being Ruminococcaceae and Lachnospiraceae the most abundant families. Lachnospiraceae was less abundant in faeces than in rectal biopsies, while no trend was observed for Ruminococcaceae. The Bacteroidetes phylum included 36.1% of the sequences and 26.5% of the phylotypes. These were almost exclusively members of the Bacteroidales order (99.3%). Large variation between subjects was observed in the relative abundance of the families within this order. The counts for Rikenellaceae were higher in faeces than in rectal biopsies, and the same occurred in general for Bacteroidaceae. All Proteobacteria classes were detected, being Alpha-, Beta-, and Gammaproteobacteria the most abundant. There was large variability between samples in the abundance of bacteria within Proteobacteria. Betaproteobacteria was the only class found in all volunteers, at least in biopsies.

Sample composition was also studied at the species level by working with phylotypes defined at 98% of sequence identity (**Figure 4.3**). A remarkable portion of the species within each volunteer was sample type-specific. An average of 17% of the phylotypes detected in each subject was found in both faeces and rectal biopsy, whereas an average of 52% of the sequences belonged to phylotypes shared between the two types of sample. Species shared between two paired samples also differed in their relative abundance. Variation was found between families in these quantitative differences. For instance, within Lachnospiraceae, most species were not shared between two paired samples, and usually those that were had a similar relative abundance in biopsies and faeces. In contrast, many Ruminococcaceae species found in biopsies were also detected in faeces, but their relative abundance was quite different in the two types of sample.

4.1.2.3 Variation between subjects and sampling sites

DCA showed a great deal of variation in the composition of the communities (**Figure 4.4**). At coarse taxonomic levels, such as phylum, no pattern was observed. However, when using intermediate levels such as family or genus, DCA plots separated faecal and biopsy samples. At the phylotype level, a new pattern emerged showing faecal and biopsy samples from the same subject to be closer to each other. We also used UniFrac to compare sample composition at this same level taking into account phylogenetic distances between phylotypes besides their abundance. UniFrac coupled with PCoA showed some clustering by sample type similar to the DCA plots at genus or family level (**Figure 4.5**).

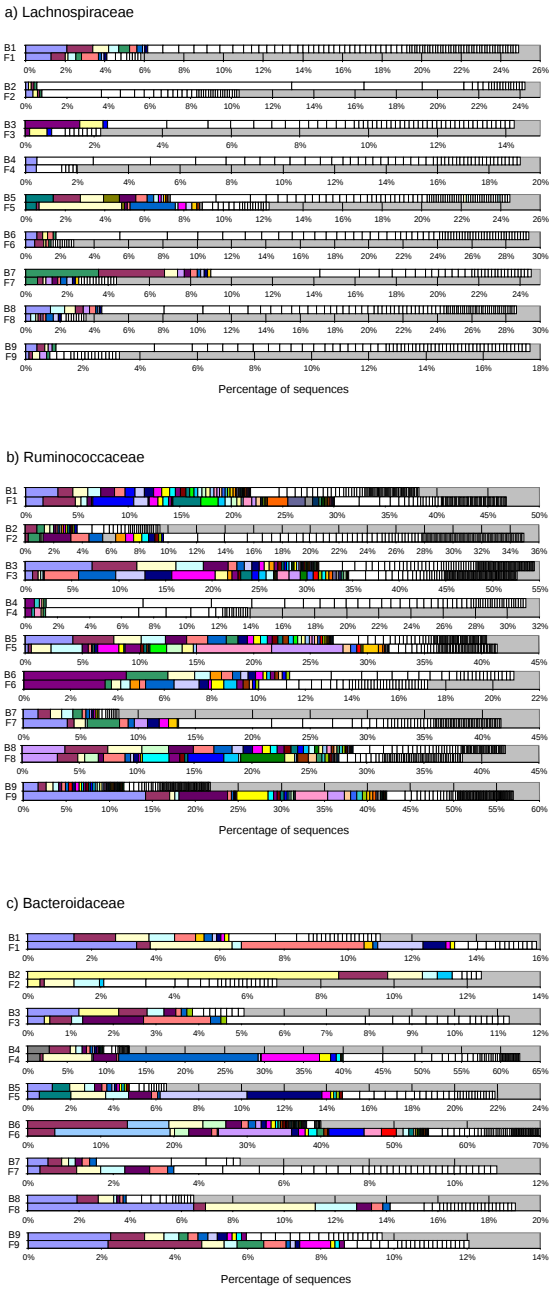


Figure 4.3: Shared phylotypes between rectal mucosa (B) and faeces (F) within the Lachnospiraceae (a), Ruminococcaceae (b), and Bacteroidaceae (c) families. In each individual, shared phylotypes are in the same colour in both sample types, while phylotypes found only in biopsies or in faeces are in white. Every rectangle represents a phylotype and its size corresponds to its relative abundance, expressed as percentage of the total number of sequences in the sample.

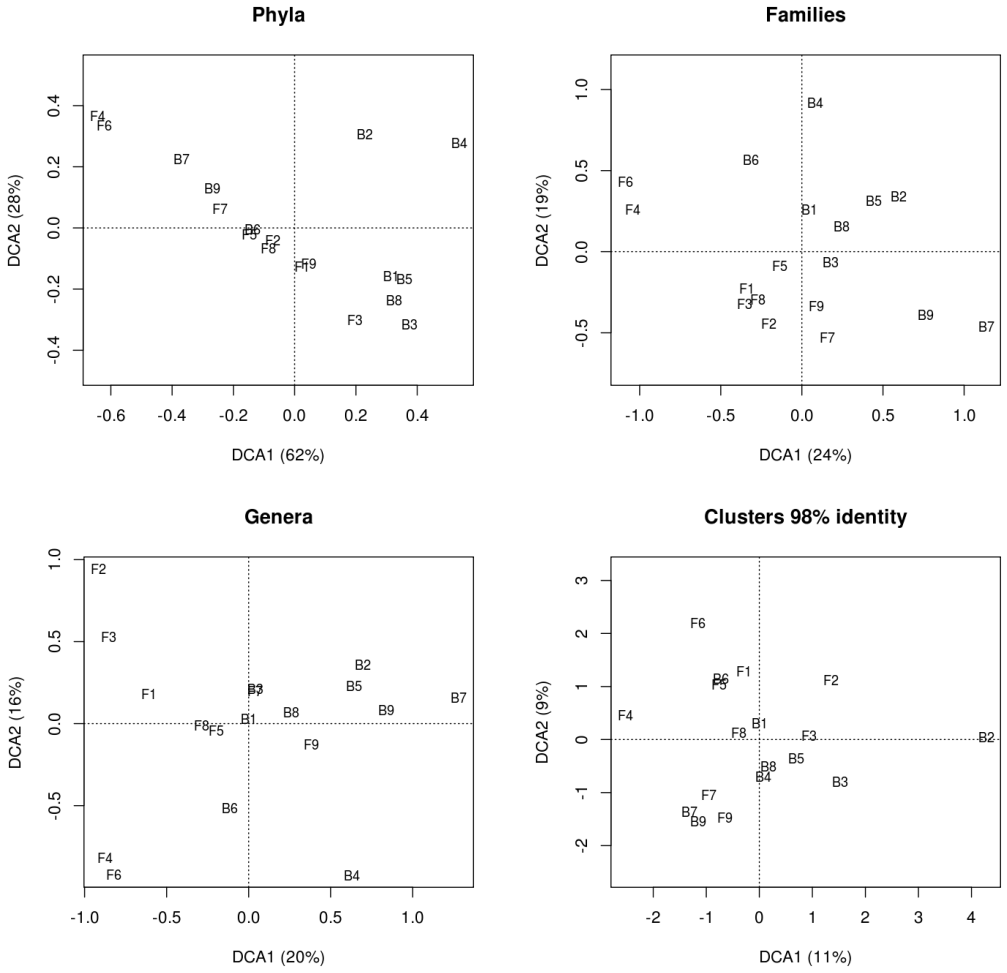


Figure 4.4: DCA at phylum, family, genus, and phylotype levels. Percentages correspond to the fraction of inertia explained by each axis. B: rectal biopsies; F: faeces.

The Bayesian model confirmed the high level of variation in the community composition between samples. It also revealed that the variation in the proportions π_{ikj} characterising the community composition was mainly due to differences between individuals (45%) and between sample types (45%), and, to a lesser extent, to differences between taxa (5%). These results were consistent across taxonomic levels.

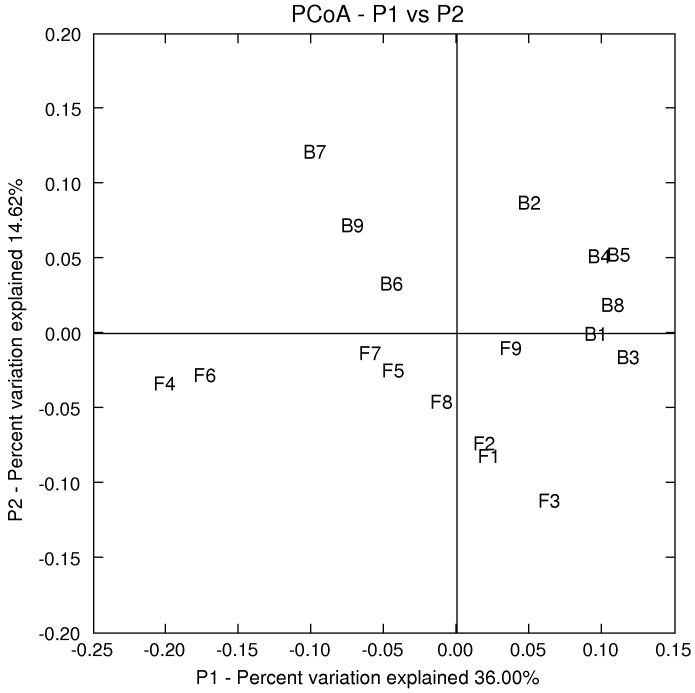


Figure 4.5: PCoA based on the UniFrac metric at the phylotype level. B: rectal biopsies; F: faeces.

A heatmap of the posterior medians of the π_{ikj} parameters (i.e., the median of the sample from the posterior distribution for each parameter provided by the MCMC methods) at the family level highlighted, as stated above, that most communities were dominated by species from just a few taxa, though there was a low prevalence of many other families too (**Figure 4.6**). We computed Euclidean distances between community distributions (on the log-odds scale) estimated with the Bayesian model to assess similarities between samples. At the family level, a hierarchical cluster based on this distance matrix revealed that the estimated community distributions grouped samples by sample type. Each column in **Table 4.2** lists all samples ranked by increasing distance to the faecal sample that appears in the column header. In most instances, we can see that for a given sample the closest ones were those of the same type, as indicated by the dendrogram in **Figure 4.6**, i.e., the closest samples to a given faecal sample were usually other faecal samples. For individuals six,

seven, and nine, the closest biopsy sample to their faecal sample turned out to be their own paired biopsy sample. These are three of the five individuals that provided both samples on the same day. It should be noted that biopsy six was the closest biopsy to seven out of the nine faecal samples, and was the second closest to the remaining two. This is partly due to the relative low diversity in biopsy six that makes its bacterial distribution similar to those found in faeces.

Table 4.2: Sample ordering based on the Euclidean distance among the logit of the taxonomic distributions $\boldsymbol{\pi}_{ik}$ at the family level. In each column, samples are ordered by increasing distance to the faecal sample in the column header. Highlighted is the paired biopsy of each faecal sample. B: rectal biopsies; F: faeces.

Order	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	F1	F2	F3	F4	F5	F6	F7	F8	F9
2	F4	F8	F8	F1	F8	F4	F8	F7	F8
3	F8	F1	F7	F3	F9	F3	F9	F9	F7
4	F3	F9	F4	F7	F7	F7	F3	F3	F5
5	F9	F3	F1	F8	F1	F8	F5	F5	F1
6	F5	F5	F9	F6	F2	F1	F4	F1	F3
7	F2	F7	F2	F9	F3	F9	F2	F2	F2
8	F7	B6	F5	F5	B6	F5	F1	F4	F4
9	B6	F4	F6	B6	F4	**B6**	**B7**	B6	**B9**
10	F6	B9	B6	F2	B9	F2	F6	B7	B6
11	**B1**	B8	B7	B7	B7	B7	B6	B9	B7
12	B9	B7	B9	B9	B8	B9	B9	**B8**	F6
13	B8	B3	**B3**	**B4**	B3	B8	B8	F6	B8
14	B7	B1	B8	B8	**B5**	B3	B3	B3	B1
15	B3	F6	B1	B1	F6	B1	B1	B1	B3
16	B4	**B2**	B4	B3	B1	B4	B4	B4	B4
17	B5	B4	B5	B5	B4	B5	B5	B5	B5
18	B2	B5	B2	B2	B2	B2	B2	B2	B2

Finally, LDA and CART did not prove useful to discriminate sample type by community composition, probably due to the large variability observed and the relatively small sample size. For LDA, we tested multivariate normality and homokedasticity of the covariance matrix of the logit π_{ikj} (the output of the Bayesian model) and did not find significant departures from the null, though this again may be a consequence of the relatively reduced sample size. CART, in contrast, is a non-parametric method and more robust in general (less sensitive to outliers, invariant to monotonic transformations of the variables, etc.), so it should be less affected by the sample size.

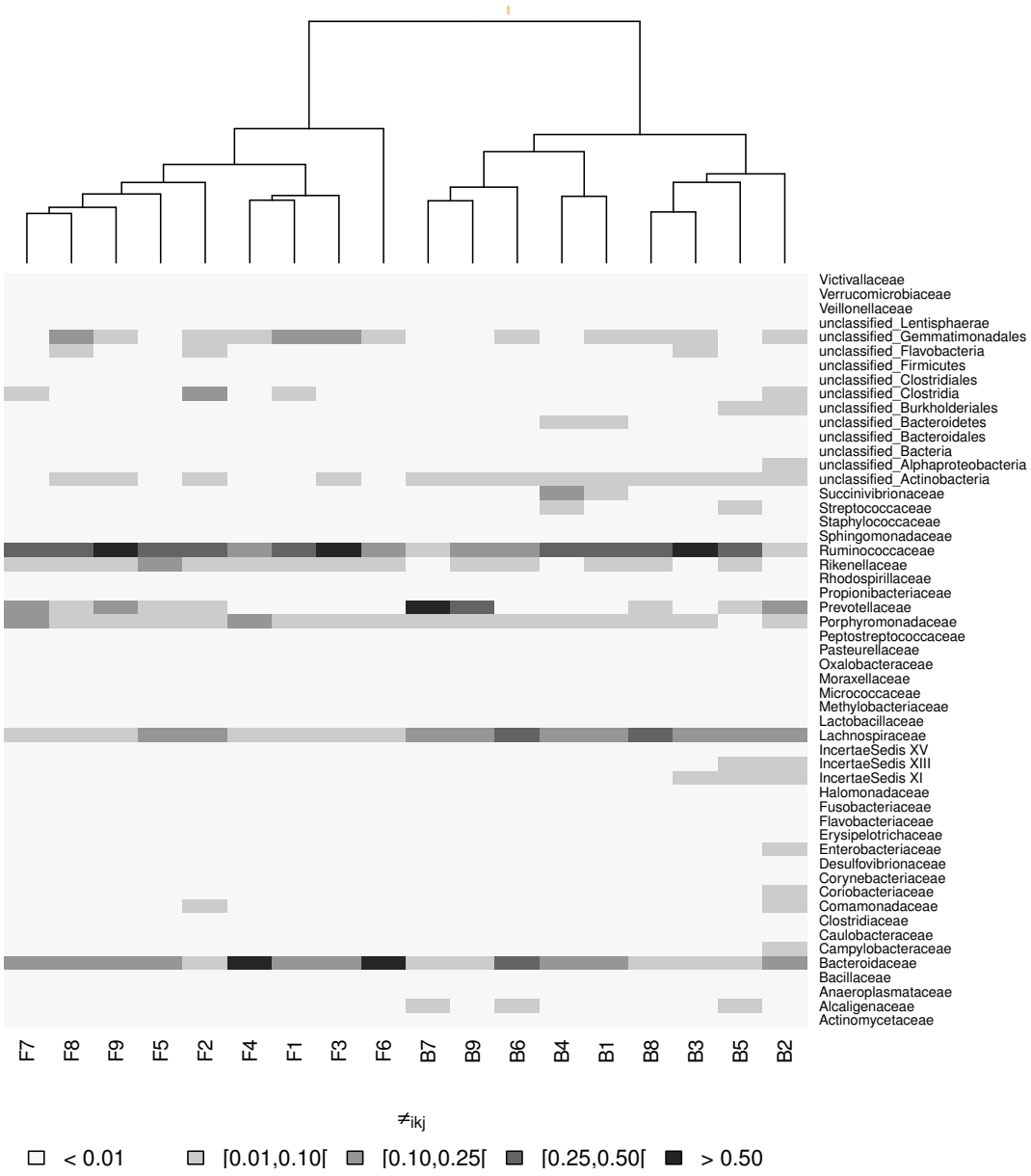


Figure 4.6: Heatmap at the family level of the posterior medians of π_{ikj} grouped into five intervals. On top, hierarchical cluster based on the Euclidean distances of the estimated distributions $\boldsymbol{\pi}_{ik} = (\pi_{ik1}, \ldots, \pi_{ikJ})'$ on the log-odds scale. B: rectal biopsies; F: faeces.

4.1.3 Discussion

Most of the studies carried out until now on gut metagenomics do not take into consideration that a faecal sample contains a microbial composition that not necessarily can be taken as a good predictor of a corresponding intestinal one. Here, we address the analysis of equivalences and/or correspondences between faecal and rectal mucosal samples from nine healthy individuals. Extrapolation of our results to other sites in the intestine should be made with caution as the microbial composition may vary along the gut.

Our work confirmed the findings from previous studies that suggested that faecal and colonic mucosal microbial diversity from the same individual are not similar (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005). However, our approach has several differences over previous studies. Firstly, we attempted to provide this diversity comparison from faecal and mucosal samples collected as close as possible in time. Secondly, we did not carry out a bowel cleansing prior to colonoscopy to avoid the potential disturbance of the mucosal microbiota associated with this procedure. This opens the possibility for rectal biopsies to contain bacteria from faeces loosely sticking to the mucus but not being actually part of the mucosal microbiota. Given the differences that we found in the microbiota between the two types of sample, we think this can hardly be a generalised situation, though the possibility can not be ruled out completely. Nevertheless, biopsies were taken in the absence of macroscopic faeces, and it is unlikely that biopsies were contaminated with faecal material because stools usually have hard consistency and are not attached to the mucosa in the rectum, where they are formed to be expelled outside the body and do not adhere to the mucosa because nothing is absorbed nor secreted there, unlike in other sections of the intestine. Finally, the results presented here have been generated through sequence analysis of clone libraries, which enabled the identification of microorganisms.

Overall, we found that two phyla, Firmicutes and Bacteroidetes, dominated those communities, accounting for nearly 95% of all sequences. However, we also observed large between-sample variability in community composition at nearly all taxonomic levels. At the phylotype level in particular, the majority of phylotypes detected were sample-specific, showing that each individual carried a particular combination of bacterial lineages, as previously reported (Eckburg et al., 2005; Tap et al., 2009; Turnbaugh et al., 2009a). Strong within-subject variability was also found in the faeces-biopsy paired samples.

Our results suggest that the community composition in faecal samples is not highly representative of the microbiota in the rectum. In fact, at family and genus levels, taxa distributions grouped samples by sample type rather than by individual, even for those sample pairs collected the same day. Evaluating the closeness between samples based on distances between their estimated compositions (on the log-odds scale), we found that any faecal sample is more similar to any other faecal sample than to a rectal biopsy sample. We also found that the closest biopsy sample to the faeces of an individual was his own paired biopsy sample in three of the five individuals that provided both samples on the same day, a finding that can not be considered as conclusive given the (statistically) small sample size, especially considering that one of the biopsies was very similar to all faecal samples. These results confirm that the intestinal microbiota is an extremely complex community, the richness and diversity of which seems to be underrepresented in faecal samples. However, it has yet to be assessed whether this impoverishment is because not all species in the intestine are susceptible to ending up in faeces or whether it is a consequence of the impact of the sudden change in the growing conditions (temperature, oxygen, nutrient availability, etc.) on leaving the body. Also, the biopsy samples were frozen immediately after collection, whereas faecal samples were not. It should be considered the possibility that this might have had an impact on reducing the biodiversity found in faeces compared with biopsies.

Faeces will continue to be used in the study of the human gut microbiota because they are easier to collect than intestinal samples and current work with biopsies is limited to certain methodologies due to the quantity of material that can be obtained from them. Actually, each sample type may provide a distinct and complementary picture of the diversity and ecology found in the human gut microbiota. However, since bacterial diversity in the colonic mucosa is underrepresented in faeces, we think it is important to caution researchers about making inference of the intestinal mucosal microbiota, or even about the entire GI one, from that found in faeces, especially when dealing with microbiota-related pathologies.

This study has been published in Microbial Ecology (Durbán et al., 2011).

4.2 Structural alterations of colonic mucosal and faecal bacteria in the irritable bowel syndrome

4.2.1 Background

Research about the aetiology and potential treatments of IBS is relevant because it is the most common functional GI disorder in Western countries, with a profound impact on patients' quality of life and high economic costs derived from it (Quigley et al., 2006). The symptoms of IBS are variable and include chronic abdominal pain, discomfort, or bloating that is relieved with defaecation and/or is associated with altered bowel habits (Longstreth et al., 2006). The cause of IBS is thought to be multifactorial but remains poorly understood. Some of the major aetiological factors that have been proposed are visceral hypersensitivity, abnormal gut motility and secretion, autonomic nervous system dysfunction, and low-grade inflammation in the intestinal mucosa (Aerssens et al., 2008; Karantanos et al., 2010). Psychosocial factors are also known to play an important role in the development and persistence of symptoms in IBS (Drossman, 1999a; Lea and Whorwell, 2003; Hood et al., 2008).

In addition, several observations point to alterations at the level of the GI microbiota in subjects suffering from IBS. First, the onset of IBS frequently follows an acute episode of infectious enteritis, which is in turn the strongest risk factor for developing IBS (Spiller and Garsed, 2009). Second, clinical trials targeting the microbiota (like antibiotics and probiotics) seem to alleviate IBS symptoms (Moayyedi et al., 2010; Basseri et al., 2011). Finally, recent molecular studies suggest an altered GI microbiota in patients, with specific features depending on the predominant bowel pattern (reviewed by Salonen et al. (2010)). Some of the findings in IBS patients compared with controls are large temporal instability and inter-subject variation in the faecal microbiota and altered abundance of specific taxa. However, the current available data do not reveal pronounced IBS-related deviations in the microbial composition, and no consensus has been reached regarding the association between specific bacteria and IBS.

Most molecular studies on IBS carried out so far have used faecal samples. Nevertheless, it is important to study the different compartments where the gut microbiota can be found when assessing its role in the aetiology of a

particular disease. This is because each intestinal habitat contains a specific microbial community (Section 4.1) (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005) and therefore can offer a distinct and complementary picture of the microbiota and its relationship with the host. Specifically, mucosal communities may be more relevant to the disorder because of their closer association with the host epithelium and thus their role in the regulation of intestinal immunity and inflammation.

The main objective of this study was to explore the potential alterations of the intestinal bacteria in IBS in several compartments within the GI tract and in faeces. To this end, mucosal biopsies of the ascending and the descending colon and faecal samples were obtained from sixteen IBS patients and nine healthy controls, and the entire bacterial communities were analysed through sequencing of PCR-amplified 16S rRNA genes. A secondary goal was to compare the bacterial composition between the three sampled sites.

4.2.2 Results

We obtained ∼268000 non-chimeric sequences with an average length of 235 nt. The bacterial composition for the group-type pooled samples is shown in **Figure 4.7**. Most sequences belonged to the Bacteroidetes, Firmicutes, and Proteobacteria phyla. Low prevalence of Actinobacteria, Fusobacteria, and other phyla was also found.

Figure 4.8 shows the cluster analyses based on the community composition at the genus level. The dendrogram of the group-type pooled samples (**Figure 4.8, top**) revealed larger differences in the microbial composition between biopsies and faeces than between IBS patients and controls. Biopsies, however, clustered together by group instead of site. These findings were reproduced in the dendrogram of the individual samples (**Figure 4.8, bottom**), where healthy controls and IBS patients appeared mixed within clusters. Samples from the same IBS-subtype did not cluster either. Both mucosal samples from the same individual clustered together at short distances in many cases. However, faecal samples formed several clusters far from the respective biopsies. DCA confirmed the previous descriptive analysis (**Figure 4.9**). The first DCA axis separated faeces from biopsies. The second axis separated IBS patients from controls in the pooled samples, though this pattern was not clear for the individual samples. Similarly, ANOSIM found significant differences in the global composition between sampling sites (ANOSIM R=0.158, p=0.001) but not between IBS patients and controls (ANOSIM R=-0.005, p=0.483).

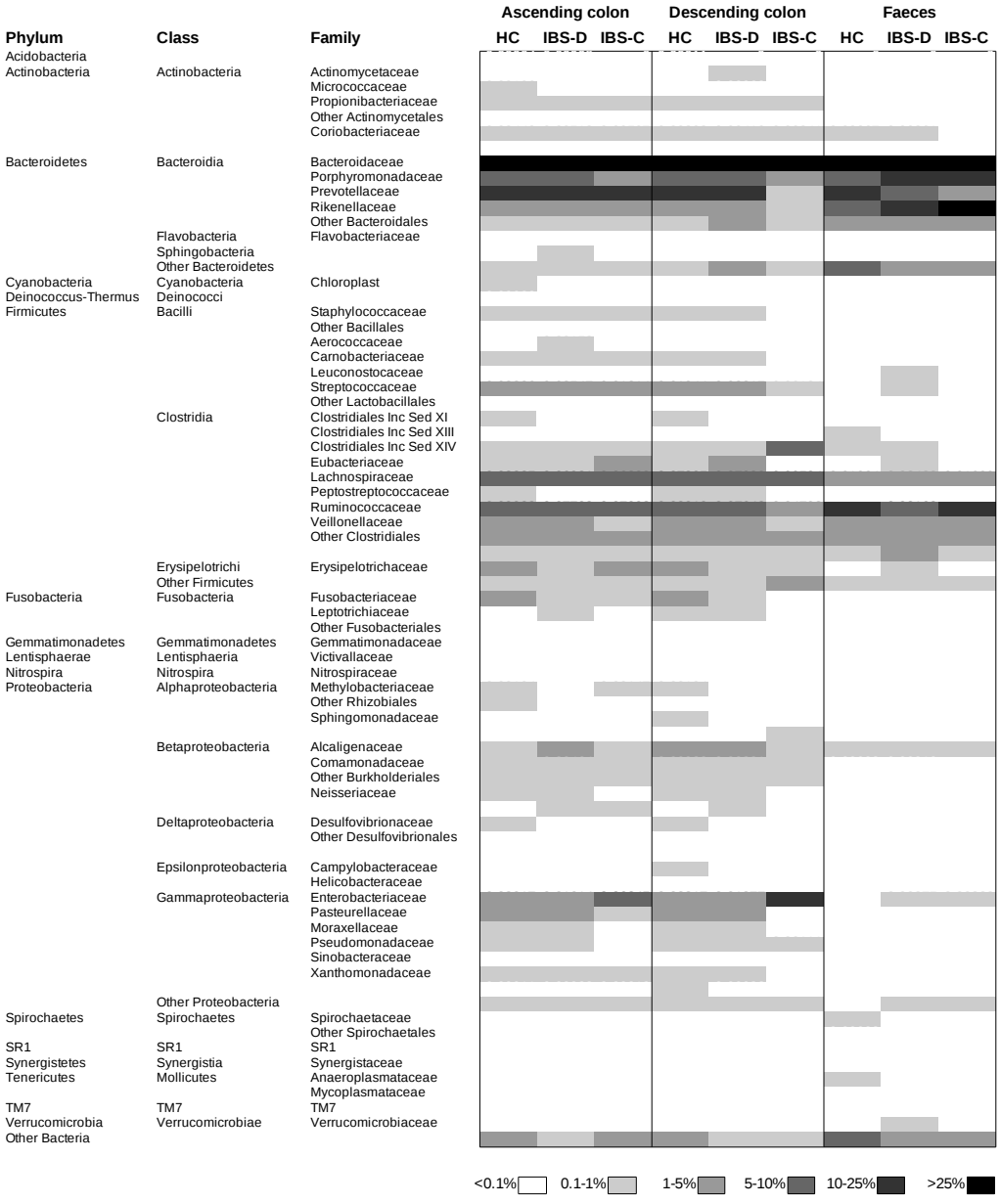


Figure 4.7: Percentage of sequences (grouped into six intervals) at phylum, class, and family levels belonging to the pooled samples of healthy controls (HC), diarrhoea-predominant IBS patients (IBS-D), and constipation-predominant IBS patients (IBS-C).

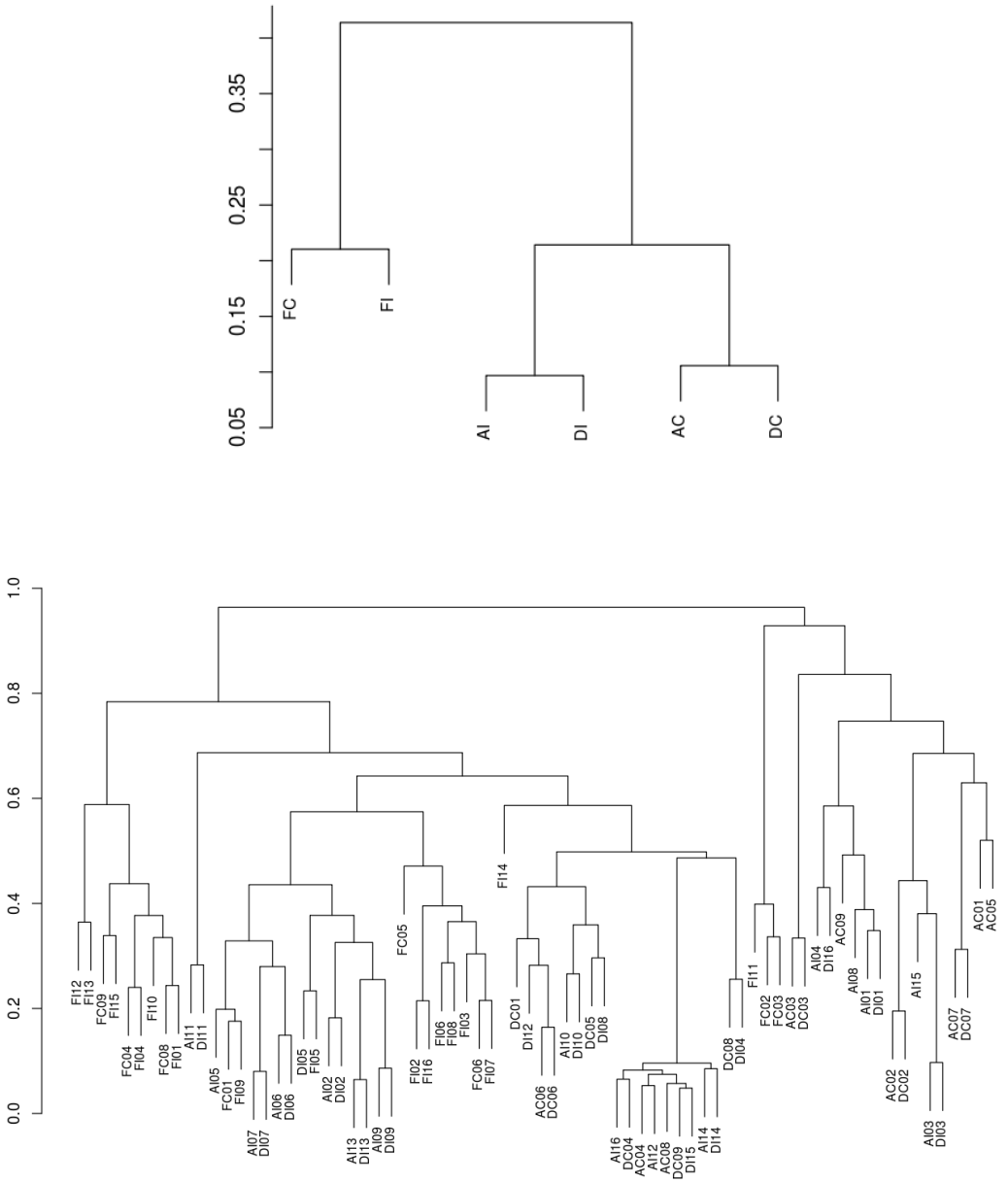


Figure 4.8: Hierarchical cluster of the group-type pooled samples (top) and the individual samples (bottom) based on the Bray-Curtis distances between the observed distributions of genera. AC, DC, and FC: ascending colon, descending colon, and faeces of healthy controls; AI, DI, and FI: ascending colon, descending colon, and faeces of IBS patients. Patients with constipation-predominant IBS: I07, I10, I12; patients with diarrhoea-predominant IBS: the rest.

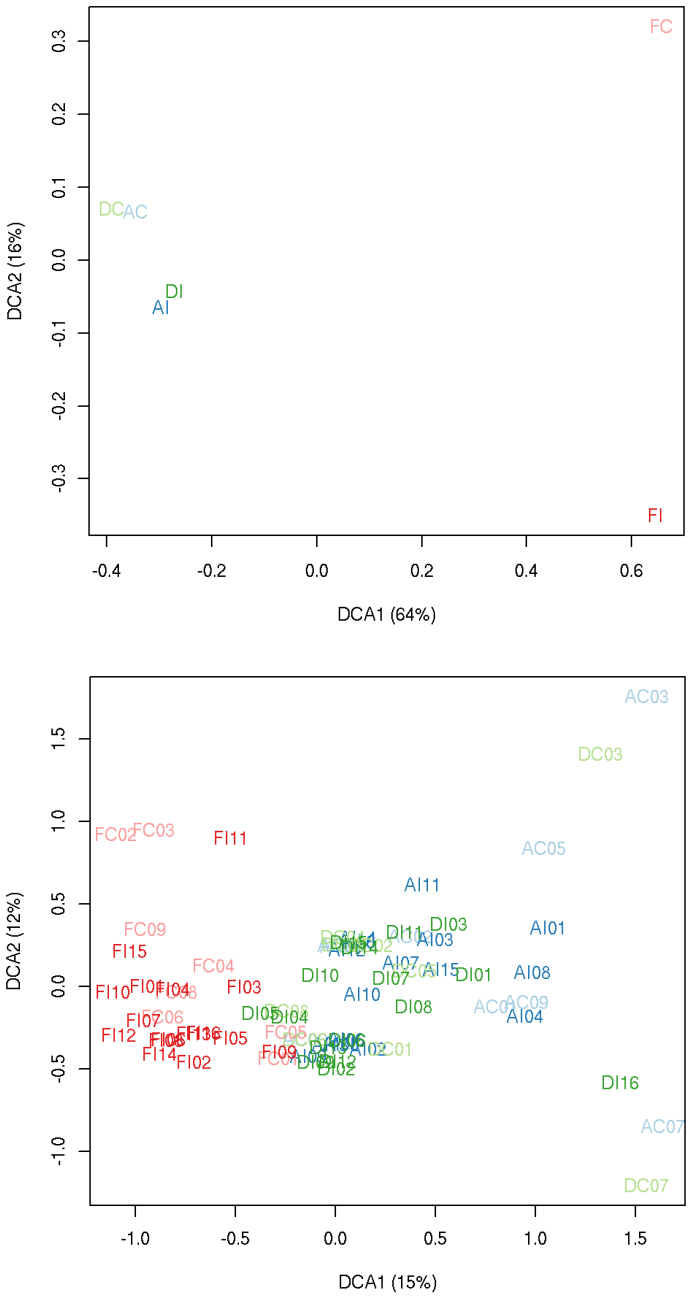


Figure 4.9: DCA at the genus level of the group-type pooled samples (top) and the individual samples (bottom). Percentages correspond to the fraction of inertia explained by each axis. AC, DC, and FC: ascending colon, descending colon, and faeces of healthy controls; AI, DI, and FI: ascending colon, descending colon, and faeces of IBS patients. Constipation-predominant IBS: I07, I10, I12; diarrhoea-predominant IBS: the rest.

4.2.2.1 Comparison of IBS patients and controls

The Shannon diversity indices (H) were on average lower in IBS cases than in healthy controls in the three intestinal compartments (**Table 4.3**). The Chao1 and ACE estimators of total richness were quite similar and indicated substantial numbers of unseen phylotypes, while the estimated numbers of genera were closer to the observed ones. The richness estimators were on average lower in IBS patients than in controls (**Table 4.3**).

Table 4.3: Average values (and standard deviations) of the Shannon diversity index (H) and the Chao1 and ACE richness estimators of the individual samples within the control (HC), IBS-diarrhoea (IBS-D), and IBS-constipation (IBS-C) groups at each sampling site (A, D, and F: ascending colon, descending colon, and faeces) at the levels of genus and phylotype. n: number of subjects.

				Genus			Phylotype 97% identity		
Site	Group	n	No. reads	Shannon H	Chao1	ACE	Shannon H	Chao1	ACE
A	HC	9	27277	2.87 (0.38)	96 (23.54)	96 (20.48)	4.74 (0.45)	671 (509.14)	684 (488.13)
	IBS-D	13	36672	2.43 (0.64)	85 (44.51)	81 (24.63)	4.47 (0.66)	597 (348.03)	615 (332.74)
	IBS-C	3	7494	2.21 (0.74)	81 (35.85)	88 (47.25)	4.61 (1.05)	658 (510.44)	669 (513.33)
D	HC	9	26750	2.71 (0.45)	100 (29.13)	98 (29.15)	4.84 (0.64)	765 (367.16)	785 (374.13)
	IBS-D	13	43896	2.38 (0.56)	89 (30.70)	85 (25.37)	4.59 (0.62)	687 (396.40)	699 (377.17)
	IBS-C	3	15338	1.95 (0.61)	58 (4.69)	60 (4.53)	4.23 (0.77)	565 (72.88)	554 (89.37)
F	HC	8	49743	2.24 (0.46)	58 (17.00)	57 (12.52)	4.49 (0.44)	639 (124.13)	636 (116.05)
	IBS-D	13	47652	2.14 (0.48)	51 (18.78)	54 (21.23)	4.25 (0.44)	469 (190.05)	474 (196.23)
	IBS-C	3	13010	2.18 (0.14)	41 (13.16)	43 (13.33)	4.36 (0.27)	403 (128.70)	425 (151.11)

The community structure of the pooled samples of IBS patients and controls looked overall quite similar (**Figure 4.7**). However, mucosal samples of patients had higher counts of Bacteroidaceae (IBS-D and ascending colon of IBS-C ≈38%, descending colon of IBS-C 54%, controls ≈28%). In addition, faeces of patients had more Rikenellaceae than controls (IBS-D 12%, IBS-C 29%, controls 7%) as well as more Porphyromonadaceae (IBS-D 11%, IBS-C 16%, controls 7%) and less Ruminococcaceae (IBS-D 8%, IBS-C 11%, controls 15%). Some other characteristics of the IBS-C patients were higher counts of Enterobacteriaceae in mucosal samples and Rikenellaceae in faecal samples when compared with the rest. Specifically, the descending colon of the IBS-C patients differed more from the rest of colonic samples.

We assessed the within-group homogeneity in the prevalence of each bacterial taxon using Gini coefficients and found large between-individual variability

in the abundance of almost all bacterial taxa within each cohort, even for the most dominating genera (**Tables 4.4** and **4.5**).

Only a few of the genera that were found altered in the microbiota of IBS patients with respect to controls turned out to be statistically significant after applying a random labelling permutation process to assess differences in the microbial composition between cases and controls taking into account the large inter-individual variation. That is, the OR for the actual pooled samples was between the most extreme ones in comparison with the ORs of simulated pooled samples obtained under the null hypothesis of no differences between cases and controls (**Tables 4.4** and **4.5**). The shape of the OR distributions built by the random labelling process also reflected the large variability between subjects (**Figure 4.10**). It was bimodal for many of the genera, thus indicating that for those genera, differences between cases and controls were due to only a few of the individuals (sometimes even just one) included in the cohorts rather than a general trend, and hence not detected as statistically significant by the random labelling process. In the IBS-D patients, we found an overrepresentation of *Acinetobacter* (OR=16.71, p=0.02), *Butyricimonas* (OR=2.29, p=0.042), *Leuconostoc* (OR=21.42, p=0.018), and *Odoribacter* (OR=6.11, p=0.003) in faeces with respect to controls, an underrepresentation of *Desulfovibrio* (OR=0.03, p=0.037) and *Oribacterium* (OR=0.17, p=0.041) in the ascending colon, and an underrepresentation of *Brevundimonas* (OR=0.09, p=0.009) and *Butyricicoccus* (OR=0.38, p=0.026) in the descencing colon. We found evidence for the following changes in IBS-C cases compared with controls: an increase in *Alistipes* (OR=5.82, p=0.01) and *Butyricimonas* (OR=3.27, p=0.004) in faeces, as well as an increase in *Bacteroides* (OR=3.15, p=0.039) and a decrease in *Coprococcus* (OR=0.03, p=0.007), *Eubacterium* (OR=0.08, p=0.044), *Fusobacterium* (OR=0.02, p=0.036), *Haemophilus* (OR=0, p=0.019), *Neisseria* (OR=0.02, p=0.037), *Odoribacter* (OR=0.14, p=0.02), *Streptococcus* (OR=0.06, p=0.007), and *Veillonella* (OR=0.03, p=0.044) in the descencing colon. Of these altered genera, only *Alistipes*, *Bacteroides*, *Butyricimonas*, *Eubacterium*, *Fusobacterium*, *Odoribacter*, and *Streptococcus* had a relative abundance greater than 1%.

To establish whether the changes seen at the genus level were due to one, a few, or many bacterial species/strains within them, we also analysed the compositional differences between patients and controls for each phylotype defined at 97% of sequence identity. **Table 4.6** lists the phylotypes for which significant differences were found after applying a random labelling permutation process

Table 4.4: Comparisons between controls (HC) and IBS-diarrhoea cases (IBS-D). It is shown the ORs for the comparison between IBS-D and control pooled samples at the genus level for genera in which significant differences were found in a chi-square test, the Gini coefficients for the abundances of each genus in the individual samples within the IBS-D and control groups, and the OR ranks, which are the ranks of the OR values for the IBS and control pooled samples when compared with the values of 999 simulated pooled samples obtained after random labelling the individual ones as IBS or control (extreme values are indicative of true association). In the comparison sample X vs. sample Y, an OR greater than one indicates overrepresentation in sample X, whereas an OR lower than one indicates underrepresentation in sample X.

		ASCENDING COLON IBS-D vs. HC				DESCENDING COLON IBS-D vs. HC				FAECES IBS-D vs. HC			
Phylum (Class)	Genus	OR	Gini HC	Gini IBS	OR rank	OR	Gini HC	Gini IBS	OR rank	OR	Gini HC	Gini IBS	OR rank
Actinobacteria	*Propionibacterium*	0.38	0.57	0.43	148	0.46	0.65	0.49	197				
Bacteroidetes	*Bacteroides*	1.70	0.65	0.58	863	1.55	0.43	0.37	876	1.22	0.61	0.47	593
(Bacteroidia)	*Barnesiella*	2.93	0.63	0.74	881	1.54	0.58	0.53	836	1.59	0.61	0.56	806
	Butyricimonas	0.41	0.73	0.63	158					2.29	0.44	0.54	958
	Odoribacter	0.66	0.73	0.67	164					6.11	0.56	0.63	997
	Parabacteroides	1.33	0.65	0.55	814	0.82	0.48	0.49	211	1.41	0.52	0.41	783
	Porphyromonas					0.22	0.85	0.85	262				
	Paraprevotella	0.23	0.85	0.71	224	2.64	0.71	0.70	796	1.63	0.74	0.68	714
	Prevotella	0.71	0.77	0.73	350	0.71	0.75	0.77	313	0.87	0.71	0.86	463
	Alistipes	0.82	0.68	0.63	262	0.84	0.57	0.54	250	1.88	0.52	0.43	887
Bacteroidetes (Sphingobacteria)	*Sphingobacterium*	5.95	0.83	0.75	914								
Firmicutes	*Gemella*					2.15	0.48	0.71	796				
(Bacilli)	*Brevibacillus*					0.00	0.89		427				
	Granulicatella					1.95	0.45	0.60	841				
	Leuconostoc									21.42	0.67	0.67	982
	Weissella									81.69	0.75	0.87	912
	Lactococcus									3.75	0.60	0.75	908
	Streptococcus					1.67	0.44	0.65	751				
Firmicutes	*Clostridium*									10.44	0.88	0.86	770
(Clostridia)	*Finegoldia*	0.04	0.88	0.92	205	0.00	0.77		52				
	Peptoniphilus					0.00	0.84		159				
	Blautia									0.50	0.56	0.73	171
	Eubacterium					1.51	0.66	0.74	531	3.96	0.70	0.79	795
	Coprococcus					0.61	0.59	0.68	113				
	Dorea	1.54	0.62	0.59	737								
	Oribacterium	0.17	0.50	0.82	41								
	Roseburia	0.60	0.72	0.69	95	0.61	0.53	0.58	86	1.86	0.46	0.72	667
	Syntrophococcus	0.33	0.80	0.65	81					10.44	0.88	0.90	727
	Sporacetigenium	0.04	0.88	0.92	187								
	Butyricicoccus					0.38	0.59	0.71	26				
	Faecalibacterium									0.75	0.60	0.65	363
	Hydrogenoanaerobacterium					2.96	0.75	0.78	800	0.09	0.80	0.85	128
	Oscillibacter	0.56	0.63	0.64	109					0.44	0.56	0.52	153
	Ruminococcus	0.22	0.86	0.65	213					0.17	0.84	0.53	356
	Acidaminococcus									12.01	0.75	0.85	804
	Dialister									2.43	0.68	0.76	822
	Phascolarctobacterium	0.57	0.85	0.80	290								
	Succinispira									4.70	0.73	0.87	785
	Veillonella									10.23	0.68	0.89	828
Firmicutes	*Catenibacterium*	0.18	0.76	0.82	121								
(Erysipelotrichi)	*Coprobacillus*					0.49	0.66	0.69	103	4.77	0.61	0.81	838
	Turicibacter									Inf		0.92	611
Fusobacteria	*Fusobacterium*	1.65	0.36	0.75	576	0.50	0.68	0.73	215	0.10	0.86	0.85	316
	Leptotrichia	5.21	0.63	0.85	780	0.28	0.77	0.79	186				
Proteobacteria	*Brevundimonas*					0.09	0.71	0.85	9				
(Alpha)	*Methylobacterium*	0.24	0.79	0.79	160	0.08	0.86	0.88	192				
	Sphingomonas					0.07	0.85	0.77	123				
Proteobacteria	*Sutterella*					2.70	0.73	0.79	738				
(Beta)	*Delftia*	0.45	0.67	0.56	262	0.35	0.56	0.51	101				
	Diaphorobacter					0.10	0.85	0.81	157				
	Neisseria	2.65	0.48	0.79	702								
Proteobacteria (Delta)	*Desulfovibrio*	0.03	0.85	0.85	37								
Proteobacteria (Epsilon)	*Campylobacter*					0.33	0.73	0.75	180				
Proteobacteria	*Escherichia/Shigella*	0.34	0.74	0.67	186	0.42	0.77	0.66	274	3.54	0.86	0.85	777
(Gamma)	*Aggregatibacter*	0.23	0.83	0.67	290	0.51	0.87	0.88	312				
	Acinetobacter									16.71	0.88	0.70	980
	Pseudomonas	6.15	0.63	0.75	899	2.46	0.75	0.82	713				
Spirochaetes	*Treponema*	Inf		0.92						0.00	0.88		386
Tenericutes	*Asteroleplasma*									0.00	0.87		158
Verrucomicrobia	*Akkermansia*									2.69	0.78	0.92	617

Table 4.5: Comparisons between controls (HC) and IBS-constipation cases (IBS-C). It is shown the ORs for the comparison between IBS-C and control pooled samples at the genus level for genera in which significant differences were found in a chi-square test, the Gini coefficients for the abundances of each genus in the individual samples within the IBS-C and control groups, and the OR ranks, which are the ranks of the OR values for the IBS and control pooled samples when compared with the values of 999 simulated pooled samples obtained after random labelling the individual ones as IBS or control (extreme values are indicative of true association). In the comparison sample X vs. sample Y, an OR greater than one indicates overrepresentation in sample X, whereas an OR lower than one indicates underrepresentation in sample X.

		ASCENDING COLON				DESCENDING COLON				FAECES			
		IBS-C vs. HC				IBS-C vs. HC				IBS-C vs. HC			
Phylum (Class)	Genus	OR	Gini HC	Gini IBS	OR rank	OR	Gini HC	Gini IBS	OR rank	OR	Gini HC	Gini IBS	OR rank
Actinobacteria	*Propionibacterium*					0.35	0.65	0.44	306				
Bacteroidetes	*Bacteroides*	1.44	0.65	0.22	940	3.15	0.43	0.37	961	0.68	0.61	0.07	444
(Bacteroidia)	*Barnesiella*					0.41	0.58	0.48	147	2.56	0.61	0.26	934
	Butyricimonas	0.20	0.73	0.67	85					3.27	0.44	0.14	996
	Odoribacter					0.14	0.53	0.11	20	4.76	0.56	0.59	878
	Parabacteroides					0.54	0.48	0.37	155	2.11	0.52	0.36	813
	Porphyromonas					0.00	0.85		250				
	Paraprevotella	0.38	0.85	0.47	737					1.30	0.74	0.37	665
	Prevotella	1.33	0.77	0.66	530	0.02	0.75	0.31	127	0.02	0.71	0.67	106
	Alistipes					0.28	0.57	0.29	103	5.82	0.52	0.23	990
Firmicutes	*Gemella*					0.12	0.48	0.50	67				
(Bacilli)	*Granulicatella*	0.22	0.55	0.60	120								
	Streptococcus	0.54	0.36	0.58	220	0.06	0.44	0.21	7				
Firmicutes	*Blautia*					9.48	0.51	0.65	786	0.18	0.56	0.44	116
(Clostridia)	*Eubacterium*	1.80	0.69	0.53	772	0.08	0.66	0.40	44				
	Coprococcus					0.03	0.59	0.44	7				
	Roseburia					2.02	0.53	0.58	775	1.53	0.46	0.42	717
	Butyricicoccus					0.11	0.59	0.56	81				
	Faecalibacterium					0.65	0.27	0.54	256	0.61	0.60	0.43	477
	Hydrogenoanaerobacterium									0.00	0.80		254
	Ruminococcus	0.06	0.86	0.50	437					0.26	0.84	0.67	592
	Subdoligranulum					0.34	0.45	0.41	115	0.34	0.54	0.33	205
	Dialister					0.03	0.76	0.67	154	1.88	0.68	0.66	671
	Phascolarctobacterium	0.27	0.85	0.45	683	1.79	0.62	0.66	731				
	Succinispira					0.00	0.84		462				
	Veillonella					0.03	0.63	0.67	44				
Firmicutes	*Catenibacterium*					0.00	0.71		150				
(Erysipelotrichi)	*Coprobacillus*	2.61	0.66	0.34	881	0.41	0.66	0.30	247				
Fusobacteria	*Fusobacterium*	0.31	0.36	0.67	121	0.02	0.68	0.44	36				
	Leptotrichia					0.00	0.77		98				
Proteobacteria	*Methylobacterium*					0.00	0.86		258				
(Alpha)	*Sphingomonas*					0.00	0.85		372				
Proteobacteria	*Parasutterella*	0.10	0.79	0.44	217	1.90	0.71	0.56	727	2.57	0.76	0.51	850
(Beta)	*Sutterella*					0.00	0.73		94				
	Delftia	0.28	0.67	0.42	299	0.27	0.56	0.57	163				
	Neisseria					0.02	0.72	0.67	37				
Proteobacteria	*Campylobacter*					0.00	0.73		98				
(Epsilon)													
Proteobacteria	*Escherichia/Shigella*	2.99	0.74	0.52	736	6.93	0.77	0.65	819				
(Gamma)	*Aggregatibacter*	0.14	0.83	0.67	409	0.01	0.87	0.67	194				
	Haemophilus	0.21	0.35	0.67	72	0.00	0.61		19				
	Acinetobacter	0.10	0.71	0.33	95	0.23	0.72	0.33	262				
Spirochaetes	*Treponema*									0.00	0.88		245
Tenericutes	*Asteroleplasma*									0.00	0.87		459

to assess differences in the composition of the microbiota between cases and controls taking into account the inter-individual variation. We found several phylotypes with an altered abundance in patients compared with controls. Differences of opposite sign could be found for phylotypes belonging to the same genus. Only a few differences between IBS cases and controls were shared

between sample types or IBS subtypes. One phylotype was consistently over-represented in the colonic mucosa and faeces of IBS-D patients. It accounted for ≈2% of the sequences in these samples and was closely related to *Bacteroides vulgatus*.

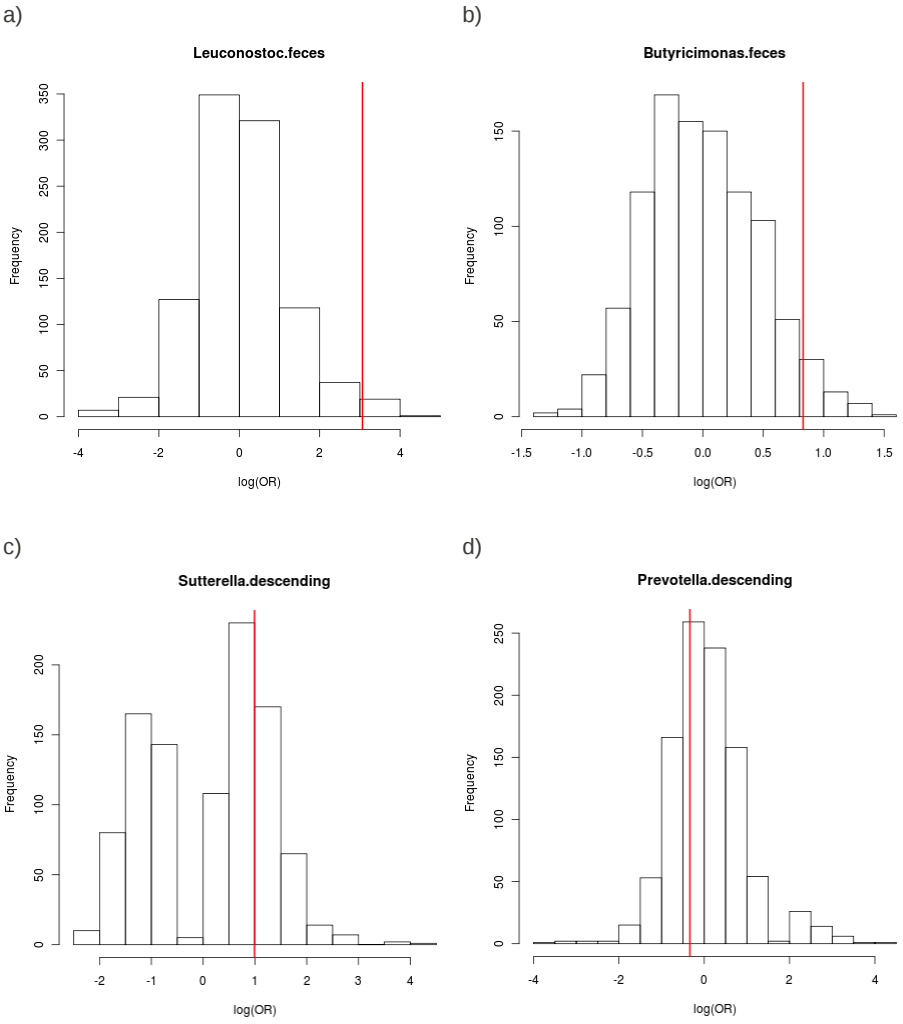


Figure 4.10: Distribution of the ORs calculated for the IBS-D subtype and the HC pooled samples (red line) and for 999 simulated pooled samples obtained after random labelling the individual ones as IBS or control. a, b: examples of genera for which the change detected in the pooled samples is well supported by the individual samples; c, d: examples of genera for which the change is not supported.

Table 4.6: Comparisons between controls (HC) and IBS cases (IBS-D or IBS-C). It is shown the ORs for the comparisons between sampling site-pooled samples for phylotypes in which significant differences were found in a chi-square test, also supported by a permutation test, and with a relative abundance greater than 0.1%. In the comparison sample X vs. sample Y, an OR greater than one indicates overrepresentation in sample X, whereas an OR lower than one indicates underrepresentation in sample X. The closest species are the species of best blast hits with more than 96% of sequence identity and more than 95% of query coverage.

Comparison	Phylotype 97% identity	OR	Genus	Closest species
IBS-D vs. HC	974650	0.01	*Bacteroides*	*B. dorei*
ASCENDING COLON	97431	0.25	*Bacteroides*	*B. ovatus*
	971776	0.11	*Bacteroides*	*B. uniformis*
	974218	3.65	*Bacteroides*	*B. vulgatus*
	974361	6.10	*Bacteroides*	*B. vulgatus*
	974516	7.33	*Bacteroides*	*B. vulgatus*
	971992	0.09	*Escherichia/Shigella*	*E. coli*
	972853	0.10	*Escherichia/Shigella*	*E. coli, S. dysenteriae*
	972296	0.20	*Faecalibacterium*	*F. prausnitzii*
	971322	0.15	*Odoribacter*	*O. splanchnicus*
	972166	0.36	*Odoribacter*	*O. splanchnicus*
	973087	0.09	*Oribacterium*	*O. sinus*
IBS-D vs. HC	971080	0.02	*Alistipes*	
DESCENDING COLON	97431	0.30	*Bacteroides*	*B. ovatus*
	97968	19.20	*Bacteroides*	*B. thetaiotaomicron*
	974516	4.06	*Bacteroides*	*B. vulgatus*
	974242	0.05	*Blautia*	
	971270	0.13	*Butyricicoccus*	
	97804	0.19	*Faecalibacterium*	
	974956	0.20	*Parabacteroides*	
	97256	0.00	*Parasutterella*	*P. excrementihominis*
	972173	0.06	*Prevotella*	*P. copri*
	975039	27.45	*Streptococcus*	*S. australis, S. oralis, S. sanguinis*
IBS-D vs. HC	97714	50.26	*Alistipes*	*A. finegoldii*
FAECES	971080	0.09	*Alistipes*	
	971132	310.19	*Bacteroides*	*B. coprocola*
	972067	605.41	*Bacteroides*	*B. coprocola*
	97495	98.70	*Bacteroides*	*B. plebeius*
	97727	8.63	*Bacteroides*	*B. plebeius*
	971028	9.03	*Bacteroides*	*B. plebeius*
	974187	19.86	*Bacteroides*	*B. vulgatus*
	974516	8.64	*Bacteroides*	*B. vulgatus*
	974527	5.40	*Bacteroides*	
	974548	61.11	*Bacteroides*	*B. vulgatus*
	974639	23.54	*Bacteroides*	*B. vulgatus*
	97134	4.97	*Bacteroides*	*B. xylanisolvens*
	97346	13.58	*Bacteroides*	*B. xylanisolvens*
	97646	9.40	*Bacteroides*	
	97106	9.40	*Butyricimonas*	
	97162	11.49	*Butyricimonas*	
	97222	12.88	*Butyricimonas*	*B. virosa*
	97804	0.10	*Faecalibacterium*	
	97903	0.04	*Faecalibacterium*	
	971559	19.78	*Faecalibacterium*	*F. prausnitzii*
	973391	20.45	*Leuconostoc*	*L. citreum*
	972166	5.56	*Odoribacter*	*O. splanchnicus*
	972776	8.43	*Odoribacter*	*O. splanchnicus*
	972979	14.13	*Odoribacter*	*O. splanchnicus*
	972053	0.16	*Oscillibacter*	
	974143	0.00	*Oscillibacter*	
	974333	0.08	*Oscillibacter*	
	974425	0.14	*Oscillibacter*	
	973220	189.66	*Parabacteroides*	*P. distasonis*
IBS-C vs. HC	971771	0.04	*Bacteroides*	*B. thetaiotaomicron*
DESCENDING COLON	974454	0.04	*Bacteroides*	*B. vulgatus*
	97271	6.49	*Parasutterella*	*P. excrementihominis*
	97924	15.74	*Roseburia*	
	973927	0.05	*Streptococcus*	*S. mitis*
IBS-C vs. HC	97714	81.97	*Alistipes*	*A. finegoldii*
FAECES	97446	88.09	*Alistipes*	*A. indistincts*
	97298	145.71	*Alistipes*	*A. indistinctus*
	971023	686.07	*Alistipes*	*A. onderdonkii*
	971529	475.36	*Alistipes*	*A. onderdonkii*
	971373	19.13	*Alistipes*	
	972546	7.36	*Alistipes*	*A. putredinis*
	973534	585.83	*Alistipes*	
	973660	288.42	*Alistipes*	
	971232	6.95	*Alistipes*	
	97642	22.00	*Bacteroides*	
	97136	33.21	*Bacteroides*	*B. cellulosilyticus*
	97646	39.92	*Bacteroides*	
	971907	918.98	*Barnesiella*	
	972044	60.05	*Barnesiella*	*B. intestinihominis*
	972369	72.51	*Barnesiella*	
	9791	6.18	*Butyricimonas*	
	97106	12.12	*Butyricimonas*	
	972297	48.64	*Faecalibacterium*	*F. prausnitzii*
	973233	7.41	*Faecalibacterium*	
	972979	38.91	*Odoribacter*	*O. splanchnicus*
	972366	70.93	*Parabacteroides*	

4.2.2.2 Comparison of sampling sites

The Shannon diversity indices (H) and the Chao1 and ACE richness estimators were on average lower in faeces than in mucosa in IBS patients (in both subtypes) and in healthy controls, with few differences between the ascending and the descending colon, except for the descending colon of the IBS-C patients (**Table 4.3**).

Differences in the composition of the microbiota between sample types can be appreciated visually, especially between the intestinal and the faecal ones (**Figure 4.7**). The Bacteroidetes phylum accounted for 58% of the sequences in mucosal samples and 72% in faeces. In contrast, 29% of the sequences in the intestinal mucosa belonged to the Firmicutes phylum compared with 21% in faeces. Proteobacteria accounted for 9.5% of sequences in mucosa and 0.9% in faeces. At the family level, faeces had more members of Rikenellaceae (11.3% in faeces versus 2.4% in mucosa) and less of Lachnospiraceae (2.1% versus 7.6%), Streptococcaceae (0.1% versus 2.3%), and Enterobacteriaceae (0.2% versus 3.1%).

Comparisons between the microbial composition of the communities found in the ascending colon, descending colon, and faeces at the genus level are shown in **Table 4.7**. No relevant differences were found between the ascending and the descending colon. Those differences in bacterial composition between colonic mucosa and faeces detected in at least two thirds of the individuals were considered statistically significant. Specifically, we found an underrepresentation in colonic mucosa with respect to faeces of *Barnesiella* (OR$\approx$0.20) and *Alistipes* (OR$\approx$0.19), and an overrepresentation of *Streptococcus* (OR$\approx$39.50) and *Dorea* (OR$\approx$7.75). Results were consistent within the IBS and the control cohorts, further confirming the shifts detected.

4.2.2.3 Analysis of clinical and lifestyle data

We found no significant association between case/control status and the collected clinical and lifestyle factors (except for age, the controls being on average older than the patients). No patterns relating bacterial composition and lifestyle factors were observed in CCA plots.

Table 4.7: Comparisons between sampling site-pooled samples. Only genera for which significant differences were detected for at least one third of the individuals in at least one sampling site are shown. In the comparison sample X vs. sample Y, an OR greater than one indicates overrepresentation in sample X, whereas an OR lower than one indicates underrepresentation in sample X. It is indicated the number of subjects for which a significant overrepresentation (↑) or underrepresentation (↓) was found in the comparison between his/her samples, showing data for controls (HC) and patients (IBS) separately. ASC.: ascending colon; DESC.: descending colon; n: number of subjects included in each group.

		ASC. vs. DESC.					ASC. vs. FAECES					DESC. vs. FAECES				
Phylum	Genus		HC n=9		IBS n=16			HC n=8		IBS n=16			HC n=8		IBS n=16	
(Class)		OR	↑	↓	↑	↓	OR	↑	↓	↑	↓	OR	↑	↓	↑	↓
Bacteroidetes	*Bacteroides*	0.97					1.09	3	3	7	6	1.11	4	1	9	5
(Bacteroidia)	*Barnesiella*	1.30				2	0.23		5	1	10	0.17		6	1	11
	Parabacteroides	0.99					0.87	1	2	3	7	0.88	2		4	7
	Paraprevotella	1.33					0.20		3	2	7	0.15		5	1	6
	Prevotella	1.06					2.26	2	4	8	2	2.13	2	2	5	2
	Alistipes	1.10					0.20	1	5	1	13	0.18		5	1	14
Firmicutes	*Gemella*	1.43	1		1	1	223.33	6		6		162.32	2		3	
(Bacilli)	*Granulicatella*	1.78	2		1	1	137.04	4		5		88.03	1		5	
	Streptococcus	1.01					38.54	8		11		41.42	7		10	
Firmicutes	*Blautia*	0.36		1	1	2	3.56	1		5		9.83	3		5	
(Clostridia)	*Eubacterium*	0.78			1		4.39	3		6		5.60	4		4	
	Dorea	1.08					7.03	5		11		6.49	8		12	
	Roseburia	1.01					2.29	3	1	4	2	2.26	4	1	7	2
	Faecalibacterium	1.25	2	3	4	4	2.23	4	1	10		1.76	5	1	9	2
	Oscillibacter	1.24					0.17		5	2	9	0.13		6	1	9
	Veillonella	1.06					2.31	6		4	1	2.26	4		2	1
Firmicutes	*Coprobacillus*	0.98					4.13	5		6	2	4.20	4		5	2
(Erysipelotrichi)																
Fusobacteria	*Fusobacterium*	0.93					26.72	5		5		29.48	5		4	
Proteobacteria	*Sutterella*	0.66	1		1	2	14.69	2		5		22.17	3		5	
(Beta)	*Delftia*	1.38	1		1	1	641.94	4		6		451.64	5		5	
Proteobacteria	*Escherichia/*	0.54	2		1	1	25.41	3		5		46.61	4		5	
(Gamma)	*Shigella*															
	Haemophilus	1.97	4	1	4		49.37	6		7		26.34	4		3	
	Pseudomonas	1.30					97.71	3		5		74.40	2		4	
	Stenotrophomonas	1.56	2	1	2		138.68	4		4		90.50	3		1	

4.2.3 Discussion

This is the first study that explored the potential alterations of the intestinal microbiota in IBS patients simultaneously in colonic mucosa and faeces through a 16S rRNA gene sequencing approach, although massive sequencing has been widely used in recent times to study microbiomes in both faeces and intestinal mucosa in relation to other GI pathologies.

Previous studies on IBS using 16S rRNA gene sequencing based their conclusions on pooled faecal samples of patients and healthy controls without accounting for the within-group variability (Kassinen et al., 2007; Krogius-kurikka et al., 2009). We found that the prevalence of nearly all bacterial groups present in the samples was highly variable between individuals, which may create spurious differences between the pooled samples due to just a few individuals dominating the composition of these bacterial taxa in the pooled

samples instead of a general trend. To avoid this problem, we carried out a random labelling procedure to compare the microbial composition of IBS patients versus to that of controls while accounting for the individual variation. After that, we found a few genera associated with IBS, with differences varying by site. Our results are in agreement with recent research on IBS that points to subtle rather than pronounced alterations in the gut microbiota of IBS patients (Salonen et al., 2010).

The bacterial communities of IBS patients in mucosal sites along the colon and in faeces were less diverse (as measured by the Shannon index) than those found in healthy subjects. Although several studies on IBS seem to contradict these findings (Carroll et al., 2011; Ponnusamy et al., 2011), a lower biodiversity has also been found in the gut microbiota in association with obesity (Turnbaugh et al., 2009a) and IBD (Seksik, 2010; Qin et al., 2010), and in the skin microbiota of allergic children (Hanski et al., 2012). The findings fit with the "biodiversity hypothesis", which holds that reduced exposure to microbes may adversely affect the composition of the human microbiota and its immunomodulatory capacity (von Hertzen et al., 2011).

Studies in healthy individuals have shown that the microbiota associated with the intestinal mucosa differs from the microbiota in faecal samples (Section 4.1) (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005). However, there is little information about the differences between these compartments in IBS, mostly because most of the studies are based on faecal samples. We noticed that the communities in the ascending colon, the descending colon, and faeces are affected differently by the disorder. The microbiota of IBS patients showed lower biodiversity than that of controls in colonic mucosa and faeces, but when analysing the differences in the abundance of particular genera, we found no statistically significant changes common to all three or even two of the sampling sites. In faeces, we found more genera overrepresented than underrepresented in the microbiota of IBS patients with respect to controls. In contrast, mucosal sites showed communities with more genera underrepresented in IBS cases compared with controls. This finding must serve to caution researchers about the risk of inferring the role of the endogenous microbiota in IBS from that found in faeces, especially considering that the mucosa-associated microbiota may be more relevant to the pathogenesis of IBS because it is closer to the host epithelial and immune cells.

Several precedent studies compared both mucosal and faecal bacteria of IBS patients and healthy controls, though not via high-throughput sequencing. In

contrast to our study, Kerckhoffs and co-workers found a decrease in *Bifidobacterium* in IBS patients by using FISH (Kerckhoffs et al., 2009), and an increase in *Pseudomonas* in IBS patients by using quantitative PCR (Kerckhoffs et al., 2011), whereas Carroll et al. (2010) detected an increase in *Lactobacillus* only in the faeces of IBS patients with quantitative PCR. Carroll et al. (2011) also characterised the faecal and sigmoid mucosal microbiota of IBS-D patients and controls by T-RFLP fingerprinting. They found differences in the composition between patients and controls, and a lower diversity in the former just in the faecal communities. However, we found a lower diversity in all sampling sites from IBS patients compared with those of controls.

The comparison between sampling sites showed a high level of similarity between the mucosal-associated communities of the ascending and the descending colon from the same subject, whereas faecal samples clustered separately. Faeces showed lower richness and diversity than biopsies, with many bacterial genera underrepresented and just a few overrepresented when compared to colonic mucosa. These patterns could be due to differences in the ecological conditions between the two environments, to the susceptibility of different bacteria in the intestine to ending up in faeces, or to the possibility that faeces better represent luminal than mucosal-associated microbiota. Our results must be interpreted with caution given that the bowel cleansing prior to endoscopy, which is the standard clinical procedure, could have partly removed the outer mucus layer, which might have attached bacteria. Nevertheless, the results of this study are overall in agreement with those of Section 4.1, where we compared faeces and rectal mucosa in healthy individuals, despite biopsies there were collected without any previous preparation of the colon and the entire 16S rRNA gene, instead of the V1-V2 region, was targeted.

The existing knowledge on IBS does not reveal pronounced and reproducible compositional deviations in the gut microbiota. There are several explanations for the lack of reproducibility in the transversal studies carried out so far. One is that dysbiosis in IBS is characterised by subtle alterations in the microbiota instead of the high level phylogenetic alterations that occur in other pathologies such as obesity or IBDs (Ley et al., 2006; Qin et al., 2010). The large inter-subject variability in the microbiota composition due to other reasons than the pathological state and the use of relatively small cohorts make difficult the detection of such subtle changes. The other is that IBS is a complex functional disorder, and patients with perhaps heterogeneous aetiology and symptoms at the moment of sampling were included in these studies, further difficulting the

detection of patterns.

IBS-related studies have been almost exclusively limited to the analysis of the community structure based on the variability of the 16S rRNA gene. Further research is needed to assess the implication of the gut microbiota in IBS from a functional perspective. Moreover, longitudinal studies in which patients are followed over time and samples are taken at moments with different symptomatology would help to mitigate the confusion caused by the inter-subject variability and the heterogeneity within IBS patients that is problematic in cross-sectional studies.

This study has been published in Environmental Microbiology Reports (Durbán et al., 2012b).

4.3 Stability and host-specificity of faecal bacteria in healthy subjects

4.3.1 Background

In the last decade, an extensive effort has been made to characterise the human GI microbiota by means of 16S rRNA gene sequencing and metagenomic analyses. Regarding temporal stability, it is known that an ecological succession takes place within the human GI microbiota over the first years of life that culminates in the complex adult pattern (Palmer et al., 2007; Vaishampayan et al., 2010; Koenig et al., 2011). Then, it is generally accepted that, in the absence of perturbation, the adult human gut microbiota is composed of stable communities inhabiting the different niches found along and across the intestine. In older people, modifications of the gut microbiota have been reported related to the physiological changes associated with ageing (Woodmansey, 2007; Claesson et al., 2011).

Many factors are known to influence the structure of the microbiota of adult humans, such as host genotype (Stewart et al., 2005; Spor et al., 2011), disease (Turnbaugh et al., 2006; Seksik, 2010), diet (De Filippo et al., 2010; Wu et al., 2011), and stochastic events such as the colonisation history (Deloris Alexander et al., 2006; Mulder et al., 2009; Ubeda et al., 2012) or external disturbances (De La Cochetière et al., 2005; Dethlefsen and Relman, 2011). Little is known about the short- and long-term effects that these factors have in the composition of the microbiota. To assess the response of the gut microbiota to external perturbations or its role in pathological states, it is important to know first the "normal" temporal dynamics of this ecosystem. However, relatively few studies have followed intestinal microbial communities over time to assess their stability in the absence of perturbation.

Human faecal communities have been considered stable because the temporal variation within individuals is smaller than the inter-individual variation. The stability has typically been examined with several samples collected at intervals of weeks or months, finding that the microbiota of healthy individuals remains fairly constant over these long intervals (Franks et al., 1998; Zoetendal et al., 1998; Vanhoutte et al., 2004). However, some variation has been found in the abundance of certain taxa, while others remain more constant (Vanhoutte et al., 2004). Recently, studies of the temporal variation using shorter time periods confirmed the stability of the community composition at a lower phy-

logenetic level, but also revealed a high degree of dynamism, with relatively large fluctuations around average abundance values (Caporaso et al., 2011; Dethlefsen and Relman, 2011).

Ecological interactions between the members of gut communities, like nutritional associations, niche adaptation, growth stimulation, resource competition, and interference mechanisms, also contribute to the shape and stability of these ecosystems. Despite their importance, little is known about the relationships between gut bacteria. A first step for the study of these interactions is searching for patterns of co-occurring bacteria. Due to the high level of interpersonal variation in the community assembly, it seems more appropiate to do this in longitudinal studies of single communities as opposed to cross-sectional ones. This is because there is no between-individual variation when following the same community over time. The same inference can be done using cross-sectional studies (see e.g. Arumugam et al. (2011)), but probably at the expense of a much larger number of samples.

The objectives of this study were to analyse the intrinsic daily variation in the community structure of the faecal bacteria from three healthy subjects over fifteen days, and to assess correlations between shifts in the relative abundance of specific bacteria to reconstruct networks of potential interactions within the communities. The global community structure was analysed through sequencing of PCR-amplified rRNA genes.

4.3.2 Results

After sequencing all 45 samples, we obtained an average of 1500 sequences per sample for A, 4165 for B, and 6510 for C. The average read length was ∼250 nt in A and ∼350 nt in B and C. This difference in read length was the result of the slightly different sequencing technologies employed for subject A and subjects B and C, but, in both cases, the variable regions V1 and V2 of the 16S rRNA gene were covered. The comparison between subjects could be affected by differential biases introduced during the PCR amplification due to the sequencing adaptors, or by the impact of read length in the taxonomic assignment. However, we think that their impact on the estimate of the variation between subjects is small, and none on the analysis within each subject, which was our main objective.

The Chao1 index of richness indicated a good coverage for most samples

at the genus level (**Table 4.8**). The coverage decreased at the phylotype level (defined at 97% of sequence identity), with larger differences between observed and expected number of phylotypes (**Table 4.9**). Although the number of sequences per sample is important for the estimation of richness, it has little impact on the estimation of biodiversity (as it is demonstrated in **Tables 4.8** and **4.9**, which show the Shannon diversity indices for the original samples and subsamples obtained with the number of sequences of the smallest sample). This is because biodiversity is based on relative abundances, the estimation of which is less sensitive to the number of sequences per sample (though the associated standard errors are obviously larger for the smaller samples).

4.3.2.1 Between-subject variation

Most members of the sampled communities belonged to a small number of genera within the Bacteroidetes phylum (61% of the sequences on average in samples from A, ∼86% in B and C) and the Firmicutes phylum (26% in A, ∼10% in B and C). The composition of the faecal microbiota of B and C was quite similar from the phylum to the genus level. At the genus level, samples from A were dominated by *Alistipes* (23%), *Bacteroides* (22%), several genera within Porphyromonadaceae, such as *Barnesiella* (8%) and *Parabacteroides* (2%), and several genera within Ruminococcaceae, such as *Faecalibacterium* (4%) and *Oscillibacter* (2%) (**Figure 4.11(a)**). Samples from B and C were dominated by *Prevotella* (76% in B, 72% in C), *Bacteroides* (5% in B, 12% in C), and Betaproteobacteria within the genus *Sutterella* (∼3%) (**Figure 4.11(b)** and **4.11(c)**).

At a finer taxonomic scale (phylotypes defined at 97% of sequence identity), the faecal microbiota was highly specific to each individual. About half of the phylotypes detected in each subject were exclusive to him, whereas 20% were shared with the other two subjects. The greater similarity between B and C was also seen at this level, as they had more phylotypes in common than each of them with A. Most of the sequences were concentrated in a small fraction of the phylotypes. Thus, 11% of the phylotypes in A, and 4% of the phylotypes in B and C, had a relative abundance of at least 5‰, and all together accounted for 86%, 80%, and 89% of the respective number of sequences. The most prevalent genera in A, *Alistipes*, *Bacteroides*, and *Barnesiella*, accounted for 9%, 10%, and 3% of the total number of phylotypes, whereas 16% of the phylotypes in B, and 10% in C, belonged to *Prevotella*.

Table 4.8: Observed richness, Chao1 richness estimator, and Shannon diversity index (H) at the genus level for each sample with the full datasets and average values for 1000 subsamples with the number of sequences of the smallest sample (*). Samples are labelled with the name of the subject (A, B, C) and the day over the follow-up.

Sample	No. reads	No. observed genera	Chao1	Average Chao1 (1000 subsamples)	Shannon H	Average Shannon H (1000 subsamples)
A.1	1386	34	39.6	35.15	3.45	3.42
A.2	1535	32	41.33	32.11	3.37	3.34
A.3	1576	27	27.75	26.45	3.26	3.24
A.4	1508	28	33	27.01	3.12	3.09
A.5	1741	29	38.33	27.87	3.03	3
A.6	2181	29	30	26.74	3.01	2.99
A.7	4312	34	52.33	25.61	3.08	3.05
A.8	1014	25	32	28.69	3.23	3.21
A.9	602	29	36.5	35.95	3.58	3.58
A.10	1547	28	35.5	27.68	3.66	3.64
A.11	1816	35	38	35.35	3.73	3.7
A.12 *	590	27	32	-	3.46	-
A.13	1451	36	38.63	37.33	3.31	3.28
A.14	2906	33	40	27.95	3.15	3.12
A.15	1216	25	31	24.4	3.26	3.24
B.1	5915	41	63	29.06	2.22	2.18
B.2	3797	32	59.5	24.09	1.51	1.48
B.3	3313	26	33	20.23	1.19	1.16
B.4	4147	33	40	26.77	1.57	1.53
B.5	3822	38	65.5	30.56	2.27	2.24
B.6	737	27	42	38	2	1.99
B.7	4902	36	58	24.86	1.9	1.86
B.8	4821	31	38	24.17	1.1	1.07
B.9	4429	31	33.5	23.98	1.18	1.14
B.10	6118	32	50	22.67	1.24	1.21
B.11	5168	30	32.5	22.19	1.08	1.05
B.12	5209	44	70.25	31.67	1.63	1.59
B.13	4816	36	47	25.89	1.68	1.65
B.14	5583	40	41.11	31.16	2.27	2.23
B.15	4879	36	38.63	26.91	1.57	1.53
C.1	8192	28	46	18.01	1.02	1
C.2	5938	32	34.5	22.9	1.08	1.05
C.3	7009	45	56	30.29	1.76	1.72
C.4	6372	40	66	27.12	2.07	2.04
C.5	6682	38	40.5	28.95	1.92	1.88
C.6	8028	41	45	27.27	2.04	2
C.7	6126	37	46	25.7	1.82	1.78
C.8	7805	38	43.6	25.7	1.86	1.82
C.9	7299	35	41	22.78	1.79	1.77
C.10	7121	36	39	24.08	1.64	1.61
C.11	8678	37	39.5	25.28	1.38	1.35
C.12	4373	29	31.5	22.51	1.26	1.24
C.13	4185	35	57.75	24.26	1.57	1.54
C.14	4675	28	29	22.94	2.14	2.11
C.15	7428	34	39.25	24.29	2.12	2.08

Table 4.9: Observed richness, Chao1 richness estimator, and Shannon diversity index (H) at the phylotype level for each sample with the full datasets and average values for 1000 subsamples with the number of sequences of the smallest sample (*). Samples are labelled with the name of the subject (A, B, C) and the day over the follow-up.

Sample	No. reads	No. observed phylotypes 97%	Chao1	Average Chao1 (1000 subsamples)	Shannon H	Average Shannon H (1000 subsamples)
A.1	1386	142	195.45	152.2	5.56	5.43
A.2	1535	149	206.24	159.99	5.52	5.39
A.3	1576	141	190	156.59	5.58	5.46
A.4	1508	132	167.29	137.87	5.73	5.61
A.5	1741	145	237.81	144.4	5.61	5.48
A.6	2181	167	222.03	161.69	5.81	5.64
A.7	4312	184	250.5	138.67	5.46	5.29
A.8	1014	133	182.21	172.94	5.58	5.49
A.9	602	127	202.14	201.5	6.04	6.03
A.10	1547	179	278.11	200.79	5.63	5.47
A.11	1816	189	256.03	196.46	5.71	5.53
A.12 *	590	130	220.05	-	5.99	-
A.13	1451	171	236.1	190.64	6.01	5.86
A.14	2906	208	298.62	176.52	5.97	5.76
A.15	1216	161	260.53	184.94	5.91	5.78
B.1	5915	281	390.28	192.06	5.69	5.43
B.2	3797	224	381.24	176.49	5.04	4.83
B.3	3313	174	256.83	146.06	4.81	4.64
B.4	4147	226	320.53	180.17	4.9	4.69
B.5	3822	247	353.95	190.86	5.55	5.31
B.6	737	121	235.07	191.75	5.26	5.22
B.7	4902	240	364.62	174.03	5.24	5
B.8	4821	225	392.38	154.1	4.33	4.14
B.9	4429	198	346.75	141.46	4.35	4.17
B.10	6118	236	355.6	154.28	4.55	4.35
B.11	5168	212	333.83	146.33	4.42	4.22
B.12	5209	263	433.63	181.3	5.05	4.8
B.13	4816	258	414.09	183.09	4.99	4.75
B.14	5583	292	481.87	199.74	5.68	5.41
B.15	4879	273	427.34	190.45	4.93	4.69
C.1	8192	144	214.71	80.62	3.1	2.97
C.2	5938	139	173.18	91.3	2.97	2.85
C.3	7009	224	388.54	133.1	3.79	3.6
C.4	6372	192	303	120.13	4.09	3.93
C.5	6682	234	321.78	149.81	4.12	3.9
C.6	8028	217	284	130.71	3.98	3.79
C.7	6126	237	409.44	150.17	3.99	3.78
C.8	7805	224	303.08	135.46	4.1	3.91
C.9	7299	208	320.86	126.43	4.03	3.85
C.10	7121	195	249.03	123.95	3.71	3.54
C.11	8678	206	285.44	117.64	3.46	3.29
C.12	4373	146	220.25	106.92	3.34	3.19
C.13	4185	141	194.26	107.63	3.52	3.39
C.14	4675	152	199.22	110.41	4.35	4.21
C.15	7428	218	289.5	138.18	4.21	4.02

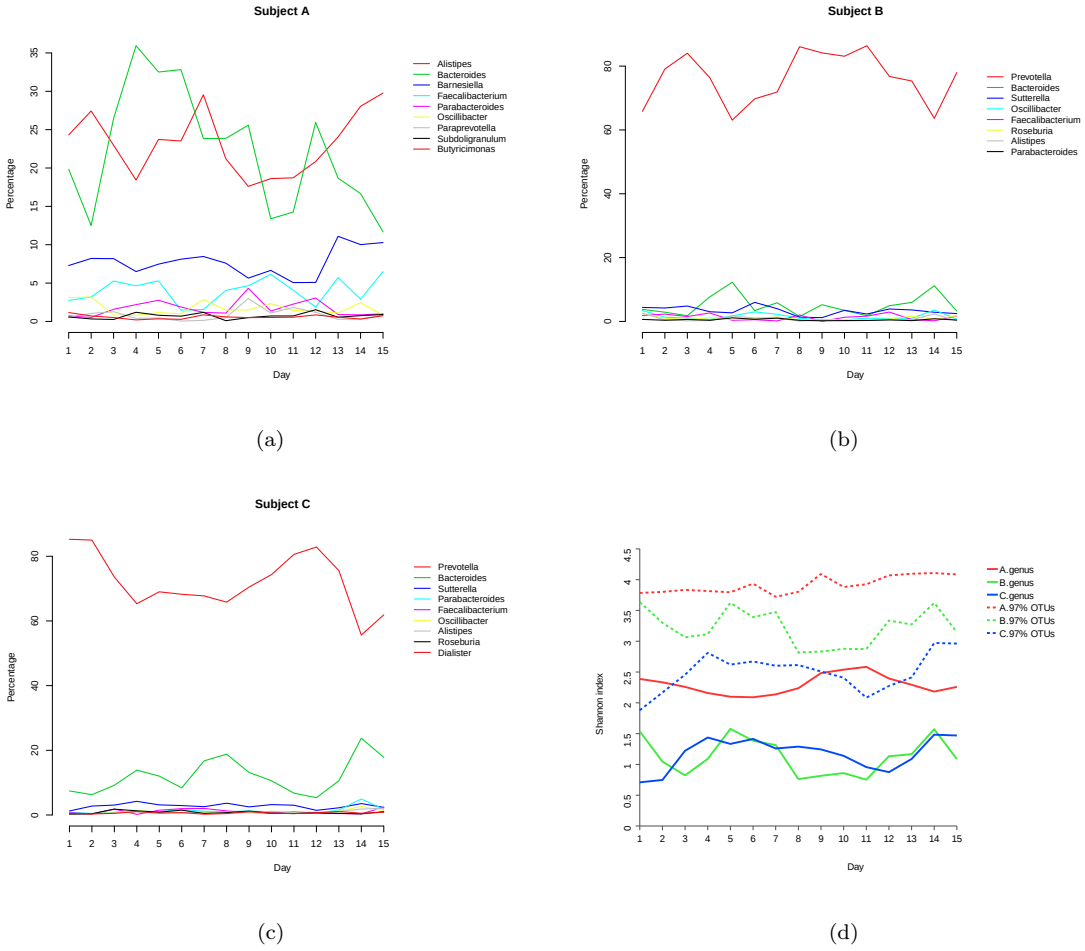
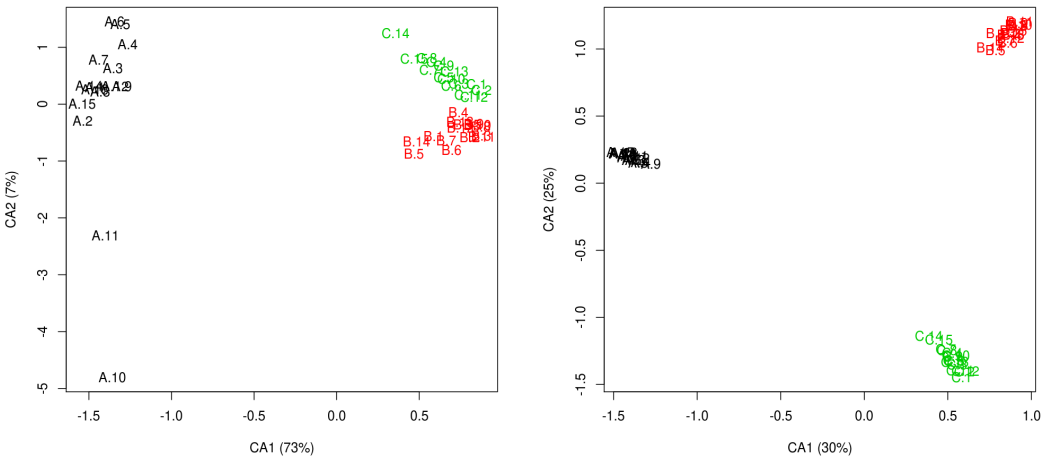


Figure 4.11: Daily fluctuations in the relative abundance of the main bacterial genera (average abundance $\geq 5‰$) in subjects A (a), B (b), and C (c). Daily variation in the bacterial diversity (Shannon index) at genus and phylotype levels (d).

Between-subject differences in the community structure were larger than dissimilarities between samples from the same subject. Therefore, CA plots clearly discriminated the samples of the three subjects at the genus and phylotype levels (**Figure 4.12**). The first CA axis separated A from B and C; the second axis separated B from C. ANOSIM found significant differences between subjects (genus level: R=0.686, p=0.001; phylotype level: R=1, p=0.001). At

the genus level, the between-subject median rank of distances was 1.5 times that of within A, and 4 times that of within B or C. At the phylotype level, the between-subject variation was 2.6, 4.5, and 8.6 times that of within A, B, and C, respectively, thus reflecting even larger differences in the composition of the microbiota between individuals at this finer taxonomic resolution.



(a) (b)

Figure 4.12: CA at genus (a) and phylotype (b) levels. Percentages correspond to the fraction of inertia explained by each axis. Samples are labelled with the name of the subject (A, B, C) and the day over the follow-up.

4.3.2.2 Within-subject variation

The communities experienced daily fluctuations in the bacterial diversity according to the Shannon index (**Tables 4.8** and **4.9**, **Figure 4.11(d)**), though these values varied around a rather constant level. This is consistent with the fact that the structure of the sampled faecal communities remained quite stable over time. Relative abundances showed daily fluctuations for all genera but no temporal trends were observed (**Figure 4.11(a)**, **4.11(b)**, and **4.11(c)**). All genera with an abundance of at least 5‰were permanent members of the communities, whereas genera with prevalence below that threshold were sometimes lost and recovered, though some were consistently detected too. The taxa with the highest prevalence showed the lowest relative fluctuations in time. In

A, *Alistipes*, *Bacteroides*, and *Barnesiella* experienced an average daily relative change of 2.9%, 3.8%, and 6.4% in their abundance, respectively, whereas in B and C, the average daily relative variation in *Prevotella* was ∼1.8%.

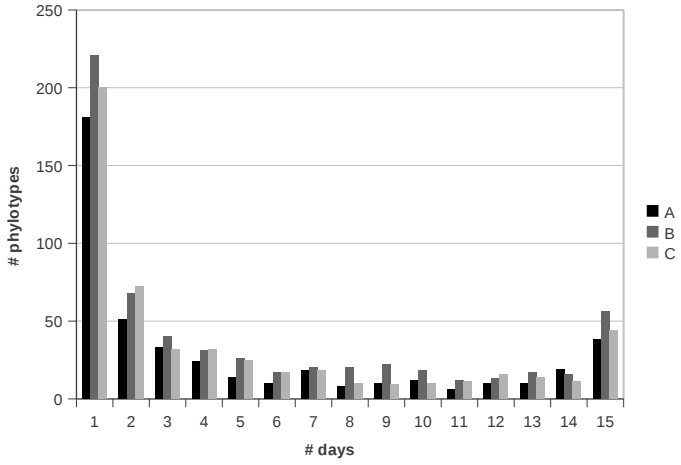
At the subgeneric level, a reduced number of phylotypes were consistently detected throughout within each subject (**Figure 4.13(a)**, which shows the number of phylotypes detected on just one day, on two days, etc.). The intra-individual core (phylotypes detected all the days during the study period) comprised 38 phylotypes in A, 56 in B, and 44 in C, which corresponded to approximately 9% of the total number of phylotypes within each subject. Conversely, most of the phylotypes detected within each subject were found in a few days only (**Figure 4.13(a)**). The core phylotypes belonged to the most prevalent genera and accounted for most of the sequences in the samples (74% of the total number of sequences in A, 90% in B, and 93% in C) (**Figure 4.13(b)**, which shows the cummulative percentage of sequences belonging to phylotypes detected on just one day, on two days, etc.). Again, phylotypes showed fluctuations in their abundance over time, but the trend was constant. Only 0.2% of the phylotypes were detected in all subjects all the days during the follow-ups, while 2.6% of the phylotypes were simultaneously detected throughout in the more similar subjects B and C. With a less restrictive definition of the bacterial core, we found that 0.5% of the phylotypes were detected in all three subjects in at least 13 out of 15 days.

4.3.2.3 Correlations between co-occurring genera

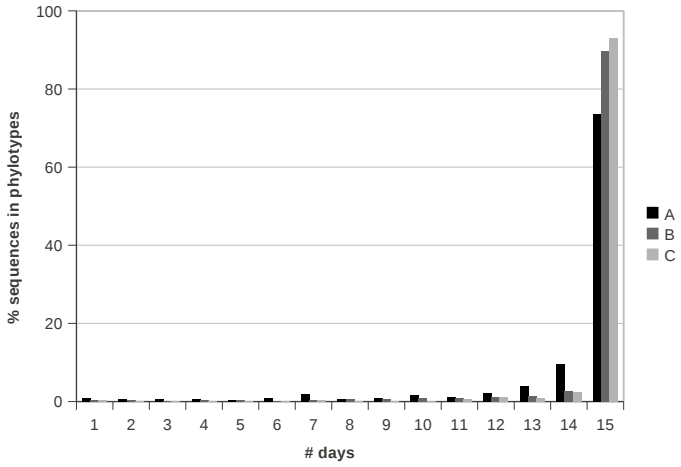
Using GGN, we built interaction networks from the statistically significant partial correlations between genera estimated with the Bayesian model. These networks are represented in a graph, where nodes correspond to genera and edges represent interactions (**Figure 4.14** shows the subgraphs of the networks including the statistically significant correlations).

Most genera showed correlations with a small number of other genera, while a small number of genera correlated with many. The main components of the faecal communities scarcely correlated with other genera. *Alistipes* correlated (positively) only with *Barnesiella* in A. *Prevotella* was negatively correlated with *Alistipes* in B, and with no genus in C. *Bacteroides*, the second genus in abundance, was correlated with five, three, and two genera in A, B, and C, respectively. The most connected genera were *Paraprevotella* in A, *Faecalibacterium*, *Alistipes*, and *Odoribacter* in B, and *Coprococcus*, *Escherichia/Shigella*, and *Blautia* in C. No significant correlations were found for many of the genera

in the samples.



(a)



(b)

Figure 4.13: Community structure at the phylotype level in subjects A, B, and C. a) Occurrence of phylotypes during the fifteen-day study period, i.e. number of phylotypes detected on just one day, on two days, etc. b) Combined average relative abundance of phylotypes detected on just one day, on two days, etc., computed over all samples where they occur.

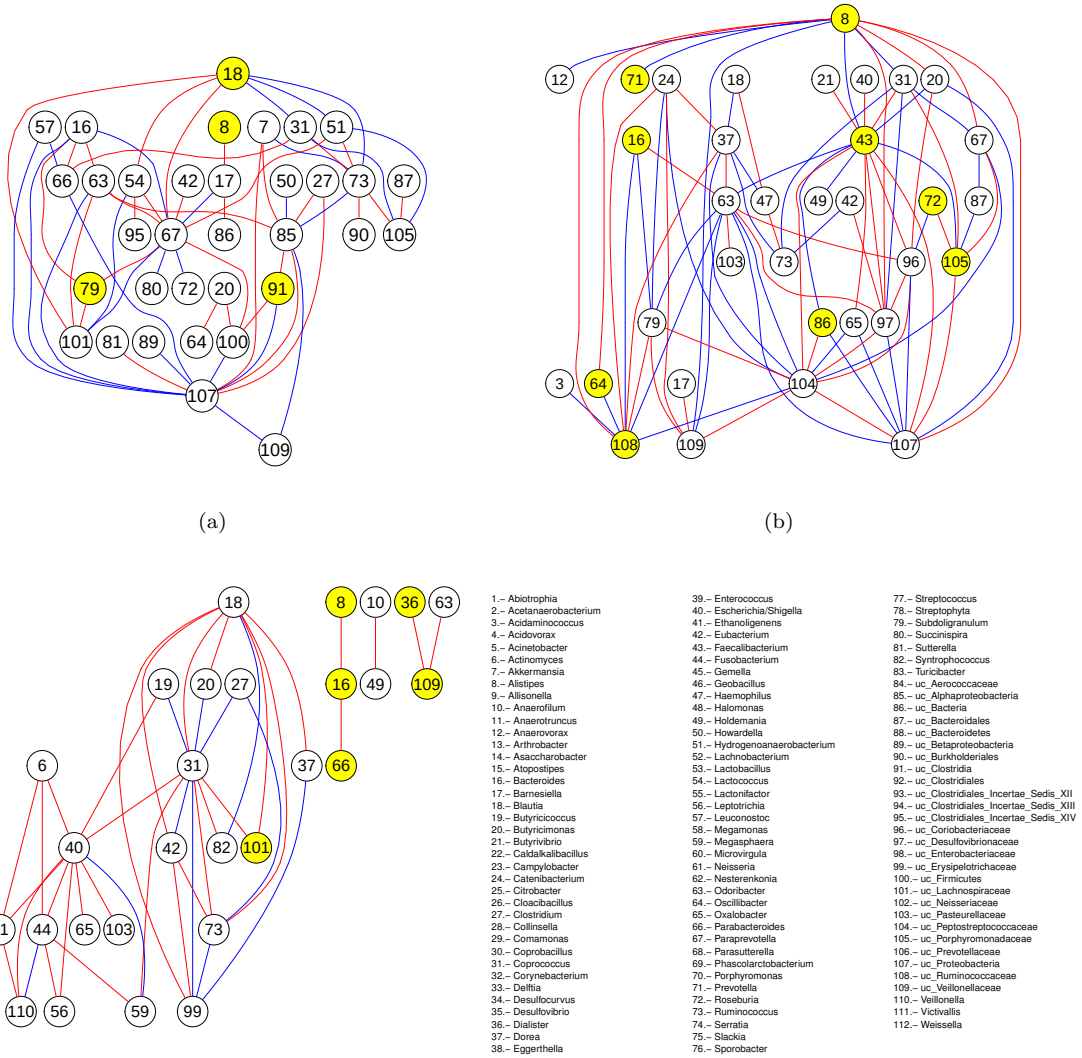


Figure 4.14: Graphical Gaussian networks representing the interactions between genera within subjects A (a), B (b), and C (c). Red and blue edges represent positive and negative partial correlations, respectively. Yellow nodes are those of genera with an average relative abundance $\geq 5‰$. Genera for which no significant partial correlations were found are not shown in the networks.

The sign of the correlation did not depend on the degree of phylogenetic relatedness. Both positive and negative correlations were found between closely and distantly related genera. For example, in subject A, *Paraprevotella*, a genus within Bacteroidetes-Bacteroidia-Bacteroidales-Prevotellaceae, had a positive correlation with *Odoribacter* and a negative correlation with *Barnesiella*, two genera within Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae; in subject C, *Butyricimonas*, a genus within Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae, had a positive correlation with *Blautia* and a negative correlation with *Coprococcus*, two genera within Firmicutes-Clostridia-Clostridiales-Lachnospiraceae.

The patterns of correlations were highly subject-specific, even for B and C, in which the community structure was quite similar. Typically, the same genus correlated with different genera in different individuals. For example, in subject B, *Bacteroides* co-occurred with *Odoribacter*, while in subject C, it co-occurred with *Parabacteroides* and *Alistipes*. Only five pairwise correlations were found in more than one individual: *Bacteroides-Odoribacter* (A and B), *Bacteroides-Parabacteroides* (A and C), *Odoribacter*-unclassified Proteobacteria (A and B), *Coprococcus-Ruminococcus* (A and C), and *Blautia*-unclassified Lachnospiraceae (A and C). Furthermore, nine pairwise correlations were positive in one subject but negative in another: *Blautia-Coprococcus*, *Blautia-Dorea*, *Blautia-Ruminococcus*, *Eubacterium-Ruminococcus*, *Coprococcus-Ruminococcus*, *Coprococcus*-unclassified Porphyromonadaceae, *Bacteroides-Subdoligranulum*, unclassified Bacteroidales-unclassified Porphyromonadaceae, and *Odoribacter*-unclassified Veillonellaceae.

We did not find a relationship between the diversity of the community and the complexity of the associated interaction network. Even though the bacterial community of A was the most diverse at the genus level (Shannon index $\sim$2.3), and the ones of B and C harboured a similar diversity (Shannon index $\sim$1.1), the highest number of genera in the network (the same as in A) and the largest average number of links per genus were observed in B, and the lowest ones in C.

4.3.3 Discussion

In this study, we monitored the faecal bacterial communities of three subjects during fifteen consecutive days. We analysed the daily variation in the composition of the microbiota and the degree of stability of specific members

of the community. The analysis of the time series also revealed patterns of correlations between the abundance profiles of specific bacterial groups, thus indicating potential cooperative and competitive interactions between them.

In accordance with current knowledge of the human gut microbiota, we found that the faecal microbiota is host-specific and fairly stable in the absence of perturbation, being the within-subject variability much smaller than that between subjects (Franks et al., 1998; Zoetendal et al., 1998; Vanhoutte et al., 2004; Caporaso et al., 2011; Dethlefsen and Relman, 2011). The level of prevalence of the predominant members of the faecal communities in our three study subjects was sustained over time, with abundances fluctuating around rather constant average values. This also held true for less abundant members, although some of them were not persistently detected in our sampling. In addition, a considerable number of bacterial groups were detected at very low abundance in one or few samples across the time series, probably representing transient members of the communities. Similar observations were made at genus and phylotype levels.

Some studies had examined the daily variation in the faecal microbiota using high-throughput technologies. Dethlefsen and Relman (2011) monitored the response of faecal microbiota to antibiotic perturbation in three subjects and took daily samples in the periods surrounding each antibiotic course. They found that just a few OTUs maintained uniform abundance, even in the antibiotic-free intervals. Caporaso et al. (2011) sampled daily three body sites, including faeces, in two healthy subjects for fifteen and six months, respectively. They reported that only a small fraction of the OTUs within a single body site was present across all or nearly all time points, and from this, they concluded that a minimal temporal core exists. In a shorter temporal window, Booijink et al. (2010) detected morning-afternoon variation in the human ileal microbiota that exceeded the fluctuations between samples collected at the same time point of the day. We also found that only a small fraction of the 97% OTUs in our samples was systematically present throughout the follow-up (9% of the total number of OTUs within each subject). However, as their combined abundance was high (74%, 90%, and 93% in subjects A, B, and C, respectively), we consider that a high-abundance core faecal microbiota exists also at the subgeneric level. In our view, this finding is indicative of a dynamic ecosystem with a stable core. In addition, it suggests that studies aimed at linking the prevalence of specific OTUs with various environmental exposures and/or disease risk should be limited to the numerically most dominant OTUs,

as the rest appears transient in the GI tract.

We used a Bayesian model to estimate the covariance matrix between relative abundances of taxa (on the log-odds scale) while accounting for the temporal autocorrelation in those. The posterior mean of the covariance matrix Σ (estimated using all the samples of a given individual) was used to estimate the partial correlation matrix, which was in turn the input for the GGN-based methods to detect associations between taxa. Although these methods cope well with sparse matrices of relatively large dimension, some of the statistically significant associations we found involve taxa with low relative abundances, which makes difficult any biological interpretation. In addition, functional studies would be needed to disentangle the biological meaning of the correlations detected between abundance profiles.

With regard to the inter-individual variation, the bacterial communities sampled in this study differed not only on the basis of their composition, but also in the correlation patterns found between their members. Barely a few of the correlations were shared between subjects. The vast majority of genera usually correlated with others that varied between subjects, even when the community composition was quite similar (subjects B and C). Furthermore, correlations of opposite sign were detected for some pairs of genera in different subjects. This disparity of community assemblies could imply that each individual may be offering somewhat different niches to the gut bacteria (regarding pH, temperature, secretions, retention time, etc.) in which the same species can establish different microbial interactions, even though the gut environment is overall similar in all subjects. At the same time, the identity of the interacting bacteria could be affected by host selection of commensal/mutualistic microbes, as well as by the order in which microbes arrive in the colonisation processes and selection by already established microbes (Dethlefsen et al., 2006; Van den Abbeele et al., 2011). Due to functional redundancy, several microorganisms can potentially occupy a specific niche within the GI habitats. The first ones to arrive can settle and then select for cooperative or non-overlapping microorganisms, as well as exclude competitors. Related to this, it has been reported a long-term impact of the early life environment on the composition of the intestinal microbiota in mice and pigs (Deloris Alexander et al., 2006; Mulder et al., 2009; Ubeda et al., 2012).

The persistent diversity and individuality of human gut communities could be explained by a combination of factors that vary between individuals, such as host genotype and diet, but also less predictable events, such as the coloni-

sation history during the community assembly and external perturbations with long-term effect on the gut microbiota, an example of which is antibiotic treatment. Several studies evaluating its effect on the gut microbiota have found an important loss of diversity followed by a rapid return to the pre-treatment community composition (De La Cochetière et al., 2005; Dethlefsen and Relman, 2011). The ability of these communities to recover their original structure suggests the existence of selective forces shaping them. The microbiota itself is thought to account for some of its diversity through the modification of the intestinal habitats and the interactions established between their members (Dethlefsen et al., 2006; Van den Abbeele et al., 2011). Our results suggest that specific microbial interactions are set within each individual, which may be an important factor contributing to the inter-individual variability, the resilience, and the temporal stability of the gut microbiota.

Recent studies have shown that despite the variation in the community structure between subjects, the normal fluctuations in the community composition detected over time, and the shifts due to disturbances, the overall community function seem to be maintained, which is indicative of some degree of functional redundancy among individual microbes and consortia within the gut microbiota (Turnbaugh et al., 2009a; Dethlefsen and Relman, 2011). Functional studies analysing the within-subject temporal variation under different conditions may help to better understand the contribution of the intestinal microorganisms to human well-being and disease.

This study has been published in FEMS Microbiology Ecology (Durbán et al., 2012a).

4.4 Follow-up of faecal microbial communities in the irritable bowel syndrome

4.4.1 Background

Over the past decade, there has been an accumulation of evidence suggesting a role of the gut microbiota in IBS, mainly coming from case-control studies based on molecular methods (Salonen et al., 2010). These have revealed qualitative and quantitative alterations in the composition of the intestinal microbiota of IBS patients when compared with that of healthy subjects, although no consensus has been reached regarding the association of specific microbes with IBS.

This lack of reproducibility may partly arise from differences in the applied molecular methods as well as in the depth and statistical power of the analyses. The detection of subtle alterations such as those that seem to characterise microbial dysbiosis in IBS is hampered by the relatively small size of the cohorts and the existence of many sources of variability in the composition of the microbiota unrelated to the intestinal disorder. Moreover, the detection of patterns may have been further difficulted by differences between studies (or even within a single study) in the nature of IBS patients regarding aetiology, symptomatology, etc. Longitudinal studies in which patients are followed over time can help to overcome the confusion introduced by the high inter-subject variation that is problematic in cross-sectional studies.

There are few longitudinal studies on IBS apart from those that address the improvement of symptoms by clinical trials targeting the GI microbiota (reviewed in Moayyedi et al. (2010); Basseri et al. (2011)). Mättö et al. (2005) assessed the long-term temporal variation in the faecal bacteria by PCR-DGGE analysis with samples taken every three months. They found more temporal instability in the predominant bacteria in IBS patients than in controls, although this result should be interpreted with caution because many of those IBS patients took antibiotics during the study. Later, Maukonen et al. (2006) re-analysed the same cohorts excluding those patients who had recently taken antibiotics. Focusing on the predominant clostridial populations and samples obtained six months apart, they found greater instability in IBS patients according to RNA-based RT-PCR-DGGE profiles, that is, in the metabolically active clostridial groups, but a similar stability to that of controls in DNA-based PCR-DGGE profiles. The authors speculated that instability might be

due to different symptomatic phases at the days of sampling (although they did not record symptom data), in line with what has been found between active and inactive phases in IBD patients (Seksik et al., 2003; Scanlan et al., 2006). In support of this idea, stabilisation of the faecal microbiota has been observed in IBS patients after a probiotic supplementation that alleviated the symptoms (Kajander et al., 2008).

The functions of the microbiota are an essential factor to consider in understanding the GI disorders. Tana et al. (2010) provided a feasible link between specific intestinal microbes, their products, and the features of IBS. They found higher levels of acetic acid and propionic acid, and also of *Veillonella* and *Lactobacillus*, bacteria that are known to produce these SCFAs, in the faecal samples of IBS patients compared with those of controls. Besides, higher levels of these acids were associated with worse GI symptoms and quality of life. Le Gall et al. (2011) applied a functional genomic approach in IBS. They explored the differences between IBS patients, UC patients, and healthy subjects in the composition and metabolites of the faecal microbiota (through PCR-DGGE and NMR spectroscopy, respectively). They also addressed the temporal stability with samples obtained at six-month intervals and detected a higher dynamics at the functional level in all groups, in line with the results reported by Maukonen et al. (2006) in IBS patients. Specific metabolites were associated with each group, but IBS condition could not be predicted from the metabolite profiles, unlike UC or control condition, similar to what happens at the compositional level (Section 4.2) (Qin et al., 2010). The paucity of data on the functional impact of the GI microbiota and its dysbiosis in IBS requires further investigation.

In this study, we explored potential alterations in composition and function of the faecal microbiota in IBS through the analysis of two female IBS patients with diarrhoea as predominant bowel habit at several points over a two-month period. An age-matched healthy control of one of them, who also shared her environment, was included. Faecal samples were collected every two days the first week and once a week thereafter. Additional samples were collected when patients reported acute symptoms. The faecal microbiota was analysed for the first time in IBS using whole-community metagenomics and metatranscriptomics. Self-reported symptom diaries allowed relating microbiological attributes to the presence and severity of symptoms.

4.4.2 Results

The global gene expression of the microbial communities in the samples of Control 1, Patient 1, and Patient 2 was explored through sequencing their metatranscriptomes. The metagenomes of some of the first samples of Patients 1 and 2 were also sequenced to evaluate the gene composition. **Table 4.10** contains information about the sequences obtained from the samples. 16S rRNA genes and 16S rRNAs identified in metagenomes and metatranscriptomes, respectively, were used to analyse the taxonomic composition, whereas protein coding genes in metagenomes and mRNAs in metatranscriptomes were used to analyse the functional composition. We tried to relate distinctive features of the microbial profiles to variation in symptoms.

4.4.2.1 Classification of samples based on symptoms

Each patient was characterised based on their bowel symptoms over the follow-up, given that they daily reported symptoms (see **Table 3.4** in Section 3). Both patients were diagnosed with IBS with diarrhoea as predominant bowel habit, but clearly differed in the symptoms they experienced. Patient 1 constantly complained about abdominal pain, abdominal distension, and defaecatory urgency, whereas Patient 2 had a high number of depositions with diarrhoea (stools of types 6 and 7 in the Bristol scale, **Figure 3.1**). This difference is illustrated by the PCA based on weekly summaries of symptoms over the follow-ups of Patient 1 and Patient 2 (**Figure 4.15**). In addition, Patient 1 remained rather stable over the follow-up, while Patient 2 went through phases of acute diarrhoea (see, for example, days 3, 43, and 55 in **Table 3.4**).

Within each patient, sampling days were classified based on symptoms (**Table 3.4**, **Figure 4.16**). Days with milder symptoms could be classified into two groups: those with no pain-distension-urgency (day 21 of Patient 1; days 1, 7, and 42 of Patient 2) and those with less stools-diarrhoea (days 1, 3, and 7 of Patient 1; day 35 of Patient 2). Days with severe symptoms were considered those in which the number of diarrhoeal stools was higher (days 14, 28, 37, and 42 of Patient 1; days 3 and 28 of Patient 2). Day 56 of Patient 2 was also classified as one with severe symptoms because the number of diarrhoeal stools in the surrounding days were among the highest.

As mentioned above, the differences between samples of Patient 1 were minimal regarding the number of diarrhoeal stools (two when symptoms were considered more severe, one when symptoms were considered mild). Patient 2,

in contrast, had several acute diarrhoeal phases (the worst on day 3, with eleven entirely liquid depositions).

This information was employed to interpret the comparison of samples based on attributes of the microbiota and to define which groups of samples would be compared.

Table 4.10: Characteristics of the metagenomic (DNA) and metatranscriptomic (cDNA) samples from the faecal microbial communities analysed in this study. Samples are labelled with the code of the volunteer (C1, P1, P2) and the sampling day over the follow-up. When there are two samples from a given individual in the same day, these are numbered after the day.

Library	Sample	No. of reads	Total base pairs (Mbp)	Average read length (nt)	No. of 16S rRNA reads	No. of reads with NCBI-nr hits	No. of reads with TIGRFAM hits	No. of reads with KEGG GENES hits
DNA	P1.1	262322	104.10	397	635 (0.24%)	173860 (66.28%)	81104 (30.92%)	68280 (26.03%)
DNA	P1.5	527021	209.22	397	1210 (0.23%)	216817 (41.14%)	137821 (26.15%)	108772 (20.64%)
DNA	P2.1	224546	89.04	397	533 (0.24%)	194123 (86.45%)	76722 (34.17%)	69637 (31.01%)
DNA	P2.3.1	294084	113.12	385	631 (0.21%)	144198 (49.03%)	79417 (27.00%)	68307 (23.23%)
DNA	P2.3.2	151553	41.29	273	264 (0.17%)	89511 (59.06%)	32757 (21.61%)	30791 (20.32%)
DNA	P2.4	284271	92.74	326	617 (0.22%)	164360 (57.82%)	82276 (28.94%)	76234 (26.82%)
DNA	P2.5	269569	106.16	394	837 (0.31%)	130164 (48.29%)	78305 (29.05%)	67767 (25.14%)
cDNA	C1.1	14382	2.37	165	2680 (18.63%)	210 (1.46%)	15 (0.10%)	10 (0.07%)
cDNA	C1.3	74136	9.74	131	18049 (24.35%)	1321 (1.78%)	69 (0.09%)	65 (0.09%)
cDNA	C1.5	39937	6.42	161	6640 (16.63%)	590 (1.48%)	65 (0.16%)	61 (0.15%)
cDNA	C1.7	94795	12.82	135	26047 (27.48%)	1808 (1.91%)	151 (0.16%)	132 (0.14%)
cDNA	C1.14	55149	7.95	144	18969 (34.40%)	545 (0.99%)	30 (0.05%)	18 (0.03%)
cDNA	C1.21	20204	2.44	121	4400 (21.78%)	644 (3.19%)	18 (0.09%)	28 (0.14%)
cDNA	C1.28	77314	10.94	142	19525 (25.25%)	1118 (1.45%)	109 (0.14%)	98 (0.13%)
cDNA	C1.37	82799	12.46	151	17383 (20.99%)	1669 (2.02%)	164 (0.20%)	149 (0.18%)
cDNA	C1.42	66079	8.34	126	15363 (23.25%)	1437 (2.17%)	291 (0.44%)	244 (0.37%)
cDNA	P1.1	35325	6.16	174	6499 (18.40%)	695 (1.97%)	67 (0.19%)	82 (0.23%)
cDNA	P1.3	6951	0.90	130	1083 (15.58%)	197 (2.83%)	34 (0.49%)	40 (0.58%)
cDNA	P1.5	62871	7.06	112	11674 (18.57%)	2688 (4.28%)	313 (0.50%)	321 (0.51%)
cDNA	P1.7	33687	4.09	122	5696 (16.91%)	1325 (3.93%)	97 (0.29%)	76 (0.23%)
cDNA	P1.14	70913	7.16	101	10739 (15.14%)	2882 (4.06%)	116 (0.16%)	110 (0.16%)
cDNA	P1.21	20877	2.37	113	4555 (21.82%)	1192 (5.71%)	47 (0.23%)	22 (0.11%)
cDNA	P1.28	51801	7.86	152	5546 (10.71%)	1706 (3.29%)	361 (0.70%)	322 (0.62%)
cDNA	P1.37	35206	5.75	163	4246 (12.06%)	1259 (3.58%)	272 (0.77%)	309 (0.88%)
cDNA	P1.42	12966	1.40	108	2460 (18.97%)	428 (3.30%)	52 (0.40%)	64 (0.49%)
cDNA	P2.1	54323	6.98	129	14948 (27.52%)	4237 (7.80%)	534 (0.98%)	609 (1.12%)
cDNA	P2.3.1	77316	11.47	148	19843 (25.66%)	12479 (16.14%)	256 (0.33%)	277 (0.36%)
cDNA	P2.3.2	149422	15.97	107	19750 (13.22%)	15410 (10.31%)	187 (0.13%)	320 (0.21%)
cDNA	P2.4	40936	4.87	119	3432 (8.38%)	3174 (7.75%)	1159 (2.83%)	1333 (3.26%)
cDNA	P2.5	20505	2.33	114	2932 (14.30%)	2899 (14.14%)	664 (3.24%)	950 (4.63%)
cDNA	P2.7	105818	11.25	106	23381 (22.10%)	7093 (6.70%)	929 (0.88%)	1170 (1.11%)
cDNA	P2.14	121556	19.16	158	47851 (39.37%)	6596 (5.43%)	874 (0.72%)	948 (0.78%)
cDNA	P2.21	30382	4.01	132	5905 (19.44%)	4192 (13.80%)	107 (0.35%)	96 (0.32%)
cDNA	P2.27	89030	12.58	141	11833 (13.29%)	3924 (4.41%)	197 (0.22%)	213 (0.24%)
cDNA	P2.28	45138	4.73	105	7043 (15.60%)	5334 (11.82%)	610 (1.35%)	694 (1.54%)
cDNA	P2.35	29424	2.68	91	4799 (16.31%)	3070 (10.43%)	379 (1.29%)	448 (1.52%)
cDNA	P2.42	15260	1.05	69	161 (1.06%)	488 (3.20%)	27 (0.18%)	41 (0.27%)
cDNA	P2.49	56892	6.61	116	8627 (15.16%)	4555 (8.01%)	495 (0.87%)	555 (0.98%)
cDNA	P2.56	38022	4.09	108	2836 (7.46%)	1885 (4.96%)	254 (0.67%)	522 (1.37%)

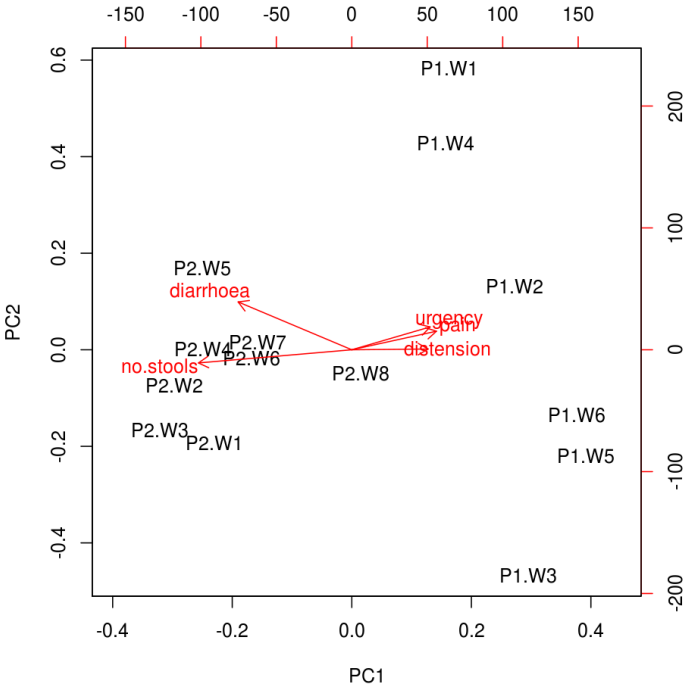


Figure 4.15: PCA on the summary of symptoms for each week over the follow-ups. Each point represents one week (W1-W8) of Patient 1 (P1) or Patient 2 (P2).

4.4.2.2 Dynamics of the microbial taxonomic profile

The similarity between metatranscriptomic samples according to the distribution of microbial families was assessed with DCA. When analysing the sample composition estimated from the 16S rRNAs (**Figure 4.17, top**), the first DCA axis separated Patient 2 from Patient 1 and Control 1. Samples of Control 1 were mixed with those of Patient 1. However, the last three samples of Patient 1 (taken on days with more severe symptoms) were slightly separated from the rest and closer to samples of Patient 2. Within each patient, samples taken on days with milder symptoms were quite similar to each other, and most samples taken when symptoms were more severe differed from those taken on days with mild/moderate symptoms. In Patient 2, the metabolically active microbiota on days with severe symptoms differed markedly from each other. We also analy-

sed the distribution of active bacteria estimated from the taxonomic affiliation of mRNAs. These profiles differed significantly from those obtained from the 16S rRNAs. These discrepancies could be due to the difference in the reference database (RDP database for 16S rRNAs, NCBI-nr protein database for mRNAs) and the assignment method (RDP Bayesian classifier for 16S rRNAs, BLAST plus LCA for mRNAs). However, the patterns of between- and within-subject variation observed with both procedures were similar (**Figure 4.17**).

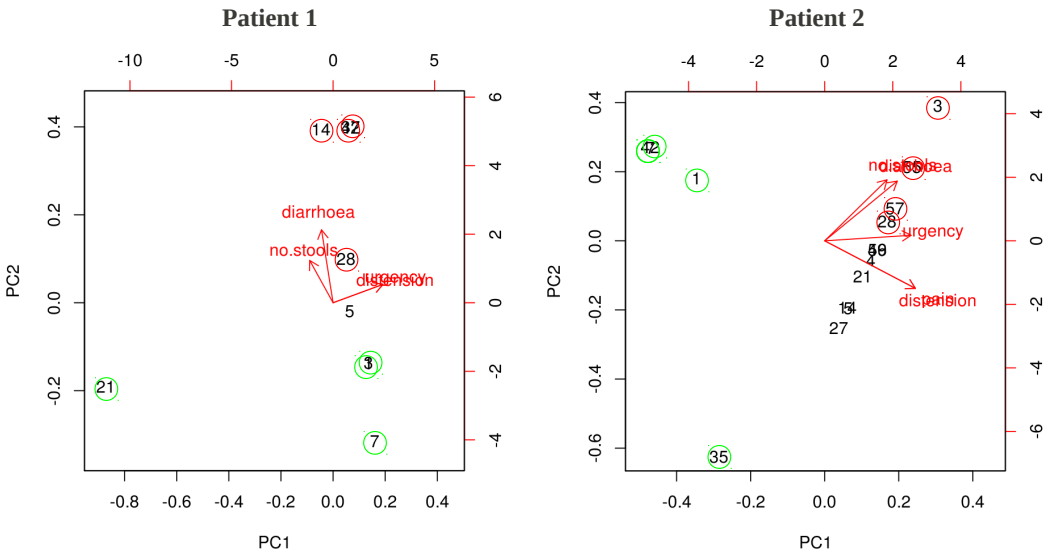


Figure 4.16: PCA on the symptoms at days of sampling. Samples are labelled with the sampling day over the follow-up. Days with milder symptoms are highlighted in green, whereas those with severe symptoms are in red.

The major groups of metabolically active faecal bacteria were similar and remained quite constant over time in Control 1 and Patient 1 (**Figure 4.18**). Firmicutes and Bacteroidetes, mainly within the Clostridia and Bacteroidia classes, provided the largest number of 16S rRNAs in the communities of these subjects (about 90% of the total number of sequences). The ratio Firmicutes/Bacteroidetes was greater than 1 in all samples (except in the last one of Patient 1). The active bacteria belonged, in decreasing order of abundance, to the Bacteroidaceae, Prevotellaceae, and Porphyromonadaceae families within the Bacteroidia class, and to Lachnospiraceae and Ruminococcaceae within the

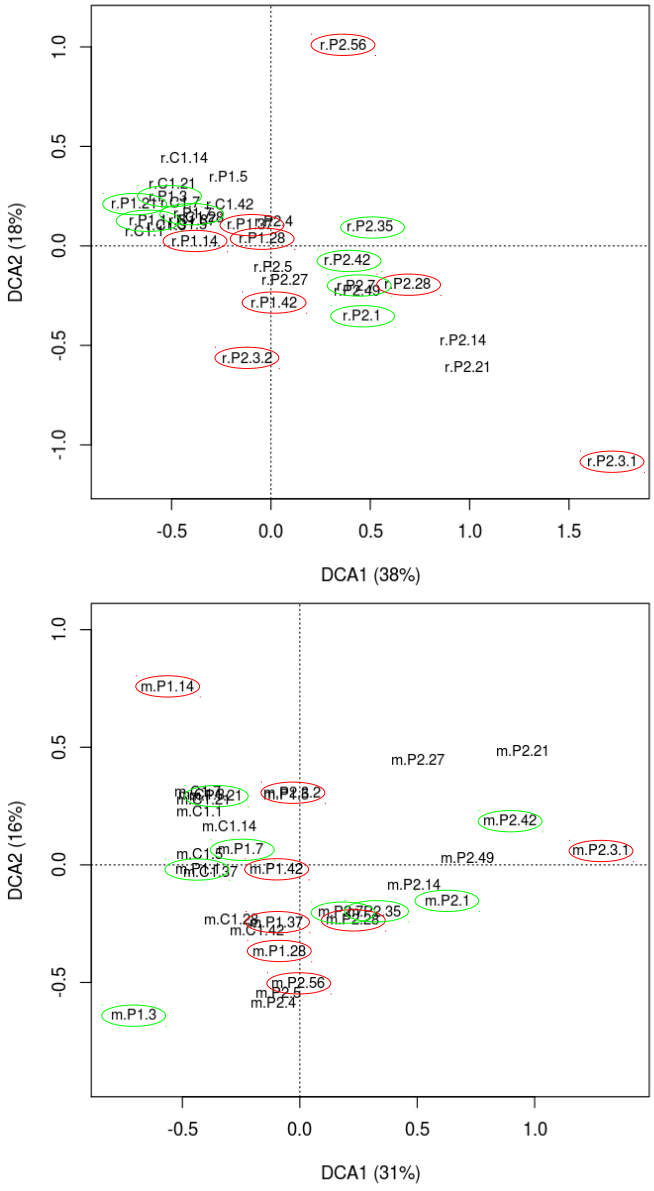


Figure 4.17: Taxonomic differences between the metatranscriptomes of Control 1, Patient 1, and Patient 2. DCA of the distribution of microbial families estimated from 16S rRNAs (top) or mRNAs (bottom). Percentages correspond to the fraction of inertia explained by each axis. Samples are labelled with the code of the volunteer (C1, P1, P2) and the sampling day over the follow-up. When there are two samples from a given individual in the same day, these are numbered after the day. Days with milder symptoms are highlighted in green, whereas those with severe symptoms are in red.

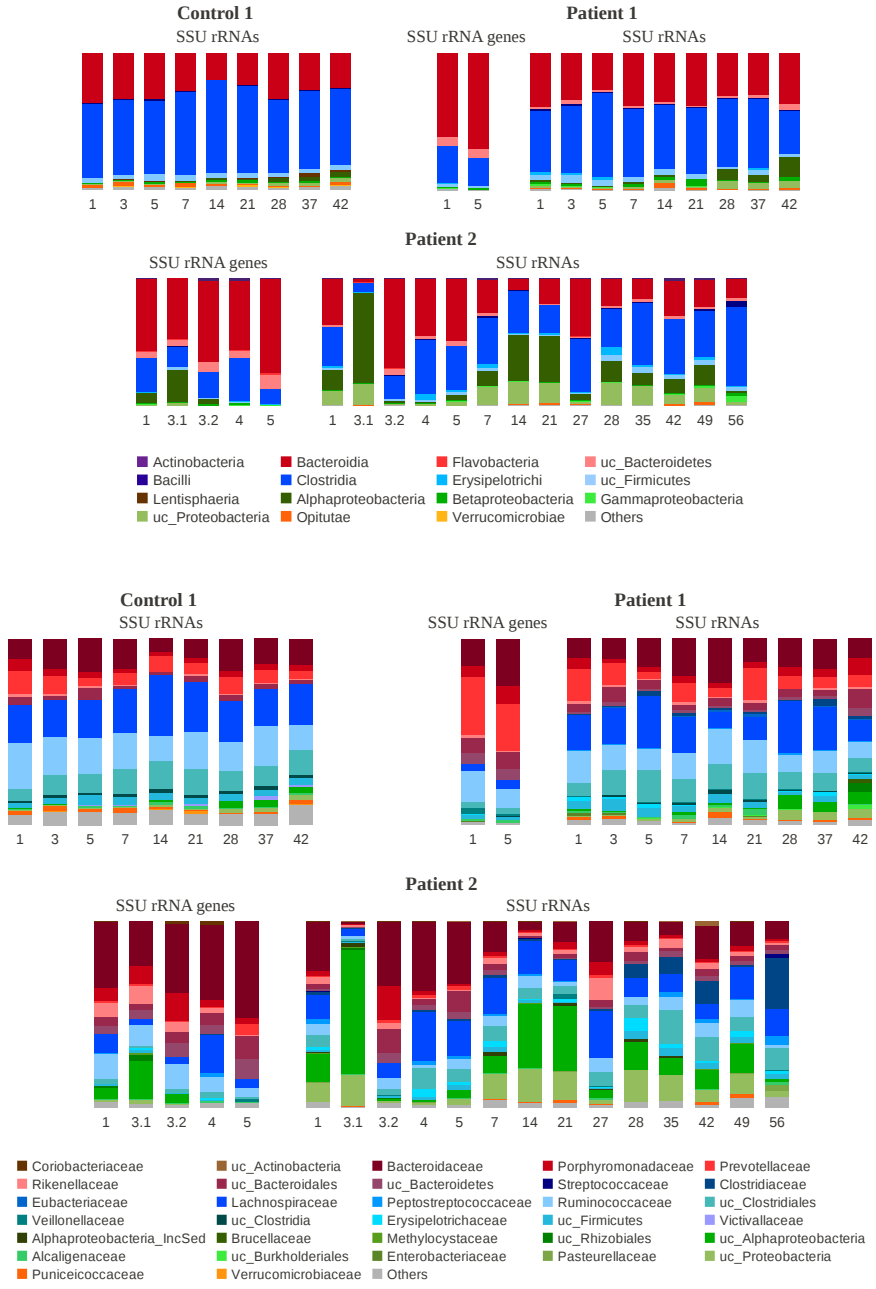


Figure 4.18: Relative abundance of microbial classes (top) and families (bottom) in the metagenomes (16S rRNA genes) and the metatranscriptomes (16S rRNAs) of Control 1, Patient 1, and Patient 2. The sampling day is indicated below each column. When there are two samples from a given individual in the same day, these are numbered after the day.

Clostridia class. Among the similarities of these subjects, there was an increase in unclassified Alphaproteobacteria and unclassified Proteobacteria the last weeks of their follow-ups, which might reflect shared environmental factors affecting the microbiota.

Conversely, the faecal microbiota of Patient 2 was characterised by great variation over time in the composition of the metabolically active bacteria (**Figure 4.18**). No temporal trends were observed in the relative abundance of active bacteria. A prominent feature of Patient 2 was the high activity of Proteobacteria some days, mainly unclassified Alphaproteobacteria and unclassified Proteobacteria. The proportion of Proteobacteria and the ratio Firmicutes/Bacteroidetes varied widely over time in the active fraction of the microbiota. Also, the fraction of each bacterial family within the Clostridia and the Bacteroidia classes experienced greater variation than in Control 1 and Patient 1 (see, for example, the amount of Clostridiaceae or Rikenellaceae over the follow-up). No correlation was found between the level of activity of Proteobacteria and the severity of symptoms.

The increases in the relative abundance of Proteobacteria were not due to a single or a few species, as demonstrated by the detection of multiple phylotypes mapped along reference 16S rRNA genes after clustering the sequences affiliated to Proteobacteria at 97% of sequence identity.

Strong and quick compositional shifts were associated with acute diarrhoea in Patient 2. She provided a faecal sample in the morning on day 3. Then, she began to feel more abdominal discomfort and increased stool frequency, which were increasingly more watery, and took another sample that afternoon. The change in the distribution of active bacteria was very pronounced between the pairs of consecutive samples taken the first days of her follow-up, which included the day with acute diarrhoea, even between the samples collected that day a few hours apart.

Thus, temporal instability in the distribution of active bacteria in faeces was associated with the IBS condition (greater instability in patients than in Control 1) and with severe diarrhoea (greater instability in Patient 2 than in Patient 1). It was reflected in the distances between samples within each subject. ANOSIM based on the distribution of bacterial classes revealed that the median rank of distances between samples of Patient 1 was 1.8 times that of between samples of Control 1, but less than one-third of that found between samples of Patient 2. At the family level, the median rank of distances within Patient 1 was four times that of within Control 1, and half of that found within

Patient 2.

The alterations in the active microbiota detected in association with worsening of symptoms were barely consistent between patients and within a single patient in different days. In Patient 2, the morning sample of day 3 had the highest level of activity of Alphaproteobacteria of all samples, the afternoon sample, the highest level of total Bacteroidia and Porphyromonadaceae, and sample of day 56, the highest level of Streptococcaceae, Clostridiaceae, Betaproteobacteria, and Gammaproteobacteria. In Patient 1, sample of day 14 had an increase in Verrucomicrobia, while samples of days 28, 37, and 42, an increase in unknown members of Alphaproteobacteria and Proteobacteria. **Table 4.11** shows all significant differences in the level of activity of bacterial families between days with more severe symptoms and days with milder symptoms within each patient.

Table 4.11: Significant differences detected in chi-square tests in the relative abundance of bacterial families between the metatranscriptomes of Patient 1 (P1) or Patient 2 (P2) on days with severe symptoms and days with milder symptoms. Some samples were grouped by similarity of symptoms. An OR greater than one indicates overrepresentation in days with severe symptoms, whereas an OR lower than one indicates underrepresentation in days with severe symptoms. *: comparison between afternoon (3.2) and morning (3.1) samples.

Patient	P1	P1	P2	P2	P2	P2	P2
Days with severe symptoms	14	28,37,42	3.1	3.2	28	56	3.2*
Days with milder symptoms	1,3,7,21	1,3,7,21	1,7,35,42	1,7,35,42	1,7,35,42	1,7,35,42	3.1*
Bacteroidaceae (Bacteroidia)	1.95		0.06	2.39	0.54	0.48	38.92
Porphyromonadaceae (Bacteroidia)	0.56	1.32	0.24	9.78		0.47	39.98
Prevotellaceae (Bacteroidia)	0.31	0.39	0.06				20.84
Rikenellaceae (Bacteroidia)	0.29	0.51	0.22			0.34	4.53
uc Bacteroidales (Bacteroidia)		1.27	0.09	4.52	1.45		47.92
uc Bacteroidetes		1.51	0.21	2.61			12.23
Streptococcaceae (Bacilli)						31.85	
Clostridiaceae (Clostridia)		2.62	0.07	0.02	2.84	13.91	
Eubacteriaceae (Clostridia)	2.24						
Lachnospiraceae (Clostridia)	0.45	1.41	0.21	0.43	0.59		2.08
Peptostreptococcaceae (Clostridia)			0.21	0.04	0.21	2.72	
Ruminococcaceae (Clostridia)	1.17	0.59	0.22		0.67	0.34	4.83
uc Clostridiales (Clostridia)	1.26	0.80	0.15	0.36		1.40	2.43
uc Clostridia (Clostridia)	2.49						
Erysipelotrichaceae (Erysipelotrichi)			0.11	0.07	3.00		
uc Firmicutes		0.61	0.14	0.49	1.82		3.52
α-proteobacteria IncSed (α-proteobacteria)			1.56	0.02		0.00	0.01
uc α-proteobacteria (α-proteobacteria)		6.44	16.15	0.09	1.44	0.10	0.01
Alcaligenaceae (β-proteobacteria)	0.60	0.12				8.72	
Pasteurellaceae (γ-proteobacteria)						171.70	
uc Proteobacteria	2.04	4.98	1.41	0.06	1.37	0.21	0.04
Puniceicoccaceae (Opitutae)	3.36						
uc Verrucomicrobia	8.66						

The distributions of bacterial taxa obtained from the 16S rRNA genes were different from those obtained from the 16S rRNA transcripts, indicating a different contribution to the active fraction of the predominant faecal microbiota (**Figure 4.18**). For example, Bacteroidia was the predominant class and accounted for more than half of the sequences in the metagenomes but not in the metatranscriptomes. As could be expected, temporal dynamics was greater in the metatranscriptomes than in the metagenomes. In our case, the study of metatranscriptomes allowed to detect changes associated with an acute phase of the syndrome that were subtler at the metagenomic level (see changes on day 3 in Patient 2).

4.4.2.3 Dynamics of the microbial functional profile

Sequences of putative protein coding genes in the metagenomes and mRNAs in the metatranscriptomes were assigned to functional roles and pathways through homology searches in the TIGRFAM and KEGG GENES databases.

Figure 4.19 shows the distribution of broad functional categories in the TIGRFAM and KEGG functional hierarchies for the metagenomes of Patient 1 and Patient 2. The encoded functions were highly conserved between subjects and over time (within a few days) within each subject. Given the different community assemblies found in each sample, it supports the concept of functional redundancy among their constituent taxa.

While most sequences in metatranscriptomes were annotated as rRNA, only 5.6% had hits in the NCBI-nr protein database, and 0.5% in the KEGG GENES and the TIGRFAM databases. 1.4% of the sequences had hits in the Rfam database, while 13% had no hits in any of the reference databases. Samples with less than 200 sequences with an assigned functional role were discarded for analysis due to the great uncertainty in the estimation of the distribution of functions.

Temporal variation in the functional profile was higher at the gene expression level than at the genomic level (**Figure 4.20**), but lower than variation in the profile of active microbes. It also points to functional redundancy among individual microbes and/or consortia in the faecal samples.

The functional profiling based on broad functional categories provides limited information, so we analysed the distributions of intermediate categories in the same functional hierarchies. CA plots separated metagenomic samples of Patient 1 and Patient 2 (**Figure 4.21(a)**, **4.21(b)**). CA plots of metatranscriptomes gave a picture similar to that found with the distribution

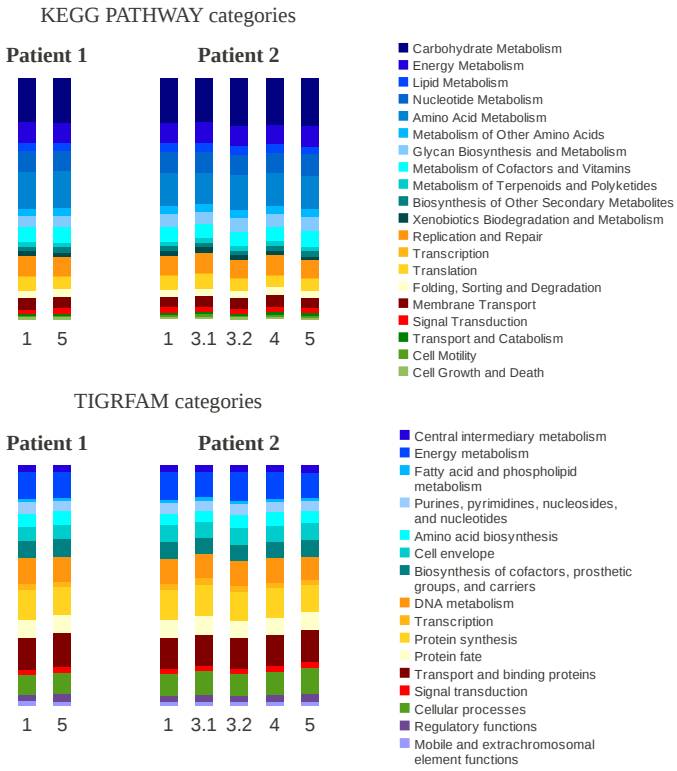


Figure 4.19: Relative abundance of broad functional categories in the KEGG PATHWAY (top) and the TIGRFAM (bottom) hierarchies in the metagenomes of Patient 1 and Patient 2. The sampling day is indicated below each column. When there are two samples from a given individual in the same day, these are numbered after the day.

of active bacteria, being samples of Patient 1 taken on days with more severe symptoms the closest to samples of Patient 2 (**Figure 4.21(d)**), and samples of Patient 2 collected during acute phases usually the most different from the rest and from each other (**Figure 4.21(c)**, **4.21(d)**).

We used the LEfSe method to identify functional roles or pathways characterising the differences between the faecal microbiota of Patient 1 and Patient 2, and the differences within each patient between the microbiota on days with severe and mild/moderate symptoms. No differentially abundant features were found between the metagenomes of Patient 1 and Patient 2, or between the metatranscriptomes of Patient 1 in association with worsening of symptoms.

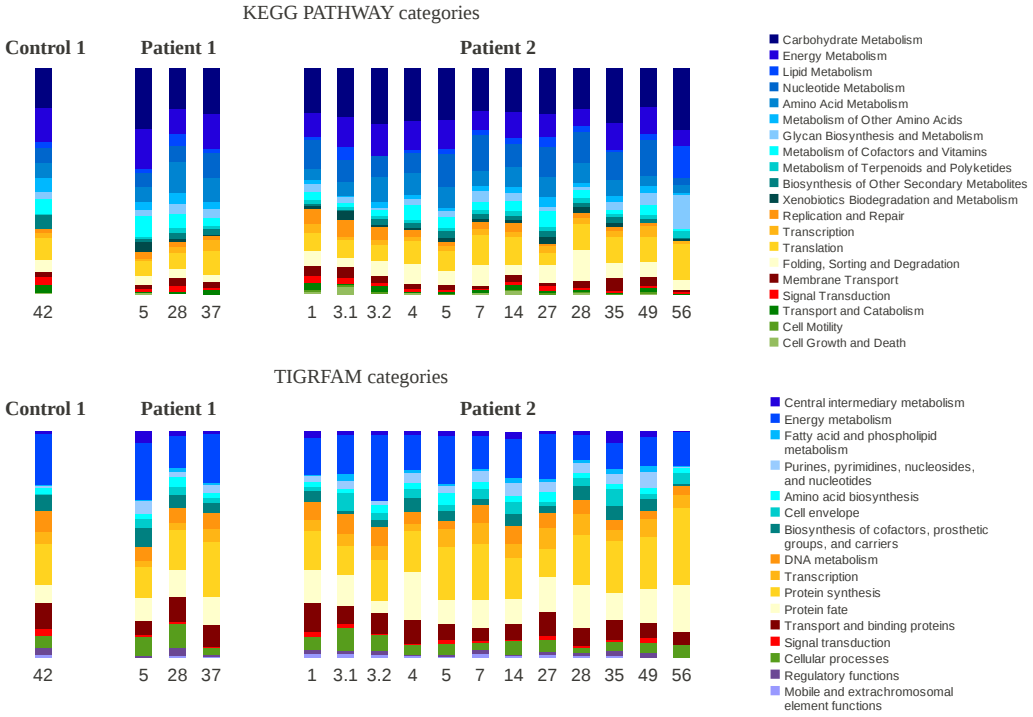


Figure 4.20: Relative abundance of broad functional categories in the KEGG PATHWAY (top) and the TIGRFAM (bottom) hierarchies in the metatranscriptomes of Control 1, Patient 1, and Patient 2. The sampling day is indicated below each column. When there are two samples from a given individual in the same day, these are numbered after the day.

Table 4.12 shows the significant differences between the metatranscriptomes of Patient 1 and Patient 2, and between days with severe and mild/moderate symptoms within Patient 2. Although statistically significant, these few differences are difficult to interpret considering the limited number of samples per group and the rather low number of sequences per sample, which also complicate the detection of significant changes. These limitations were greater when we tried to detect temporal changes in the expression profile of specific taxa, since the number of annotated sequences within each taxon was really small (except for *Bacteroidaceae*, in which only the reduction in "prophage functions" was associated with acute symptoms in Patient 2).

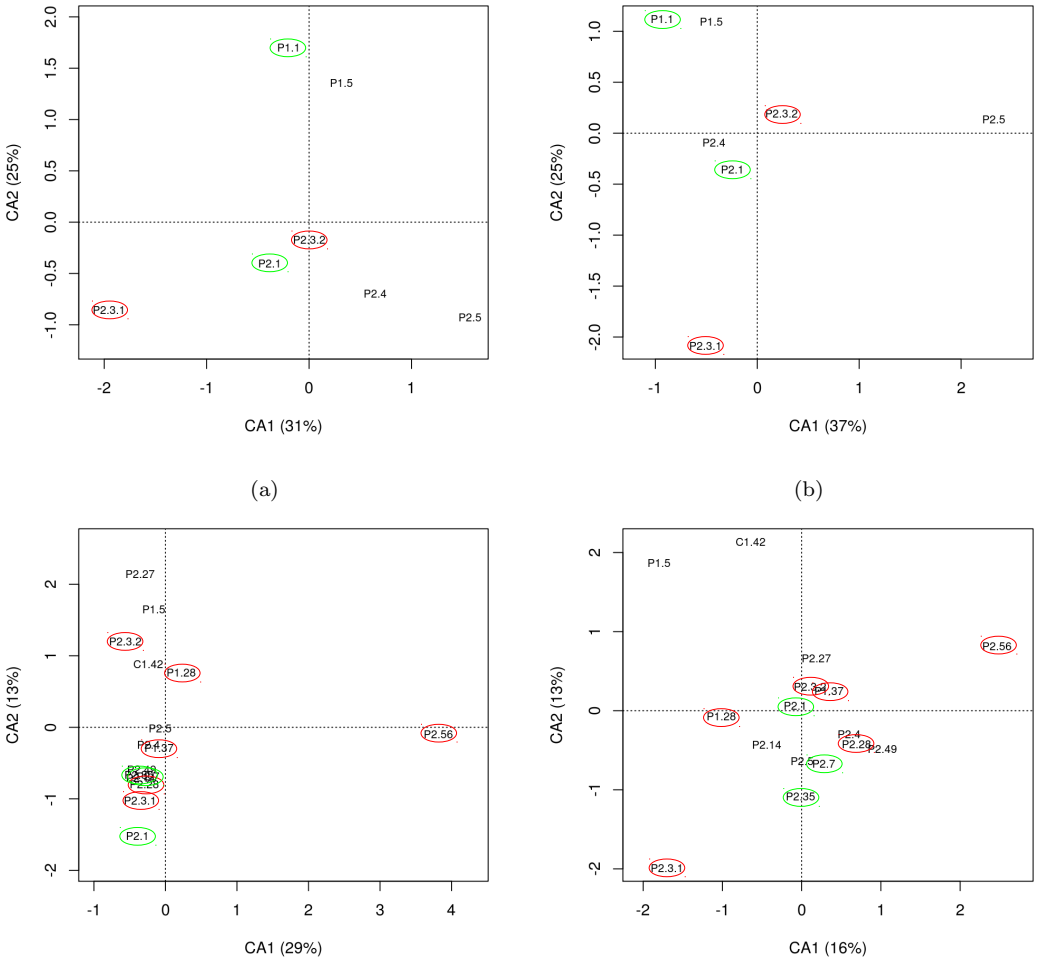


Figure 4.21: Functional differences between samples of Control 1, Patient 1, and Patient 2. CA of the distribution of KEGG metabolic pathways (a, c) and TIGRFAM functional roles (b, d) in the metagenomes (a, b) and the metatranscriptomes (c, d). Percentages correspond to the fraction of inertia explained by each axis. Samples are labelled with the code of the volunteer (C1, P1, P2) and the sampling day over the follow-up. When there are two samples from a given individual in the same day, these are numbered after the day. Days with milder symptoms are highlighted in green, whereas those with severe symptoms are in red.

Table 4.12: Significant differences detected in LEfSe analyses in the relative abundance of functional categories between the metatransciptomes of Patient 1 (P1) and Patient 2 (P2), and between the metatranscriptomes of Patient 2 on days with severe and mild/moderate symptoms.

	KEGG metabolic pathways
P1 > P2	Energy metabolism/Nitrogen metabolism
	Metabolism of cofactors and vitamins/Riboflavin metabolism
	Metabolism of terpenoids and polyketides/Biosynthesis of ansamycins
	Metabolism of terpenoids and polyketides/Limonene and pinene degradation
	Xenobiotics biodegradation and metabolism/Ethylbenzene degradation
	Xenobiotics biodegradation and metabolism/Naphthalene and anthracene degradation
P1 < P2	Folding, sorting, and degradation/RNA degradation
	Xenobiotics biodegradation and metabolism/Drug metabolism - other enzymes
mild > severe	Biosynthesis of other secondary metabolites/Phenylpropanoid biosynthesis
(within P2)	Carbohydrate metabolism/Amino sugar and nucleotide sugar metabolism
	Glycan biosynthesis and metabolism/Glycosaminoglycan degradation
	Glycan biosynthesis and metabolism/Glycosphingolipid biosynthesis
	Metabolism of other amino acids/Taurine and hypotaurine metabolism
	Xenobiotics biodegradation and metabolism/Drug metabolism - other enzymes

	TIGRFAM functional roles
P1 > P2	Biosynthesis of cofactors, prosthetic groups, and carriers/Riboflavin, FMN, and FAD
	Purines, pyrimidines, nucleosides, and nucleotides/Salvage of nucleosides and nucleotides
	Transport and binding proteins/Anions
P1 < P2	Transcription/Degradation of RNA
mild > severe	Amino acid biosynthesis/Histidine family
(within P2)	Cell envelope/Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides
	Central intermediary metabolism/Nitrogen metabolism

4.4.3 Discussion

We analysed the temporal changes in structure and function of faecal microbial communities and the evolution of GI symptoms and bowel habits in two IBS patients with diarrhoea and one healthy control. We detected a greater temporal variation in the patients, and even larger instability associated with acute diarrheoa in Patient 2. Previously, Maukonen et al. (2006) found greater instability in IBS patients than in controls in the 16S rRNA levels of some clostridial groups. It is confirmed in our study, where, in addition, the stability of the whole community was evaluated. The abnormal instability could reflect a loss of homeostasis, where the community is unable to maintain its struc-

ture. Moreover, a high degree of temporal variation is typical of disturbed and re-establishing communities, as seen in gut microbiota subject to treatment with antibiotics (De La Cochetière et al., 2005; Dethlefsen and Relman, 2011) or short-term dietary alterations (Section 4.5) (Wu et al., 2011). It has also been found in the faecal communities of patients with recurrent *Clostridium difficile*–associated diarrhoea (CDAD) (Chang et al., 2008).

Previous studies reported differences between IBS patients and healthy controls in the abundance of specific bacterial groups in the gut (Salonen et al., 2010), including our study presented in Section 4.2, but could not clearly discriminate IBS cases from controls. In contrast, IBDs can be linked to pronounced dysbioses, and patients can be distinguished from healthy individuals based on the composition of the gut microbiota (Qin et al., 2010). In this study, the differences in the distribution of active microorganisms in the samples discriminated Patient 2 from Control 1 and Patient 1. The last two shared environmental factors and were of similar age, which undoubtedly contributed to their differentiation from Patient 2. However, the mild symptoms experienced by Patient 1 compared with Patient 2, together with the large alterations in the microbiota of Patient 2 concomitant with changes in the intensity of symptoms, support that the singularities of Patient 2 were also related to the disorder.

We focused on the study of microbial activity (at the level of gene expression) because there are few functional studies on IBS and because we found greater variation between and within subjects at this level than at the genomic level. Similarly, Le Gall et al. (2011) reported greater temporal variation in the profile of metabolites of the gut microbiota than in its taxonomic composition. It is not surprising that a given community exhibits temporal variation in its *in situ* activity in a changing environment such as the GI tract in response to the available resources, growing conditions, environmental stressors, host physiology, etc. Therefore, to minimise the sources of variation and to make the samples more comparable, our volunteers kept their routine lifestyle habits and took samples at a defined time slot each sampling day during the follow-up.

We observed compositional shifts in the active fraction of the microbiota associated with worsening of symptoms, although with low reproducibility between and within patients. Changes in the pattern of gene expression also characterised days with acute symptoms in Patient 2. Unfortunately, we only had a small number of sequences with a functional annotation in the metatranscriptomes due to methodological limitations (because of the difficulty to

enrich mRNAs prior to sequencing and the small length of non-rRNA sequences that makes difficult a confident assignment in homology searches). This leads to considerable uncertainty in the estimated distributions of functional categories and to an overestimation of the variation between samples, so these results should be interpreted with caution.

Another limitation of our study is that patients experienced little variation in the severity of symptoms throughout. Further studies should follow more patients and over longer periods to increase the chance of collecting samples in acute phases. Moreover, sampling should be adapted to collect samples in phases of relapse and remission of symptoms, which can occur within a short time and can be associated with equally quick changes in the microbiota, as happenned in this study.

Cohort studies on IBS usually take into account the classification of patients into subtypes according to the primary symptom that patients experience (diarrhoea, constipation, or an alternation between both). In addition, it would be helpful to have a finer classification of patients based on the pattern and intensity of symptoms, which can vary widely among patients, as occurred between the four patients initially included in this study. Furthermore, our results reveal the importance of knowing if there are symptoms and their severity at the time of sampling due to the characteristic fluctuation in IBS-related symptoms, as it is done in IBDs when active and inactive phases of the disease are distinguished (Seksik et al., 2003; Scanlan et al., 2006; Sokol et al., 2006; Martinez et al., 2008). Otherwise, heterogeneity within patients could complicate the detection of alterations in the gut microbiota.

Interestingly, some episodes of diarrhoea were associated with a significant increase in Proteobacteria in faeces, which are normally found in small proportions in the faeces of healthy individuals, as assessed by sequencing of metagenomes and metatranscriptomes (Arumugam et al., 2011; Gosalbes et al., 2011). There are no estimates of the microbial distribution in the intestinal mucosa by direct sequencing, but our studies based on 16S rRNA gene sequencing show that the prevalence of Proteobacteria in mucosal biopsies is much higher than in faeces (Sections 4.1 and 4.2). One possible explanation is that the Proteobacteria that we found in these faecal samples, and maybe other bacteria, represent, at least in part, bacteria detached from the mucosa by the acute diarrhoea.

Diversity analyses, using both culture and molecular methods, have demonstrated an expansion of members of the Enterobacteriaceae family in faeces and

intestinal mucosa of IBD patients (Seksik, 2010), as well as in faecal samples of IBS patients (Si et al., 2004; Krogius-kurikka et al., 2009), when compared with those of healthy controls. Specific pathogenic Proteobacteria (adherent-invasive *Escherichia coli*, *Campylobacter concisus*, enterohepatic *Helicobacter*) have also been found in association with IBD (Mukhopadhya et al., 2012). However, we did not detect these abnormalities in our IBS patients, but the levels of unknown members of Alphaproteobacteria and Proteobacteria were unusually high.

Cross-sectional studies on IBS have not found consistent alterations in the gut microbiota. Similarly, the alterations we found in this study associated with acute symptoms are not systematic within a single patient. We consider it unlikely that dysbiosis is the underlying cause of the development of symptoms in IBS. Other mechanisms may trigger the acute phases, for example, stress. Animal studies suggest that psychological stress can change the composition of the microbiota via perturbation of the normal GI habitat (Collins and Bercik, 2009). Besides, IBS sufferers tend to have a low threshold for coping with stressful situations and a high incidence of psychiatric comorbidity (Drossman, 1999a; Hungin et al., 2003; Lea and Whorwell, 2003; Hood et al., 2008). Thus, stress may promote the alterations of the GI function observed in IBS, and, subsequently, the alterations of the GI microbiota. Taken as a whole, our results suggest that the association of the intestinal microbiota with IBS is rather weak. We have proved that IBS is associated with a decrease in the stability (this study) and a decrease in the biodiversity (Section 4.2) of the gut microbiota. Regardless of whether the community imbalances are a cause or a consequence of symptom development, treatments that promote the stabilisation of the gut microbiota would be helpful in IBS.

4.5 Response of a *Prevotella*-dominated human faecal microbiota to a ketogenic diet

4.5.1 Background

Recent evidence suggests that the gut microbiota may play an important role in the pathogenesis of obesity (see Section 1.2.2.5). We found that the obese subjects included in Section 4.3 had a low bacterial diversity compared to other individuals in our surveys. Such a decrease in the level of diversity had been previously associated with obesity (Turnbaugh et al., 2009a). In our cases, we found a remarkable and unusually high prevalence of *Prevotella*. This led us to think of a potential link between bacteria within this genus and some type of obesity. Some observations extracted from the literature available at that time appeared to support our hypothesis.

Firstly, Zhang et al. (2009) found higher levels of Prevotellaceae and Methanobacteriales, which are respectively H_2-producing fermentative bacteria and H_2-consuming methanogenic archaea, in obese individuals than in normal-weight individuals. They suggested that such interspecies H_2 transfer is an important mechanism for increasing fermentation and energy uptake from food, thus contributing to obesity.

De Filippo et al. (2010) compared the faecal microbiota of African and European children and found that the former showed communities very rich in *Prevotella*. In addition, higher amounts of SCFAs, which are end-products of bacterial fermentation, were detected in the faecal samples of the African children. The authors considered diet as the main discriminating factor between both populations. The African children ate a diet rich in starch, fibre, and plant polysaccharides, and low in fat and animal protein, whereas the European children ate a typical Western diet high in fat, animal protein, sugar, and starch, but low in fibre. Also, the African children consumed only 2/3 of the calories consumed by the Europeans. The authors hypothesised that *Prevotella* co-evolved with the polysaccharide-rich food of Africans because these bacteria are more efficient in extracting energy from complex polysaccharides (Flint et al., 2008). Going beyond, we postulate that this capability of *Prevotella* may be useful on a low-calorie diet to survive, but otherwise it may lead to obesity.

Arumugam et al. (2011) identified three main types of microbial assemblies in the human gut, named enterotypes, after clustering faecal samples

obtained in different studies based on their species composition or gene pools. Subjects B and C in Section 4.3, the obese ones, could be included in the enterotype enriched in *Prevotella*. Looking further into Arumugam's analysis of the pyrosequencing-based 16S rRNA samples from Turnbaugh et al. (2009a), we noticed that 17 of the 20 individuals of the *Prevotella*-enriched enterotype were obese, whereas 87 of the 134 individuals classified in the other enterotypes were so. This gives a statistically non-significant OR=3.04 (p=0.1221) of association between this enterotype and obesity. Adding our three subjects in Section 4.3, there are 19 obese out of 22 in the *Prevotella* enterotype, and 87 out of 135 in the other ones. This gives an OR=3.47 (p=0.0498), which begins to be supportive of a potential link between the *Prevotella* enterotype and some type of obesity.

Finally, Wu et al. (2011) linked the human microbial enterotypes suggested by Arumugam et al. (2011) to long-term dietary patterns. They clustered the faecal communities of healthy volunteers into two enterotypes, driven primarily by the levels of *Bacteroides* and *Prevotella*. Then, communities with relatively high levels of the *Bacteroides* genus were associated with high intake of animal protein, amino acids, and saturated fats, whereas relatively high levels of the *Prevotella* genus were associated with low values for these nutrients, but high values for CHOs and simple sugars.

In this study, we monitored one obese volunteer (individual B in Section 4.3) subject to a ketogenic diet (high in fat and protein and virtually without CHOs) to assess the changes in the composition of his faecal microbiota associated with CHO-restriction, and the ability of this community to return to its original composition after restoration of CHOs back into the diet.

4.5.2 Results

The volunteer provided faecal samples before, during, and after a dietary intervention lasting for 24 days. Samples at days -6, -5, -2, -1 (pre-diet), 1, 2, 3, 4, 7, 8, 9, 13, 14, 15, 19, 24 (during diet), and +1, +2, +3, +4, +27, +57 (post-diet) were analysed. An average of 3900 sequences was obtained per sample. Sequences covered the V1 and V2 variable regions of the 16S rRNA gene, with an average read length of 350 nt.

The normal profile of the faecal microbial community in this subject (based on the samples taken before initiating the dietary intervention) was dominated by bacteria within the Bacteroidetes phylum (92% of the sequences), mostly

(85%) within *Prevotella* (Bacteroidia-Bacteroidales-Prevotellaceae) (**Figure 4.22**). *Bacteroides* (Bacteroidia-Bacteroidales-Bacteroidaceae) included 4.5% of the sequences, and other taxa within Bacteroidetes, less than 1%. Apart from those genera, only *Sutterella* (Betaproteobacteria-Burkholderiales-Alcaligenaceae), *Roseburia* (Clostridia-Clostridiales-Lachnospiraceae), and *Faecalibacterium* (Clostridia-Clostridiales-Ruminococcaceae) had a prevalence greater than 1%. This profile is in agreement with that of the previous survey done for subject B (see Section 4.3).

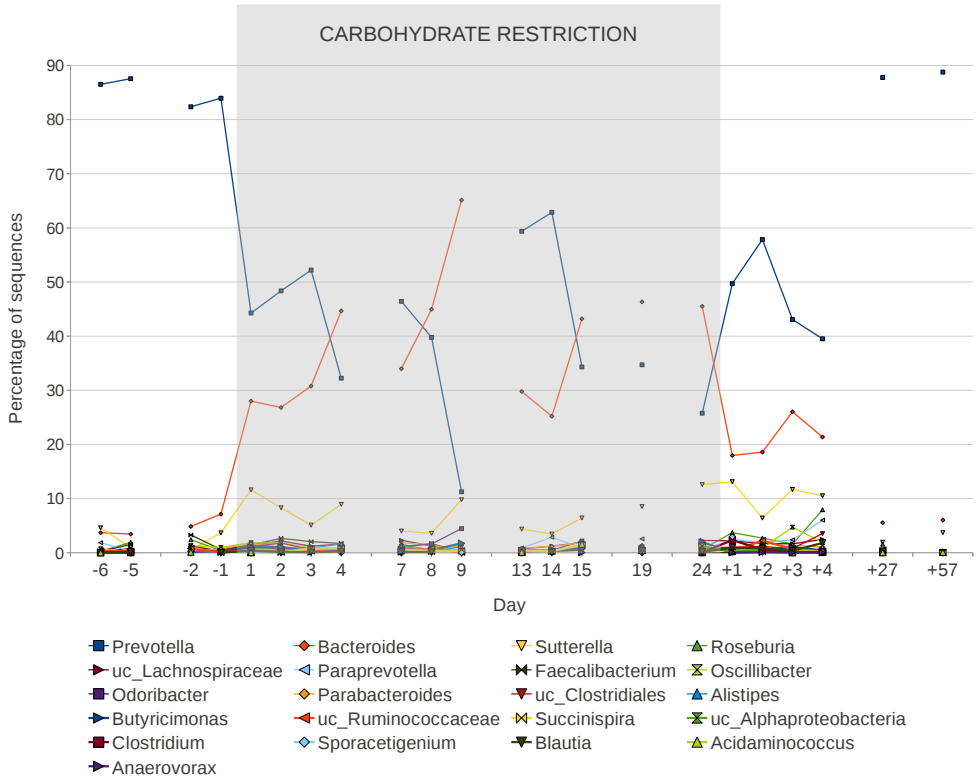


Figure 4.22: Relative abundance of genera with an average abundance $\geq 5‰$ in the faecal microbiota of subject B before (days -6 to -1), during (days 1 to 24), and after (days +1 to +57) a CHO-restricted diet.

The community structure experienced marked alterations immediately after the introduction of a CHO-restricted diet (**Figure 4.22**). Within 24 hours, the

prevalence of *Prevotella* was reduced by half. This space was filled mainly by the increase in *Bacteroides* and *Sutterella*, although most of the genera also increased their abundance on the CHO-restricted diet. The largest relative increases (9-fold on average) were those in *Bacteroides*, *Butyricimonas*, and *Odoribacter*, all genera within Bacteroidetes. These changes were reflected in the increase in the bacterial diversity of this community (as measured by the Shannon index) produced by the change in diet (**Figure 4.23**).

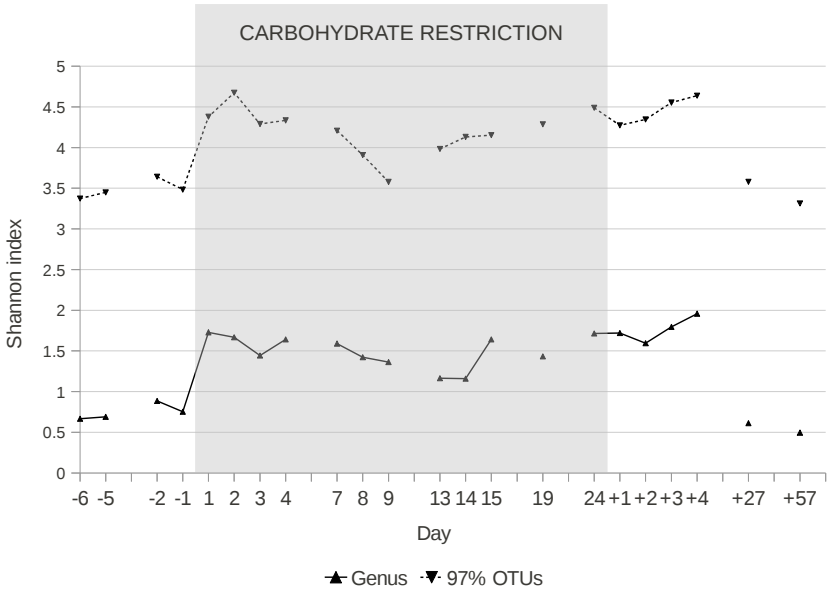


Figure 4.23: Bacterial diversity (Shannon index) in the faecal microbiota of subject B at genus and phylotype levels before (days -6 to -1), during (days 1 to 24), and after (days +1 to +57) a CHO-restricted diet.

Change in diet increased the temporal instability in the structure of the community. That is, faecal bacteria showed large fluctuations in their relative abundances, but without temporal trends. This instability persisted the first days after the subject switched to normal feeding. We could not correlate abrupt shifts in the community structure during the ketogenic diet with specific nutrients or events (for example, on day 9 during diet).

No later than 4 weeks after restoring the normal diet, the faecal microbiota moved back to its original structure. DCA plots at the genus and phylo-

type (defined at 97% of sequence identity) levels clearly showed the similarity between the samples taken before changing the diet and those of a few weeks after restoring the normal diet, and the strong shifts and instability in the community composition introduced by the change in diet (**Figure 4.24**).

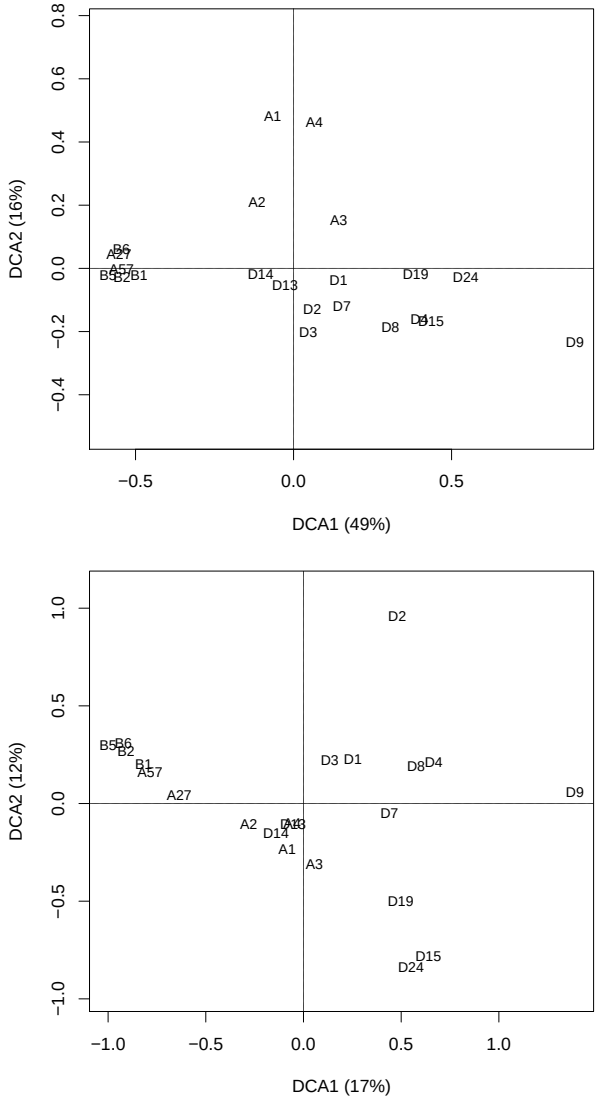


Figure 4.24: DCA at genus (top) and phylotype (bottom) levels. Percentages correspond to the fraction of inertia explained by each axis. Samples were taken before (days B6-B1), during (days D1-D24), and after (days A1-A57) a CHO-restricted diet.

The few remarkable changes between the microbial composition before the ketogenic diet and eight weeks after restoring the normal diet were the decrease in *Paraprevotella*, *Roseburia*, *Faecalibacterium*, and *Oscillibacter*, though we could not discriminate whether these differences corresponded to normal fluctuations in the community structure or were due to long-term effects of the dietary intervention.

The dynamics at the phylotype level was consistent with that detected at the genus level. The same phylotypes predominated within each genus and period (before, during, and after the change in diet). Thus, the ketogenic diet did not change the identity of the members of the community, but altered their relative abundances: phylotypes within *Prevotella* decreased in abundance while those within *Bacteroides*, *Sutterella*, etc., increased in abundance during the CHO-restricted diet, and then returned to their previous levels (**Figure 4.25**).

4.5.3 Discussion

In this study, we assessed the response of the faecal microbiota to a short-term dietary intervention. Specifically, we monitored the *Prevotella*-dominated bacterial community of an obese subject fed on a diet rich in protein and fat and very low in CHOs. Our results add to the increasing evidence of the impact of diet on the composition of the gut microbiota. Wu et al. (2011) already showed both the long- and short-term effects of diet on the gut microbiota. They monitored subjects within the *Bacteroides* enterotype when fed on a high-fat/low-fibre or low-fat/high-fibre diet. Both studies revealed rapid (detectable within 24 hours) changes in the composition of the microbiota in response to the alteration of diet. In our study, the prevalence of *Prevotella* under CHO-restriction was on average half that before CHO-restriction, but still relatively high. This decrease in the major genus was compensated by the increase in other genera, notably the 8-fold increase in *Bacteroides*. Faecal microbiota moved back to its previous structure no later than four weeks after reintroducing CHOs. Our results are consistent with the nutrient preferences reported for *Prevotella* and *Bacteroides* (Wu et al., 2011) and confirm the strong dependence on CHOs of bacteria within the *Prevotella* genus. In turn, Wu et al. (2011) found greater compositional changes in the low-fat/high-fibre diet group, as could be expected since it is the diet that correlated negatively with the *Bacteroides* enterotype, but it was not reported whether the high-fibre diet led to an increase in *Prevotella* and, if so, its mag-

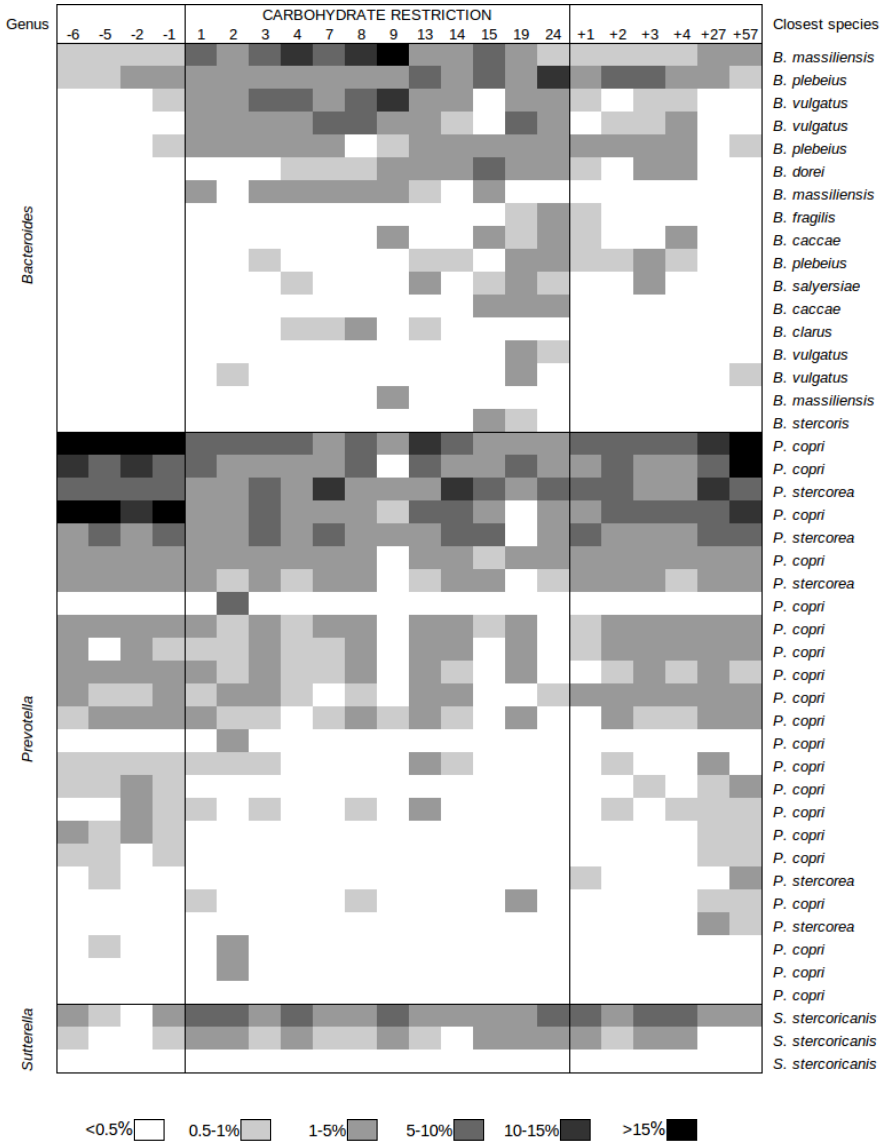


Figure 4.25: Relative abundance (grouped into six intervals) of phylotypes within the most abundant genera in the faecal samples of subject B before (days -6 to -1), during (days 1 to 24), and after (days +1 to +57) a CHO-restricted diet. Each row represents a phylotype defined at 97% of sequence identity. It is shown the genus and the species of the closest isolate for each phylotype.

nitude. Similar to our study, no subject switched stably to another enterotype.

Both studies confirm the significant structural resilience of the gut microbiota of adult humans to short-term dietary perturbations. The rapid restoration of the community composition is indicative of selective forces imposed by the environment (composition of the diet) and intrinsic to the community (ecological interactions between the gut microbes). Non-selective forces could also be implied, such as recolonisation after perturbation of the gut lumen from the outmost of the mucus layer (Van den Abbeele et al., 2011).

A reduced intake of dietary CHOs is expected to impact both the activity and the abundance of the different bacterial groups that populate the colon. The effect of low-CHO weight loss diets on the faecal bacterial populations had been previously evaluated in obese and overweight humans. Duncan et al. (2007) found a reduction in *Roseburia* spp., the *Eubacterium rectale* group, and bifidobacteria in obese subjects as CHO intake decreased, whereas Walker et al. (2011) found a reduction in *Collinsella aerofaciens* and an increase in the *Oscillibacter* group on the weight loss diet, together with a marked inter-individual variation in the responses to the dietary change. Interestingly, those studies reported no significant change in the proportion of Bacteroidetes or in the relative counts of the *Bacteroides-Prevotella* group as CHO intake decreased. However, the *Bacteroides/Prevotella* ratio was not explored, since a probe targeting the *Bacteroides-Prevotella* group was used for FISH and qPCR. Therefore, it is unknown whether changes similar to the ones observed by us in these genera occurred in those trials. Anyway, it is likely that the responses to dietary changes depend on the initial composition of the individual's gut microbiota, as it is suggested by the clustering of samples by individual rather than by diet observed by Walker et al. (2011).

It should be noted that the actual prevalence of genera within Bacteroidetes in the faecal samples of our volunteer migth be not so high, since the choice of primers and reaction conditions for the PCR amplification are known to impact the estimated relative abundance of bacterial taxa present in the samples (von Wintzingerode et al., 1997; Hongoh et al., 2003; Sipos et al., 2007; Hong et al., 2009). This is indeed a general limitation of PCR-based analyses of microbial diversity. However, relative abundances similar to the ones we found for genera within Bacteroidetes have been previously reported in the literature, even when using different methodologies (De Filippo et al., 2010; Caporaso et al., 2011; Wu et al., 2011). Also, we performed a metatranscriptomic analysis (PCR-independent) of a sample of this subject and found a large proportion of

Prevotella and Bacteroidetes (see Gosalbes et al. (2011), Figure 1, individual E).

An intriguing question is how such communities dominated by *Prevotella* are established. Several factors (and their interaction) are likely involved, including the host genotype, especially in immunity-related genes (Spor et al., 2011), and environmental factors such as the parental inocula and other early exposures to microbes (Deloris Alexander et al., 2006; Mulder et al., 2009; Ubeda et al., 2012), the type of feeding (Penders et al., 2006; De Filippo et al., 2010; Wu et al., 2011), and long-term effects of antibiotics on the microbiota (De La Cochetière et al., 2005; Dethlefsen and Relman, 2011).

Our study warrants further research into the contribution of *Prevotella* to obesity. Although the initial factors that shape the *Prevotella*-enriched communities are difficult to define, the causal relationship between high levels of *Prevotella* in the gut and subsequent development of obesity can be tested through studies in animal models. In this respect, we are currently conducting an experiment in which the main strains of *Prevotella* identified in obese humans in our surveys (*P. copri* DSM-18205 and *P. stercorea* DSM-18206) will be administered to mice fed on two different diets. The energy yield from diet, weight gain, and other obesity-related phenotypes will be compared between manipulated and non-manipulated mice, and the effect of a diet rich in fibre in obesity induction will be tested. We expect that this study will provide new insights about the link between this particular dysbiosis and obesity, similarly to the reports of transmission of obesity (Turnbaugh et al., 2006), colitis (Garrett et al., 2007), and metabolic syndrome (Vijay-Kumar et al., 2010) by transplantation of a dysbiotic microbiota from diseased to normal mice.

5 Conclusions

- Faeces contain bacterial communities that are less diverse than those associated with the mucus layer along the colon. The structural differences found between these sample types were reproduced in several studies in spite of variations in the experimental and analytical procedures.
- The variation in community composition between faeces and colonic mucosa is as large as the variation between samples of the same compartment taken from different individuals. In contrast, the mucosa-associated communities along the colon are very similar.
- The different microbial habitats of the distal gut should be analysed before discarding the involvement of the colonic microbiota in a pathological state or extrapolating the findings in one compartment to the entire GI microbiota.
- Faecal communities do not appear to be fully representative subsets of mucosa-associated communities in the colon. Moreover, mucosal biopsies do not seem contaminated by faecal material during sampling. This warrants the validity of each compartment to analyse the local communities.
- The composition of the faecal microbiota is stable over time. The prevalence of each bacterial taxon experiences fluctuations over short-term intervals (e.g. daily). However, the community diversity and the relative abundance of predominant bacteria do not change significantly in the absence of major perturbations.
- Within each host, there is a large core made of few bacterial taxa. Only a small number of taxa within the faecal bacteria, from phylum to phylotype level, is consistently present over time. However, these quantitatively comprise the majority of faecal microbes. Therefore, the intra-individual

core of gut bacteria is very large when considering not only the percentage of taxa it represents, as it is usually measured, but also the fraction of the microbiota it accounts for.

- The search for dysbiosis in pathological states should be better focused on the high-abundance core microbes as they are undoubtedly part of the resident microbiota.
- Large inter-subject variation is found in the patterns of co-occurrence between bacterial taxa in faeces. This suggest that specific ecological interactions established between resident microbes within each host may contribute to the maintenance of a stable host-specific microbiota.
- IBS patients harbour bacterial communities in mucosal sites along the colon and in faeces that are less diverse than those found in healthy subjects.
- The global distribution of gut bacteria does not allow to distinguish IBS patients from healthy subjects, or IBS patients with diarrhoea from IBS patients with constipation. However, subtle alterations in the relative abundance of specific gut bacteria are found in IBS patients compared with healthy subjects. They differ between gut compartments, especially between colonic mucosal sites and faeces, and between IBS patients with diarrhoea and IBS patients with constipation.
- An increase in the temporal variation in the distribution of active faecal bacteria is associated with the IBS condition and with fluctuation of bowel symptoms.
- Strong and quick structural and functional changes in the faecal microbiota are associated with acute symptomatology in IBS patients with diarrhoea. The specific changes are barely reproduced between and within patients, which makes it difficult to assess whether dysbiosis contributes to the initiation and/or perpetuation of the pathogenesis of IBS or appears as a result of the development of symptoms.
- IBS symptoms may be mitigated by therapeutic interventions aimed at increasing the diversity and stability of the GI microbiota.
- An accurate classification of IBS patients by type and severity of symptoms must be considered before evaluating suspected aetiological factors and alterations of the GI microbiota.

- Functional redundancy between individual members or consortia within the gut microbiota is suggested by the host-specificity in community structure and microbial co-occurrence patterns. This is also confirmed by the overall similarity in the genetic potential and gene expression of different community assemblies.

- A high prevalence of bacteria within the *Prevotella* genus is found in faeces of a subset of obese individuals. They experience a significant reduction in abundance under a CHO-restricted diet, compensated by the increase in other bacteria, and a rapid recovery after restoration of CHO intake.

- *Prevotella* is associated with an increase in bacterial fermentation in the colon that may contribute to obesity in developed countries. Therefore, the potential inductive role of *Prevotella* in obesity should be tested in animal models.

- The presence of selective forces that maintain the structure of the faecal microbiota is indicated by a stable composition with relatively small fluctuations over time and the rapid recovery after disturbance by short-term dietary interventions.

A Resumen en castellano

A.1 Introducción

Los humanos viven en simbiosis con complejas comunidades microbianas que habitan distintas superficies y cavidades corporales, como la piel, la vagina, las vías respiratorias superiores, la cavidad oral, y el tracto gastrointestinal (Aas et al., 2005; Zoetendal et al., 2006; Grice et al., 2009; Charlson et al., 2010; Fredricks, 2011). Estas comunidades se denominan microbiota humana. Cuantitativamente, el número de microorganismos en el cuerpo humano es diez veces el de células humanas, encontrándose la mayoría en el tracto gastrointestinal, y, dentro de este, en el intestino grueso, donde se alcanza una de las mayores concentraciones microbianas registradas en cualquier ambiente (Luckey, 1972; Whitman et al., 1998).

Tradicionalmente, los microorganismos asociados a humanos han sido considerados comensales o patógenos oportunistas, y el interés en su estudio se ha centrado en el potencial patógeno de microorganismos concretos. Sin embargo, las investigaciones llevadas a cabo desde hace más de un siglo han revelado que la mayoría mantiene una relación mutualista con el hospedador de la que ambos partícipes extraen beneficio: el hospedador proporciona nutrientes y un ambiente protegido a los simbiontes microbianos mientras que estos asisten al normal desarrollo y fisiología del hospedador, al tiempo que lo protegen de potenciales patógenos (Guarner and Malagelada, 2003; Dethlefsen et al., 2007).

Recientemente, la aplicación de métodos moleculares en ecología microbiana ha permitido el acceso al conjunto de la microbiota humana. Estos incluyen la caracterización de la diversidad de las comunidades microbianas mediante el análisis de la variabilidad de los genes de la subunidad pequeña del RNA ribosomal (rRNA 16S) y la secuenciación del DNA o el RNA totales aislados de las comunidades microbianas, denominados metagenoma y metatranscriptoma, respectivamente (Zoetendal et al., 2004, 2008). Con anterioridad, sólo se podían

caracterizar aquellos simbiontes que pueden ser cultivados en el laboratorio, los cuales representan una fracción minoritaria de estas comunidades (Wilson and Blitchington, 1996; Suau et al., 1999).

Estos estudios han revelado que la microbiota humana está constituida principalmente por bacterias, la mayoría pertenecientes a los filos Firmicutes, Bacteroidetes, Actinobacteria, y Proteobacteria. También incluye varias especies de arqueas, levaduras, y otros eucariotas microscópicos. Cada hábitat corporal contiene una microbiota característica que es relativamente consistente entre individuos en cuanto a la abundancia de filos, aunque la composición en especies o cepas puede diferir considerablemente (Dethlefsen et al., 2007; Costello et al., 2009; Spor et al., 2011).

La microbiota gastrointestinal está dominada por bacterias anaerobias estrictas de los filos Firmicutes y Bacteroidetes. Algunos otros filos, principalmente Proteobacteria y Actinobacteria, se pueden detectar en proporciones reducidas. Por contra, la diversidad dentro de estos pocos linajes profundos es muy elevada, con estimaciones de miles de especies en el colon de cada individuo (Dethlefsen et al., 2007; Zoetendal et al., 2008). También aparecen varias especies minoritarias de arqueas metanógenas (Scanlan et al., 2008). Estos microorganismos pueden ser residentes o transitorios en los hábitats gastrointestinales. Algunos criterios para distinguir los microorganismos indígenas es que estén presentes con niveles estables en la población normal, aparezcan asociados a la mucosa, o se observe su actividad *in situ*.

La microbiota cambia a lo largo de las estructuras del tracto gastrointestinal debido a la variación longitudinal en las condiciones de crecimiento (pH, concentración de oxígeno, tiempo de retención, secreciones del hospedador, disponibilidad de nutrientes, etc.) (Savage, 1977). También puede variar transversalmente, entre el lumen y la mucosa epitelial. Se hipotetiza que los microorganismos asociados a la mucosa son aquellos que escapan a los anticuerpos y péptidos antimicrobianos producidos por el hospedador, atrapados en la capa mucosa, los que se adhieren o se alimentan de mucina, y/o los que resisten la mayor concentración de oxígeno cerca del epitelio (Van den Abbeele et al., 2011).

La estructura de las comunidades microbianas asociadas a los humanos adultos está influenciada por múltiples factores (Dethlefsen et al., 2006; Spor et al., 2011). Algunos son intrínsecamente aleatorios, como la historia de colonización durante el ensamblaje de la comunidad o la respuesta a infecciones. Otros son más deterministas, como la selección impuesta por el genotipo del

hospedador, especialmente por polimorfismos en genes relacionados con el sistema inmune, y por factores ambientales, como el tipo de dieta o las terapias con efectos duraderos sobre la microbiota, por ejemplo, el tratamiento con antibióticos. Además, las interacciones ecológicas establecidas entre microorganismos concretos pueden contribuir a la estabilidad e individualidad de la microbiota.

Los estudios realizados en modelos animales han sido clave para revelar los beneficios que aporta la microbiota gastrointestinal al hospedador (Guarner and Malagelada, 2003; O'Hara and Shanahan, 2006). Las funciones metabólicas de la microbiota incluyen la extracción de energía a partir de la dieta, principalmente a través de la hidrólisis de polisacáridos complejos no digeribles por el propio hospedador y su fermentación hasta la generación de ácidos grasos de cadena corta, la producción de vitaminas y aminoácidos esenciales, y la mejora en la absorción de iones por parte del epitelio intestinal. La microbiota cumple funciones tróficas tan importantes como la estimulación de la proliferación y diferenciación de las células del epitelio intestinal y la regulación del desarrollo y homeostasis del sistema inmune. Además, la microbiota limita la colonización por patógenos compitiendo por los recursos disponibles, produciendo compuestos antimicrobianos, y regulando la respuesta inmune del hospedador.

Alteraciones en la composición de la microbiota humana, conocidas como disbiosis, pueden afectar a cualquiera de las interacciones mutualistas entre la microbiota y el hospedador y, por ello, contribuir al inicio, mantenimiento o empeoramiento de ciertos estados patológicos. De hecho, estudios recientes han encontrado disbiosis en desórdenes tan diversos como obesidad, enfermedad celíaca, enfermedad de Crohn y colitis ulcerosa (comúnmente denonimadas enfermedad inflamatoria intestinal), síndrome del intestino irritable, diarrea asociada a antibióticos, cáncer colorrectal, vaginosis bacteriana, fibrosis quística, o psoriasis (Frank et al., 2011; Gerritsen et al., 2011; Tlaskalová-Hogenová et al., 2011).

A.2 Objetivos

El objetivo principal de la presente tesis era estudiar la relación entre la microbiota humana del intestino distal y el síndrome del intestino irritable (SII), uno de los trastornos intestinales más comunes en los países desarrollados (Longstreth et al., 2006; Quigley et al., 2006). El SII es un desorden funcional,

es decir, sin causa orgánica aparente, caracterizado por síntomas intestinales crónicos como dolor o malestar abdominal y alteración de los hábitos intestinales. Los afectados se clasifican en función del hábito predominante (diarrea, estreñimiento, o alternancia de ambos). Su diagnóstico y tratamiento son complicados debido a su etiología, probablemente multifactorial y en gran medida desconocida, y a la heterogeneidad en la sintomatología de los pacientes (Talley and Spiller, 2002; Karantanos et al., 2010). Se sospechaba desde hacía tiempo de alteraciones en la microbiota intestinal como uno de los factores implicados en el desencadenamiento y persistencia del SII, aunque los estudios de casos y controles no habían revelado cambios marcados en la composición de la microbiota relacionados con el SII y no se había alcanzado un consenso respecto a la implicación de bacterias específicas (Salonen et al., 2010).

Antes de estudiar las alteraciones potenciales de la microbiota en el SII, consideramos necesario evaluar otras fuentes de variación de la microbiota intestinal. En primer lugar, las diferencias entre la microbiota aislada de distintos hábitats intestinales (heces y mucosa del colon) de distintos individuos sanos (**Estudio 1**). Las heces son el tipo de muestra escogido en la mayoría de estudios de la microbiota gastrointestinal debido a la facilidad de su recolección. Sin embargo, la microbiota fecal podría no ser enteramente representativa de las comunidades asociadas a la mucosa del colon, como apuntaban estudios previos (Zoetendal et al., 2002; Ott et al., 2004; Eckburg et al., 2005; Lepage et al., 2005).

Dado que los distintos tipos de muestra podrían revelar información distinta y complementaria sobre la microbiota intestinal, nuestro primer estudio sobre el SII consistió en comparar la composición microbiana de pacientes con SII y controles sanos en heces y varios puntos de la mucosa del colon (**Estudio 2**).

Simultáneamente, analizamos la dinámica temporal de la microbiota fecal de individuos sanos para evaluar la variación normal a corto plazo en la composición de las comunidades y la extensión del *core* de bacterias dentro de cada individuo, y para inferir interacciones potenciales entre grupos de bacterias a partir de los patrones de co-ocurrencia (**Estudio 3**).

Los resultados del Estudio 2 y de otros estudios transversales realizados en el SII, junto con los datos acerca de la variación temporal de la microbiota fecal, nos llevaron a plantear un estudio longitudinal en el que se tomaran muestras fecales de pacientes con SII a lo largo de cierto período de tiempo y se registrara la fluctuación en los síntomas intestinales (**Estudio 4**). Monitorizar a pacientes individuales permite superar la confusión introducida por la

gran variación entre individuos que es problemática en los estudios de casos y controles. Además, analizamos por primera vez el metagenoma y el metatranscriptoma de pacientes con SII, lo que nos permitió evaluar los cambios en el potencial genético y la expresión génica de la microbiota relacionados con la presentación de síntomas.

Adicionalmente, detectamos que algunos de los voluntarios obesos incluidos en nuestras cohortes albergaban comunidades microbianas con una abundancia inusualmente elevada de bacterias del género *Prevotella*. Esta observación, junto con otras extraídas de la literatura, nos llevó a sugerir una vinculación de estas bacterias con la obesidad, que fue testada, en primer lugar, mediante el análisis de la respuesta de la microbiota de uno de los voluntarios a una dieta pobre en carbohidratos (**Estudio 5**).

Así pues, esta tesis se estructura en los siguientes estudios:
Estudio 1. Diversidad bacteriana en heces y mucosa rectal de individuos sanos.
Estudio 2. Alteraciones estructurales de las bacterias fecales y de la mucosa del colon asociadas al SII.
Estudio 3. Estabilidad y especificidad de las bacterias fecales de individuos sanos.
Estudio 4. Seguimiento de la microbiota fecal de afectados por el SII.
Estudio 5. Respuesta de una microbiota fecal dominada por *Prevotella* a la restricción de carbohidratos.

A.3 Metodología

Toma de muestras

Todos los participantes dieron consentimiento informado al protocolo de estudio, aprobado por el comité ético del Hospital Universitario La Fe, de Valencia. Los voluntarios cumplimentaron cuestionarios sobre su estilo de vida y datos clínicos relevantes. Los voluntarios sanos no presentaban un historial previo de enfermedad gastrointestinal o comorbididad sistémica, tratamiento reciente con antibióticos, terapia inmunomodulatoria, antidiarreicos, o laxantes. Los pacientes con SII cumplían los criterios diagnósticos Roma II (Drossman, 1999b) y fueron clasificados en los subtipos SII-diarrea y SII-estreñimiento (Longstreth et al., 2006).

La **Tabla A.1** resume los participantes incluidos en cada estudio y las muestras recogidas.

Tabla A.1: Muestreo.

Estudio	**1**	**2**	**3**	**4**	**5**
Voluntarios	9 sanos	13 SII-diarrea 3 SII-estreñimiento 9 controles sanos	3 sanos	2 SII-diarrea 1 control sano	1 sano (obeso)
Tipo de muestra	Heces Mucosa rectal	Heces Mucosa del colon ascendente y del colon descendente	Heces	Heces	Heces
Número de muestras	Una de cada tipo	Una de cada tipo	Una diaria (15 días consecutivos)	Paciente 1/Control 1: días 1,3,5,7,14,21,28, 37,42 Paciente 2: días 1,3 (mañana y tarde),4, 5,7,14,21,27,28,35,42, 49,56	Antes de la dieta: días -6,-5,-2,-1 Durante la dieta: días 1,2,3,4,7,8,9,13, 14,15,19,24 Tras la dieta: días +1, +2,+3,+4,+27,+57
Datos adicionales				Síntomas gastrointestinales	Registro de la dieta y de la pérdida de peso

Las muestras de la mucosa del colon se obtuvieron mediante endoscopia, sin (Estudio 1) o con (Estudio 2) preparación previa del colon. Las biopsias fueron preservadas inmediatamente a -80 °C. Las muestras fecales fueron recogidas en tampón fosfato salino (PBS), previamente a la toma de biopsias en los Estudios 1 y 2. Las muestras fecales se conservaron a 4 °C durante 1-2 horas antes de ser almacenadas a -80 °C.

Extracción de ácidos nucleicos

En los Estudios 1 y 2, el DNA fue extraído de las biopsias del colon utilizando el kit QIAamp DNA Mini (QIAGEN). En los Estudios 1, 2, y 3, el DNA fue extraído de las muestras fecales utilizando el kit QIAamp DNA Stool Mini (QIAGEN). En los Estudios 4 y 5, el DNA y el RNA fueron aislados simultáneamente a partir de las muestras fecales utilizando el kit AllPrep DNA/RNA Mini (QIAGEN). Antes de la extracción, las muestras fecales fueron resuspendidas en PBS y centrifugadas a 4000 rpm para descartar el material particulado grande.

Amplificación y secuenciación de los genes del rRNA 16S

Los genes del rRNA 16S fueron amplificados a partir del DNA extraído de las muestras de los Estudios 1, 2, 3, y 5 mediante reacción en cadena de la polimerasa (PCR) utilizando cebadores de amplio rango para bacterias.

Las condiciones de amplificación (selección de cebadores, número de ciclos de amplificación, temperatura de alineamiento) se ajustaron para minimizar los sesgos potencialmente introducidos durante la PCR en la estimación de la distribución en las muestras de los distintos taxones bacterianos (von Wintzingerode et al., 1997; Hongoh et al., 2003; Sipos et al., 2007; Hong et al., 2009).

Las muestras del Estudio 1 fueron secuenciadas mediante el método de Sanger y analizadas en un secuenciador ABI 3730 (Applied Biosystems). Las muestras de los Estudios 2, 3, y 5 fueron secuenciadas mediante pirosecuenciación en un secuenciador FLX utilizando la química GS FLX o GS FLX Titanium (454 Life Sciences, Roche). Las diferencias en el método de secuenciación se deben a que adoptamos los avances en la tecnología de secuenciación conforme estuvieron disponibles.

La **Figura A.1** esquematiza la estrategia seguida en los distintos estudios.

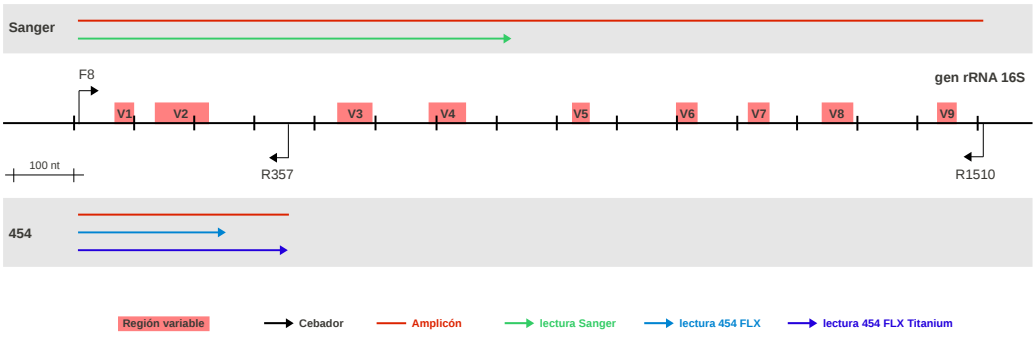


Figura A.1: Cebadores, amplicones, y lecturas obtenidas mediante secuenciación por el método de Sanger o pirosecuenciación (454), mapeados a lo largo del gen del rRNA 16S.

Secuenciación de metagenomas y metatranscriptomas

El RNA extraído de las muestras del Estudio 4 fue tratado con DNasa I (Ambion), y, posteriormente, amplificado con el kit MessageAmp II-Bacteria (Ambion), que favorece el enriquecimiento en RNA mensajero (mRNA). El RNA resultante fue retro-transcrito a DNA complementario (cDNA) utilizando el kit High Capacity cDNA Reverse Transcription (Applied Biosystems) y las enzimas DNA Polimerasa I de *E. coli*, DNA Ligasa de *E. coli*, RNasa H, y DNA

Polimerasa del fago T4 (New England Biolabs).

El cDNA resultante y el DNA extraído de las muestras del Estudio 4 fueron secuenciados mediante pirosecuenciación en un secuenciador FLX con la química GS FLX Titanium (454 Life Sciences, Roche).

Procesamiento de las secuencias

Las lecturas con baja calidad y/o longitud fueron excluidas del análisis. Adicionalmente, se eliminaron las posibles quimeras detectadas en las lecturas de amplicones del gen del rRNA 16S con la herramienta *chimera.slayer* del paquete *mothur* (Schloss et al., 2009).

Asignación taxonómica de los rRNAs 16S

La asignación taxonómica (filo, clase, orden, familia, género) de las secuencias de los genes y los cDNAs del rRNA 16S fue determinada con el clasificador del *Ribosomal Database Project (RDP) II* (Cole et al., 2007, 2009). Las secuencias fueron agrupadas en filotipos asimilables a especies en base a su identidad de secuencia utilizando los programas *cd-hit-est* del paquete *cd-hit* (Li and Godzik, 2006) o *cluster* del paquete *usearch* (Edgar, 2010).

Anotación de metagenomas y metatranscriptomas

Los genes y cDNAs de los rRNAs 16S fueron identificados mediante búsqueda de homologías con la herramienta *BLASTN* (Altschul et al., 1990) en la base de datos SSUrdb (Urich et al., 2008). En las lecturas restantes, se identificaron los genes y cDNAs de los rRNAs 23S mediante búsqueda *BLASTN* en la base de datos LSUrdb (Urich et al., 2008).

Las lecturas de cada metagenoma fueron ensambladas usando el programa *runAssembly* del paquete *Newbler* (454 Life Sciences, Roche). Las lecturas de los metatranscriptomas no asignadas a rRNAs fueron alineadas con los ensamblajes metagenómicos procedentes del mismo voluntario usando el programa *runMapping* del paquete *Newbler* (454 Life Sciences, Roche).

Las regiones codificantes de proteínas fueron identificadas en los metagenomas mediante búsqueda *BLASTX* (Altschul et al., 1990) en la base de datos no redundante de proteínas NCBI-nr (ftp://ftp.ncbi.nlm.nih.gov/blast/db). Además, se utilizó el programa de predicción génica *Glimmer* (Salzberg et al., 1998) (http://www.cbcb.umd.edu/software/glimmer).

La asignación funcional de los supuestos genes codificantes se llevó a cabo comparándolos con la base de datos KEGG GENES (Kanehisa et al.,

2004) (http://www.genome.ad.jp/kegg) mediante búsqueda *BLASTX*, y con la base de datos de familias de proteínas TIGRFAM (Selengut et al., 2007) (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi) utilizando el programa *HMMER* (Eddy, 1998) (http://hmmer.janelia.org). Las lecturas de los metatranscriptomas que mapearon sobre regiones de los metagenomas adoptaron sus anotaciones funcionales. Las lecturas de los metatranscriptomas restantes fueron comparadas con la base de datos de familias de RNAs no codificantes Rfam (Griffiths-Jones et al., 2003) (http://rfam.sanger.ac.uk) utilizando el programa *INFERNAL* (Nawrocki et al., 2009) (http://infernal.janelia.org).

Análisis estadísticos

La mayor parte de los análisis estadísticos se realizaron utilizando el entorno estadístico R (R Development Core Team, 2010) (http://www.R-project.org) y el paquete de R *vegan* (Oksanen et al., 2011).

La diversidad microbiana en las muestras fue estimada con el índice de diversidad de Shannon (Shannon, 1948) y los estimadores de riqueza Chao1 y *Abundance-based Coverage (ACE)* (Chao, 1987; Chazdon et al., 1998; Chao et al., 2000). La riqueza estimada y las curvas de rarefacción permitieron evaluar la cobertura de las muestras.

La similitud de las muestras con respecto a la distribución de taxones o funciones fue evaluada aplicando análisis de correspondencias, análisis de conglomerados, y análisis de similitudes (ANOSIM) (Quinn and Keough, 2002, p 459, 488 y 484, respectivamente).

En el Estudio 1, se estimó la abundancia relativa de cada taxón en cada muestra aplicando un modelo Bayesiano que tenía en cuenta las interacciones entre individuos y taxones, y entre tipos de muestra y taxones. La similitud entre las muestras se calculó en base a las distribuciones estimadas con dicho modelo.

En el Estudio 2, se aplicaron test chi-cuadrado para detectar cambios significativos en la abundancia relativa de cada taxón entre las muestras de pacientes con SII y las de controles sanos. Para ello, se agruparon las muestras según la cohorte (SII-diarrea, SII-estreñimiento, control) y el tipo de muestra (colon ascendente, colon descendente, heces). La robustez de los resultados obtenidos con las muestras agrupadas frente a la variación interindividual se evaluó comparando los resultados con los de muestras agrupadas artificiales obtenidas tras etiquetar aleatoriamente las muestras individuales como procedentes de casos o controles.

En el Estudio 3, se analizó la serie temporal de cada individuo con un modelo Bayesiano que tenía en cuenta la posible correlación entre taxones y la autocorrelación temporal dentro de cada taxón. La asociación potencial entre taxones fue inferida aplicando *graphical Gaussian networks (GGNs)* (Schäfer and Strimmer, 2005a,b) para estimar la correlación parcial entre los perfiles de abundancia a partir de la matriz de covarianzas obtenida con el modelo Bayesiano.

En el Estudio 4, se aplicaron análisis de componentes principales para analizar la variación temporal en la presentación de síntomas (Quinn and Keough, 2002, p 443), test chi-cuadrado para detectar cambios significativos en la abundancia relativa de cada taxón entre los días con síntomas agudos y los días con síntomas leves, y el algoritmo *LEfSe (linear discriminant analysis (LDA) effect size)* (Segata et al., 2011) para identificar categorías funcionales que caracterizaran los días con síntomas agudos y leves.

A.4 Conclusiones

Diferencias en la composición de las comunidades bacterianas de la mucosa del colon y de las heces:

- Las comunidades bacterianas que se encuentran en heces son menos diversas que las asociadas a la mucosa intestinal en los distintos tramos del colon estudiados (colon ascendente, colon descendente, y recto). Las diferencias en la composición bacteriana entre heces y mucosa del colon han sido reproducidas en dos estudios a pesar de las variaciones en la metodología aplicada.

- La variación en la composición de la microbiota entre heces y mucosa del colon es tan grande como la variación entre individuos en muestras del mismo compartimento intestinal. En cambio, las comunidades microbianas asociadas a la mucosa intestinal en distintos puntos del colon son similares.

- Dadas estas diferencias, los distintos hábitats del intestino distal deberían ser analizados antes de poder descartar la implicación de la microbiota en una patología o de extrapolar los hallazgos hechos en un compartimento a la totalidad de la microbiota intestinal.

- Las comunidades microbianas presentes en heces no parecen ser total-

mente representativas de las asociadas a la mucosa del colon. Además, las muestras de mucosa obtenidas por endoscopia no parecen contaminadas por material fecal. Por tanto, ambos tipos de muestra resultan válidos para analizar las correspondientes comunidades microbianas.

Variación temporal normal en la estructura de la microbiota fecal:

- La composición de la microbiota fecal es estable a lo largo del tiempo. La abundancia relativa de los distintos taxones bacterianos fluctúa en ausencia de perturbaciones externas en intervalos cortos de tiempo (p. ej. diariamente). Sin embargo, la diversidad de la comunidad y la prevalencia de las bacterias predominantes no cambian singnificativamente.

- Dentro de cada individuo, sólo un número reducido de taxones bacterianos es detectado en heces de forma constante a lo largo del tiempo (visto a distintos niveles, desde filo hasta especie). Sin embargo, en términos cuantitativos, estos engloban la mayoría de las bacterias fecales. Por tanto, el *core* de bacterias dentro de cada individuo es grande si se tiene en cuenta la fracción de la microbiota que representa, no sólo el número relativo de grupos, que es como se evalúa habitualmente.

- Podría ser conveniente restringir la búsqueda de disbiosis en estados patológicos a los microorganismos abundantes ya que estos son parte de la microbiota residente, no microorganismos que se detectan transitoriamente en el intestino.

Interacciones entre microorganismos intestinales:

- La microbiota intestinal difiere entre distintos individuos no sólo en su composición, también en los patrones de co-ocurrencia entre grupos específicos de bacterias. Esto sugiere que las interacciones ecológicas (de cooperación, competencia, etc.) establecidas dentro de cada hospedador podrían contribuir a la estabilidad e individualidad de la microbiota.

Alteraciones de la microbiota intestinal asociadas al SII:

- Las comunidades bacterianas de las heces y de la mucosa del colon ascendente y descendente de los afectados por el SII son menos diversas que las respectivas de controles sanos.

- La distribución global de bacterias intestinales no permite distinguir entre pacientes con SII y controles sanos, ni entre pacientes de distinto subtipo

(diarrea o estreñimiento como síntoma predominante). Sin embargo, se detectan alteraciones sutiles en la abundancia relativa de grupos concretos de bacterias en los afectados por el SII con respecto a los controles sanos. Las alteraciones son distintas en los diferentes compartimentos intestinales analizados, especialmente entre las heces y la mucosa intestinal. También son distintas según el subtipo (SII con diarrea o SII con estreñimiento).

- La variación temporal en el perfil de bacterias activas en heces es mayor en los pacientes con SII que en los controles sanos, y está asociada positivamente a la severidad y fluctuación de los síntomas intestinales.

- Las fases agudas de diarrea en afectados por SII están asociadas a cambios marcados y rápidos en la estructura y el funcionamiento de la microbiota fecal. Las alteraciones detectadas no se reproducen en distintos pacientes ni dentro del mismo paciente en distintas fases, lo que dificulta evaluar si la disbiosis contribuye a la presentación de síntomas o se produce a consecuencia de la alteración del hábitat gastrointestinal debida a otros factores etiológicos.

- Nuestros estudios sugieren que el papel de la microbiota en el SII es más bien débil. Los rasgos más destacados de la microbiota intestinal en pacientes con SII son una menor diversidad y una mayor inestabilidad, por lo que terapias que potencialmente contrarrestan estas características (p. ej. probióticos) pueden resultar adecuadas para mitigar los síntomas del SII.

- Se requiere una clasificación detallada de los afectados por el SII según el tipo y la severidad de los síntomas que presentan antes de evaluar la implicación de factores etiológicos potenciales y alteraciones de la microbiota intestinal. De otra forma, la heterogeneidad entre los pacientes puede dificultar la detección de patrones.

Relación de las bacterias del género *Prevotella* con la obesidad y los carbohidratos de la dieta:

- Un subconjunto de humanos obesos presenta comunidades microbianas fecales ricas en bacterias del género *Prevotella*.

- La restricción de carbohidratos en la dieta produce cambios inmediatos en la microbiota fecal dominada por bacterias del género *Prevotella* de

un individuo obeso, que consisten en la reducción significativa de estas bacterias compensada por el aumento de otras presentes en la comunidad. Estos cambios revierten tras la restauración de la dieta normal.

- La relación del género *Prevotella* con el aumento en la capacidad de extracción de energía a partir de polisacáridos complejos de la dieta puede explicar su asociación con la obesidad en los países desarrollados. La posible contribución al desarrollo de la obesidad de las bacterias del género *Prevotella* encontradas en obesos debería ser testada *in vivo*.

Observaciones generales:

- Las diferencias entre individuos en la estructura de la microbiota intestinal sugiere que existe redundancia funcional entre microorganismos o consorcios de distintas comunidades microbianas. Esta redundancia se confirma por el parecido en los perfiles funcionales obtenidos a partir de metagenomas y metatranscriptomas de comunidades fecales que difieren significativamente en el perfil de microorganismos presentes y activos.

- La estabilidad en la composición de la microbiota fecal, con fluctuaciones relativamente pequeñas en la abundancia relativa de los distintos taxones, y la capacidad de recuperación de la comunidad después de su perturbación por una intervención dietética, son indicativas de la existencia de fuerzas selectivas que mantienen la estructura de la microbiota instestinal en adultos.

B Bibliography

Aas, J. r. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *Journal of Clinical Microbiology*, 43(11):5721–5732.

Adeyemi, E. O., Desai, K. D., Towsey, M., and Ghista, D. (1999). Characterization of autonomic dysfunction in patients with irritable bowel syndrome by means of heart rate variability studies. *The American Journal of Gastroenterology*, 94(3):816–823.

Aerssens, J., Camilleri, M., Talloen, W., Thielemans, L., Göhlmann, Hinrich W H Wyngaert, I. V. A. N. D. E. N., Thielemans, T., Hoogt, R. D. E., Andrews, C. N., Bharucha, A. E., Carlson, P. J., Busciglio, I., Burton, D. D., Smyrk, T., Urrutia, R., and Coulie, B. (2008). Alterations in mucosal immunity identified in the colon of patients with irritable bowel syndrome. *Clinical Gastroenterology and Hepatology*, 6(2):194–205.

Aggarawal, A., Cutts, T. F., Abell, T. L., Cardoso, S., Familoni, B., Bremer, J., and Karas, J. (1994). Predominant symptoms in irritable bowel syndrome correlate with specific autonomic nervous abnormalities. *Gastroenterology*, 106(4):945–950.

Aguas, M., Garrigues, V., Bastida, G., Nos, P., Ortiz, V., Fernandez, A., and Ponce, J. (2011). Prevalence of irritable bowel syndrome (IBS) in first-degree relatives of patients with inflammatory bowel disease (IBD). *Journal of Crohn's and Colitis*, 5(3):227–233.

Alakomi, H. L., Skyttä, E., Saarela, M., Mattila-Sandholm, T., Latva-Kala, K., and Helander, I. M. (2000). Lactic acid permeabilizes gram-negative bacteria by disrupting the outer membrane. *Applied and environmental microbiology*, 66(5):2001–5.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.

Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., and Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE*, 3(7):e2836.

Antunes, L. C. M., Han, J., Ferreira, R. B. R., Lolic, P., Borchers, C. H., and Finlay, B. B. (2011). Effect of antibiotic treatment on the intestinal metabolome. *Antimicrobial Agents and Chemotherapy*, 55(4):1494–1503.

Aranki, A., Syed, S. A., Kenney, E. B., and Freter, R. (1969). Isolation of anaerobic bacteria from human gingiva and mouse cecum by means of a simplified glove box procedure. *Applied Microbiology*, 17(4):568–576.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D., and Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346):174–180.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., Semenkovich, C. F., and Gordon, J. I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, 101(44):15718–15723.

Bäckhed, F., Manchester, J. K., Semenkovich, C. F., and Gordon, J. I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3):979–984.

Baidoo, L. R., Kundu, R., and Berenbaum, P. L. (2006). Rifaximin is effective for small intestinal bacterial overgrowth. *Gastroenterology*, 128(Suppl 2):W1732.

Baker, G. C., Smith, J. J., and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods*, 55(3):541–555.

Balsari, A., Ceccarelli, A., Dubini, F., Fesce, E., and Poli, G. (1982). The fecal microbial population in the irritable bowel syndrome. *Microbiologica*, 5(3):185–194.

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall.

Banner, S. E. and Sanger, G. J. (1995). Differences between 5-HT3 receptor antagonists in modulation of visceral hypersensitivity. *British Journal of Pharmacology*, 114(2):558–562.

Barman, M., Unold, D., Shifley, K., Amir, E., Hung, K., Bos, N., and Salzman, N. (2008). Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infection and Immunity*, 76(3):907–915.

Basseri, R. J., Weitsman, S., Barlow, G. M., and Pimentel, M. (2011). Antibiotics for the treatment of irritable bowel syndrome. *Gastroenterology and Hepatology*, 7(7):455–493.

Bates, J. M., Mittge, E., Kuhlman, J., Baden, K. N., Cheesman, S. E., and Guillemin, K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Developmental Biology*, 297(2):374–386.

Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E., and Flint, H. J. (2006). Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. *Applied and Environmental Microbiology*, 72(5):3593–3599.

Bengston, M. B., Ronning, T., and Vatn, M. H. (2006). Irritable bowel syndrome in twins: genes and environment. *Gut*, 55(12):1754–1759.

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., and Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44):18933–18938.

Bernet, M. F., Brassart, D., Neeser, J. R., and Servin, A. L. (1994). Lactobacillus acidophilus LA 1 binds to cultured human intestinal cell lines and inhibits cell attachment and cell invasion by enterovirulent bacteria. *Gut*, 35(4):483–489.

Bibiloni, R., Tandon, P., Vargas-Voracka, F., Barreto-Zuniga, R., Lupian-Sanchez, A., Rico-Hinojosa, M. A., Guban, J., Fedorak, R., and Tannock, G. W. (2008). Differential clustering of bowel biopsy-associated bacterial profiles of specimens collected in Mexico and Canada: what do these profiles represent? *Journal of Medical Microbiology*, 57(Pt 1):111–117.

Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J., and Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences of the United States of America*, 103(3):732–737.

Booijink, C. C. G. M., El-Aidy, S., Rajilic-Stojanovic, M., Heilig, H. G. H. J., Troost, F. J., Smidt, H., Kleerebezem, M., De Vos, W. M., and Zoetendal, E. G. (2010). High temporal and inter-individual variation detected in the human ileal microbiota. *Environ Microbiol*, 12(12):3213–3227.

Borody, T. J. and Khoruts, A. (2011). Fecal microbiota transplantation and emerging applications. *Nature Reviews Gastroenterology and Hepatology*, 9(2):88–96.

Bouin, M., Plourde, V., Boivin, M., Riberdy, M., Lupien, F., Laganière, M., Verrier, P., and Poitras, P. (2002). Rectal distention testing in patients with irritable bowel syndrome: sensitivity, specificity, and predictive values of pain sensory thresholds. *Gastroenterology*, 122(7):1771–1777.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*, 185(20):6220–6223.

Brook, I. (1999). Bacterial interference. *Critical Reviews in Microbiology*, 25(3):155–172.

Bueno, L. and Fioramonti, J. (2002). Visceral perception: inflammatory and non-inflammatory mediators. *Gut*, 51(Suppl 1):i19–i23.

Burton, M. B. and Gebhart, G. F. (1995). Effects of intracolonic acetic acid on responses to colorectal distension in the rat. *Brain Research*, 672(1-2):77–82.

Butler, J. E., Sun, J., Weber, P., Navarro, P., and Francis, D. (2000). Antibody repertoire development in fetal and newborn piglets , III. Colonization of the gastrointestinal tract selectively diversifies the preimmune repertoire in mucosal lymphoid tissues. *Immunology*, 100(1):119–130.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5):335–336.

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., Gordon, J. I., and Knight, R. (2011). Moving pictures of the human microbiome. *Genome Biology*, 12(5):R50.

Cario, E. and Podolsky, D. K. (2000). Differential alteration in intestinal epithelial cell expression of Toll-Like Receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. *Infection and Immunity*, 68(12):7010–7017.

Carroll, I. M., Chang, Y.-h., Park, J., Sartor, R. B., and Ringel, Y. (2010). Luminal and mucosal-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Gut Pathogens*, 2(1):19.

Carroll, I. M., Ringel-Kulka, T., Keku, T. O., Chang, Y. H., Packey, C. D., Sartor, R. B., and Ringel, Y. (2011). Molecular analysis of the luminal- and mucosal-associated intestinal microbiota in diarrhea-predominant irritable bowel syndrome. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 301(5):799–807.

Cebra, J. J., Periwal, S. B., Lee, G., Lee, F., and Shroff, K. E. (1998). Development and maintenance of the gut-associated lymphoid tissue (GALT): the roles of enteric bacteria and viruses. *Developmental Immunology*, 6(1-2):13–18.

Chadwick, V. S., Chen, W., Shu, D., Paulus, B., Bethwaite, P., Tie, A., and Wilson, I. (2002). Activation of the mucosal immune system in irritable bowel syndrome. *Gastroenterology*, 122(7):1778–1783.

Chang, J. Y., Antonopoulos, D. a., Kalra, A., Tonelli, A., Khalife, W. T., Schmidt, T. M., and Young, V. B. (2008). Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. *The Journal of Infectious Diseases*, 197(3):435–438.

Chang, L. and Heitkemper, M. M. (2002). Gender differences in irritable bowel syndrome. *Gastroenterology*, 123(5):1686–1701.

Chao, A. (1987). Estimating the population size for capture- recapture data with unequal catchability. *Biometrics*, 43(4):783–791.

Chao, A., Hwang, W.-H., Chen, Y.-C., and Kuo, C.-Y. (2000). Estimating the number of shared species in two communities. *Statistica Sinica*, 10:227–246.

Charlson, E. S., Chen, J., Custers-Allen, R., Bittinger, K., Li, H., Sinha, R., Hwang, J., Bushman, F. D., and Collman, R. G. (2010). Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE*, 5(12):e15216.

Chazdon, R. L., Colwell, R. K., Denslow, J. S., and Guariguata, M. R. (1998). Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. In *Forest biodiversity research, monitoring and modeling: conceptual background and Old World case studies*, pages 285–309. Parthenon Publishing.

Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J. R., Falush, D., Dinan, T., Fitzgerald, G., Stanton, C., van Sinderen, D., O'Connor, M., Harnedy, N., O'Connor, K., Henry, C., O'Mahony, D., Fitzgerald, A. P., Shanahan, F., Twomey, C., Hill, C., Ross, R. P., and O'Toole, P. W. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences of the United States of America*, 108(Suppl 1):4586–4591.

Cohen, R., Roth, F. J., Delgado, E., Ahearn, D. G., and Kalser, M. H. (1969). Fungal flora of the normal human small and large intestine. *The New England Journal of Medicine*, 280(12):638–641.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Bandela, A. M., Cardenas, E., Garrity, G. M., and Tiedje, J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research*, 35(Database issue):D169–D172.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., and Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37(Database issue):D141–D145.

Collins, S. M. and Bercik, P. (2009). The relationship between intestinal microbiota and the central nervous system in normal gastrointestinal function and disease. *Gastroenterology*, 136(6):2003–2014.

Conly, J. M., Stein, K., Worobetz, L., and Rutledge-Harding, S. (1994). The contribution of vitamin K2 (metaquinones) produced by the intestinal microflora to human nutritional requirements for vitamin K. *The American Journal of Gastroenterology*, 89(6):915–923.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697.

Cummings, J. H., Pomare, E. W., Branch, H. W. J., Naylor, C. P. E., and Macfarlane, T. (1987). Short chain fatty acids in human large intestine , portal , hepatic and venous blood. *Gut*, 28(10):1221–1227.

Daniel, R. (2005). The metagenomics of soil. *Nature Reviews Microbiology*, 3(6):470–478.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33):14691–14696.

De La Cochetière, M. F., Durand, T., Lepage, P., Bourreille, A., Galmiche, J. P., and Doré, J. (2005). Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. *Journal of Clinical Microbiology*, 43(11):5588–5592.

Delong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N.-u., Martinez, A., Sullivan, M. B., Edwards, R., Rodriguez-Brito, B., Chisholm, S. W., and Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, 311(5760):496–503.

Deloris Alexander, A., Orcutt, R. P., Henry, J. C., Baker, J. J., Bissahoyo, A. C., and Threadgill, D. W. (2006). Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. *Mammalian Genome: official journal of the International Mammalian Genome Society*, 17(11):1093–1104.

Derrien, M., van Passel, M. W., van de Bovenkamp, J. H. B., Schipper, R. G., de Vos, W. M., and Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes*, 1(4):254–268.

Derrien, M., Vaughan, E. E., Plugge, C. M., and de Vos, W. M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 54(Pt 5):1469–1476.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069–5072.

Dethlefsen, L., Eckburg, P. B., Bik, E. M., and Relman, D. A. (2006). Assembly of the human intestinal microbiota. *Trends in Ecology and Evolution*, 21(9):517–523.

Dethlefsen, L., Mcfall-Ngai, M., and Relman, D. A. (2007). An ecological and evolutionary perspective on human – microbe mutualism and disease. *Nature*, 449(7164):811–818.

Dethlefsen, L. and Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(Suppl 1):4554–4561.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M., Li, L., Mcdaniel, L., Moran, M. A., Nelson, K. E., Nilsson, C., Olson, R., Paul, J., Brito, B. R., Ruan, Y., Swan, B. K., Stevens, R., Valentine, D. L., Thurber, R. V., Wegley, L., White, B. A., and Rohwer, F. (2008). Functional metagenomic profiling of nine biomes. *Nature*, 452(7187):629–632.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11971–11975.

Drossman, D. A. (1999a). Do psychosocial factors define symptom severity and patient status in irritable bowel syndrome? *The American Journal of Medicine*, 107(5A):41S–50S.

Drossman, D. A. (1999b). The functional gastrointestinal disorders and the Rome II process. *Gut*, 45(Suppl II):1–5.

Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A. M., Flint, H. J., and Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Applied and Environmental Microbiology*, 73(4):1073–1078.

Duncan, S. H., Lobley, G. E., Holtrop, G., Ince, J., Johnstone, A. M., Louis, P., and Flint, H. J. (2008). Human colonic microbiota associated with diet, obesity and weight loss. *International Journal of Obesity*, 32(11):1720–1724.

Duncan, S. H., Louis, P., Thomson, J. M., and Flint, H. J. (2009). The role of pH in determining the species composition of the human colonic microbiota. *Environmental Microbiology*, 11(8):2112–2122.

Durbán, A., Abellán, J. J., Jiménez-Hernández, N., Latorre, A., and Moya, A. (2012a). Daily follow-up of bacterial communities in the human gut reveals stable composition and host-specific patterns of interaction. *FEMS Microbiology Ecology*, 81(2):427–437.

Durbán, A., Abellán, J. J., Jiménez-Hernández, N., Ponce, M., Ponce, J., Sala, T., D'Auria, G., Latorre, A., and Moya, A. (2011). Assessing gut microbial diversity from feces and rectal mucosa. *Microbial Ecology*, 61(1):123–133.

Durbán, A., Abellán, J. J., Jiménez-Hernández, N., Salgado, P., Ponce, M., Ponce, J., Garrigues, V., Latorre, A., and Moya, A. (2012b). Structural alterations of faecal and mucosa-associated bacterial communities in irritable bowel syndrome. *Environmental Microbiology Reports*, 4(2):242–247.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308(5728):1635–1638.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9):755–763.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461.

Elsenbruch, S. and Orr, W. C. (2001). Diarrhea- and constipation-predominant IBS patients differ in postprandial autonomic and cortisol responses. *The American Journal of Gastroenterology*, 96(2):460–466.

Engelbrektson, A., Korzenik, J. R., Pittler, A., Sanders, M. E., Klaenhammer, T. R., Leyer, G., and Kitts, C. L. (2009). Probiotics to minimize the disruption of faecal microbiota in healthy subjects undergoing antibiotic therapy. *Journal of Medical Microbiology*, 58(5):663–670.

Escherich, T. (1885). Die Darmbakterien des Neugeborenen und Säuglings. *Fortschritte der Medizin*, 3:515–522,547–554.

Escherich, T. (1886). Die Darmbakterien des Säuglings und ihre Beziehungen zur Physiologie der Verdauung.

Faith, J. J., Rey, F. E., O'Donnell, D., Karlsson, M., Mcnulty, N. P., Kallstrom, G., Goodman, A. L., and Gordon, J. I. (2010). Creating and characterizing communities of human gut microbes in gnotobiotic mice. *The ISME Journal*, 4(9):1094–1098.

Fallingborg, J. (1999). Intraluminal pH of the human gastrointestinal tract. *Danish Medical Bulletin*, 46(3):183–196.

Falony, G., Vlachou, A., Verbrugghe, K., and De Vuyst, L. (2006). Cross-feeding between Bifidobacterium longum BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. *Applied and Environmental Microbiology*, 72(12):7835–7841.

Felsenstein, J. (1989). PHYLIP - phylogeny inference package (version 3.2). *Cladistics*, 5:164–166.

Finegold, S. M., Sutter, V. L., and Mathisen, G. E. (1983). Normal indigenous intestinal flora. In *Human intestinal microflora in health and disease*, pages 3–31. Academic Press.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, 8:376–386.

Flint, H. J. (2004). Polysaccharide breakdown by anaerobic microorganisms inhabiting the mammalian gut. *Advances in Applied Microbiology*, 56:89–120.

Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R., and White, B. A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology*, 6(2):121–131.

Fooks, L. J. and Gibson, G. R. (2002). Probiotics as modulators of the gut flora. *British Journal of Nutrition*, 88(Suppl 1):39–49.

Frank, D. N., Zhu, W., Sartor, R. B., and Li, E. (2011). Investigating the biological and clinical significance of human dysbioses. *Trends in Microbiology*, 19(9):427–34.

Frank, L., Kleinman, L., Rentz, A., Ciesla, G., Kim, J. J., and Zacker, C. (2002). Health-related quality of life associated with irritable bowel syndrome: comparison with other chronic diseases. *Clinical Therapeutics*, 24(4):675–689.

Frankel, W. L., Zhang, W., Singh, A., Klurfeld, D. M., Don, S., Sakata, T., Modlin, I., and Rombeau, J. L. (1994). Mediation of the trophic effects of short-chain fatty acids on the rat jejunum and colon. *Gastroenterology*, 106(2):375–380.

Franks, A. H., Harmsen, H. J., Raangs, G. C., Jansen, G. J., Schut, F., and Welling, G. W. (1998). Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Applied and Environmental Microbiology*, 64(9):3336–3345.

Fredricks, D. N. (2011). Molecular methods to describe the spectrum and dynamics of the vaginal microbiota. *Anaerobe*, 17(4):191–195.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., and Delong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10):3805–3810.

Fukudo, S. and Kanazawa, M. (2011). Gene, environment, and brain-gut interactions in irritable bowel syndrome. *Journal of Gastroenterology and Hepatology*, 26(Suppl 3):110–115.

Gareau, M. G., Sherman, P. M., and Walker, W. A. (2010). Probiotics and the gut microbiota in intestinal health and disease. *Nature Reviews Gastroenterology and Hepatology*, 7(9):503–514.

Garrett, W. S., Lord, G. M., Punit, S., Lugo-Villarino, G., Mazmanian, S. K., Ito, S., Glickman, J. N., and Glimcher, L. H. (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell*, 131(1):33–45.

Gaya, D. R., Russell, R. K., Nimmo, E. R., and Satsangi, J. (2006). New genes in inflammatory bowel disease: lessons for complex diseases? *Lancet*, 367(9518):1271–1284.

Gelfand, A. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–25.

Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F., Sarkar, D., and Hansen, K. (2008). Rgraphviz: provides plotting capabilities for R graph objects.

Gerritsen, J., Smidt, H., Rijkers, G. T., and de Vos, W. M. (2011). Intestinal microbiota in human health and disease: the impact of probiotics. *Genes and Nutrition*, 6(3):209–240.

Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J., and Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *Journal of Immunology*, 167(4):1882–1885.

Gibson, G. R., Macfarlane, G. T., and Cummings, J. H. (1993). Sulphate reducing bacteria and hydrogen metabolism in the human large intestine. *Gut*, 34(4):437–439.

Gibson, P. R., Moeller, I., Kagelari, O., Folino, M., and Young, G. P. (1992). Contrasting effects of butyrate on the expression of phenotypic markers of differentiation in neoplastic and non-neoplastic colonic epithelial cells in vitro. *Journal of Gastroenterology and Hepatology*, 7(2):165–172.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., and Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778):1355–1359.

Gini, C. (1912). *Variabilità e mutabilità. Reprinted in 'Memorie di metodologica statistica'*. Libreria Eredi Virgilio Veschi (1955).

Girardin, S. E., Tournebize, R., Mavris, M., Page, A.-l., Li, X., Stark, G. R., Bertin, J., Distefano, P. S., Yaniv, M., Sansonetti, P. J., and Philpott, D. J. (2001). CARD4/Nod1 mediates NF-kB and JNK activation by invasive Shigella flexneri. *EMBO Reports*, 2(8):736–742.

Gonzalez, A., Clemente, J. C., Shade, A., Metcalf, J. L., Song, S., Prithiviraj, B., Palmer, B. E., and Knight, R. (2011). Our microbial selves: what ecology can teach us. *EMBO Reports*, 12(8):775–784.

Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellán, J. J., Jiménez-Hernández, N., Pérez-Cobas, A. E., Latorre, A., and Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS ONE*, 6(3):e17447.

Gosalbes, M. J., Llop, S., Vallès, Y., Moya, A., Ballester, F., and Francino, M. P. (2012). Meconium microbiota types dominated by lactic acid or enteric

bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical and Experimental Allergy*, In press.

Gough, E., Shaikh, H., and Manges, A. R. (2011). Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection. *Clinical Infectious Diseases*, 53(10):994–1002.

Grehan, M. J., Borody, T. J., Leis, S. M., Campbell, J., Mitchell, H., and Wettstein, A. (2010). Durable alteration of the colonic microbiota by the administration of donor fecal flora. *Journal of Clinical Gastroenterology*, 44(8):551–561.

Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., Bouffard, G. G., Blakesley, R. W., Murray, P. R., Green, E. D., Turner, M. L., and Segre, J. A. (2009). Topographical and temporal diversity of the human skin microbiome. *Science*, 324(5931):1190–1192.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Research*, 31(1):439–441.

Grönlund, M. M., Lehtonen, O. P., Eerola, E., and Kero, P. (1999). Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *Journal of Pediatric Gastroenterology and Nutrition*, 28(1):19–25.

Guarner, F., Bourdet-Sicard, R., Brandtzaeg, P., Gill, H. S., McGuirk, P., van Eden, W., Versalovic, J., Weinstock, J. V., and Rook, G. A. (2006). Mechanisms of disease: The hygiene hypothesis revisited. *Nature Clinical Practice. Gastroenterology and Hepatology*, 3(5):275–284.

Guarner, F. and Malagelada, J.-R. (2003). Gut flora in health and disease. *Lancet*, 361(9356):512–519.

Halldén, G. and Aponte, G. W. (1997). Evidence for a role of the gut hormone PYY in the regulation of intestinal fatty acid-binding protein transcripts in differentiated subpopulations of intestinal epithelial cell hybrids. *The Journal of Biological Chemistry*, 272(19):12591–12600.

Handelsman, J. (2004). Metagenomics : application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4):669–685.

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Mäkelä, M. J., Vartiainen, E., Kosunen, T. U., Alenius, H., and Haahtela, T. (2012). Environmental biodiversity, human microbiota, and allergy are interrelated. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21):8334–8339.

Harmsen, H. J., Wildeboer-Veloo, A. C. M., Raangs, G. C., Wagendorp, A. A., Klijn, N., Bindels, J. G., and Welling, G. W. (2000). Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *Journal of Pediatric Gastroenterology and Nutrition*, 30(1):61–67.

Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M., and Benno, Y. (2005). Molecular analysis of jejunal , ileal , caecal and recto- sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *Journal of Medical Microbiology*, 54(Pt 11):1093–1101.

Hill, M. J. (1997). Intestinal flora and endogenous vitamin synthesis. *European Journal of Cancer Prevention*, 6(Suppl 1):43–45.

Högenauer, C., Hammer, H. F., Krejs, G. J., and Reisinger, E. C. (1998). Mechanisms and management of antibiotic-associated diarrhea. *Clinical Infectious Diseases*, 27(4):702–710.

Hong, S., Bunge, J., Leslin, C., Jeon, S., and Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, 3(12):1365–1373.

Hongoh, Y., Yuzawa, H., Ohkuma, M., and Kudo, T. (2003). Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiology Letters*, 221(2):299–304.

Hood, S. D., Shufflebotham, J. Q., Hendry, J., Hince, D. A., Rich, A. S., Probert, C. S. J., and Potokar, J. (2008). Irritable bowel syndrome patients exhibit depressive and anxiety scores in the subsyndromal range. *The Open Psychiatry Journal*, 2:12–22.

Hooper, L. V., Wong, M. H., Thelin, A., Hansson, L., Falk, P. G., and Gordon, J. I. (2001). Molecular analysis of commensal host-microbial relationships in the intestine. *Science*, 291(5505):881–884.

Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T., and Gordon, J. I. (1999). A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17):9833–9838.

Houten, S. M., Watanabe, M., and Auwerx, J. (2006). Endocrine functions of bile acids. *The EMBO Journal*, 25(7):1419–1425.

Hungate, R. E. (1950). The anaerobic mesophilic cellulolytic bacteria. *Bacteriology Reviews*, 14:1–49.

Hungate, R. E. (1966). *The rumen and its microbes*. Academic Press.

Hungin, A. P. S., Chang, L., Locke, G. R., Dennis, E. H., and Barghout, V. (2005). Irritable bowel syndrome in the United States: prevalence, symptom patterns and impact. *Alimentary Pharmacology and Therapeutics*, 21(11):1365–1375.

Hungin, A. P. S., Whorwell, P. J., Tack, J., and Mearin, F. (2003). The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40000 subjects. *Alimentary Pharmacology and Therapeutics*, 17(5):643–650.

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3):377–386.

Hütt, P., Shchepetova, J., Lõivukene, K., Kullisaar, T., and Mikelsaar, M. (2006). Antagonistic activity of probiotic lactobacilli and bifidobacteria against entero- and uropathogens. *Journal of Applied Microbiology*, 100(6):1324–32.

Iannitti, T. and Palmieri, B. (2010). Therapeutical use of probiotic formulations in clinical practice. *Clinical Nutrition*, 29(6):701–725.

Inadomi, J. M., Fennerty, M. B., and Bjorkman, D. (2003). Systematic review: the economic impact of irritable bowel syndrome. *Alimentary Pharmacology and Therapeutics*, 18(7):671–682.

Inagaki, T., Moschetta, A., Lee, Y.-K., Peng, L., Zhao, G., Downes, M., Yu, R. T., Shelton, J. M., Richardson, J. A., Repa, J. J., Mangelsdorf, D. J., and Kliewer, S. A. (2006). Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3920–3925.

Ingham, C. J., Sprenkels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J. E. T., and de Vos, W. M. (2007). The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46):18217–18222.

Islam, K. B. M. S., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T., and Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology*, 141(5):1773–1781.

Jernberg, C., Löfmark, S., Edlund, C., and Jansson, J. K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME Journal*, 1(1):56–66.

Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernández, L., and Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology*, 159(3):187–193.

Johansson, M. E. V., Larsson, J. M. H., and Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. 108(Suppl 1):4659–4665.

Kajander, K., Myllyluoma, E., Rajilic-Stojanovic, M., Kyrönpalo, S., Rasmussen, M., Järvenpää, S., Zoetendal, E. G., de Vos, W. M., Vapaatalo, H., and Korpela, R. (2008). Clinical trial: multispecies probiotic supplementation alleviates the symptoms of irritable bowel syndrome and stabilizes intestinal microbiota. *Alimentary Pharmacology and Therapeutics*, 27(1):48–57.

Kalantar, J. S., Locke III, G. R., Zinsmeister, A. R., Beighley, C. M., and Talley, N. J. (2003). Familial aggregation of irritable bowel syndrome: a prospective study. *Functional Bowel Disease*, 52(12):1703–1707.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research*, 32(Database issue):D277–280.

Kankainen, M., Paulin, L., Tynkkynen, S., von Ossowski, I., Reunanen, J., Partanen, P., Satokari, R., Vesterlund, S., Hendrickx, A. P. a., Lebeer, S., De Keersmaecker, S. C. J., Vanderleyden, J., Hämäläinen, T., Laukkanen,

S., Salovuori, N., Ritari, J., Alatalo, E., Korpela, R., Mattila-Sandholm, T., Lassig, A., Hatakka, K., Kinnunen, K. T., Karjalainen, H., Saxelin, M., Laakso, K., Surakka, A., Palva, A., Salusjärvi, T., Auvinen, P., and de Vos, W. M. (2009). Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human-mucus binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40):17193–17198.

Karantanos, T., Markoutsaki, T., Gazouli, M., Anagnou, N. P., and Karamanolis, D. G. (2010). Current insights in to the pathophysiology of irritable bowel syndrome. *Gut Pathogens*, 2(1):3.

Kassinen, A., Krogius-kurikka, L., Mäkivuokko, H., Teemu, R., Paulin, L., Corander, J., Malinen, E., Apajalahti, J., and Palva, A. (2007). The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology*, 133(1):24–33.

Katz, J. A. (2006). Probiotics for the prevention of antibiotic-associated diarrhea and Clostridium difficile diarrhea. *Journal of Clinical Gastroenterology*, 40(3):249–255.

Kerckhoffs, A. P. M., Samsom, M., van der Rest, M. E., de Vogel, J., Knol, J., Ben-Amor, K., and Akkermans, L. M. A. (2009). Lower Bifidobacteria counts in both duodenal mucosa-associated and fecal microbiota in irritable bowel syndrome patients. *World Journal of Gastroenterology*, 15(23):2887.

Kerckhoffs, P. M., Ben-Amor, K., Samsom, M., van der Rest, M. E., de Vogel, J., Knol, J., and Akkermans, L. M. A. (2011). Molecular analysis of faecal and duodenal samples reveals significantly higher prevalence and numbers of Pseudomonas aeruginosa in irritable bowel syndrome. *Journal of Medical Microbiology*, 60(Pt 2):236–245.

Khachatryan, Z. a., Ktsoyan, Z. a., Manukyan, G. P., Kelly, D., Ghazaryan, K. a., and Aminov, R. I. (2008). Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE*, 3(8):e3064.

Kim, H., Sun, D., and R.Tsutakawa (2001). A bivariate Bayes method for improving the estimates of mortality rates with 2-fold CAR model. *Journal of the American Statistical Association*, 96(456):1506–1521.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., and Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 108(Suppl 1):4578–4585.

Kolida, S., Tuohy, K., and Gibson, G. R. (2007). Prebiotic effects of inulin and oligofructose. *British Journal of Nutrition*, 87(Suppl 2):193–197.

Kolmeder, C. a., de Been, M., Nikkilä, J., Ritamo, I., Mättö, J., Valmu, L., Salojärvi, J., Palva, A., Salonen, A., and de Vos, W. M. (2012). Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS ONE*, 7(1):e29913.

Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology*, 10(5):323–335.

Krogius-kurikka, L., Lyra, A., Malinen, E., Aarnikunnas, J., Tuimala, J., Paulin, L., Mäkivuokko, H., Kajander, K., and Palva, A. (2009). Microbial community analysis reveals high level phylogenetic alterations in the overall gastrointestinal microbiota of diarrhoea-predominant irritable bowel syndrome sufferers. *BMC Gastroenterology*, 9:95.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrlich, D. S., Itoh, K., Takagi, T., Sakaki, Y., Hayashi, T., and Hattori, M. (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research*, 14(4):169–181.

Le Gall, G., Noor, S. O., Ridgway, K., Scovell, L., Jamieson, C., Johnson, I. T., Colquhoun, I. J., Kemsley, E. K., and Narbad, A. (2011). Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *Journal of Proteome Research*, 10(9):4208–4218.

Lea, R. and Whorwell, P. J. (2003). New insights into the psychosocial aspects of irritable bowel syndrome. *Current Gastroenterology Reports*, 5(4):343–350.

Lee, A. (1999). Helicobacter pylori: opportunistic member of the normal microflora or agent of communicable disease. In *Medical importance of the normal nicroflora*, pages 128–163. Kluwer Academic Publishers.

Lee, H. R. and Pimentel, M. (2006). Bacteria and irritable bowel syndrome: the evidence for small intestinal bacterial overgrowth. *Current Gastroenterology Reports*, 8(4):305–311.

Lepage, P., Seksik, P., Sutren, M., de la Cochetière, M. F., Jian, R., Marteau, P., and Doré, J. (2005). Biodiversity of the mucosa-associated microbiota Is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflammatory Bowel Diseases*, 11(5):473–480.

Levitt, M. D. (1971). Volume and composition of human intestinal gas determined by means of an intestinal washout technic. *The New England Journal of Medicine*, 284(25):1394–1398.

Levy, R. L., Jones, K. R., Whitehead, W. E., Feld, S. I., Talley, N. J., and Corey, L. A. (2001). Irritable bowel syndrome in twins: heredity and social learning both contribute to etiology. *Gastroenterology*, 121(4):799–804.

Lewis, S. J. and Heaton, K. W. (1997). Stool form scale as a useful guide to intestinal transit time. *Scandinavian Journal of Gastroenterology*, 32(9):920–924.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. a., Knight, R. D., and Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):11070–11075.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., and Gordon, J. I. (2008). Evolution of mammals and their gut microbes. *Science*, 320(5883):1647–1651.

Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122):1022–1023.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659.

Longstreth, G. F., Thompson, W. G., Chey, W. D., Houghton, L. A., Mearin, F., and Spiller, R. C. (2006). Functional bowel disorders. *Gastroenterology*, 130(5):1480–1491.

Lozupone, C., Hamady, M., and Knight, R. (2006). UniFrac - an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*, 7:371.

Luckey, T. D. (1972). Introduction to intestinal microecology. *The American Journal of Clinical Nutrition*, 25(12):1292–1294.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.

Lupp, C., Robertson, M. L., Wickham, M. E., Sekirov, I., Champion, O. L., Gaynor, E. C., and Finlay, B. B. (2007). Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host and Microbe*, 2(3):119–129.

Lyra, A., Rinttilä, T., Nikkilä, J., Krogius-kurikka, L., Kajander, K., Malinen, E., Mättö, J., Mäkelä, L., and Palva, A. (2009). Diarrhoea-predominant irritable bowel syndrome distinguishable by 16S rRNA gene phylotype quantification. *World Journal of Gastroenterology*, 15(47):5936–5945.

Macfarlane, G. T., Cummings, J. H., and Allison, C. (1986). Protein degradation by human intestinal bacteria. *Journal of General Microbiology*, 132(6):1647–1656.

Macfarlane, G. T., Gibson, G. R., and Cummings, J. H. (1992). Comparison of fermentation reactions in different regions of the human colon. *Journal of Applied Bacteriology*, 72(1):57–64.

Macfarlane, S. and Macfarlane, G. T. (2003). Regulation of short-chain fatty acid production. *Proceedings of the Nutrition Society*, 62(1):67–72.

Macpherson, A., Khoo, U. Y., Forgacs, I., Philpott-Howard, J., and Bjarnason, I. (1996). Mucosal antibodies in inflammatory bowel disease are directed against intestinal bacteria. *Gut*, 38(3):365–375.

Macpherson, A. J. and Uhr, T. (2004). Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science*, 303(5664):1662–1665.

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., Verberkmoes, N. C., Hettich, R. L., and Gordon, J. I. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5859–5864.

Malinen, E., Rinttilä, T., Kajander, K., Mättö, J., Kassinen, A., Krogius, L., Saarela, M., Korpela, R., and Palva, A. (2005). Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. *The American Journal of Gastroenterology*, 100(2):373–382.

Manichanh, C., Reeder, J., Gibert, P., Varela, E., Llopis, M., Antolin, M., Guigo, R., Knight, R., and Guarner, F. (2010). Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Research*, 20(10):1411–1419.

Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009). Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *The Journal of Biological Chemistry*, 284(37):24673–24677.

Martin, F.-P. J., Dumas, M.-e., Wang, Y., Legido-Quigley, C., Yap, I. K. S., Tang, H., Zirah, S., Murphy, G. M., Cloarec, O., Lindon, J. C., Sprenger, N., Fay, L. B., Kochhar, S., van Bladeren, P., Holmes, E., and Nicholson, J. K. (2007). A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Molecular Systems Biology*, 3:112.

Martinez, C., Antolin, M., Santos, J., Torrejon, A., Casellas, F., Borruel, N., Guarner, F., and Malagelada, J.-R. (2008). Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *The American Journal of Gastroenterology*, 103(3):643–648.

Martínez, V., Wang, L., Rivier, J., Grigoriadis, D., and Taché, Y. (2004). Central CRF, urocortins and stress increase colonic transit via CRF1 receptors while activation of CRF2 receptors delays gastric transit in mice. *The Journal of Physiology*, 556(Pt 1):221–234.

Mättö, J., Maunuksela, L., Kajander, K., Palva, A., Korpela, R., Kassinen, A., and Saarelaa, M. (2005). Composition and temporal stability of gas-

trointestinal microbiota in irritable bowel syndrome - a longitudinal study in IBS and control subjects. *FEMS Immunology and Medical Microbiology*, 43(2):213–222.

Maukonen, J., Satokari, R., Mättö, J., Söderlund, H., Mattila-Sandholm, T., and Saarela, M. (2006). Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria. *Journal of Medical Microbiology*, 55(Pt 5):625–633.

Maxwell, P. R., Rink, E., Kumar, D., and Mendall, M. A. (2002). Antibiotics increase functional abdominal symptoms. *The American Journal of Gastroenterology*, 97(1):104–108.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, 122(1):107–118.

Mazmanian, S. K., Round, J. L., and Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*, 453(7195):620–625.

McKenzie, Y. A., Alder, A., Anderson, W., Wills, A., Goddard, L., Gulia, P., Jankovich, E., Mutch, P., Reeves, L. B., Singer, A., and Lomer, M. C. (2012). British Dietetic Association evidence-based guidelines for the dietary management of irritable bowel syndrome in adults. *Journal of Human Nutrition and Dietetics*, 25(3):260–274.

McNeil, N. I. (1984). The contribution of the large intestine to energy supplies in man. *The American Journal of Clinical Nutrition*, 39(2):338–342.

Mendall, M. A. and Kumar, D. (1998). Antibiotic use, childhood affluence and irritable bowel syndrome (IBS). *European Journal of Gastroenterology and Hepatology*, 10(1):59–62.

Meslin, J. C., Fontaine, N., and Andrieux, C. (1999). Variation of mucin distribution in the rat intestine, caecum and colon: effect of the bacterial flora. *Comparative Biochemistry and Physiology. Part A, Molecular and Integrative Physiology*, 123(3):235–239.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11(1):31–46.

Meyer, D. and Stasse-Wolthuis, M. (2009). The bifidogenic effect of inulin and oligofructose and its consequences for gut health. *European Journal of Clinical Nutrition*, 63(11):1277–1289.

Meyer-Hoffert, U., Hornef, M. W., Henriques-Normark, B., Axelsson, L.-G., Midtvedt, T., Pütsep, K., and Andersson, M. (2008). Secreted enteric antimicrobial activity localises to the mucus surface layer. *Gut*, 57(6):764–771.

Moayyedi, P., Ford, A. C., Talley, N. J., Cremonini, F., Foxx-Orenstein, A. E., Brandt, L. J., and Quigley, E. M. M. (2010). The efficacy of probiotics in the treatment of irritable bowel syndrome: a systematic review. *Gut*, 59(3):325–332.

Momozawa, Y., Deffontaine, V., Louis, E., and Medrano, J. F. (2011). Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. *PLoS ONE*, 6(2):e16952.

Moore, W. E. C. and Holdeman, L. V. (1974). Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Applied Microbiology*, 27(5):961–979.

Mukhopadhya, I., Hansen, R., El-Omar, E. M., and Hold, G. L. (2012). IBD-what role do Proteobacteria play? *Nature Reviews Gastroenterology and Hepatology*, 9(4):219–230.

Mulder, I. E., Schmidt, B., Stokes, C. R., Lewis, M., Bailey, M., Aminov, R. I., Prosser, J. I., Gill, B. P., Pluske, J. R., Mayer, C.-D., Musk, C. C., and Kelly, D. (2009). Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. *BMC Biology*, 7:79.

Müller, C. A., Autenrieth, I. B., and Peschel, A. (2005). Innate defenses of the intestinal epithelial barrier. *Cellular and Molecular Life Sciences*, 62(12):1297–1307.

Nakamura, N., Lin, H. C., McSweeney, C. S., Mackie, R. I., and Gaskins, H. R. (2010). Mechanisms of microbial hydrogen disposal in the human colon and implications for health and disease. *Annual Review of Food Science and Technology*, 1:363–395.

Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics*, 25(10):1335–1337.

Neal, K. R., Barker, L., and Spiller, R. C. (2002). Prognosis in post-infective irritable bowel syndrome: a six year follow up study. *Gut*, 51(3):410–413.

Nell, S., Suerbaum, S., and Josenhans, C. (2010). The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nature Reviews Microbiology*, 8(8):564–577.

Neutra, M. R., Mantis, N. J., and Kraehenbuhl, J. P. (2001). Collaboration of epithelial cells with organized mucosal lymphoid tissues. *Nature Immunology*, 2(11):1004–1009.

Ng, S. C., Hart, A. L., Kamm, M. A., Stagg, A. J., and Knight, S. C. (2009). Mechanisms of action of probiotics: recent advances. *Inflammatory Bowel Diseases*, 15(2):300–310.

O'Hara, A. M. and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7):688–693.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., and Wagner, H. (2011). vegan: community ecology package.

Opgen-Rhein, R. and Strimmer, K. (2007). From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, 1:37.

Ott, S. J., Musfeldt, M., Timmis, K. N., Hampe, J., Wenderoth, D. F., and Schreiber, S. (2004). In vitro alterations of intestinal bacterial microbiota in fecal samples during storage. *Diagnostic Microbiology and Infectious Disease*, 50(4):237–245.

Ouwehand, A. C., Salminen, S., Arvola, T., Ruuska, T., and Isolauri, E. (2004). Microbiota composition of the intestinal mucosa: association with fecal microbiota? *Microbiology and Immunology*, 48(7):497–500.

Pace, F., Molteni, P., Bollani, S., Sarzi-Puttini, P., Stockbrügger, R., Bianchi-Porro, G., and Drossman, D. A. (2003). Inflammatory bowel disease versus irritable bowel syndrome: a hospital-based, case-control study of disease impact on quality of life. *Scandinavian Journal of Gastroenterology*, 38(10):1031–1038.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. a., and Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biology*, 5(7):e177.

Parkes, G. C., Brostoff, J., Whelan, K., and Sanderson, J. D. (2008). Gastrointestinal microbiota in irritable bowel syndrome: their role in its pathogenesis and treatment. *The American Journal of Gastroenterology*, 103(6):1557–1567.

Pasteur, L. (1860). Mémoire sur la fermentation alcoolique. *Annales de Chimie et de Physique*, 58:323–426.

Pei, Z., Bini, E. J., Yang, L., Zhou, M., Francois, F., and Blaser, M. J. (2004). Bacterial biota in the human distal esophagus. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12):4250–4255.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P. a., and Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, 118(2):511–521.

Peris-Bondia, F., Latorre, A., Artacho, A., Moya, A., and D'Auria, G. (2011). The active human gut microbiota differs from the total microbiota. *PLoS ONE*, 6(7):e22448.

Phillips, M. L. (2009). Gut reaction: environmental effects on the human microbiota. *Environmental Health Perspectives*, 117(5):A198–A205.

Pignatelli, M., Moya, A., and Tamames, J. (2009). EnvDB, a database for describing the environmental distribution of prokaryotic taxa. *Environmental Microbiology Reports*, 1(3):191–197.

Pimentel, M., Lin, H. C., Enayati, P., van den Burg, B., Lee, H.-R., Chen, J. H., Park, S., Kong, Y., and Conklin, J. (2006). Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 290(6):G1089–G1095.

Ponnusamy, K., Choi, J. N., Kim, J., Lee, S.-y., and Lee, C. H. (2011). Microbial community and metabolomic comparison of irritable bowel syndrome faeces. *Journal of Medical Microbiology*, 60(Pt 6):817–827.

Posserud, I., Stotzer, P.-O., Björnsson, E. S., Abrahamsson, H., and Simrén, M. (2007). Small intestinal bacterial overgrowth in patients with irritable bowel syndrome. *Gut*, 56(6):802–8.

Price, L. B., Liu, C. M., Johnson, K. E., Aziz, M., Lau, M. K., Bowers, J., Ravel, J., Keim, P. S., Serwadda, D., Wawer, M. J., and Gray, R. H. (2010). The effects of circumcision on the penis microbiome. *PLoS ONE*, 5(1):e8422.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B. r., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S. r., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.

Quigley, E. M. M., Bytzer, P., Jones, R., and Mearin, F. (2006). Irritable bowel syndrome: the burden and unmet needs in Europe. *Digestive and Liver Disease*, 38(10):717–723.

Quinn, G. P. and Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.

R Development Core Team (2010). *R: a language and environment for statistical computing*. Vienna, Austria.

Raes, J., Foerstner, K. U., and Bork, P. (2007). Get the most out of your metagenome: computational analysis of environmental sequence data. *Current Opinion in Microbiology*, 10(5):490–498.

Rajilic-Stojanovic, M. (2007). *Diversity of the human gastrointestinal microbiota - Novel perspectives from high troughput analyses*. PhD thesis, University of Wageningen, the Netherlands.

Rajilic-Stojanovic, M., Heilig, H. G. H. J., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., and de Vos, W. M. (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis

of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environmental Microbiology*, 11(7):1736–1751.

Ram, R. J., Verberkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C. n., Shah, M., Hettich, R. L., and Banfield, J. F. (2005). Community proteomics identifies key activities in a natural microbial biofilm. *Science*, 308(5730):1915–1920.

Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F., and Gordon, J. I. (2010). Viruses in the fecal microbiota of monozygotic twins and their mothers. *Nature*, 466(7304):334–338.

Ridlon, J. M., Kang, D.-J., and Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, 47(2):241–259.

Roediger, W. E. (1982). Utilization of nutrients by isolated epithelial cells of the rat colon. *Gastroenterology*, 83(2):424–429.

Roediger, W. E. and Moore, A. (1981). Effect of short-chaim fatty acid on sodium absorption in isolated human colon perfused through the vascular bed. *Digestive Diseases and Sciences*, 26(2):100–106.

Roos, S. and Jonsson, H. (2002). A high-molecular-mass cell-surface protein from Lactobacillus reuteri 1063 adheres to mucus components. *Microbiology*, 148(Pt 2):433–442.

Ruas-Madiedo, P., Gueimonde, M., Fernández-García, M., de los Reyes-Gavilán, C. G., and Margolles, A. (2008). Mucin degradation by Bifidobacterium strains isolated from the human intestinal microbiota. *Applied and environmental microbiology*, 74(6):1936–40.

Salonen, A., de Vos, W. M., and Palva, A. (2010). Gastrointestinal microbiota in irritable bowel syndrome: present state and perspectives. *Microbiology*, 156(Pt 11):3205–3215.

Salyers, A. A., West, S. E. H., Vercellotti, J. R., and Wilkins, T. D. (1977). Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Applied and Environmental Microbiology*, 34(5):529–533.

Salzberg, S. L., Delcher, A. L., Kasif, S., and White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, 26(2):544–548.

Samuel, B. S., Shaito, A., Motoike, T., Rey, F. E., Bäckhed, F., Manchester, J. K., Hammer, R. E., Williams, S. C., Crowley, J., Yanagisawa, M., and Gordon, J. I. (2008). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43):16767–16772.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual Review of Microbiology*, 31:107–133.

Savage, D. C. (2001). Microbial biota of the human intestine: a tribute to some pioneering scientists. *Current Issues in Intestinal Microbiology*, 2(1):1–15.

Savage, D. C., Siegel, J. E., Snellen, J. E., and Whitt, D. D. (1981). Transit time of epithelial cells in the small intestines of germfree mice and ex-germfree mice associated with indigenous microorganisms. *Applied and Environmental Microbiology*, 42(6):996–1001.

Scanlan, P. D. and Marchesi, J. R. (2008). Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. *The ISME Journal*, 2(12):1183–1193.

Scanlan, P. D., Shanahan, F., and Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiology*, 8:79.

Scanlan, P. D., Shanahan, F., O'Mahony, C., and Marchesi, J. R. (2006). Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *Journal of Clinical Microbiology*, 44(11):3980–3988.

Schaefer, J., Opgen-Rhein, R., and Strimmer, K. (2009). GeneNet: modeling and inferring gene networks.

Schafer, J., Opgen-Rhein, R., and Strimmer, K. (2010). corpcor: efficient estimation of covariance and (partial) correlation.

Schäfer, J. and Strimmer, K. (2005a). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4:Article 32.

Schäfer, J. and Strimmer, K. (2005b). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., and Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23):7537–7541.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. a., Donus, C., and Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity*, 18(1):190–195.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6):R60.

Seksik, P. (2010). Gut microbiota and IBD. *Gastroentérologie Clinique et Biologique*, 34(Suppl 1):44–51.

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., Jian, R., and Doré, J. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut*, 52(2):237–242.

Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., Richter, A. R., and White, O. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Research*, 35(Database issue):D260–264.

Serra, J., Azpiroz, F., and Malagelada, J.-R. (2001). Impaired transit and tolerance of intestinal gas in the irritable bowel syndrome. *Gut*, 48(1):14–19.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiology Reviews*, 28(4):405–440.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Shi, Y., Tyson, G. W., and Delong, E. F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature*, 459(7244):266–269.

Shin, R., Suzuki, M., and Morishita, Y. (2002). Influuence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic Escherichia coli O157:H7. *Journal of Medical Microbiology*, 51(3):201–206.

Si, J.-M., Yu, Y.-C., Fan, Y.-J., and Chen, S.-J. (2004). Intestinal microecology and quality of life in irritable bowel syndrome patients. *World Journal of Gastroenterology*, 10(12):1802–1805.

Siavoshian, S., Segain, J.-P., Kornprobst, M., Bonnet, C., Cherbut, C., Galmiche, J. P., and Blottière, H. M. (2000). Butyrate and trichostatin A effects on the proliferation/differentiation of human intestinal epithelial cells: induction of cyclin D3 and p21 expression. *Gut*, 46(4):507–514.

Sime-Ngando, T. and Colombet, J. (2009). Virus and prophages in aquatic ecosystems. *Canadian Journal of Microbiology*, 55(2):95–109.

Simrén, M. and Stotzer, P.-O. (2006). Use and abuse of hydrogen breath tests. *Gut*, 55(3):297–303.

Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., and Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, 60(2):341–350.

Sjölund, M., Wreiber, K., Andersson, D. I., Blaser, M. J., and Engstrand, L. (2003). Long-term persistence of resistant Enterococcus species after antibiotics to eradicate Helicobacter pylori. *Annals of Internal Medicine*, 139(6):483–487.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored rare biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32):12115–12120.

Sokol, H., Seksik, P., Rigottier-Gois, L., Lay, C., Lepage, P., Podglajen, I., Marteau, P., and Doré, J. (2006). Specificities of the fecal microbiota in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 12(2):106–111.

Sonnenburg, J. L., Chen, C. T. L., and Gordon, J. I. (2006). Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biology*, 4(12):e413.

Spiller, R. and Garsed, K. (2009). Postinfectious irritable bowel syndrome. *Gastroenterology*, 136(6):1979–1988.

Spiller, R. C. (2009). Overlap between irritable bowel syndrome and inflammatory bowel disease. *Digestive Diseases*, 27(Suppl 1):48–54.

Spiller, R. C. (2011). Targeting the 5-HT(3) receptor in the treatment of irritable bowel syndrome. *Current Opinion in Pharmacology*, 11(1):68–74.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology*, 9(4):279–290.

Staden, R., Beal, K. F., and Bonfield, J. K. (2000). The Staden package, 1998. *Methods in Molecular Biology*, 132:115–130.

Stappenbeck, T. S., Hooper, L. V., and Gordon, J. I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15451–15455.

Stephen, A. M. and Cummings, J. H. (1980). The microbial contribution to human faecal mass. *Journal of Medical Microbiology*, 13(1):45–56.

Stevens, C. and Hume, I. (2004). *Comparative physiology of the vertebrate digestive system*. Cambridge University Press.

Stevens, C. E. and Hume, I. D. (1998). Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological Reviews*, 78(2):393–427.

Stewart, J. A., Chadwick, V. S., and Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *Journal of Medical Microbiology*, 54(Pt 12):1239–1242.

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *British Medical Journal*, 299(6710):1259–1260.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16.

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D., and Doré, J. (1999). Direct analysis of genes encoding 16S rRNA from

complex communities reveals many novel molecular species within the human gut. *Applied and Environmental Microbiology*, 65(11):4799–4807.

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology*, 28(2):237–243.

Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P., and Lochs, H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *Journal of Clinical Microbiology*, 43(7):3380–3389.

Szentkuti, L., Riedesel, H., Enss, M. L., Gaertner, K., and Von Engelhardt, W. (1990). Pre-epithelial mucus layer in the colon of conventional and germ-free rats. *The Histochemical Journal*, 22(9):491–497.

Talley, N. J., Phillips, S. F., Haddad, A., Miller, L. J., Twomey, C., Zinsmeister, A. R., MacCarty, R. L., and Ciociola, A. (1990). GR 38032F (Ondansetron), a selective 5HT receptor antagonist, slows colonic transit in healthy man. *Digestive Diseases and Sciences*, 35(4):477–480.

Talley, N. J. and Spiller, R. (2002). Irritable bowel syndrome: a little understood organic bowel disease? *Lancet*, 360(9332):555–564.

Tana, C., Umesaki, Y., Imaoka, A., Handa, T., Kanazawa, M., and Fukudo, S. (2010). Altered profiles of intestinal microbiota and organic acids may be the origin of symptoms in irritable bowel syndrome. *Neurogastroenterology and Motility*, 22(5):512–519.

Tang, L. Y. L., Nabalamba, A., Graff, L. A., and Bernstein, C. N. (2008). A comparison of self-perceived health status in inflammatory bowel disease and irritable bowel syndrome patients from a Canadian national population survey. *Canadian Journal of Gastroenterology*, 22(5):475–483.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-p., Ugarte, E., Muñoz tamayo, R., Paslier, D. L. E., Nalin, R., Dore, J., and Leclerc, M. (2009). Towards the human intestinal microbiota phylogenetic core. *Environmental Microbiology*, 11(10):2574–2584.

Thompson, W. G. (2006). The road to Rome. *Gastroenterology*, 130(5):1552–1556.

Tillisch, K., Mayer, E. a., Labus, J. S., Stains, J., Chang, L., and Naliboff, B. D. (2005). Sex specific alterations in autonomic function among patients with irritable bowel syndrome. *Gut*, 54(10):1396–1401.

Tlaskalová-Hogenová, H., Stěpánková, R., Kozáková, H., Hudcovic, T., Vannucci, L., Tucková, L., Rossmann, P., Hrncír, T., Kverka, M., Zákostelská, Z., Klimesová, K., Pribylová, J., Bártová, J., Sanchez, D., Fundová, P., Borovská, D., Srutková, D., Zídek, Z., Schwarzer, M., Drastich, P., and Funda, D. P. (2011). The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cellular and Molecular Immunology*, 8(2):110–120.

Torrallardona, D., Harris, C. I., and Fuller, M. F. (2003). Pigs' gastrointestinal microflora provide them with essential amino acids. *The Journal of Nutrition*, 133(4):1127–1131.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. a., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., and Gordon, J. I. (2009a). A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. a., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122):1027–1031.

Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunenko, T., Niazi, F., Affourtit, J., Egholm, M., Henrissat, B., Knight, R., and Gordon, J. I. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7503–7508.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009b). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, 1(6):6ra14.

Ubeda, C., Lipuma, L., Gobourne, A., Viale, A., Leiner, I., Equinda, M., Khanin, R., and Pamer, E. G. (2012). Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *The Journal of Experimental Medicine*, 209(8):1445–1456.

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., and Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE*, 3(6):e2527.

Vaishampayan, P. A., Kuehl, J. V., Froula, J. L., Morgan, J. L., Ochman, H., and Francino, M. P. (2010). Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biology and Evolution*, 2:53–66.

Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L., and Hooper, L. V. (2008). Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52):20858–20863.

Van den Abbeele, P., Van de Wiele, T., Verstraete, W., and Possemiers, S. (2011). The host selects mucosal and luminal associations of co-evolved gut microbes: a novel concept. *FEMS Microbiology Reviews*, 35(4):681–704.

Vanhoutte, T., Huys, G., Brandt, E., and Swings, J. (2004). Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiology Ecology*, 48(3):437–446.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-tillson, H., Pfannkoch, C., Rogers, Y.-h., and Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667):66–74.

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L., and Jansson, J. K. (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal*, 3(2):179–189.

Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A., Cullender, T. C., Mwangi, S., Srinivasan, S., Sitaraman, S. V., Knight, R., Ley, R. E., and Gewirtz, A. T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*, 328(5975):228–231.

Vinolo, M. A., Rodrigues, H. G., Nachbar, R. T., and Curi, R. (2011). Regulation of inflammation by short chain fatty acids. *Nutrients*, 3(10):858–876.

von Hertzen, L., Hanski, I., and Haahtela, T. (2011). Natural immunity. Biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO Reports*, 12(11):1089–1093.

von Wintzingerode, F., Göbel, U. B., and Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviewsicrobiology reviews*, 21(3):213–229.

Wagner, R. (1994). The regulation of ribosomal RNA synthesis and bacterial cell growth. *Archives of Microbiology*, 161(2):100–109.

Walker, A. W., Duncan, S. H., Harmsen, H. J. M., Holtrop, G., Welling, G. W., and Flint, H. J. (2008). The species composition of the human intestinal microbiota differs between particle-associated and liquid phase communities. *Environmental Microbiology*, 10(12):3275–3283.

Walker, A. W., Duncan, S. H., Mcwilliam Leitch, E. C., Child, M. W., Flint, H. J., and Leitch, E. C. M. (2005). pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Applied and Environmental Microbiology*, 71(7):3692–3700.

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., and Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, 5(2):220–230.

Wallace, T. C., Guarner, F., Madsen, K., Cabana, M. D., Gibson, G., Hentges, E., and Sanders, M. E. (2011). Human gut microbiota and its relationship to health and disease. *Nutrition Reviews*, 69(7):392–403.

Wang, M., Ahrné, S., Jeppsson, B., and Molin, G. (2005). Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiology Ecology*, 54(2):219–231.

Wehkamp, J., Harder, J., Weichenthal, M., Mueller, O., Herrlinger, K. R., Fellermann, K., Schroeder, J. M., and Stange, E. F. (2003). Inducible and

constitutive beta-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflammatory Bowel Diseases*, 9(4):215–223.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schäffeler, E., Schlee, M., Herrlinger, K. R., Stallmach, A., Noack, F., Fritz, P., Schröder, J. M., Bevins, C. L., Fellermann, K., and Stange, E. F. (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut*, 53(11):1658–1664.

Weichselbaum, E. (2010). Potential benefits of probiotics-main findings of an in-depth review. *British Journal of Community Nursing*, 15(3):110–114.

Wendisch, V. F., Zimmer, D. P., Khodursky, A., Peter, B., Cozzarelli, N., and Kustu, S. (2001). Isolation of Escherichia coli mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. *Analytical Biochemistry*, 290:205–213.

Wheater, D. M., Hirsch, A., and Mattick, A. T. (1952). Possible identity of lactobacillin with hydrogen peroxide produced by lactobacilli. *Nature*, 170(4328):623–624.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12):6578–6583.

Wilson, K. H. and Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Applied and Environmental Microbiology*, 62(7):2273–2278.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, 51(2):221–271.

Woese, C. R., Kandlert, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12):4576–4579.

Woodmansey, E. J. (2007). Intestinal bacteria and ageing. *Journal of Applied Microbiology*, 102(5):1178–1186.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R.,

Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D., and Lewis, J. D. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–108.

Xavier, R. J. and Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, 448(7152):427–434.

Xiong, Y., Miyamoto, N., Shibata, K., Valasek, M. a., Motoike, T., Kedzierski, R. M., and Yanagisawa, M. (2004). Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):1045–1050.

Yoshioka, H., Iseki, K.-i., and Fujita, K. (1983). Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants. *Pediatrics*, 72(3):317–321.

Younes, H., Coudray, C., Bellanger, J., Demigné, C., Rayssiguier, Y., and Rémésy, C. (2001). Effects of two fermentable carbohydrates (inulin and resistant starch) and their combination on calcium and magnesium balance in rats. *British Journal of Nutrition*, 86(04):479–485.

Young, V. B. and Schmidt, T. M. (2004). Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *Journal of Clinical Microbiology*, 42(3):1203–1206.

Zengler, K., Toledo, G., Rappé, M., Elkins, J., Mathur, E. J., Short, J. M., and Keller, M. (2002). Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15681–15686.

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell, M. D., Wing, R., Rittmann, B. E., and Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7):2365–2370.

Zoetendal, E. G., Akkermans, A. D. L., and de Vos, W. M. (1998). Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Applied and Environmental Microbiology*, 64(10):3854–3859.

Zoetendal, E. G., Cheng, B., Koike, S., and Mackie, R. I. (2004). Molecular microbial ecology of the gastrointestinal tract: from phylogeny to function. *Current Issues in Intestinal Microbiology*, 5(2):31–48.

Zoetendal, E. G., Plugge, C. M., Akkermans, A. D. L., and de Vos, W. M. (2003). Victivallis vadensis gen. nov., sp. nov., a sugar-fermenting anaerobe from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, 53(1):211–215.

Zoetendal, E. G., Rajilic-Stojanovic, M., and de Vos, W. M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*, 57(11):1605–1615.

Zoetendal, E. G., Vaughan, E. E., and de Vos, W. M. (2006). A microbial world within us. *Molecular Microbiology*, 59(6):1639–1650.

Zoetendal, E. G., Wright, A. V., Vilpponen-salmela, T., Ben-Amor, K., Akkermans, A. D. L., and de Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Applied and Environmental Microbiology*, 68(7):3401–3407.

C Glossary and list of abbreviations

16S rRNA. 16S ribosomal RNA. RNA component of the small subunit of prokaryotic ribosomes. See **rRNA**.

ANOSIM. Analysis of similarities.

Antibiotic. Any substance produced by a microorganism that can destroy or inhibit the growth of other microorganisms. Synthetic antibiotics, usually chemically related to natural antibiotics, have been produced to accomplish comparable tasks.

CA. Correspondence analysis.

CD. Crohn's disease. See **IBD**.

cDNA. Complementary DNA. DNA synthesised from an RNA template in a reaction catalysed by the enzymes reverse-transcriptase and DNA polymerase.

CHO. Carbohydrate.

Colonisation. The spreading of one or more species into a new habitat.

Colonisation resistance. When applied to host-associated microorganisms, the mechanisms whereby the resident microbiota protects against the colonisation by new and potentially harmful microorganisms.

Commensalism. Symbiotic relationship between two organisms where one benefits without affecting the other.

DCA. Detrended correspondence analysis.

DGGE. Denaturing gradient gel electrophoresis. See **Fingerprinting**.

Dysbiosis. A state of imbalance among the microorganisms naturally occurring in a body site, often resulting in health problems.

Ecological succession. The process by which an ecological community undergoes more or less orderly and predictable changes following disturbance or initial colonisation of a new habitat.

Fermentation. Anaerobic energy-yielding process in which ATP is formed by substrate-level phosphorylation through association with redox transformations between organic compounds. Primary fermentation consumes monomers

released by the hydrolysis of polysaccharides and proteins; secondary fermentation further transforms the products of primary fermentation.

Fingerprinting. When applied to a microbial community, set of techniques in molecular biology that are used to quickly profile the diversity of the community by showing how many variants of a gene are present. They include denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), and terminal restriction fragment length polymorphism (T-RFLP).

Functional redundancy. The ability of one microbial taxon, consortium, or community to carry out a process at a similar rate as another under the same environmental conditions, regardless of differences in taxonomic composition.

GF animal. Germ-free animal. An animal lacking a microbiota, born and reared under sterile conditions for research purposes.

GI. Gastrointestinal (GI tract, GI microbiota, etc.)

Gnotobiotic animal. Specially reared laboratory animal in which only certain known microorganisms are present. The term also includes germ-free animals, since the status of their microbial community is known.

Host. When a distinction is made between two organisms in a symbiotic relationship, the host is the larger of the two and may or may not derive a benefit, while the smaller organism is the symbiont and is always a beneficiary in the relationship.

IBD. Inflammatory bowel disease. Group of chronic conditions that cause inflammation of the intestines. The most common forms of IBD are ulcerative colitis (UC) and Crohn's disease (CD). UC is restricted to the colon and the rectum, while CD can affect any segment of the gastrointestinal tract, although it most commonly affects the terminal ileum. The inner lining of the intestine (mucosa) becomes inflamed in UC, while CD causes inflammation that extends much deeper into the layers of the bowel wall and generally involves the entire bowel wall.

IBS. Irritable bowel syndrome. Functional disorder of the intestine that is characterised by chronic abdominal pain, discomfort, and alteration of bowel habits in the absence of any detectable organic cause. Diarrhoea or constipation may predominate, or they may alternate (classified as IBS-D, IBS-C, or IBS-A, respectively). Post-infectious IBS (PI-IBS) is a special subtype of IBS in which chronic symptoms appear following an episode of acute gastroenteritis.

Metagenome. The genetic material in a given environment, consisting of the genomes of different organisms.

Metagenomics. Study of the genetic material directly isolated from an environmental sample (i.e. the metagenome) to assess the genetic potential of the whole community, bypassing the need to isolate and culture individual species. The field may also be referred to as environmental genomics, community genomics, or ecogenomics.

Metatranscriptome. The set of RNAs produced in a given environment, consisting of the transcriptomes of different organisms.

Metatranscriptomics. Study of the RNA isolated from an environmental sample (i.e. the metatranscriptome) to assess the gene expression of the whole community.

Microbiota. The microbial community that inhabits an ecosystem, such as some part of the body of an animal host (commonly referred to according to the habitat that it occupies, e.g. the gastrointestinal microbiota).

mRNA. Messenger RNA. RNA that carries coding information for protein synthesis.

Mutualism. Symbiotic relationship between two organisms that is mutually beneficial. Also referred to as symbiosis by some authors; however, symbiosis is a broad category, defined to include any type of persistent biological interaction.

OR. Odds ratio.

OTU. Operational taxonomic unit. Any of the groups (taxa) being examined. An OTU can refer to any level of the taxonomic hierarchy, including individuals of a species, or different species, or different genera, and so on.

Pathogen. Organism that causes acute or chronic disease following infection of the host. Obligate pathogens must cause disease in order to be transmitted from one host to another. Opportunistic pathogens cause disease when the host is immunocompromised or when normal microbial antagonism is affected; otherwise, they can be transmitted from one host to another without having to cause disease. Opportunistic pathogens are distinct from pathobionts, which are symbionts that do not normally cause disease, but are able to promote pathology when specific conditions are altered in the host.

PCA. Principal components analysis.

PCoA. Principal coordinates analysis.

PCR. Polymerase chain reaction. Technique in molecular biology to amplify a single or a few copies of a DNA fragment across several orders of magnitude. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the

DNA.

Phylogenetics. Branch of biology that reconstructs the evolutionary relationships among organisms. In molecular phylogenetics, the relationships among organisms or genes are studied by analysing hereditary differences in DNA or protein sequences.

Phylotype. Biological type that classifies an organism by its evolutionary relationship to other organisms. The term is most often used in microbiology, where groups of organisms are defined by its sequence similarity, typically using the 16S ribosomal RNA genes. A sequence identity of 97–99% indicates approximately a species-level taxon.

Prebiotic. Non-digestible food ingredient that stimulates specific changes in the composition and/or activity of the gastrointestinal microbiota in ways claimed to be beneficial to host health (e.g. inulin, fructo-oligosaccharides, galacto-oligosaccharides).

Probiotic. Live microorganism that, when administered in adequate amounts, confers a health benefit to the host. Lactic acid bacteria and bifidobacteria are the most common types of microbes used as probiotics. Probiotics are commonly consumed as part of fermented food.

Resilience. When applied to a microbial community, the ability to return to its original composition after being disturbed.

rRNA. Ribosomal RNA. RNA component of the ribosome, the cell structure that is the site of protein synthesis. Ribosomes can be broken down into two subunits, the small subunit (SSU) and the large subunit (LSU). In prokaryotes, the SSU contains the 16S rRNA, and the LSU contains the 5S and 23S rRNAs. The genes coding for the 16S rRNA are the most extensively used in reconstructing phylogenies.

SCFA. Short-chain fatty acid. Fermentation end-product, formed from polysaccharide, oligosaccharide, protein, peptide, and glycoprotein precursors by anaerobic bacteria. In quantitative terms, dietary polysaccharides resistant to hydrolysis by the host enzymes are the most important SCFA precursors in the colon, and the main SCFAs are acetate, propionate, and butyrate.

Symbiont. An organism in a symbiotic relationship. When a distinction is made between two interacting organisms, the symbiont is the smaller of the two and is always a beneficiary in the relationship, while the larger organism is the host and may or may not derive a benefit.

Symbiosis. An intimate and often long-term relationship between different species, which was originally defined as “living together”. Although it is often

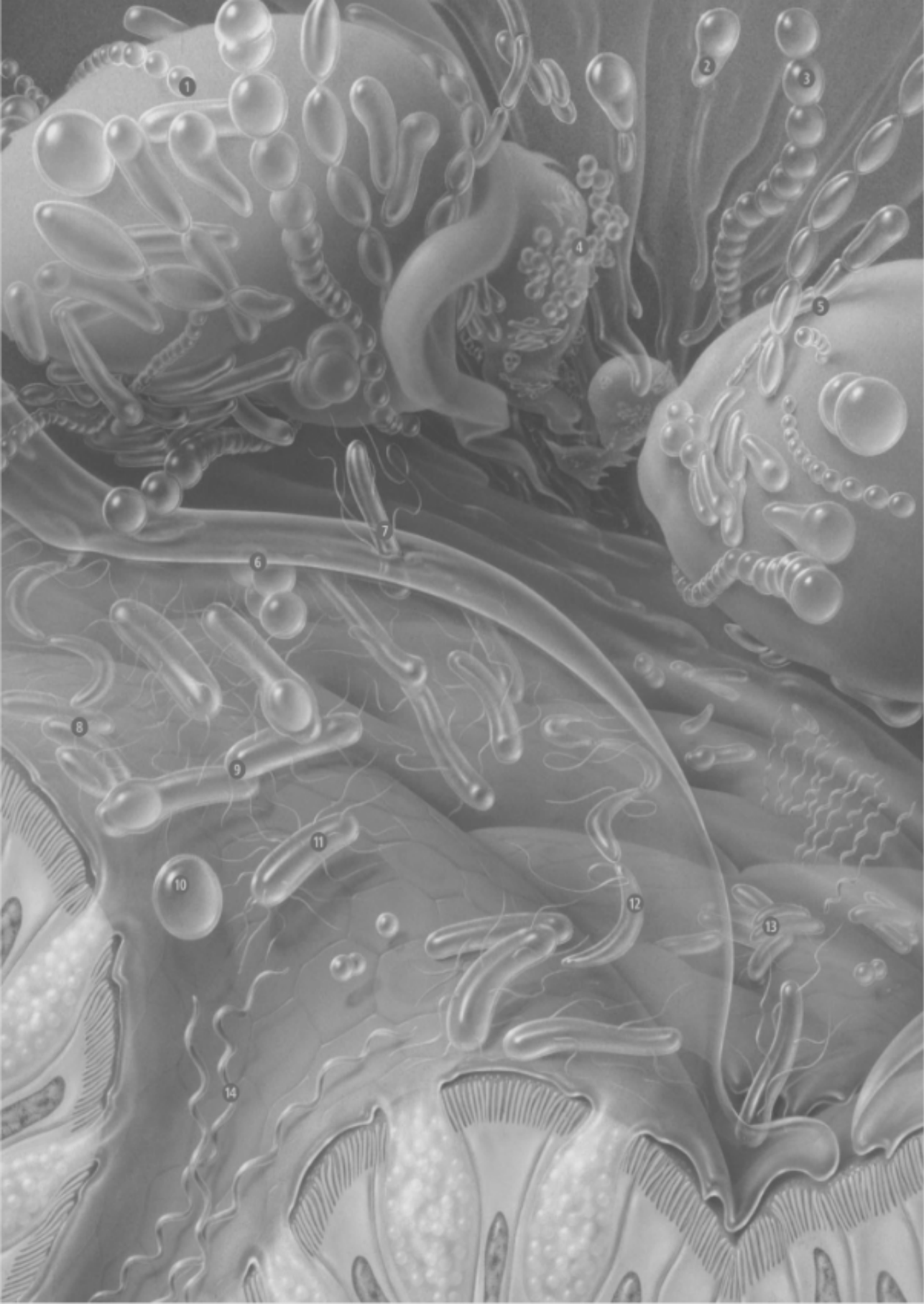
referred to beneficial relationships (mutualisms), symbiosis does not necessarily imply that either partner gains an advantage. Symbiotic relationships include those associations in which one organism lives on or inside the other. They may be obligate, i.e. necessary for the survival of the organisms, or facultative, when the relationship is not essential for the survival of the organisms.

TGGE. Temperature gradient gel electrophoresis. See **fingerprinting**.

T-RFLP. Terminal restriction fragment length polymorphism. See **fingerprinting**.

UC. Ulcerative colitis. See **IBD**.

Xenobiotic. Chemical compound that is found in an organism but that is not normally produced or expected to be present in it. It can be toxic, even at low concentrations.


1
2
3
4
5
6
7
8
9
10
11
12
13
14