Ecology meets human health: studies on human gut microbiota in health and disease

Pinto, S.

Ecology meets human health

Studies on human gut microbiota in health and disease

Susanne Pinto

Voor oma

Ecology meets human health
Studies on human gut microbiota in health and disease

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 20 november 2025
klokke 14:30 uur

door

Susanne Pinto
geboren te Dordrecht
in 1993

Promotor
prof.dr. E. W. Steyerberg (LUMC)

Copromotoren
dr. J. A. Bogaards (Amsterdam UMC)
dr. E. Benincà (RIVM)

Leden promotiecommissie
prof.dr. H. Putter (LUMC)
dr. W. K. Smits (LUMC)
dr. K. Faust (KU Leuven)
prof.dr. J. H. H. M. van de Wijgert (UMC Utrecht)

Aanvullende leden van de oppositiecommissie
prof.dr. E. Zaura (ACTA)
prof.dr. R. M. H. Merks (Universiteit Leiden)

Het werk in dit proefschrift werd gefinancierd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) via het programma *Complexiteit in Gezondheid en Voeding* (NWO-subsidie 645.001.002), met cofinanciering door het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) in Nederland.

Table of contents

Voorwoord

De *studie van het leven*, biologie, fascineerde mij zodanig dat ik erdoor gedreven raakte. Hierbij speelde de enthousiaste begeleiding van docenten aan de Universiteit Leiden een belangrijke rol. De mogelijkheid om tijdens mijn studie andere vakgebieden te verkennen, heeft mijn enthousiasme verder aangewakkerd en mij verder gevormd tot de wetenschapper die ik nu ben.

Als wetenschapper ben ik het meest geboeid door datgene waar ik het minst van weet. Tijdens mijn studie biologie leidde dit mij naar de wereld van geleedpotigen en die van cleanrooms, en tijdens mijn promotieonderzoek bestudeerde ik de ecologie van het humane microbioom. Dit traject gaf mij de kans om veel inspirerende mensen te ontmoeten en degenen die ik al kende nog meer te gaan waarderen. Dit proefschrift zou niet bestaan zonder de gezamenlijke inspanning van deze diverse groep geweldige mensen.

Allereerst wil ik Hans Bogaards en Elisa Benincà bedanken. Hans was mijn mentor en begeleidde mij in mijn ontwikkeling als onderzoeker. Hij stelde cruciale vragen die mij verder hielpen, zowel over het onderwerp en mijn werk als over mijn persoonlijke groei. Elisa was naast een fijne supervisor ook een goede vriendin tijdens mijn promotie, iemand met wie ik in vertrouwen kon praten. Van haar heb ik geleerd kritisch te zijn over mijn werk, het draait niet om kwantiteit maar om kwaliteit. Onze reis naar Kaapstad, Zuid-Afrika, behoort tot een van de hoogtepunten van mijn leven, dus ook daarvoor wil ik haar bedanken.

Mijn promotor Ewout Steyerberg wil ik bedanken omdat hij mij, eerst als leidinggevende en later als promotor, het vertrouwen gaf dat ik van waarde ben. Ook maakte hij mijn promotietraject eenvoudiger door snel te reageren en zaken effectief op te lossen. Hierbij werd hij ondersteund door de secretaresses Lies, Leonie en Aubry, die ik wil noemen vanwege hun vriendelijkheid en hulpvaardigheid.

Ik wil in het bijzonder enkele collega's bedanken voor hun waardevolle bijdragen aan dit proefschrift. Allereerst Marten Scheffer en Egbert van Nes, die voor mij een grote bron van inspiratie waren. Zij lieten mij zien hoe plezierig wetenschap kan zijn en hoe waardevol het is wanneer onderzoekers elkaar aanvullen, want wetenschap doe je nooit alleen. Daisy Jonkers, John Penders, Liz Terveer, Josbert Keller en Andrea van der Meulen-de Jong wil ik bedanken voor het delen van hun kennis en data. Tijdens mijn promotie besefte ik hoe kostbaar data is en hoeveel werk eraan voorafgaat. Dominika Šajbenová en Sam Nooij dank ik voor onze samenwerking, waarin ik ondervond hoe leerzaam het is om samen te groeien terwijl we de uitdagingen van ons onderzoek aangingen. Bernard Cazelles, Susana Fuentes, Sudarshan Shetty en Gianluca Galazzo ben ik dankbaar voor de inzichten, feedback en methodologische kennis die zij met mij hebben gedeeld.

Door de beperking in het aantal woorden kan ik helaas niet al mijn andere betrokken collega's bij naam noemen. Toch wil ik mijn waardering uiten voor iedereen die ik ontmoette aan de Universiteit Leiden, het LUMC, het RIVM, Wageningen University & Research, Maastricht UMC+, Amsterdam UMC, UMC Utrecht en via cursussen en congressen. Ik hoop dat eenieder die betrokken is bij mijn academische pad zich aangesproken voelt. Ik wil iedereen nadrukkelijk bedanken voor het verrijken van mijn kennis en vaardigheden en voor het hartelijke welkom in de fascinerende wereld van wetenschap en microbioom.

Graag wil ik de onderzoeksgroepen CMAT (LUMC) en Modellering (RIVM) bedanken voor de mogelijkheid om aan de leerzame overleggen deel te nemen en voor de prettige samenwerking.

Daarnaast wil ik mijn medepromovendi bedanken voor de inspiratie die ik vond in hun proefschriften en voor hun warme gezelschap. Dat gaf mij het gevoel er niet alleen voor te staan.

Voordat mijn PhD-onderzoeken en daarmee dit proefschrift waren voltooid, vond ik een nieuwe baan in het UMC Utrecht. In de groep van Maaike van Mourik kreeg ik de kans om aan een interessant en belangrijk project te werken. Het combineren van mijn nieuwe baan met het afronden van het proefschrift was een uitdaging. Ik ben Maaike dankbaar voor haar begrip in deze periode.

Dankzij de liefdevolle aanmoedigingen van mijn oma, en de steun van mijn moeder, mijn zusjes Eveline, Marianne, Rosaline en Caroline, en ook van Marianne Vysma, die met zorg de hoofdstukken nalazen en mij hielpen fouten te verbeteren en de samenvatting te verhelderen, kon ik dit proefschrift tot een goed einde brengen. En natuurlijk dankzij mijn vader, die de opmaak verzorgde en mij daarmee niet alleen veel tijd bespaarde, maar ook een prachtig proefschrift schonk. Aan mijn familie: hartelijk dank voor alle steun, betrokkenheid en hulp.

Mijn vrienden, onder wie paranimfen Dominique en Danielle, bleven altijd betrokken en boden hun hulp aan. Bovendien gaven de gezellige feestjes en etentjes mij nieuwe energie. Daardoor kon ik weer lange uren achter mijn bureau doorbrengen, samen met mijn kat Sam als aangenaam gezelschap.

Jasper, jouw vertrouwen en onvoorwaardelijke steun waren mijn houvast en jouw liefde betekent alles voor mij. Dank je dat je er altijd voor me bent.

Chapter 1

General introduction and thesis outline

General introduction and thesis outline

Complex systems

Complex systems encompass a diverse array of phenomena and processes, from financial markets and climate patterns to the microbial communities in our gut. Their challenges involve apparently intractable and often unpredictable problems, such as organizational transformation, political conflict, climate change, disruptions in infrastructure, and recurring infections.[1, 2] Complex systems evolve over time, and changes can manifest as gradual trends or fast fluctuations.[3, 4] Occasionally, the system might undergo a complete transformation into a new state. For example, a pathogenic species in the Caribbean coral reef caused a mass mortality event in the sea urchin *Diadema antillarum*. This loss had dramatic consequences: without the grazing activity of the urchins, the reef was quickly overgrown by brown fleshy algae, fundamentally altering the entire structure of the community.[2] Similarly, in the Sahel-Sahara region, a gradual change in solar irradiation triggered an abrupt shift, transforming the landscape with dense vegetation into a desert environment.[2] Moreover, interactions among species can lead to oscillations and even sometimes chaotic dynamics, by themselves[5] or in response to environmental conditions.[6] Consequently, in such systems, slight differences in initial conditions can lead to different outcomes with extinctions of varying magnitudes due to non-linear dynamics (Box 1.1). In contrast, systems may display resilience by recovering from disturbances and reverting to their previous state.[7]

Box 1.1 - Tipping points in ecosystems. In the context of ecology, ecosystems experience shifts when confronted with alterations in factors such as food sources, climate fluctuations, or human interventions. When an ecosystem encounters an environmental change, there may be a noticeable shift in species composition and overall biodiversity. Similar to a game of Jenga, where removing individual blocks may not immediately affect the stability of the tower, small changes in a system might not have noticeable consequences until a critical tipping point is reached. However, once that tipping point is crossed, the system can experience a sudden and significant transformation, resembling the collapse of a Jenga tower when a crucial block is removed. This phenomenon is closely tied to the system's high connectivity, where the failure of one element can impact the entire system, often leading to irreversible changes.

The individual components of a complex system often represent relatively simple processes. However, synchronization of activities among individual components can lead them to act as a cohesive unit with additional functionalities (Figure 1.1). A greater diversity of these components can display richer properties, functions, or behaviours, and enhanced resilience.[1, 3, 4] The theory of complex systems seeks to infer the underlying models and properties of their patterns and behaviours, as well as to develop tools and concepts for effectively modeling their interactions and dynamics. Because if we can understand the behaviour of complex systems, we can develop solutions to address their challenges, aiming for a resilient and adaptive future for our society and health. Achieving this requires interdisciplinary collaborations, where experts from diverse fields offer their perspectives.

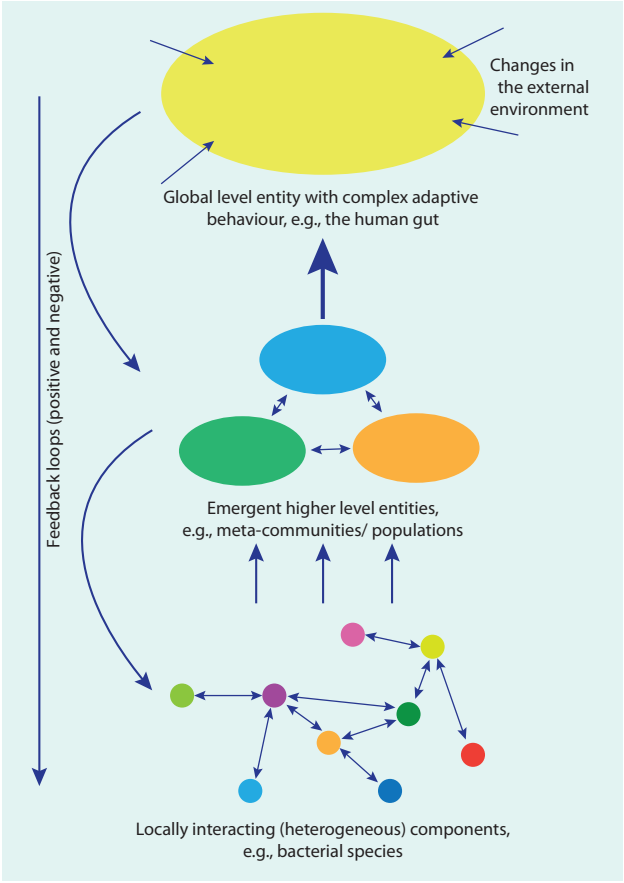


Figure 1.1 - Diagram illustrating the interactions and relationships within a complex system across different scales. In the human gut microbiota, emergence refers to the phenomenon where the overall functionality of the microbial community arises from the interactions among individual microbial species. The gut is home to trillions of microbial cells, including bacteria, viruses, fungi, and other microbes that interact with each other (bottom figure). The interactions between these diverse agents, that mutually affect each other, lead to the emergence of various functionalities and behaviours (middle figure) that contribute to digestion, nutrient absorption, and overall host health (top figure).[8] The science of complexity shows that insights at one level (e.g., health outcomes) are influenced at another level (e.g., the interactions within a system), and that one cannot be fully understood without knowledge of the other, as they are interconnected in a continuous feedback loop. Therefore, complex systems such as the human gut microbiota are difficult to comprehend using traditional scientific analyses. Traditionally, experimental scientists have primarily focused on reducing complex systems to their individual elements, providing essential knowledge of the system's components, but overlooking the significance of interactions between them. Interestingly, the global system is often not fully explained by or predictable with the knowledge of the component parts. However, the inverse is also true; without an understanding of the dynamics of the component parts, understanding of the whole system is unattainable. By considering the dynamics of a complex system as a whole, with both the internal and external forces, rather than merely the sum of its parts, new insights and theories can be developed.[1, 9] This figure is based on Lewin (1999) and Parrott and Lange (2013).[1, 10]

The research for this thesis was conducted by a multidisciplinary team of ecologists, microbiologists, bioinformaticians, statisticians, epidemiologists, and medical specialists who collaborated to explore new perspectives on the complex ecosystem of the gut microbiota and its relationship with human health and disease.

The human microbiome

The human body serves as an ecosystem for a multitude of microorganisms, with the gastro-intestinal (GI) tract being a particularly rich and diverse habitat.[11-13] In 2022, it was estimated that there are about ten times more bacterial genomes in the human gut than there are genes in our own genome.[14, 15] Actually, the body is not a single ecosystem; instead, it comprises multiple habitats, each with its own unique environment, which are likely interconnected with one another. The entire collection of microorganisms (commensals, mutualists, pathogens, and opportunists), encompassing bacteria, viruses, protozoa, archaea, and fungi, along with their cumulative genetic content, is collectively referred to as the microbiome, a concept introduced by Nobel Prize laureate Joshua Lederberg in 2001.[16, 17] A distinct term, the metagenome, encapsulates the combined genetic makeup of the microbes. The microbiota, in a narrower sense, refers to the assorted microbial species occupying specific niches, such as the 'oral microbiota' or the 'gut microbiota'.[17, 18] This thesis focuses on the bacteria in the human gut microbiota.

Our understanding of the composition and functions of the microbiome has increased exponentially over the last 15 years. This has been mainly due to the new 'omics' technologies that have facilitated large-scale analyses of the phylogenetic and metabolic profiles of microbial communities.[19-23] These insights have revealed the vital role that microbial communities play in human health, as they coexist symbiotically with the human host and contribute significantly to maintaining physiological balance. The human gut, for example, serves as a unique ecosystem, providing a nutrient-rich environment for its microbial communities. Many benefits of the human microbiome for the human host have already been identified, including the prevention of pathogenic bacteria and viruses through competition for metabolic resources, maintenance of metabolic balance, processing of nutrients (such as fiber digestion and vitamin synthesis), drug modification (affecting drug efficacy), and the maturation and regulation of gastrointestinal immune responses.[20, 24-30] Moreover, the relation between microbes and various human health conditions has been shown for, among others: obesity, cardiovascular disease, *Clostridioides difficile* colitis, inflammatory bowel diseases (IBD), irritable bowel syndrome (IBS), non-alcoholic fatty liver disease, dental caries, asthma, autoimmune diseases (such as celiac disease, inflammatory arthritis, and primary sclerosing cholangitis), and sepsis.[31-44]

Every person harbors distinct and relatively stable microbial communities in and on their body.[45] Stability means that samples collected over time from an individual exhibit greater similarity to each other compared to samples obtained from other individuals.[22, 45-51] Certain host factors, e.g., host genetics, age, diet, and medication use, cumulatively explain about 20% of the gut microbiota compositional variation.[52-54] Despite the individual variability, a shared core microbiota with notably similar functional gene profiles can be detected in most healthy adults.[21, 22] Stability appears to be an important ecosystem trait, persisting over several months or even years.[22, 45, 46, 48-51]

However, natural fluctuations in community composition, featuring sporadic blooming of species, are normal in gut microbiota dynamics (Figure 1.2), reinforcing homeostatic interactions with the host.[52, 55, 56] Environmental stimuli influence these fluctuations and the microbiome typically shows autoregressive dynamics, allowing it to recover after disturbances.[46, 47, 57-59] An example of such stimuli is variation in nutrient availability, especially in the small intestine, as the colonic microbiota thrives on the breakdown of complex carbohydrates.[60, 61] Additionally, significant factors such as antibiotic administration, travelling, or drastic dietary changes can prompt bacterial population levels to shift within one day.[45, 46, 59, 62, 63] The extent to which the human gut microbiota subsequently absorbs disturbances, adapts to the changing conditions, and maintains its essential functions, characteristics, and structure depends on the resilience of the system.[22] Interestingly, substantial commonalities are found among seemingly divergent responses to disturbances.[64-69]

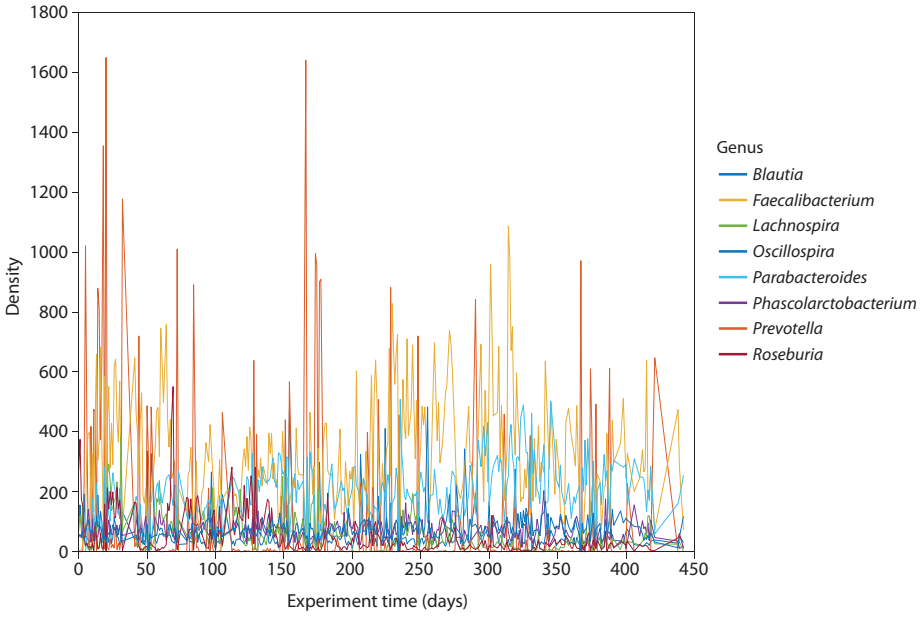


Figure 1.2 - Time series of the gut microbiota of one healthy male individual.[46] The time series shows temporal fluctuations on shorter timescales and overall stability over extended periods.[23]

Ecology of the human gut microbiota

A child is born with almost no microbiome. Colonization by maternal and environmental bacteria occurs within days of birth, influenced by factors such as delivery mode, antibiotic exposure, and ecological drivers (Box 1.2).[55, 64, 67, 68, 70, 71] Breastfeeding contributes directly to neonatal microbiota establishment through providing living bacteria (from the skin and milk of the mother) and indirectly through prebiotic nutrients and bioactive components.[67, 72-75] Human milk oligosaccharides (HMOs) promote the growth of beneficial species and strains of *Bifidobacterium* (a key early life microbe associated with improved development of the immune system) that produce enzymes to break down these complex sugars.[71]

Box 1.2 - Gut microbiota shaped by early colonizers and community dynamics. For the gut microbiota, it has been shown that the temporal development is not purely random; rather, it is partly deterministic (and to some extent predictable, i.e., succession), partially stochastic, and often contingent on the community's previous states.[81-87] This implies that the initial conditions, including environmental factors and the early colonizers (founder effect or pioneer species), have an impact on the later community dynamics as well as the time span needed to reach the adult state.[23, 68, 88-95] The microbiota are built upon these early colonizers, as they facilitate the growth of certain species, while impeding the growth of others.[86, 96] For example, the first colonizers entering the infant's gut are facultative aerobic bacteria such as Proteobacteria members. They alter the environment through metabolic byproducts, creating new ecological niches that promote diversification.[97] They pave the way by decreasing the oxygen concentration for subsequent colonization by anaerobic bacteria, such as Bacteroidota (formerly Bacteroidetes), Actinobacteria, and Bacillota (formerly Firmicutes) phyla.[23, 70] Critical ecological drivers such as community interactions, immigration, niche filtering, stochasticity, environmental conditions (such as oxygen, moisture, and pH) and host characteristics (such as age, diet, and medication use) keep continuously shaping the patterns of microbial community dynamics.[21, 49, 57, 67, 69, 75, 98-101]

The introduction of solid foods at four to six months after birth further shapes microbial composition, with effects varying based on dietary habits across different geographical regions.[67, 75, 76] Also, the child's living environment, including pets and siblings, impacts microbial development.[64, 77-79] After colonization and the stabilization of the gut microbiota, individuals can maintain distinct core microbial communities for extended periods of time.[22, 45, 46, 48-51] These stable physiological states are sustained by negative feedback loops, preserving homeostasis even when the gut environment undergoes changes (Figure 1.3).[22, 23, 80] The ability to adapt while being robust against changing environments may seem contradictory, but most complex systems are clearly adaptive and robust at the same time.[3]

However, if a system cannot recover from a significant perturbation, it might shift to an alternative stable state with distinct characteristics (Figure 1.3B). When this happens in the gut microbiota, the new state might have severe health implications for the human host.[48, 102-106] Bistable abundance distributions, i.e., arising from species with population sizes going back and forth between high and low abundances with moderate abundances being underrepresented in sampling, can be indicative of alternative stable states.[48, 102-107] For example, the bimodal abundance patterns of *Prevotella melaninogenica*, *Bacteroides fragilis*, and two groups of uncultured Clostridiales were verified in independent sets of sampled individuals, who varied in dietary patterns, geographic regions, and DNA extraction methods. These bimodal patterns appeared unaffected by these factors; rather, they were associated with factors such as aging or weight loss.[102] The discovery of bistable bacteria led them to be labelled as 'tipping elements' and possibly keystone species, i.e., organisms that have a disproportionate effect on community structure and function relative to their abundance. This prompted questions about whether the significant shifts in microbiota composition and function are associated with changes in the abundances of specific taxa or with a broader dysbiosis across the community.[102, 108-110]

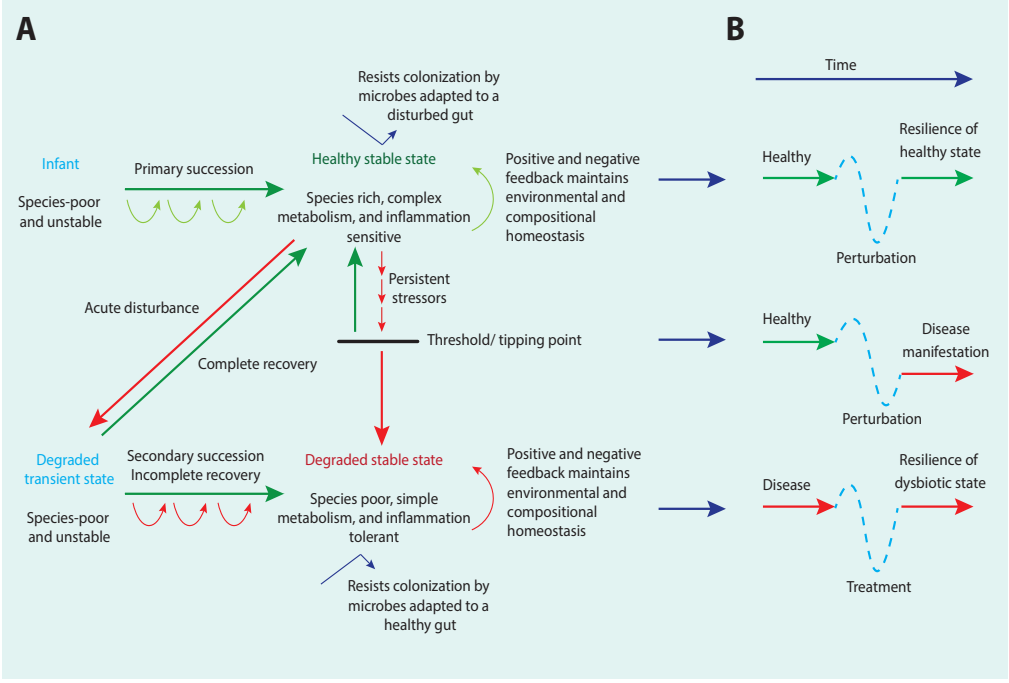


Figure 1.3 - Transitions in the composition of the human gut microbiota.
A) Positive and negative feedback loops probably have a role in driving succession and maintaining equilibrium states (resilience). B) Perturbations are often followed by an unstable state, which may return to a previous stable state or shift to an alternative stable state. This figure is adapted from Lozupone et al. (2012) and Sommer et al. (2017).[22, 23]

Consequently, detecting the keystone species has become a focus in microbiota research.[111-114] Mainly (specialist) primary degraders have the potential to manipulate and regulate community states as keystone species.[115] For example, despite their low abundance, highly active sulfate reducing bacteria, in wetland ecosystems, as well as in the human gut, play a crucial role in important biogeochemical processes.[116, 117] However, very few proposed hub taxa, suggested by statistical techniques, such as network analysis, have been experimentally confirmed as keystone species; therefore, the reliability of methods used to detect keystones remains uncertain.[118]

There are a multitude of (dynamic) species-species interactions within the gut microbiome, rooted in metabolic processes, such as cross-feeding (Figure 1.4).[119-121] Interactions among species in human microbiota tend to repel potential invaders and prevent outgrowth of certain species. For example, genetically diverse *Escherichia coli* populations produce secondary carbon sources sustaining other community members and preventing colonization of species that could outcompete them.[122] Moreover, antimicrobial production, space and nutrient competition, predation, and the trade-off between growth-maximizing organisms (*r*-strategists) and those adapted for resource competition (*K*-strategists) are mechanisms that reflect how organisms maximize nutrient uptake, often at the expense of other organisms.[123-125] The cumulative outcome involves the reconstruction of a network within the gut microbial ecosystem, facilitating the coexistence of a diverse bacterial community.

Interactions within ecological networks can engender diverse outcomes, encompassing positive impacts ('win'), negative impacts ('loss'), or no discernible impact on the participating species (Figure 1.4B). The interaction conferring benefits to both participants, such as two species that engage in the exchange of metabolic products (exhibit complementary auxotrophies), is called mutualism.[126, 136] This win-win relationship also occurs, for instance, when bacteria from disparate taxonomic groups collaboratively construct a biofilm, bestowing antibiotic survival upon its constituents and facilitating co-colonization.[137] The prominence of such interdependencies is underscored by their heightened relative abundance when both species are present.[8] Commensalistic relationships denote scenarios where one partner accrues benefits without inducing either harm or assistance to the other. Such relationships frequently manifest in biodegradation contexts, where commensals derive sustenance from compounds generated by fellow community members, as evidenced in cellulose degradation processes (Figure 1.4A).[138]

Conversely, antagonistic relationships may stem from amensalism, parasitism, and competition.[8] The inhibition of other species can occur through direct competition for resources (niche preemption) or by altering the habitat to reduce its suitability for other species (niche modification).[96] Bacteria use effectors of direct antagonism, including quorum sensing molecules, quenching molecules, antibiotics, and toxic substances such as bacteriocins and metal ion binding proteins, to inhibit the growth of competitors, especially in dense cellular environments.[139, 140] Classical loss-win dynamics, as materialized in parasitic relationships, are observed in the relation between bacteria and their bacteriophages.[23] Many bacterial species may exhibit predatory behaviour to some extent.[141] *Pseudomonas fluorescens*, for example, has been used as a biocontrol agent to control plant pathogens by antagonizing other microbes, including *Myxococcus xanthus*.[142-144] This species secretes various antibiotics and produces toxic volatile compounds such as cyanide.[143, 145] Because *Pseudomonas fluorescens* can then grow on nutrients derived from the cells it has killed, it can be categorized as a predator.[141]

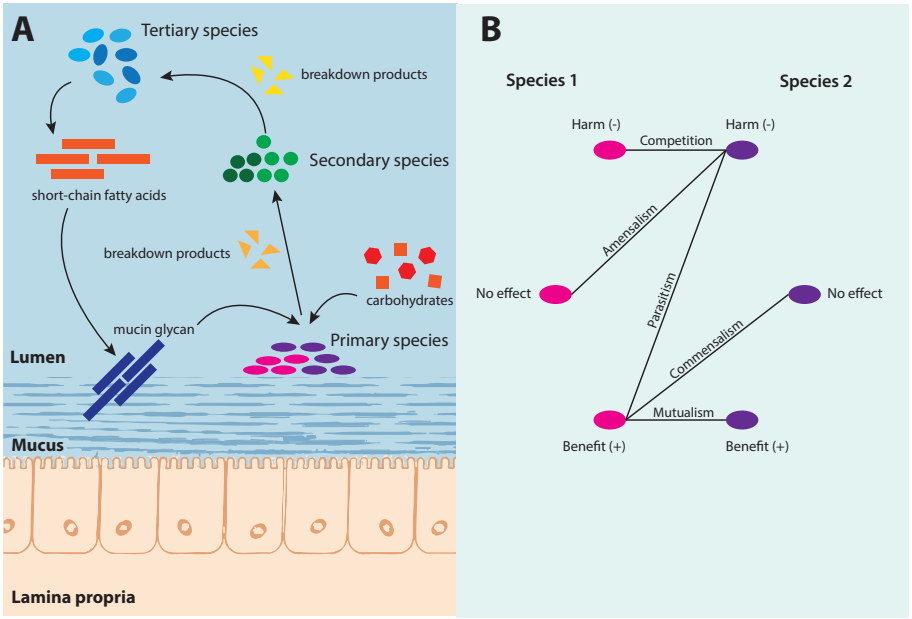


Figure 1.4 - A schematic representation of the gut, showing primary, secondary, and tertiary consumers and their potential interactions. A) At the start of the bacterial food chain are the bacteria that consume the primary nutrient sources, such as polysaccharides, oligosaccharides, proteins, sugars, and mucins secreted by the colonic epithelium.[126-130] Primary species effectively colonize the epithelial mucosa due to their ability to degrade mucin. They can also break down dietary plant- and animal-derived carbohydrates, initiating a series of cross-feeding interactions that support the growth of other bacteria, particularly those that rely on the breakdown of complex carbohydrates into simpler sugars for energy.[131-135] They facilitate the growth of secondary species and indirectly promote the growth of tertiary species. Some tertiary species produce short-chain fatty acids, which are subsequently utilized by colonocytes for their growth, leading to increased mucin production. This positive feedback loop may enhance ecological recovery in terms of diversity and biomass. It is important to note that there is likely no strict distinction between primary producers and secondary cross-feeders, as many microorganisms may function as both and will probably take the opportunity to cross-feed or degrade nutrients whenever possible, depending on the available substrates. This figure is adapted from Chng et al. (2020).[132] B) Cartoon illustrating the different interaction mechanisms. In competitive interactions, both species experience a negative effect. An example is when one or both species produce toxic compounds that are harmful to the other species as well as to themselves. Amensalism is a one-sided negative interaction. Amensalism occurs when a species causes harm to another species, without benefit or harm to itself. Parasitism occurs when one species benefits from another species at the expense of the other. Commensalism is a one-sided positive interaction. This type of interaction occurs when one species benefits from another without affecting it. In mutualistic interactions, both species experience a positive effect. An example is when one species feeds on the metabolites excreted by another species, thereby cleaning the 'waste' from the environment.

An illustration of microorganism competition (a loss-loss relationship) was given by Gause, already in the 1930s when he conducted a series of co-culturing experiments.[146] In his observations, he found that some species pairs, which thrived on their own, cannot coexist with constant population values. He showed one species (*Paramecium aurelia*) taking control over the other species (*Paramecium caudatum*) when they were grown together. Even if one organism ultimately 'wins' by securing more resources, the energy and resources spent in the competitive process could have been used for growth, reproduction, or other survival functions. Therefore, both species suffer initially during the competition, and eventually, the less competitive species is driven to extinction in that environment. This formed the foundation for Gause's law of competitive exclusion, asserting that species with similar ecological niches mutually preclude each other's survival.[146, 147] The deterministic nature of competitive dynamics of microbial communities, particularly within newly established ecosystems, has long been a topic of debate among ecologists. One theoretical framework that has emerged in this context is neutral niche theory. The neutral niche theory assumes that communities in certain niches are built only by random draws, driven by stochastic colonization, where the gut niches are likely to be filled by random 'winners', as in a lottery scenario, instead of predictable winners.[84, 87]

Amensalism, a situation where one partner is harmed without benefitting the other, can be seen in scenarios when a microbial species produces metabolic by-products that change the environment to the detriment of other microorganisms, such as the acidification caused by lactobacilli activity.[148, 149] Previous experimental investigations have substantiated that antagonistic interactions are more likely among closely related species sharing analogous metabolic pathways.[137, 150]

Gut microbiota associations with health and disease

The interplay between humans and gut microbiota has been shaped over more than a billion years of coevolution, resulting in a symbiotic relationship similar to a holobiont or superorganism. As a result, the intestinal microbiota contribute to various health functions, including the maturation and ongoing training of the host immune response.[20, 151, 152] Detrimental changes in the gut microbiota's characteristics (abundance, metagenomic function, diversity, and composition), collectively referred to as 'dysbiosis', can weaken the intestinal barrier, leading to the colonization or outgrowth of organisms, increased inflammation, immune dysregulation, and metabolic issues, thus compromising human health (Figure 1.3).[19, 22, 92, 153-159] Note that dysbiosis remains poorly defined, largely due to significant interindividual variability within patients and across different diseases, which complicates the establishment of a clear definition for a healthy and unhealthy gut microbiota. To measure dysbiosis, several indices have been proposed.[160] However, the proposed measures are not widely adopted and may not fully capture the complexities of dysbiosis.

One of the early milestone papers on the relation between the microbiota and disease is a study by Turnbaugh et al. published in 2009.[161] Here, the authors showed that obese mice had a gut microbiota with increased capability for energy harvest from the diet. Also, they linked the gut microbiota to the pathophysiology of obesity through a series of experiments. This included transplanting feces from obese mice into gnotobiotic mice, which led to a greater increase in body fat than when gnotobiotic mice received a fecal microbiota transplantation

from lean mice. This study not only found a correlation between the gut microbiota and disease, but also showed a causal link between the two. Subsequently, the study triggered a global interest in the role of the gut microbiome in human health and disease.[161]

Advancements driven by initiatives such as the Human Microbiome Project (HMP) and European Metagenomics of the Human Intestinal Tract (MetaHIT) have harnessed vast sequencing datasets to illustrate the structure and function of the healthy core microbiota.[21, 162] Defining the healthy microbiota is extremely difficult, as healthy gut microbiota are characterized by substantial interindividual variation. In the gut, however, healthy microbiota are associated with bacterial diversity, as they exhibit lower susceptibility to invasion, suppress the outgrowth of harmful species, and demonstrate greater resilience to perturbations.[91, 163] Intriguingly, while the human gut microbiota's compositional diversity is substantial, functional gene profiles remain strikingly similar across individuals.[22] This similarity was first reported in a study of 18 females who shared more than 93% of the enzyme-level functional groups, and was later confirmed in a much larger population by the HMP and MetaHIT data.[21, 162, 164] This functional similarity among distinct microbiota profiles underscores the significance of function over species identity. However, variations in species could impact functional effectiveness, as seen with variations in short chain fatty acid synthesis.[19, 22, 165, 166] Understanding the dynamics of the gut microbiota can guide strategies to increase the resilience of healthy states or counteract unhealthy ones (Figure 1.3B). Overall, the idea is that it is beneficial to have a diverse gut microbiome, which provides metabolic flexibility while reducing the risk of infections and the development of inflammatory diseases (Box 1.3).

Box 1.3 - Gastrointestinal diseases and microbial dysbiosis. A proposed hypothesis for the development of gastrointestinal diseases delineates a multi-step mechanism involving factors that trigger mucosal abnormalities and inflammation, microbial dysbiosis, morphological and functional changes, and interindividual microbial transfer as a continuous pathogenic cycle.[20, 151] For example, *Clostridioides difficile*, the main causative agent of nosocomial diarrhea, is an anaerobic, gram-positive, spore-forming bacillus.[167] *Clostridioides difficile* may outcompete other species, especially in a dysbiotic microbiome after antibiotic use, leading to colonization of the gut and subsequently to disease.[34] Recurrence of infections is not solely attributed to the reduction in diversity following antibiotic use, but there are also distinct bacterial signatures linked to recurrent colitis. These include a decrease in beneficial bacteria (e.g., *Faecalibacterium prausnitzii*) and an increase in strains from for example Lachnospiraceae, *Coprococcus*, *Ruminococcus*, and several *Clostridium* species.[168]

Microbial shifts have been associated with disease activity in gastrointestinal inflammatory disorders such as IBD, encompassing Crohn's disease (CD) and ulcerative colitis (UC). Most IBD patients suffer from periods of flares of inflammation with a severe impact on patients' quality of life. Although the exact cause of the disease and its exacerbation remain unclear, it is considered to result from complex interactions between an altered intestinal immune response to commensal bacteria, shifts in the intestinal microbiota, and external environmental factors in a genetically susceptible host.[169, 170] The gut microbiota of individuals with ileal CD shares similarities with that of infants: both are characterized by reduced diversity, elevated levels of *Ruminococcus gnavus* and Enterobacteriaceae, and an under-representation of the genera that are prevalent in healthy adults, including *Faecalibacterium prausnitzii* and *Roseburia*.[22, 33, 171-174]

Additionally, microbial variations have been observed relating to heightened *Bacteroides* spp. and diminished *Clostridium coccoides*.[175-180] However, these associations vary among studies, likely due to the heterogeneity of CD, differences in sequencing technologies, and the interindividual microbiota variability.[154, 181, 182]

Microbiome-related therapies, including prebiotics, probiotics, and fecal microbiota transplantation (FMT), aim to transition the patients' microbiome from a dysbiotic to a healthy state.[183-185] Although many probiotic strains demonstrate strong survival during passage through the gastrointestinal tract and retain metabolic activity, most human studies indicate they have very short-term persistence and minimal influence on the resident microbiota composition. In contrast, FMTs (transplanting healthy donor fecal matter into the patient's gut) seem to be more effective at changing an existing gut microbiota, yet the underlying processes leading to recovery remain largely unexplored and not well understood.[95, 186-188] The current thought is that the succession in the recovery process seems to start with an increase in facultative anaerobes and aerotolerant bacteria (similar to the development of the microbiota in an infant's gut), possibly because of temporary changes in redox potential, and then the re-establishment of obligate anaerobes.[22] FMT has demonstrated success in treating recurrent *Clostridioides difficile* infection, curing up to 85% of the patients, but its application in other diseases yields contrasting results.[189, 190] For IBD, the remission rate after FMT is 45%, though relapses occur in a certain proportion of patients.[191] Repeated FMT administrations seem to be needed to alter the chronic dysbiosis in the IBD patients' microbiota and allow for lasting changes.[192-195] Also, associated factors such as age, sex, donor characteristics (e.g., donor gut microbiota diversity), pretreatment, and antibiotic use influence FMT outcomes, underscoring the interplay between the host, the host microbiota, and the donor microbiota.[189, 193-200]

Approaches and challenges in analysing microbiota datasets

Samples from the gut microbiota provide a glimpse into the abundant diversity within the colon, revealing the multifaceted microbial ecosystem of the gastrointestinal tract.[61, 201] The most commonly used sample type for analysis of the gut microbiota is feces. Alternative sampling methods include taking biopsies during endoscopy or rectal swabbing. The advantage of rectal swabbing is that it relies on standardized protocols, whereas fecal sample collection often depends on individuals collecting the feces samples themselves at home, which can introduce variability. Both fecal sampling and rectal swabbing are also much less invasive than taking a biopsy. Moreover, a lower microbiota diversity is often found in samples obtained by a biopsy compared to fecal or rectal samples, which is probably caused by the bowel preparation beforehand, making this the least preferred method. Still, fecal samples or rectal swabs may miss specific microbial communities found in other (earlier) parts of the colon. For example, differences in microbial composition between rectal swabs and biopsies from the sigmoid colon suggest that distinct microbial communities exist in these areas. Rectal swabs may capture species suited to the transitional zone between anaerobic and more aerobic environments, while the squamous epithelium near the anal canal may host different microbes than the columnar epithelium further in the colon. Interestingly, UC often begins in this transitional zone, advancing inward from there.[202]

Driven by the challenge that over 99% of gut microbes are difficult to culture in a laboratory setting, researchers developed methods to study these microorganisms directly within their natural environment, primarily through sequencing the 16S ribosomal RNA (rRNA) gene. The advent of high-throughput sequencing has revolutionized the study of microbial communities, providing valuable insights into their compositions. Its relatively low cost has made it a widely used method for assessing gut microbiota.[203] This approach targets a specific region of the 16S rRNA gene that is unique to bacteria and present in all bacterial species containing multiple conserved and variable regions. The more conserved regions are useful to determine the higher-ranking taxa, whereas the more variable regions can help in identifying lower-ranking taxa, such as genera.[204] In short, after samples are collected, Polymerase Chain Reaction (PCR) amplification of the rRNA genes is applied, with primers amplifying the target gene for a wide range of microorganisms. Next, the PCR products are sequenced. The resultant sequence reads can be clustered into, for example, operational taxonomical units (OTUs), amplicon sequence variants (ASVs), or metagenomic-based operational taxonomic units (mOTUs). These units are then aligned to a reference database and annotated into taxonomic names.[205-209]

Note that a lot of bias originates from the sequencing technique and the misclassification of sequencing reads.[118, 210] Therefore, positive and negative controls are commonly processed along with the real samples.[111, 211-213] Negative controls allow assessing potential contamination, and positive controls (mock communities) allow the assessment of bias and variability among different runs (batch effects).[214] Taxonomy annotation employs the Linnaean classification system, encompassing three domains: Bacteria, Archaea, and Eukaryota, with prokaryotic microorganisms largely categorized within Bacteria and Archaea. The specificity increases through kingdom, phylum, class, order, family, genus, and species classifications. The technique of 16S rRNA gene sequencing allows accurate taxonomic classification up to the genus level, but lacks reliable species-level or functional information.[19] For a comprehensive assessment, to species or even strain or genotype level, deeper exploration through whole genome (shotgun) sequencing (WGS) is imperative. This higher-resolution approach uncovers the functional genes of microbial communities but is considerably more expensive compared to amplicon sequencing. Even further, for a more detailed understanding, proteomics and metabolomics can determine the biochemical associations between microbial taxa (and human host). Proteomics provides information on the proteins present, including their structures and functions, while metabolomics offers insights into the metabolites in the sample.

Microbiota data are often manifested in matrices with the samples as rows and the taxa as columns. It is important to note that the interpretation of these data is complicated by several statistical challenges.[215] First of all, most datasets are comprised of more features (columns) than objects (rows), which makes classical statistics challenging. Secondly, species-abundance distributions exhibit a pronounced long-tail pattern, with many low-abundance taxa appearing in only a small fraction of samples.[216] Consequently, microbiota abundance data also frequently faces zero-inflation (i.e., the matrices are highly sparse) due to true absences or undetected presences when the abundance falls below detection limits.[215, 217, 218] However, possibly the biggest challenge is that the count measurements obtained are not viewed as 'true' count data, instead only relative abundances are available.[215, 219] Because, regardless of the amount of information available in the DNA sample, the output of a sequencing analysis is constrained by the limitations and sequencing depth of the platform used.[219, 220]

Sequencing instruments are limited to delivering reads up to their capacity, with each sample constrained by the available slots and the molar concentration loaded in the sequencing machine.[221] Therefore, the total read count observed in a high-throughput sequencing run is a fixed size, resulting in a random sample of the relative abundance of the molecules in the sample. This is explicitly acknowledged when microbiota datasets are mathematically transformed or converted to relative abundance values (Box 1.4).[201, 215, 219, 222]

Box 1.4 - The impact of data transformations in microbial ecology research. Rarefaction aims to rectify discrepancies in total reads per sample. However, rarefaction sacrifices statistical power and fails to really address the compositionality issues, as it involves subsampling to the lowest read depth across samples.[214, 222] Alternatives to rarefaction all involve some type of transformation, the most common of which are scaling, log-ratio transformations, or converting the abundance count of each taxon into proportions or relative abundances that sum up to one for each sample.[201, 215, 219, 220, 222-224] However, this brings another challenge, as it is quite possible that a significant change in the relative abundance of a species is observed, while the absolute number does not change. In microbial ecology studies, this phenomenon is important to consider when analysing shifts in species composition within a population or ecosystem. Imagine a simplified scenario with only two species, A and B, in a microbial community. Initially, there are 100 individuals of species A and 100 individuals of species B, making the total population size 200. This results in a 50% relative abundance for both species (100/200). Now, an environmental change or intervention occurs that favors the growth of species A, causing it to double in number to 200 individuals. Species B, however, remains at 100 individuals. The total population size is now 300 (200 of A and 100 of B). Despite the absolute number of species B remaining the same, the relative abundance of species B has decreased to 33% (100/300), while the relative abundance of species A has increased to 67% (200/300).

A common goal in microbiome research is to understand the relationships, ecological stability, and dynamic behaviours of the microbiota communities and to unravel their impact on health and disease. An important decision in study design involves whether to gather repeated measurements from the same individuals or to allocate resources to sample from more subjects at a single time point. Often, it is not possible to collect repeated samples from many subjects. This is due to the high costs associated with longitudinal sampling and, particularly in medical studies, the burden it places on patients to return for follow-up visits. The choice to gather repeated samples or not should hinge on the study's objective. Cross-sectional designs, with one sample per subject, are suited for examining differences in microbiota composition in association with health or disease.[22, 62] In contrast, longitudinal designs are preferred for studying disease-course dynamics, treatment effects in randomized controlled studies, and temporal fluctuations within the microbial community.[46, 52, 225, 226] Consequently, to distinguish intra-individual gut microbiota fluctuations from disease or treatment specific signals, robust assessment of microbial features demands repeated sampling.

Working with longitudinal microbiota data is challenged by many difficulties, including inconsistent sampling frequencies, varying numbers of subjects per phenotype, or varying numbers of samples per subject.

Realizing the importance of the gut microbiome for health and disease has encouraged the development of methods and tools for its analysis and modeling. Techniques encompass, among others, visualization, (temporal) clustering, network analyses, longitudinal and time series models. Co-occurrence based methods, based on e.g., Pearson's and Spearman's correlation measures, are quite popular for network inference due to their ease of use.[227, 228] While measures of co-occurrence, such as correlations, are powerful tools for generating hypotheses, caution is advised when assigning biological meaning to them.[216, 220, 229] Graph theory has gained prominence for its ability to depict microbial community structures, capturing the potential interrelations among a multitude of species, possibly highlighting potential keystone species and subcommunities.[8] In these graphical representations, nodes typically represent biological features, such as microbial taxa, genes, metabolites, or even environmental and host factors.[112, 223] Edges signify correlations between nodes, but they are often too easily interpreted as biological relationships. Edges between microbial taxa might result from direct interactions, such as competition, secretion of substances, immune modulation, or from mere co-occurrence without any direct biological meaning, e.g., due to shared preferences, nutrient availability, or similar responses to environmental factors.[8, 230]

Ordination analysis (e.g., principal component analysis (PCA), principal coordinates analysis (PCoA), and non-metric multidimensional scaling (NMDS)) reduces data with many variables (high dimensionality) to a set of two or three dimensions.[220, 231] PCA identifies linear relationships and projects data onto orthogonal axes, PCoA uses distance matrices for non-linear relationships, and NMDS preserves the rank order of distances for data visualization. Ordination analyses are tools used for visualizing and comparing microbial community differences. In ordination plots, microbial communities are depicted as points, with sequential samples linked by arrows. These arrows illustrate the system's trajectory through the phase space.[232, 233] Samples with similar bacterial communities tend to cluster closer together, whereas those with distinct compositions are positioned further apart.[234] Clustering techniques can then be applied to identify groups of points that share greater similarity with each other compared to points in other clusters.[235] Note that applying a clustering technique after dimension reduction by ordination analysis neglects significant information. Therefore, it is recommended to cluster samples based on the original data, as, for example, by Dirichlet multinomial mixture models. This is a clustering technique that is well-suited for multivariate relative indices and establishes relationships between patient samples by identifying similarities among them. This method has been used before in, for example, uncovering patterns within the microbiota development of infant cohorts.[236]

An alternative method for analysing microbiota data is to parameterize mathematical models of community dynamics using longitudinal data. However, the high-dimensional aspect of the microbial communities remains challenging for fitting dynamical models to data. Moreover, constructing such models requires substantial prior knowledge of the system, which is seldom available. An alternative approach is to construct a system that captures the core characteristics of the system's elements. The time development of a dynamical system will in this case be described by a set of ordinary differential equations (ODEs) that define the principles governing the system's dynamics (Box 1.5).[232] However, high-dimensionality is again a challenge for these types of models. Most existing modeling approaches consider a few species at a time and fail to capture the true multivariate nature of the data. Also, they have high computational costs and low prediction accuracy.[217] In general, when selecting a model, one must choose a balance between realism (the complexity of the system) and the ability to systematically and comprehensively analyse the microbial system with regard to the study's objectives.[232]

Box 1.5 - Models for microbial community dynamics. Dynamical systems theory is a well-established mathematical framework used to describe behaviour and evolution of complex systems over time.[237] The development and analyses of dynamical models allows a better understanding and prediction of community dynamics and engineering of community properties.[232, 237] The generalized Lotka-Volterra (gLV) framework is a popular choice, benefitting from a deep theoretical understanding.[100, 114, 238, 239] However, the validity of this approach is under debate, due to the model's reliance on strong assumptions, such as leveraging quasi-linearity in interaction terms. While pairwise models, such as the gLV models, focus on the increase and decrease of abundance of local species, mechanistic models consider interaction mediators as state variables.[240] For example, if a certain species releases a compound which stimulates another species growth upon consumption, then a mechanistic model tracks abundances of both species and also the concentrations of the compound. Genome-scale metabolic models (GEMs) or constraint-based reconstruction and analysis (COBRA) models show great potential for modeling the metabolism of microbial communities.[237, 241-244] Note that mechanistic models often exclude molecular details, such as the processes by which chemical signals are received and processed by recipients, as well as the subsequent effects these signals have on the recipients' behaviour or function.[240] Ideally, models could also include the physical and chemical environment, as this is a very important part of the species' environment.

Aim and outline of the thesis

In this thesis, we aim to bridge the gap between microbiological, ecological, and clinical concepts, which may help to better understand microbial dynamics, microbial involvement in inflammatory bowel disease (IBD), and treatment success in fecal microbiota transplantation (FMT). Specific aims addressed in this thesis are:

1 Characterize ecological structure in the human gut microbiota

Here we aim to unravel the correspondence between correlation-based networks and the underlying network of ecological interactions. Human microbiota networks are often characterized by pairwise correlation-based methods, applied to a few sampling points in time. Such characterization implicitly assumes that the microbial system tends towards a stable equilibrium. However, temporal ecological microbiota dynamics challenge the assumptions of prevailing correlation-based methods and provide leads for alternative characterizations.

2 Describe associations between gut microbial abundances and IBD

For this aim, we analyse fecal samples derived from Crohn's disease (CD) patients. CD, a type of IBD, has been associated with atypical microbiota composition and metagenomic function. However, results from the literature on microbial associations with IBD have not been consistent, especially with respect to disease activity. This could be because the process of changing from a healthy to an unhealthy microbiota may not always follow a deterministic pattern. It could be unique per patient. We provide a possible solution by studying associations across a spectrum of individual patient responses to disease activity.

3 Examine ecological microbiota determinants associated with FMT treatment success

FMT has emerged as a promising treatment for microbiota-related intestinal disorders, but its effectiveness in patients with ulcerative colitis (UC), another type of IBD, is still limited. To characterize microbiota determinants of clinical remission, we examined longitudinal associations between bacterial families and clinical response to FMT. It was previously assumed that successful grafting of donor-derived microbes is associated with clinical remission, but this donor-centric view has recently been questioned. Therefore, we also investigate whether donor-derived, newly emerging, or host-associated species are linked to patients achieving clinical remission after FMT treatment.

This thesis starts with methodological considerations dedicated to the characterization of microbial interactions and communities. Thereafter, our studies have a more clinical application.

In Chapter 2 we use a mathematical model as a ground truth to simulate bacterial communities. We specifically investigate how microbial network inference is related to interindividual variation in population-dynamic parameters and different types of networks of microbial interactions. In addition, we assess the impact of sample size or measurement noise on the performance of correlation-based network reconstruction.

In Chapter 3, we apply a technique that clusters time series based on similarities in their dynamical patterns, so-called wavelet clustering analysis. This technique, almost unknown in the microbiota field, provides insight into the dynamic relationships between members of the microbial community. This allows for an alternative characterization of community structures as compared to the commonly used correlation-based methods.

In Chapter 4 we apply quantile regression, an extension of the general linear model that allows for investigation of relationships across different quantiles of the distribution of a response variable. The idea behind this method is that not all individuals are equally responsive to disease-induced changes in terms of abundance of specific bacterial groups. We test especially whether associations between relative abundances of specific families with CD can be found relative to healthy controls and for different disease courses (i.e., remission vs. exacerbation).

In Chapters 5 and 6 we investigate a longitudinal dataset of UC patients who underwent FMT treatment. In Chapter 5 we employ several multivariate analyses to examine associations between bacterial families and FMT treatment success: a Dirichlet multinomial mixture model, longitudinal mixed models, and PCA with Aitchison distances. In Chapter 6 we map the ecological dynamics in the gut microbiota during and after the FMT treatment. We categorize all the species in ecological groups based on their origin (already present in the host pre-FMT, derived from the donor, or introduced as a novel species that was neither present in the host nor donor) and investigate their patterns of presence and absence, as well as their relative abundance over time.

All findings are summarized and placed in the broader context of existing literature in Chapter 7.

Part I

Ecological structure in the human gut microbiota

Chapter 2

Correlations carry limited information about network interactions

Susanne Pinto[1], Elisa Benincà[2], Egbert H. van Nes[3], Marten Scheffer[3], Johannes A. Bogaards[4]

1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
2 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
3 Department of Environmental Sciences, Wageningen University and Research, Wageningen, the Netherlands
4 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

PLOS Comput Biol. 2022 Sep 9;18(9):e1010491. doi: 10.1371/journal.pcbi.1010491.

Species abundance correlations carry limited information about microbial network interactions

Abstract

Unraveling the network of interactions in ecological communities is a daunting task. Common methods to infer interspecific interactions from cross-sectional data are based on co-occurrence measures. For instance, interactions in the human microbiome are often inferred from correlations between the abundances of bacterial phylogenetic groups across subjects. We tested whether such correlation-based methods are indeed reliable for inferring interaction networks. For this purpose, we simulated bacterial communities by means of the generalized Lotka-Volterra model, with variation in model parameters representing variability among hosts. Our results show that correlations can be indicative of the presence of bacterial interactions, but only when measurement noise is low relative to the variation in interaction strengths between hosts. Indication of interaction was affected by type of interaction network, process noise, and sampling under non-equilibrium conditions. The sign of a correlation mostly coincided with the nature of the strongest pairwise interaction, but this is not necessarily the case. For instance, under rare conditions of identical interaction strength, we found that competitive and exploitative interactions can result in positive as well as negative correlations. Thus, cross-sectional abundance data carry limited information on specific interaction types. Correlations in abundance may hint at interactions but require independent validation.

Introduction

The human body harbors an exceptional bacterial diversity.[21] The composition of these bacterial communities is generally shaped by characteristics of the host and by the ecological dependencies among bacterial species themselves.[8, 13, 245] These dependencies often occur through competitive or synergistic interactions, which may lead to a (mutual) decrease or increase in the abundance of interacting species.[123] For instance, it is known that bacteria can interact with each other through excreted metabolites, which can function as an antimicrobial or as a food source.[8, 246] Among other mechanisms, for example negative interactions take place when toxic compounds produced by one species harm other bacteria, whereas positive interactions occur when bacteria feed on the nutrients that are produced by others. Besides, many different forms of interactions exist, depending on the effects experienced by the species involved. Knowledge of interspecific interactions in the human microbiome is paramount to understand ecological processes and compositional changes in relation to health and disease.[49, 190]

Most human microbiome studies are limited to only a few samples in time, presenting mere 'snapshots' of the microbial ecosystem, even if these samples are derived from hundreds of human hosts.

A common way to infer microbial networks from such cross-sectional data is by quantifying co-occurrence, e.g., through (partial) correlations, between bacterial phylogenetic groups. Several different conclusions have been derived from such endeavors, for example on species associations that reflect shared or overlapping niche preferences,[247] microbial community structure,[248, 249] the resilience of microbial communities to perturbations,[100] and keystone species in microbial networks.[114] Currently there are several correlation-based network tools available that can deal with the difficulties of microbiome data, such as compositionality.[230, 250, 251] The potential of correlation-based approaches for uncovering microbial networks has been highlighted in previous research.[252]

Whether correlation-based networks represent meaningful ecological structure in microbial communities is, however, debated. Carr et al. (2019) showed that spurious correlations may occur due to the use of sequencing methods, data transformations, and the large number of unmeasured variables.[216] Berry & Widder (2014) and Hirano & Takemoto (2019) assessed the performance of different co-occurrence methods for inferring interaction structure and found that their performance strongly depends on the underlying network properties, such as network size and density, and the number of samples used to construct the network.[114, 229] Apart from the challenges of metagenomic-based abundance data and disagreement between various network tools, here we question whether correlations themselves are at all useful to distinguish between different ecological interaction types. Resource competition and metabolic cooperation have been successfully inferred within environmental microbiomes, by linking ecological distribution data to multi-species metabolic models and subsequent verification of putative interactions by means of experimental co-growth analysis.[119] However, host-associated microbiomes often include non-culturable organisms, without information on nutrient requirements or metabolic function. Likewise, performance of correlation analysis in relation to alternative interaction types in the human microbiota is not well understood and deserves further investigation.

Correspondence of correlations with ecological interactions needs to be studied against a known ground truth, which can be achieved by means of simulation. Mathematical models have been used as ground truth in assessment of correlation network techniques before,[253] but correlation networks have not been systematically investigated against distinct interaction types in dynamic models. This requires elucidation especially as the 'true' ecological networks governing microbiome dynamics are still unknown. For this purpose, we assessed the performance of correlation-based network reconstruction by simulating abundance data based on the generalized Lotka-Volterra (gLV) model. The gLV model describes the collective dynamics of multiple species by means of an interaction matrix that can modulate different types of interactions.[232] The model is commonly used in microbiome studies for different aims: to simulate microbial communities under various interaction structures,[232] to infer interaction structure from time series data,[100] to forecast population dynamics after a perturbation,[254] to infer the network topology from steady state samples,[255] and to identify the efficiency of intervention protocols in altering the state of a system via the addition or subtraction of microbial species.[256] In ecology, gLV-type models have been questioned for their reliance on pairwise additive interactions, as well as for the strictly linear effects imposed on interspecific interactions. Nonetheless, from the perspective of network inference, it makes sense to first investigate gLV-type models, as their first-order description of ecological dependencies, specified through a pairwise interaction matrix, resembles the objective of correlation analysis and most network models.[8]

In addressing how gLV-type interactions can be inferred from cross-sectional data, we mainly focus on the correspondence between the obtained correlation-based networks and the underlying network of ecological interactions. We specifically investigate how the inference of microbial interaction types is enabled by interindividual variation in population-dynamic parameters (e.g., species-specific carrying capacities, intrinsic growth rates, and strength of interspecific interactions) and how network reconstruction is affected by gLV model assumptions. We highlight several situations where correlations cannot distinguish microbial interaction types and therefore recommend careful interpretation and validation when inferring networks from cross-sectional abundance data.

Methods

Two-species Lotka-Volterra model with self-limitation

First, we investigated how interactions between two species of microbial populations are displayed in terms of correlations of abundances in the Lotka-Volterra model. For the sake of convenience, we use the term 'species', although in studies with real microbiota data it is often not possible to characterize the taxonomic abundances at species-level and therefore genera or higher taxonomic levels are often used instead.

The two-species Lotka-Volterra model is given by the following set of ordinary differential equations:

$$\frac{dN_1}{dt} = r_1 N_1 (1 - K_1^{-1} N_1 + \alpha_{12} N_2) \qquad \textbf{Eq. 2.1}$$

$$\frac{dN_2}{dt} = r_2 N_2 (1 - K_2^{-1} N_2 + \alpha_{21} N_1) \qquad \textbf{Eq. 2.2}$$

Here, N_i is the abundance of either species 1 or species 2 (with $i = 1$ or $i = 2$). The term r_i is the intrinsic growth rate of each species, here normalized to 1 and 2 per time unit for species 1 and 2, respectively. The effect of each species' abundance on its own growth is defined in terms of the species-specific carrying capacities K_i, with $\alpha_{ii} = -K_i^{-1}$ denoting intraspecific competition. We arbitrarily chose the carrying capacity for the first species to be higher than the carrying capacity for the second species ($K_1 = 1.5$; $K_2 = 1.1$), meaning intraspecific competition is less strong for species 1 compared to species 2. Furthermore, α_{ij} ($i = 1, 2$; $j = 1, 2$; $i \neq j$) indicates the interspecific interactions (the effect of one species abundance on the growth of the other species). A positive α_{ij} (e.g., as in the case of mutualism) denotes a positive effect of species j on the growth of species i, a negative α_{ij} (e.g., as in the case of competition) means a negative effect of species j on the growth of species i (Appendix Figure 2.1). We assessed the effect of variation in the interspecific interaction parameters on correlation in equilibrium abundance between both species. For this purpose, the interspecific interaction strengths (α_{12} and α_{21}) were drawn randomly from two normal distributions with similar or different mean and similar or different standard deviations (σ_α). Moreover, we also investigated the situation where $|\alpha_{12}| = |\alpha_{21}|$. Note that it was not possible to achieve stable coexistence for every combination of α_{12} and α_{21}. More information on the conditions for coexistence can be found in Box 2.1.

Box 2.1 - Coexistence in a two-species Lotka-Volterra model with self-limitation.
The conditions for coexistence in the two-species Lotka-Volterra model with self-limitation can be derived by setting both growth equations to zero and investigating what parameter combinations yield $\overline{N_1} > 0$ as well as $\overline{N_2} > 0$. Here, $\overline{N_1}$ denotes the equilibrium abundance of species 1 and $\overline{N_2}$ denotes the equilibrium abundance of species 2. Writing these conditions in terms of N_1 as functions of N_2 gives the following:

$$\frac{dN_1}{dt} = 0 \wedge N_1 > 0 \rightarrow N_1 = f_1(N_2) = K_1 + \alpha_{12} K_1 N_2 \qquad \textbf{Eq. 2.A}$$

$$\frac{dN_2}{dt} = 0 \wedge N_2 > 0 \rightarrow N_1 = f_2(N_2) = -\frac{1}{\alpha_{21}} + \frac{1}{(\alpha_{21} K_2)} N_2 \qquad \textbf{Eq. 2.B}$$

The joint equilibrium abundance of both species ($\overline{N_1}$, $\overline{N_2}$) is determined by $f_1(N_2) = f_2(N_2)$. Equation 2.A shows that species 1 grows to its carrying capacity K_1 in the absence of interspecific interactions, i.e., if $\alpha_{12} = 0$. Likewise, $\alpha_{12} > 0$ allows species 1 to grow to higher abundance in the presence of species 2 than determined by its own carrying capacity, whereas $\alpha_{12} < 0$ leads to a reduced abundance of species 1 in the presence of species 2. Similar relations hold for the abundance of species 2 in the presence of species 1, depending on α_{21}. From Equation 2.B, it can also be derived that $\overline{N_1} > 0$ is only compatible with $\overline{N_2}$ being above its carrying capacity K_2 if at the same time $\alpha_{21} > 0$, whereas $\overline{N_2}$ being below K_2 requires $\alpha_{21} < 0$.

Joint inspection of Equations 2.A and 2.B also establishes the following, more subtle, conditions for coexistence:

1 If $\alpha_{12} > 0$ and $\alpha_{21} > 0$, e.g., in case of mutualism, f_2 has a negative intercept in the Cartesian (N_2, N_1) coordinate system (Figure 2.A - panel A). As both functions have a positive slope in this situation, and f_1 always has a positive intercept, f_2 must have a stronger slope than f_1 for both to intersect in the positive quadrant. This boils down to $\frac{1}{\alpha_{21} K_2} > \alpha_{12} K_1$, or equivalently $\alpha_{21}\alpha_{12} < \alpha_{11}\alpha_{22}$, as $K_i = -\frac{1}{\alpha_{ii}}$ by definition. This means that the product of interspecific mutualism needs to be smaller than the product of intraspecific competition for both species to co-exist, otherwise there is no control of population growth.

2 If $\alpha_{12} < 0$ and $\alpha_{21} < 0$, e.g., in case of competition, both functions have positive intercept and negative slope (Figure 2.A - panel B). Intersection in the positive quadrant requires the function with the larger intercept to intersect the abscissa, i.e., the N_2 axis where $N_1 = 0$, at a smaller value than the function with the smaller intercept. Thus, this requires $|\alpha_{21}| > \frac{1}{K_1}$ and $|\alpha_{12}| > \frac{1}{K_2}$, with f_2 having the larger intercept, or alternatively, $|\alpha_{21}| < \frac{1}{K_2}$ and $|\alpha_{12}| < \frac{1}{K_1}$, with f_2 having the larger intercept. In the first instance, interspecific competition is stronger than intraspecific competition, whereas in the second instance, interspecific competition is less strong than intraspecific competition. It turns out that only the last of these conditions yields a stable equilibrium, meaning that the abundances of both species return to equilibrium after small displacements.

3 If $\alpha_{12} < 0$ and $\alpha_{21} > 0$, e.g., in case of exploitation of species 1 by species 2, f_1 has a positive intercept and negative slope, whereas f_2 still has a negative intercept and positive slope (Figure 2.A - panel C). Intersection in the positive quadrant requires f_1 to intersect the abscissa at a larger value than K_2, the point where f_2 intersects the abscissa. The condition for coexistence thus becomes $|\alpha_{12}| < \frac{1}{K_2}$, or equivalently $\alpha_{12} < \alpha_{22}$, meaning that the parasite should exert stronger inhibitory effect on its own growth than on that of the exploited species.

4 Conversely, in case of exploitation of species 2 by species 1, i.e., if $\alpha_{12} > 0$ and $\alpha_{21} < 0$, both f_1 and f_2 have a positive intercept, but f_1 now has a positive slope whereas f_2 has a negative slope (Figure 2.A - panel D). Intersection in the positive quadrant then requires f_1 to have a smaller intercept than f_2. The condition for coexistence thus becomes $\frac{1}{K_1}|\alpha_{21}| <$, or equivalently $\alpha_{21} < \alpha_{11}$, again meaning that the parasite should exert stronger inhibitory effect on its own growth than on that of the exploited species.

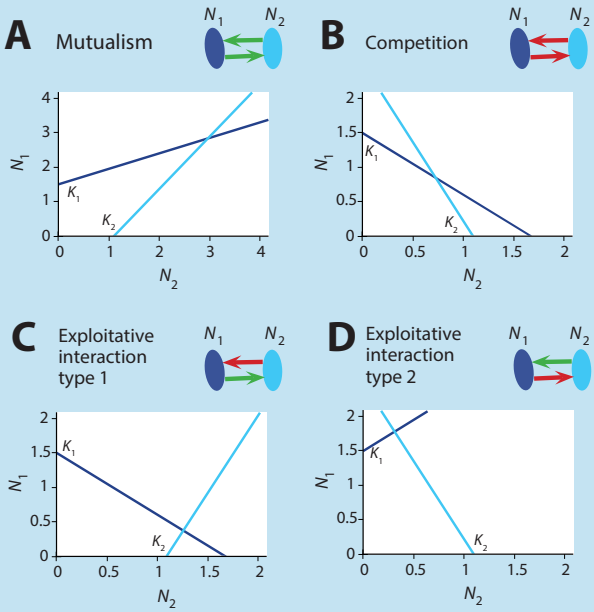


Figure 2.A - Zero-growth isoclines ('null-clines') in the two-species Lotka-Volterra model. Visualization of the effect of species 1 and 2 abundances on each other in the Cartesian (N_2, N_1) coordinate system. Here, f_1 denotes the isocline of zero growth of species 1, i.e., $f_1(N_1)$ (in dark blue), and f_2 denotes the isocline of zero growth of species 2, i.e., $f_2(N_1)$ (in light blue). Their point of intersection represents the joint equilibrium abundance of both species, i.e., $(\overline{N_1}, \overline{N_2})$. Throughout $K_1 = 1.5$ and $K_2 = 1.1$. Parameters for the various scenarios: $\alpha_{12} = 0.3$ and $\alpha_{21} = 0.6$ under mutualism; $\alpha_{12} = -0.6$ and $\alpha_{21} = -0.4$ under competition; $\alpha_{12} = -0.6$ and $\alpha_{21} = 0.4$ under exploitative interaction type 1; and $\alpha_{12} = 0.6$ and $\alpha_{21} = -0.4$ under exploitative interaction type 2.

The additional requirement for stable coexistence is that the two-species system should be locally stable around the equilibria $(\overline{N_1}, \overline{N_2})$, which can be formalized in terms of the Jacobian matrix of the Lotka-Volterra model evaluated at $(\overline{N_1}, \overline{N_2})$. This amounts to determining the trace and determinant of the matrix of the partial derivatives of the growth equations regarding either species, i.e.,

$$\begin{bmatrix} r_1 - 2r_1\overline{N_1}/K_1 + r_1\alpha_{12}\overline{N_2} & r_1\alpha_{12}\overline{N_1} \\ r_2\alpha_{21}\overline{N_2} & r_2 - 2r_2\overline{N_2}/K_2 + r_2\alpha_{21}\overline{N_1} \end{bmatrix} \quad \textbf{Eq. 2.C}$$

It can be verified that the conditions for coexistence stated under mutualism and exploitative interactions yield equilibria that are locally stable, just as the last of the conditions under competition. We will not derive these conditions here, as these are covered by textbooks on theoretical ecology.[257]

In summary, the two-species Lotka-Volterra model with self-limitation has the following possibilities for stable coexistence (Table 2.A):

Table 2.A - Conditions for stable coexistence in the two-species Lotka-Volterra model.

Type of interaction	Condition	Outcome
Mutualism $\alpha_{12} > 0 \wedge \alpha_{21} > 0$	$\alpha_{12}\alpha_{21} < \frac{1}{(K_1K_2)}$	$\overline{N_1} > K_1 \wedge \overline{N_2} > K_2$
Competition $\alpha_{12} < 0 \wedge \alpha_{21} < 0$	$\lvert\alpha_{12}\rvert < \frac{1}{K_2} \wedge \lvert\alpha_{21}\rvert < \frac{1}{K_1}$	$\overline{N_1} < K_1 \wedge \overline{N_2} < K_2$
Exploitative interaction type 1[a] $\alpha_{12} < 0 \wedge \alpha_{21} > 0$	$\lvert\alpha_{12}\rvert < \frac{1}{K_2}$	$\overline{N_1} < K_1 \wedge \overline{N_2} > K_2$
Exploitative interaction type 2[b] $\alpha_{12} > 0 \wedge \alpha_{21} < 0$	$\lvert\alpha_{21}\rvert < \frac{1}{K_1}$	$\overline{N_1} > K_1 \wedge \overline{N_2} < K_2$

[a] Exploitative interaction type 1: species 1 is being exploited by species 2

[b] Exploitative interaction type 2: species 2 is being exploited by species 1

The condition for stable coexistence of competitors requires both species to have less effect on the growth of the other species than on itself. In case of an unstable equilibrium, either species will eventually outcompete the other; the species with an initial advantage will drive the other species to extinction, a condition referred to as competitive exclusion.[258, 259] This will occur, for instance, when each species produces a substance which is toxic to the other species but relatively harmless to itself.

Generalized host-specific Lotka-Volterra model

Microbial abundance is not only shaped by intra- and interspecific interactions, but also by host characteristics, for example lifestyle, diet, and age.[260] Therefore, we investigated the performance of correlation-based network inference of microbial networks for a host-specific version of the gLV model. The host-specific gLV model is given by:

$$\frac{dN_{i,m}}{dt} = r_{i,m} N_{i,m} (1 - K_{i,m}^{-1} N_{i,m} + \sum_{\substack{j=1 \\ j \neq i}}^{S} a_{ij,m} N_{j,m}) \qquad \textbf{Eq. 2.3}$$

Here, $N_{i,m}$ is the abundance of each species i in host m, with $i = 1, ..., s$ (s being the total number of bacterial species) and $m = 1, ..., 300$ (the total number of hosts). The terms $r_{i,m}$ and $K_{i,m}$ are the intrinsic growth rates and the carrying capacities of each species i in host m. The carrying capacities are kept separated from the interaction matrix A which only contains interspecific interactions (namely, the pairwise terms a_{ij}), facilitating a one-to-one comparison with the correlation matrix.

Parameterization of the base case simulations

We started with a base case, and we added step by step variation to this case. Note that the base case parametrization does not reflect any particular real-world system. Rather, parameters were chosen in such a way to facilitate computation and promote coexistence among species. Variations to the base case parameters are shown later on, but also here, findings should be appreciated from a qualitative rather than quantitative viewpoint. In the base case, the number of bacteria equals ten. The species-specific growth rate r_i and the species-specific carrying capacity K_i were randomly drawn from uniform distributions, respectively U(0.05, 0.1) and U(0, 1). The density of the interaction matrix A in the base case was chosen such that both sparsity of the interaction network and coexistence of the species was promoted in all simulations; in the base case, density was ¼ meaning that three out of four possible interactions were set to zero. Moreover, to ensure coexistence between species in the model we chose stronger intraspecific interactions than pairwise interspecific interactions. The species-specific parameters a_{ij} were drawn from a Gaussian mixture distribution, as follows. Half of the interactions were drawn from a negative normal distribution: $a_{ij} \sim N(-0.25, 0.1)$; and the other half of the interactions were drawn from a positive normal distribution: $a_{ij} \sim N(0.25, 0.1)$. All interactions were restricted to lie between –0.5 and 0.5, i.e., the normal distributions were truncated at –0.5 and 0.5. The parameters r_i, K_i, and the interaction matrix A were randomly drawn 1000 times from the aforementioned distributions to obtain 1000 different parameter combinations. Hereafter, host-specific parameters were drawn from log-normal distributions around species-specific parameters, as follows:

$$\begin{cases} \ln(|a_{ij,m}|) \sim N(\ln(|a_{ij}|), \sigma_a) \\ \ln(r_{i,m}) \sim N(\ln(r_i), \sigma_r) \\ \ln(K_{i,m}) \sim N(\ln(K_i), \sigma_K) \end{cases} \qquad \textbf{Eq. 2.4}$$

Here, σ_a denotes the interindividual variability in interspecific interactions among the 300 hosts (with $\sigma_a = 0.25$ in the base case), and $|a_{ij,m}|$ denotes the absolute strength of interaction from species j on the growth of species i for each host m. Note that, for the sake of simplicity, the use of log-normal distributions was adopted to induce fold-changes around population means, where both the presence and the sign of interspecific interactions are kept constant

across hosts. However, this may be untrue in real microbiota as many microbes can change metabolic pathways and therefore may switch between interaction types and interaction partners. In the base case model, the carrying capacities and growth rates were kept constant across hosts, meaning σ_r and σ_K were set equal to 0.

The simulation process yielded 300,000 time series (300 host-specific time series for each of the 1000 ten species networks). The running time of the model was chosen such that all species reached their equilibrium abundance. If at least one species did not survive (i.e., when its abundance dropped below 0.001), we rejected the simulation in favor of another randomly drawn parameter set. After sampling the abundances at equilibrium, we added independent and identically distributed noise υ to mimic uncertainty in measurements (with $\upsilon \sim U(-0.01, 0.01)$ in the base case). This measurement noise can be thought of as representing, for example, sampling errors, environmental contamination, batch effects during sequencing, or annotation errors in reference genomes.[261] Simulations were performed in R (R version 3.6.0; www.r-project.org). The gLV model was solved with the 'lsoda' function from the 'deSolve' R package (version 1.24) which uses a FORTRAN ODE solver written by Petzold & Hindmarsh (1995).[262, 263] R code is available on the GitHub repository (susannepinto/gLV_microbiome). A general overview of the base case simulation design is given in Figure 2.1.

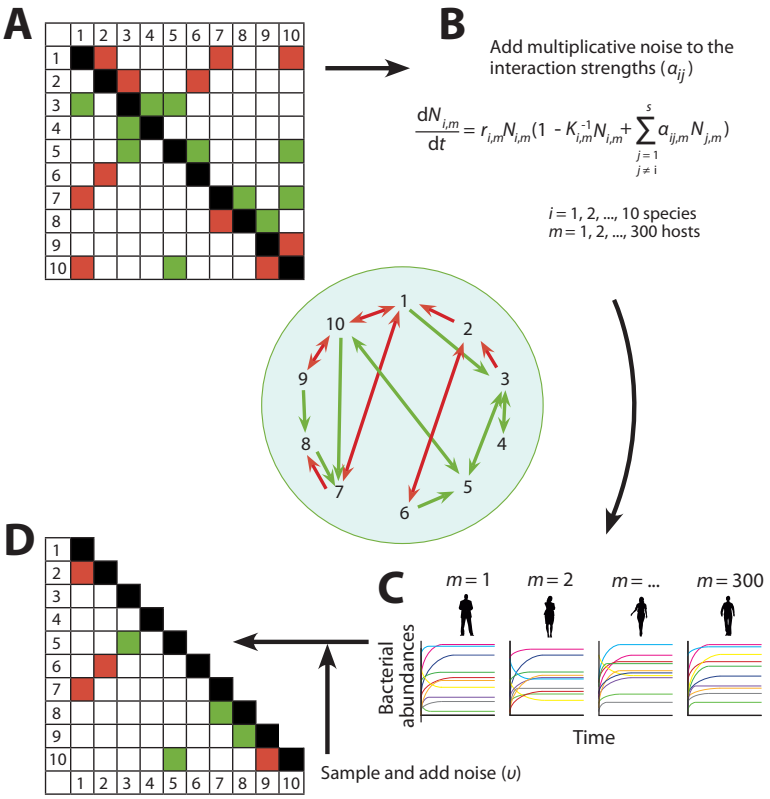


Figure 2.1 - Representation of the workflow. In an interaction network, singular green and red arrows represent a commensalistic interaction and an amensalistic interaction, respectively, whereas double green arrows represent mutualism and double red arrows competition. A combination of a green and red arrow signifies an exploitative interaction.

See Appendix Figure 2.1 for more details. A) A random interaction matrix *i*. This interaction matrix is implemented in the gLV model, B) together with the intrinsic growth rates and carrying capacities of the species. C) All time series are (slightly) different due to the variation in the interaction strengths. D) The partial correlations are calculated from the abundances per species sampled from the 300 different hosts at equilibrium. Only the significant correlations and the lower part of the matrix are used for the comparison with the original interaction matrix *i*. Variations to the workflow were studied by adding for example a perturbation or process noise.

Variations to the base case model

We studied multiple variations to the base case model. Similar to the base case simulations, we did 1000 simulations per variation. As a first variation, we added host-specific variability to the species-specific parameters r_i and K_i using Equation 2.4, with $\sigma_r = 0.25$ and $\sigma_K = 0.25$.

Second, we varied the amount of measurement noise, from $\upsilon \sim U(-0.01, 0.01)$ (medium noise in the base case) to $\upsilon \sim U(-0.001, 0.001)$ (low noise), and to $\upsilon \sim U(-0.1, 0.1)$ (high noise). We also simulated time series with a different type of noise, namely varying magnitudes of process noise *W* (Appendix Figure 2.2). In contrast to measurement noise, which was added only to the sampled abundances, process noise was added to the gLV model such that within-host population dynamics were perturbed at discrete time intervals Δt ($\Delta t = 1$ time unit). The time-varying process noise was drawn from a log-normal distribution to prevent the abundances from dropping below zero, i.e., $\Delta W_i = \ln(N_{i,m(\Delta t)}) - \ln(N_{i,m(t)}) \sim N(\ln(N_{i,t}), \sigma_W)$ (with $\sigma_W \sim N(0, 1)$ for high process noise and $\sigma_W \sim N(0, 0.1)$ for low process noise).

Further, we simulated data with interaction strengths drawn from a uniform ($\alpha_{ij} \sim U(-0.5, 0.5)$) or unimodal ($\alpha_{ij} \sim N(0, 0.15)$) distribution. As in the base case, the interaction strengths were restricted to lie between –0.5 and 0.5 (Appendix Figure 2.3).

We also analysed three different structures of microbial networks. First, we increased the number of species *s* from 10 to 30. To promote coexistence, we also reduced the density of the interaction matrix to ⅙. Secondly, we simulated a network based on a producer consumer relation between the species (Appendix Figure 2.4). Instead of random interaction networks (Appendix Figure 2.4A), the producer-consumer networks are based on a cross-feeding structure between producers and consumers (with equal numbers of producers and consumers) (Appendix Figure 2.4B). Producers excrete metabolites which are consumed by the consumers. Because consumers remove the 'waste' from the producers, the presence of a consumer can also be beneficial for the producers. Therefore, between producers and consumers positive interactions are more likely to occur than negative interactions. For this purpose, we drew the consumer-producer interactions from the positive side of the Gaussian mixture distribution ($\alpha_{ij} \sim N(0.25, 0.1)$). In contrast, among producers and consumers themselves, the interactions are predominantly negative as these species are more likely to compete for similar resources. For this purpose, we drew the interactions among producers and among consumers from the negative side of the Gaussian mixture distribution ($\alpha_{ij} \sim N(-0.25, 0.1)$). Third, we simulated a microbial network with interaction hubs, i.e., a network containing species with unusually high numbers of ecological interactions compared to other species in the network (Appendix Figure 2.4C).[264] Hub-species networks were created according to the Barabási-Albert model[265] and implemented with the 'barabasi.game' function from the 'igraph' R package (version 1.2.11). In the network-generating algorithm, interactions are distributed according to a mechanism of preferential attachment.

Thus, species with interactions obtain a higher chance of getting more interactions, resulting in a few 'hub-species' with many interactions. We constructed two scale-free directed graphs (with power = 2), denoting 'incoming' and 'outgoing' interactions, and combined these to obtain a bidirected graph. Density was kept similar to the base case model (¼).

Next, we also investigated how network inference is affected by sample size by considering a scenario with 3000 instead of 300 hosts. We did this for the base case model with random interaction networks, as well as for the producer-consumer and hub-species networks described before.

Last, we investigated the effect of a perturbation on the performance of network inference. The populations were perturbed after 175 time units, with a perturbation that lasted for 50 time units. The perturbation was modelled by taking a new set of random carrying capacities per species per sample. Due to the simulated perturbation, the equilibrium distribution shifted. After the perturbation, the species grew back to their original equilibrium. Sampling occurred before, during, or after the perturbation.

Assessment of correlation-based network inference

With the simulated data at hand, we created a dataset with the abundances of the model species sampled at equilibrium for each host *m*. After adding measurement noise to the data, we inferred the correlations between species by calculating the Pearson's partial correlation coefficients ρ between all abundances N_i across the *m* different hosts (Figure 2.1). We did not use plain correlations, because partial correlations have the advantage of controlling for confounding interactions (e.g., interactions between bacterial species affecting the abundance of a third species).[227] Agreement between the partial correlation matrix and the interaction matrix *A* from the gLV model was assessed qualitatively, i.e., we only considered whether significant entries in the partial correlation matrix agreed with the interaction matrix in terms of nonzero entries with the same sign. We used the Benjamini-Hochberg procedure to control for the expected proportion of 'false discoveries' after calculating partial correlations between each pair of species.[266] The results (true positives, true negatives, false positives, and false negatives) were stored in a confusion matrix (Table 2.1). Because a correlation matrix is symmetric and an interaction matrix *A* is not, we only used half of the partial correlation matrix (Figure 2.1D) to construct the confusion matrix. For a correctly classified interaction, either one or both interactions in the upper and lower part of the *A* matrix must have the same sign as in the lower part of the partial correlation matrix. This can produce a bias, because asymmetric interactions can result in a true positive result for correspondence of the correlation coefficient (ρ) with either interaction. For example, for exploitative interactions, both negative and positive correlations were classified as true positive results. Therefore, we tested the effect of this bias on the success of network inference by specifying the intended sign in correlation analysis, as the sign of the strongest interaction in each pair of species. Hence, for an exploitative interaction, only a positive or a negative correlation is correct, depending on the weights of the asymmetric interactions. We also tested the effect of this bias on the success of network inference by setting the rule that the sign of both interactions must be matched by the inferred correlation coefficient. Therefore, only mutualism and competition can be inferred correctly, as amensalism, commensalism, and exploitative interactions are asymmetric.

Table 2.1. The confusion matrix as used in this study.
The inferred partial correlation coefficient ρ (from the lower part of the partial correlation matrix) must have the same sign as one of the interactions in the interaction matrix A to be considered as a true positive finding in base case analysis.

Interaction in the *A* matrix from the model		Inferred partial correlation		
		Negative[a]	Not significant	Positive[a]
No interaction	0, 0	false positive	true negative	false positive
Mutualism	+, +	false positive	false negative	true positive
Competition	–, –	true positive	false negative	false positive
Commensalism	+, 0 | 0, +	false positive	false negative	true positive
Amensalism	–, 0 | 0, –	true positive	false negative	false positive
Exploitative interaction	+, – | –, +	true positive	false negative	true positive

[a] Only significant partial correlations (with p-value < 0.05) are considered after correction for multiple testing with Benjamini-Hochberg procedure.

Performance of network inference was evaluated using precision and recall, as well as a combination of both measures, called the F_1-score.[267] The precision is the fraction of correctly classified interactions among the total number of significantly predicted interactions (i.e., significant partial correlations) and the recall is the fraction of correctly classified interactions among the total number of non-zero interactions in the interaction matrix A. The F_1-score (on a scale from 0 (no agreement) to 1 (perfect agreement)) is obtained as the harmonic mean of precision and recall, weighted equally, as given in the following equation:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \qquad \textbf{Eq. 2.5}$$

Results

Inference of asymmetric and symmetric interactions in a two-species system

Correlations in abundances of the species in a two-species Lotka-Volterra model are shaped by the type of interaction involved. Figure 2.2 shows scatterplots of the abundances of two bacterial species for different interaction mechanisms over a range of different combinations of α_{12} and α_{21}. Mutualistic interactions clearly yielded a positive correlation in abundance between the two species involved (Figures 2.2A and Appendix Figure 2.5). Competitive interactions generally yielded negative correlations (Figures 2.2B and Appendix Figure 2.5). However, under perfectly symmetric competition (when $\alpha_{12} = \alpha_{21}$) we did find a positive correlation depending on interaction strength and carrying capacities of the species involved (Appendix Figure 2.5D - second panel). In the situation where one of the two species does not experience any benefits or limitations in growth from the other species, as is the case with commensalism and amensalism (i.e., $\alpha_{12} = 0$ or $\alpha_{21} = 0$), correlations are zero because one of the species will grow to its carrying capacity irrespective of the abundance of the other species (Figure 2.2C and 2.2D).

Correlations under exploitative interactions among bacteria, benefitting one but harming the other species, generally yielded positive correlations (Figures 2.2E and 2.2F, and Appendix Figure 2.5), but negative correlations were also found.

This happened when the exploitative benefit was of equal magnitude as the harm done to the other species (Appendix Figure 2.5D), or of similar mean magnitude but with more variation (e.g., species 1 is exploited by species 2; $-\alpha_{12} = \alpha_{21}$ and $\sigma_{\alpha 12} << \sigma_{\alpha 21}$ (exploitative interaction type 1) or species 2 is exploited by species 1; $\alpha_{12} = -\alpha_{21}$ and $\sigma_{\alpha 21} << \sigma_{\alpha 12}$ (exploitative interaction type 2) (Appendix Figure 2.5B). However, if the exploitative benefit outweighs the harm done to the other species, exploitative interactions will generally yield positive correlations. It should also be noted that the two species were not exchangeable, because species 1 was given a weaker intraspecific interaction strength than species 2. Thus, in the absence of interspecific interactions, species 1 can reach a higher abundance at equilibrium. This means that, for the same interspecific interaction strength, the species with the higher carrying capacity exerts a stronger (negative) effect on the growth of the other species.

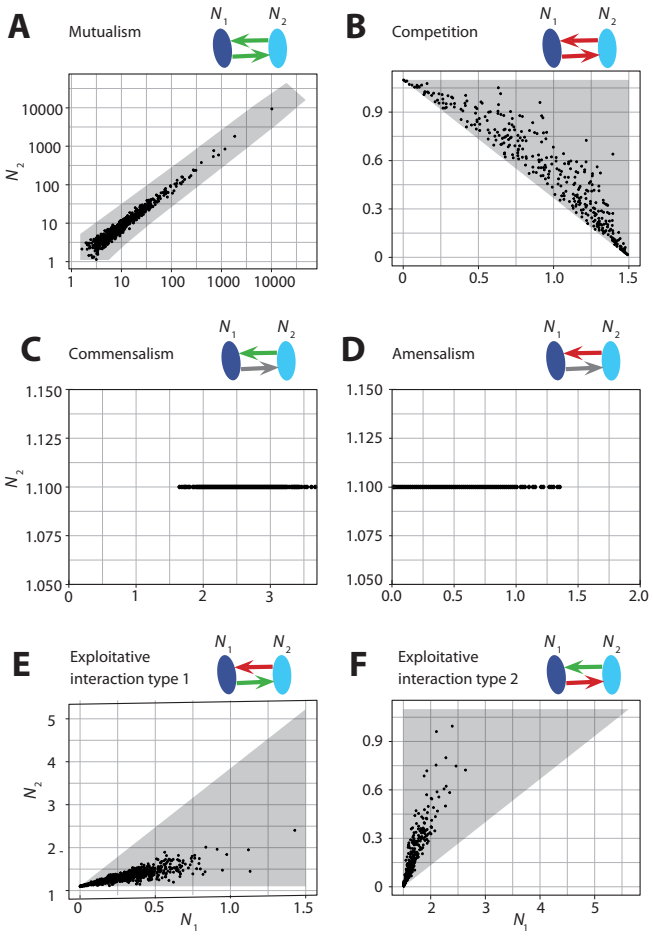


Figure 2.2 - Scatter plots between the abundances of two bacterial species for different interaction mechanisms. A) mutualism, B) competition, C) commensalism, D) amensalism, and E, F) exploitative interactions. The abundances of the two species N_1 and N_2 at equilibrium are shown as scatterplots and have been obtained by running the two-species Lotka-Volterra model, with $K_1 = 1.5$; $K_2 = 1.1$; $r_1 = 1$; $r_2 = 2$ and a_{ij} drawn randomly from normal

distributions with identical means and standard deviations ($\alpha_{12} \sim N(|0.7|, 0.2)$, $\alpha_{21} \sim N(|0.7|, 0.2)$). In the case of commensalism and amensalism: $\alpha_{12} \sim N(|0.7|, 0.2)$ and $\alpha_{21} = 0$. The two species can co-exist under certain combinations of α_{ij} (Box 2.1). The grey polygon indicates the area where coexistence is possible. Note that the axes have different ranges in each subplot. Because the two species have different carrying capacities, the two situations of exploitative interactions are different; i.e., in case of exploitative interaction type 1 (species 1 is exploited by species 2) and in case of exploitative interaction type 2 (species 2 is exploited by species 1).

Network inference under various interaction types

Here, we used the base case model to assess the success rate of recovering a particular interaction type between pairs of species: amensalism, commensalism, exploitative interactions, mutualism, and competition (Appendix Figure 2.1). Figure 2.3A shows that correlations were more often found in mutualistic and competitive interactions, where interacting species experience the same qualitative effects from each other, than in amensalistic and commensalistic interactions, where only one species experiences an effect from the presence of another species. For exploitative interactions among bacteria, either a positive or negative correlation coefficient ρ could be found, with a success rate comparable to amensalistic and commensalistic interactions. Contrary to the results that included symmetric interactions, there was no difference between the successful inference of positive interactions over negative interactions in any interaction type (Figure 2.3B). For all interaction types, the sign of the significant correlation coefficient ρ found, mostly agreed with the sign of the interaction type (Figure 2.3). However, with the inferred correlations neither the type nor direction of the original interaction could be recovered.

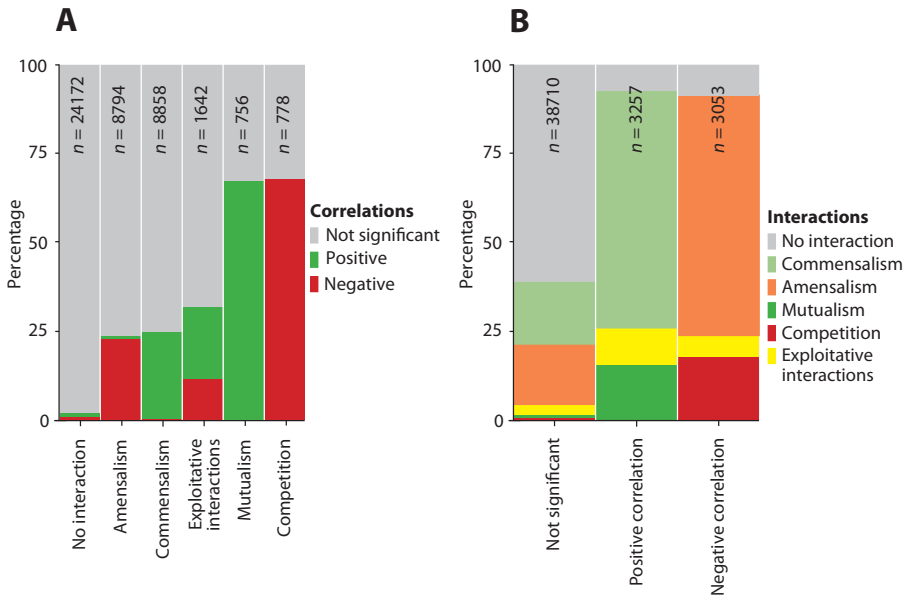


Figure 2.3 - The percentage of significant partial correlations (with sign matching interaction in either direction), as recovered from the base case model.

A) For different types of pairwise interactions and B) for the different correlations.

Network inference under various sources of process variability

Next, we investigated how correct network inference was affected by several variations to the base case model (Figure 2.4 and Appendix Table 2.1). In all cases considered, interactions were recovered with precision exceeding recall. This means that the likelihood of missing an interaction (i.e., 1 – recall) was higher than the likelihood of finding a false interaction (i.e., 1 – precision), illustrating the effect of false discovery rate control.

Partial correlations corresponded to non-zero entries in the interaction matrix only when interindividual variation existed in the interaction parameters (α_{ij}) and/or carrying capacities (K_i) (Figure 2.4A and 2.4B). These parameters directly influence microbial abundance patterns, as interspecific interactions and carrying capacities determine the equilibrium of the gLV model. The intrinsic growth rate only determines the speed at which species reach their equilibrium, and this parameter is not informative for the equilibrium abundances. In fact, performance under interindividual variation in growth rates was just as bad as the performance under pure measurement noise with no variation in model parameters (Figure 2.4B).

Performance of correlation-based network inference was robust to measurement noise, if measurement noise was small compared to interindividual variation in process parameters (Figure 2.4C). When measurement noise became of the same magnitude as the variation in interspecific interactions, the F_1-score deteriorated, and it was no longer possible to use correlations as a proxy for interactions (Figure 2.4C). We also checked whether adding process noise would affect the inference. We did observe a significant improvement of the inference from a model with process noise relative to only measurement noise (Figure 2.4C and Appendix Table 2.1).

Hereafter, we investigated the effect of drawing the interaction strengths from different types of distributions (Figure 2.4D and Appendix Figure 2.3). We did not observe a difference between the success rate of network inference under a Gaussian mixture distribution or uniform distribution, which were conditioned to have similar variances (Appendix Table 2.1). However, successful inference deteriorates with reduced interaction strength; success rates were better under a Gaussian mixture distribution or uniform distribution compared to a unimodal distribution around zero (with smaller variance) (Figure 2.4D). The weaker interactions have a smaller effect on equilibrium abundances of other species, which makes them harder to detect with correlation analysis.

Figure 2.4E shows the results for different network types. Increasing the number of species from 10 to 30 had a significant negative effect on the success of the inference (Appendix Table 2.1), which was mainly due to reduced precision. Conversely, F_1-scores were improved as compared to the base case when assuming a producer-consumer based network (Appendix Figure 2.4 and Appendix Table 2.1), on account of an improved recall. Inference in a network with interaction hubs (as explained in Appendix Figure 2.4) was significantly worse than in a random network, which could be attributed to a somewhat reduced recall.

Note that problems may arise with asymmetric relationships. When using the rule that pairwise correlations should match the strongest interaction between both species involved as the intended sign, we found only a slight non-significant reduction in F_1-score as compared to the base case scenario (Figure 2.4F and Appendix Table 2.1).

Figure 2.4 - Inference under various sources of process variability. For the different scenarios we show the precision, recall, and the F_1-score. A) The base case model. B) Host-specific variation in the carrying capacities and intrinsic growth rates. C) Decreased and increased amount of measurement noise (v) and the effect of process noise (W) (Appendix Figure 2.2). D) Interaction strengths drawn from a uniform and unimodal distribution (Appendix Figure 2.3). E) The results for a 30-species system, a network based on a producer-consumer structure and a network with interaction hubs (Appendix Figure 2.4). F) The effect of network inference when specifying the intended sign in correlation analysis, as the sign of the strongest interaction in each pair of species, or by setting the rule that the sign of both interactions must be matched by the inferred correlation coefficient (strict inference). G) Three scenarios with 3000 hosts, for the base case with random interaction networks as well as for the scenarios with structured (i.e., producer-consumer and hub-species) networks. Network inference was assessed by the F_1-score, which measures agreement between the interaction matrix in the gLV model and the inferred partial correlation matrix on a scale from 0 (no agreement) to 1 (perfect agreement) (according to the rules of Table 2.1). The dashed line indicates the median result from the base case model. The bars of the boxplots indicate the variability of the data outside the middle 50% (i.e., the lower 25% of scores and the upper 25% of scores). All p-values are given in Appendix Table 2.1.

Thus, pairwise interactions wherein the net effect on population growth is positive or negative are mostly picked up as such in correlation analysis. However, under the rule that mutual interactions must both be reflected in the sign of the correlations, asymmetric interactions cannot be recovered as correlations are symmetric. We indeed found much lower F_1-scores when detection of asymmetric interactions was no longer considered as a true positive result after inferring a significant correlation coefficient ρ (either positive or negative) (Figure 2.4F).

Finally, we verified that network inference improved with increasing sample size. This applied to models with random as well as structured interaction networks (Figure 2.4G). In the base case, precision was somewhat reduced at increased sample size notwithstanding Benjamini-Hochberg control. However, this was compensated by substantially improved recall, resulting in significantly increased F_1-scores. Interestingly, precision stayed more or less constant at increased sample size in producer-consumer and hub-species networks, whereas recall improved but remained somewhat behind that of random networks.

Network inference under non-equilibrium conditions

Figure 2.5 shows that the equilibrium assumption is not necessary for successful correlation-based network inference. In fact, our results even suggest that a perturbation can positively affect the performance of network inference. Variation in the growth rates becomes significantly informative outside the equilibrium (Appendix Table 2.2). Also, variation in the interactions becomes even more informative when the population is still growing towards the equilibrium. Network inference is impaired only right after the start of a perturbation, when the population is still far from a new equilibrium, unless the interindividual variation is in the carrying capacities (Figure 2.5B). We also assessed the success of correlation-based inference when the sampling occurred randomly in time in relation to the perturbation. We found that the F_1-score resembled an average of F_1-scores across various sampling time points.

Discussion

Correlation-based network inference has been used in many studies and for many different types of human and environmental microbial communities.[227] The reliability of the results with regards to true ecological dependencies has been criticized, to the extent that correlation analysis has been suggested to almost never reveal anything substantive about the biotic relationships between bacteria.[216] However, the theoretical basis that enables ecological interactions to be inferred from cross-sectional abundance data remains poorly understood. Most of the previous research has focused on the reconstructed network properties or the difficulties pertaining to metagenomics-based abundance patterns, e.g., the compositionality of the data and the high proportion of zeros.[216, 227, 268] While these difficulties are pervasive and merit further consideration, here, we question whether correlations are at all useful in distinguishing different interaction types in microbial networks.

We demonstrated multiple pitfalls when using correlation-based methods for inferring interactions. Some of those pitfalls are well known, as they relate to the inherent symmetry of correlation-based metrics and the frequent asymmetry of ecological interactions.[216] As a result, asymmetric interaction types (commensalism, amensalism, and exploitative interactions) cannot be recovered with an indication of the direction of interaction, which agrees with prior work done by Weiss et al. (2016).[253]

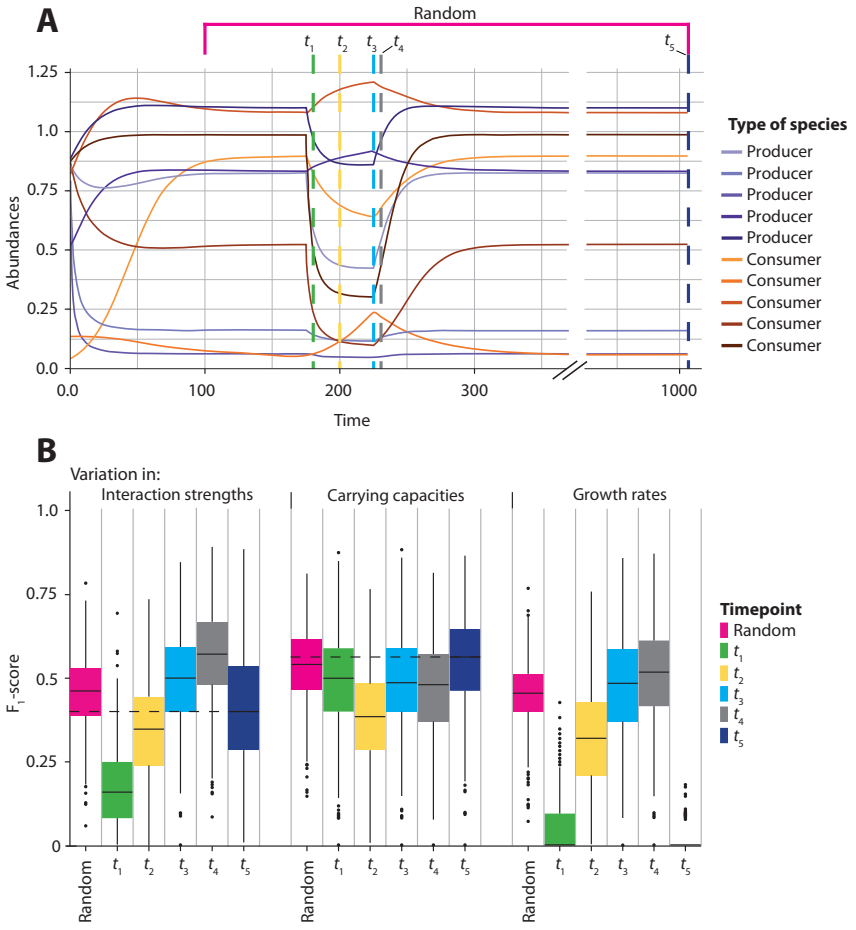


Figure 2.5 - The effect of a perturbation on correlation-based network inference. A) Example of a time series. Dashed lines represent sampling time points. Sampling was performed during the perturbation (t_1 = green, t_2 = yellow, t_3 = blue, and t_4 = grey) and at equilibrium (t_5 = dark blue). Alternatively, sampling was performed randomly between t = 100 and t = 1000 (random = pink). B) Results (F_1-scores) of network inference for sampling at various time points. After a perturbation all species grow back to their original equilibrium. The bars of the boxplots indicate the variability outside the middle 50% (i.e., the lower 25% of scores and the upper 25% of scores). Dashed lines represent median results of sampling during equilibrium. All *p*-values are given in Appendix Table 2.2.

Symmetric interaction types, where species involved affect each other's growth in a qualitatively similar way (competition and mutualism) can be recovered, although competitive interactions may also result in positive correlations, albeit in very rare cases where species have identical competitive strength. Likewise, we found that exploitative interactions generally induce positive correlations, especially in the likely circumstance where the exploitative benefit outweighs the harm to the exploited species. These findings might explain why empirical correlation-based networks have a relative shortage of negative correlations.[119, 268, 269] It remains to be investigated whether the high frequency of positive edges in reconstructed networks is caused by methodologic limitations or whether the interspecific interactions in host-associated microbiota are primarily mutualistic.[112, 136, 270, 271]

Still, as illustrated by our analysis, correlations in microbial abundance across independently sampled hosts can be indicative of underlying ecological interactions under host-specific variation in microbial population dynamics. That is, if microbial groups of interest are omnipresent and their interactions are appropriately captured by generalized Lotka-Volterra (gLV) dynamics, the variation in population abundances should be driven by interindividual variability in population-dynamic parameters. In the context of the gLV model, the informative parameters are primarily related to intrinsic growth rates, carrying capacities, and strength of between-species interactions of microbial groups considered. A change in species abundances can be informative for the interactions among those species, as was also previously shown by Stone and Roberts (1991).[272] It remains to be determined how much variability across individual hosts is driven by external forcing and by gradual differences in process related parameters relative to measurement noise. On one hand, it is well known that microbes adapt to host-specific environments, shaped by, among others, diet, lifestyle, hormonal regulation, and the immune system.[260] As an example, increased abundance of a particular bacterial species at increased glucose intake levels might be reflective of increased resource availability (affecting carrying capacity and growth rate) or superior competitive strength (affecting interactions with other species).[246] On the other hand, environmental drivers of bacterial growth can operate over different spatial and temporal scales and correlations in abundance can be reflective of shared environmental niches that have no meaning in terms of direct biotic interactions.[21]

Therefore, a correlation between the abundance of two species does not imply that those species are interacting.[111] Many of the detected correlations may be caused by shared environmental preferences rather than species interactions.[273] Such environmental filtering can mask putative between-species interactions as well as induce spurious correlations.[216] Also, co-occurring species may appear to be dependent on each other, while their co-occurrence can be explained by them actually sharing a similar dependency on a third species so that co-occurrence, and hence apparent dependencies drawn from that, may also be explained by higher-order interactions.[274] Berry and Widder (2014) suggested that network interpretation is only possible if samples are derived from similar environments.[114] Our analysis suggests that network inference partially depends on a degree of heterogeneity in population-dynamic parameters. If differences in bacterial abundances between hosts are mainly due to measurement noise, their correlations are not informative of underlying interactions. In our simulations, with relative standard deviation in process-related parameters between hosts of about 25%, inference performed well as long as measurement noise had coefficients of variation well below 10% of the mean bacterial abundances. Strikingly, the inference of interactions was even improved when process noise was added. More research is needed to delineate the extent to which correlation analyses can be confounded by latent

environmental drivers of microbial population dynamics, and how strongly one should condition on environmental or host homogeneity.

Our results have been obtained using the gLV model. While the gLV model has been very popular in microbiome research because of its manageability, it has several drawbacks. In ecology, the gLV model has been criticized for the absence of trophic levels within the model.[275] This is in contrast to most classical ecological (e.g., plant-herbivore or predator-prey) systems, where direct consumption and predation offer more opportunity for top-down regulation, possibly obscuring interactions in co-occurrence patterns.[276] But trophic levels are probably not so relevant in the human microbiome as bacteria mainly interact with each other through excreted metabolites.[8] Furthermore, the interactions between bacteria might be much more complex than the additive and pairwise interactions that the gLV model assumes. Momeni et al. (2017) claimed that pairwise modeling will often fail to predict microbial dynamics, as many interactions occur through chemical production pathways (such as cross-feeding and nutrient competition) involving more than two species.[240] Correlation analysis fails to capture the resulting higher-order interactions, for which more advanced techniques, e.g., graphical models, might be more appropriate.[277] It is unclear how well directed links predicted by these methods recover true ecological interaction types. Often, they require more prior knowledge of the network of microbial interactions, time series, or more fine-grained data on the pathways of interaction. Moreover, microbial networks can be bidirected and cyclic,[119] which poses problems for inference of directionality and type of interactions from mere cross-sectional data. More classical methods of separating direct from indirect interactions, e.g., path analysis,[278] rely on testing of specific alternative causal hypotheses, which can only be considered as a next step in network inference. To shed more light on causal pathways, there is a need in microbial ecology for models that can describe the full set of metabolite concentrations, metabolic fluxes, and species abundances within a community.[120] Based on metabolic modeling, Freilich et al. (2011) concluded that cooperative interactions are relatively rare among free-living bacteria and, if present, are often unidirectional. Machado et al. (2021) suggested that mutualistic interactions are much more common among host-associated bacteria, that often form highly cooperative communities and have smaller genomes and fewer metabolic genes compared to other species. Cooperative communities are resilient to nutrient change and adaptable to a wide variety of different environments, including the human body.[119, 274] Metabolic modeling is still challenging and heavily based on a priori assumptions, but is also a rapidly developing field that may prove useful for computational validation of correlation-based interaction networks.[279]

In addition, the gLV model disregards important biological processes, such as adaptation (for instance, switching of mutualistic partners due to for example horizontal gene transfer[280]), that may affect the topology of ecological networks, rather than the strength of ecological interactions in a network. Furthermore, the gLV model displays dynamics that are characterized by strong equilibrium attractors. Many studies have shown the occurrence of complex dynamics as alternative stable states,[105] oscillations, and chaos in experimental,[6, 281, 282] but also in field studies,[6] with ecological communities. Whether this also applies to the bacterial communities inhabiting the human body is still unknown, due to the paucity of long-term human microbiome studies. However, a study among a thousand Western individuals has suggested the existence of tipping elements in the intestinal microbiome[102] indicating the possible presence of alternative attractors in the dynamics of gut microbiome communities.[97, 283]

As a general critique, the use of simulated data based on gLV dynamics raises the question to what extent the necessary model assumptions (and therefore the results) are representative for the human microbiome. Of course, real data are much more complex than simulated data. To reiterate, our base case parametrization does not reflect any particular real-world system, and findings should be appreciated from a qualitative rather than quantitative viewpoint. Even so, while models can only serve as very crude approximations, the main features of model-based analysis might still hold, as demonstrated by Freilich et al. (2018).[273] They compared a well-resolved, empirically defined interaction network of species in the rocky intertidal zone in central Chile to a reconstructed network based on the co-occurrence of those species. There are similarities in their findings to our results. For example, they found that weak interactions are missed more often than interactions above a certain threshold. They also concluded that the ability to correctly detect a true link varies across different interaction types, and that positive interactions are better detected than negative interactions. Interestingly, in line with our results, they also found that negative interactions are misclassified as positive interactions more often than vice versa.

In our simulation studies, the chance of finding false interactions was well under control using partial correlations with adjustment for multiple testing. It should be noted that application of correlation-based network reconstruction to real-world high-throughput microbial abundance data typically requires additional constraints for control of false discovery rates. Real-world microbiome data have some specific challenges which may negatively affect the success of correlation-based network inference. The compositionality of the data, the diversity of species (with many rare species) and the density of interactions make these networks harder to predict and apparent correlations more likely to appear.[229, 250] Various correlation-based methods, often free of charge and provided as pre-programmed packages are available to handle these challenges. However, Weiss et al. (2016) showed that with the same data, there is much disagreement between the inferred networks generated by different tools.[253] Thus, even if correlations are a useful proxy of microbial interactions, performance of network inference in high-dimensional settings will also strongly depend on the specific network modeling approach taken.

To summarize, correlation-based methods are particularly insensitive for the detection of asymmetric interactions (such as exploitative interactions, amensalism, or commensalism), as direction of interaction cannot be recovered from co-occurrence data. Still, they may perform well when applied to networks that are dominated by mutualistic and competitive interactions, as in producer-consumer systems. Applicability of correlation-based network inference to readily available microbiome data thus depends on the type of interactions that govern microbiome dynamics, which likely depends on each application. To conclude, our study suggests that hypotheses about microbial interactions, generated with correlation-based methods, should be questioned with domain-specific knowledge. We highlight again the careful interpretation and validation that is required.

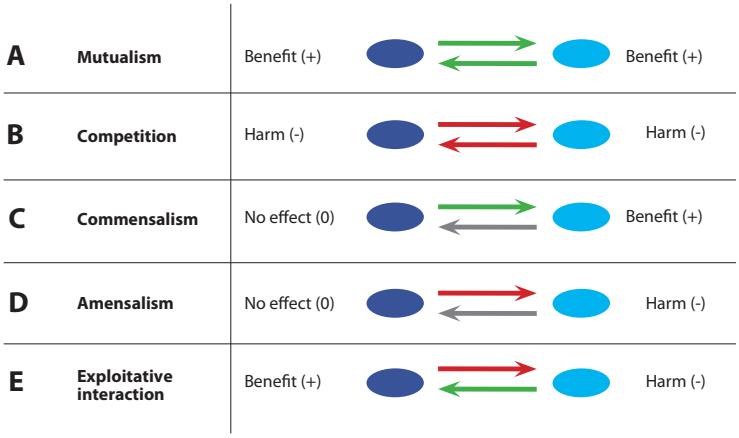
Appendices of Chapter 2

Appendix Table 2.1 - Mann-Whitney U test results for the F_1-scores of the base case model and for the F_1-scores of the model with different sources of process variability.
Significant results are highlighted in bold and blue.

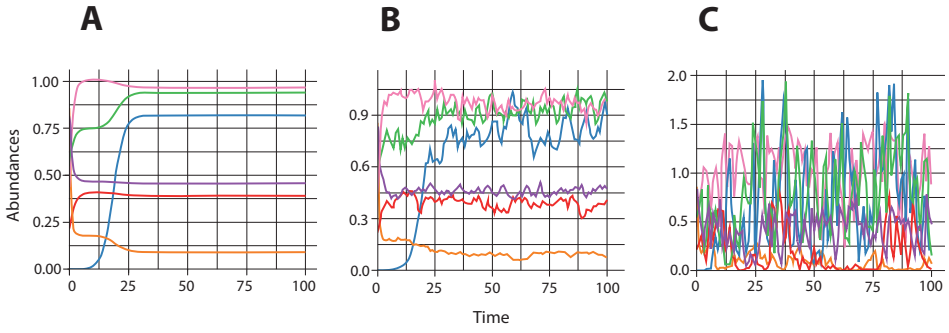
Scenario	*p*-value
Variation in all parameters	**< 0.001**
Variation in carrying capacities	**< 0.001**
Variation in growth rates	**< 0.001**
No variation in parameters	**< 0.001**
Low measurement noise	**< 0.001**
High measurement noise	**< 0.001**
Low process noise	**< 0.001**
High process noise	**< 0.001**
Uniform distribution	> 0.05
Unimodal distribution	**< 0.001**
30 species system	**< 0.001**
Producer-consumer network	**< 0.001**
Hub-species network	**< 0.001**
Inference of intended sign	> 0.05
Strict inference	**< 0.001**
Base case (3000 hosts)	**< 0.001**
Producer-consumer network (3000 hosts)	**< 0.001**
Hub-species network (3000 hosts)	**< 0.001**

Appendix Table 2.2 - Mann-Whitney U test results for the F_1-scores of the samples taken outside equilibrium relative to those taken at equilibrium (t_5 in Figure 2.5) and for the F_1-scores of the samples taken randomly.
Significant results are highlighted in bold and blue.

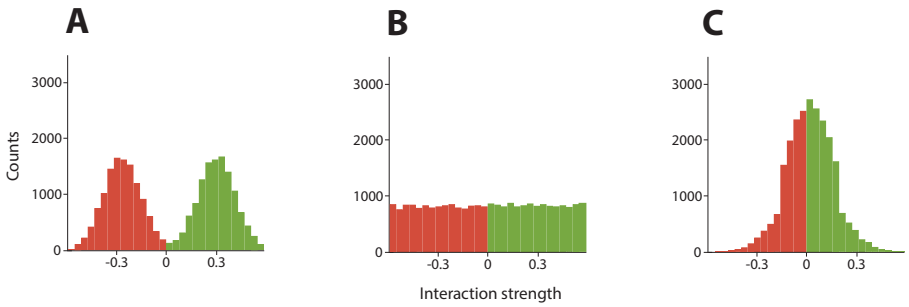
Variation in interactions	
Time point	***p*-value**
Random	**< 0.001**
t_1	**< 0.001**
t_2	**< 0.001**
t_3	**< 0.001**
t_4	**< 0.001**
Variation in carrying capacities	
Time point	***p*-value**
Random	> 0.05
t_1	**< 0.001**
t_2	**< 0.001**
t_3	**< 0.001**
t_4	**< 0.001**
Variation in growth rates	
Time point	***p*-value**
Random	**< 0.001**
t_1	**< 0.001**
t_2	**< 0.001**
t_3	**< 0.001**
t_4	**< 0.001**



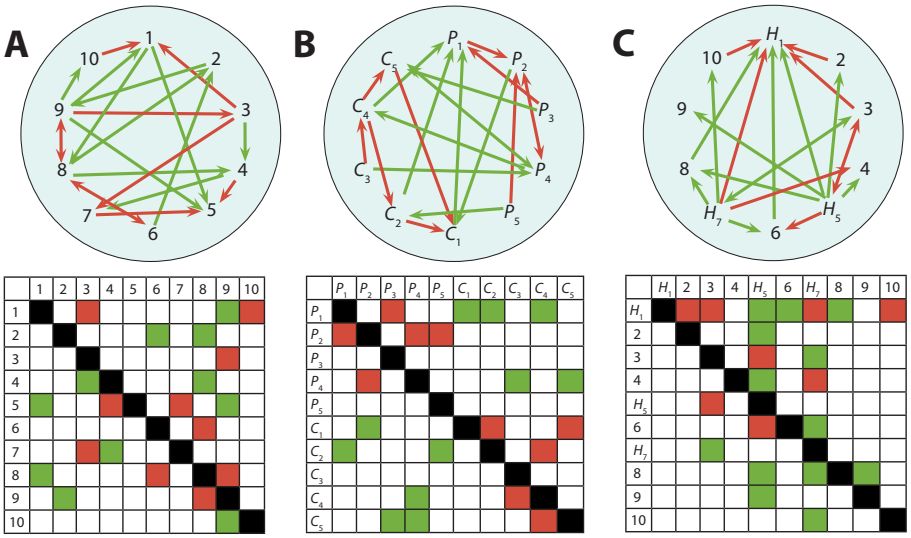
Appendix Figure 2.1 - Cartoon illustrating the different interaction mechanisms. A) In mutualistic interactions, both species experience a positive effect. An example is when a species feeds on the metabolites excreted by the other species. B) In competitive interactions both species experience a negative effect. An example is when both species produce toxic compounds that are harmful to the other species as well as to themselves. C) Commensalism is a one-sided positive interaction. This type of interaction occurs when one species is beneficial to another species, without benefit or harm to itself. D) Amensalism is a one-sided negative interaction. Amensalism occurs when a species causes harm to another species, without benefit or harm to itself. E) Exploitative interactions occur when one species derives a benefit from another species at the expense of the latter, such as when one species kills and subsequently consumes the other. Red arrows represent negative interactions, green arrows represent positive interactions, and grey arrows indicate no interactions.



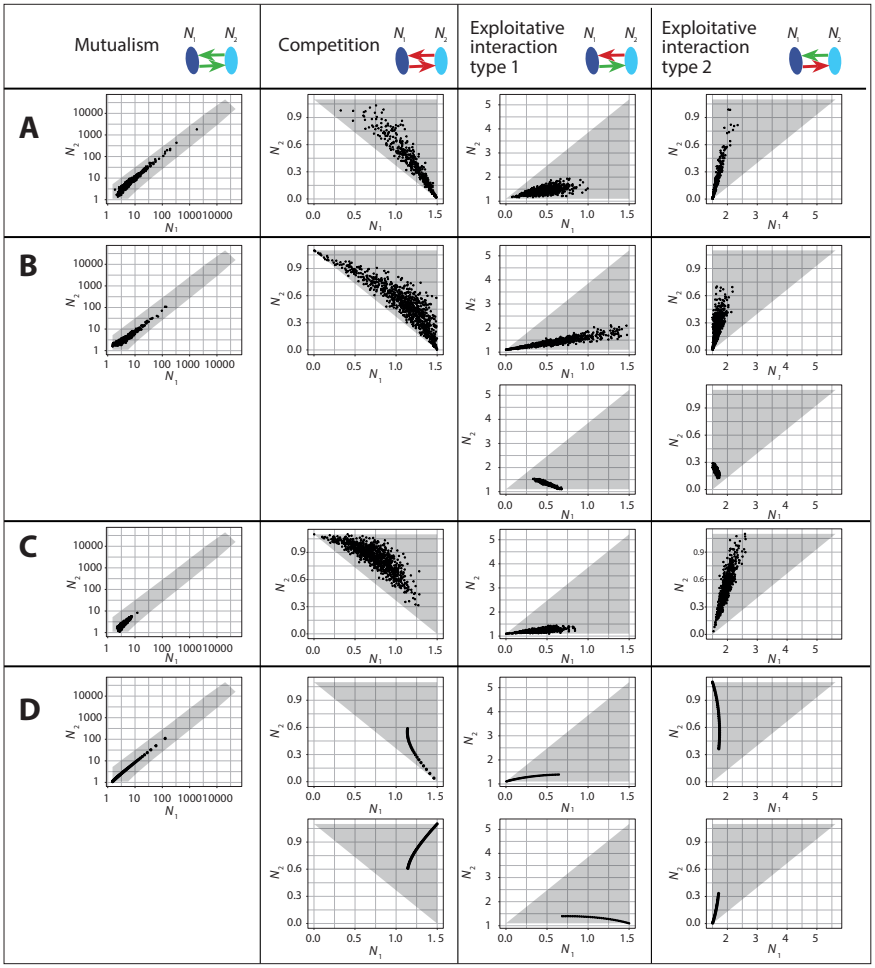
Appendix Figure 2.2 - The effect of process noise (*W*) on the within host population dynamics. Process noise was added by means of the 'events' function from the 'deSolve' R package.[263] The time-varying noise was drawn from a log-normal distribution to prevent the abundances from dropping below zero, i.e., $\Delta W_i = \ln(N_{i,m(\Delta t)}) - \ln(N_{i,m(t)}) \sim N(\ln(N_{i,t}), \sigma_W)$ at every timestep, $\Delta t = 1$. A) Simulated time series without process noise, B) with low process noise ($\sigma_W \sim N(0, 0.1)$), and C) high process noise ($\sigma_W \sim N(0, 1)$).



Appendix Figure 2.3 - Distributions of interaction strengths in three different scenarios. A) The interaction strengths in the base case follow a Gaussian mixture distribution. Half of the interactions were drawn from a negative normal distribution: $\alpha_{ij} \sim N(-0.25, 0.1)$; and the other half of the interactions were drawn from a positive normal distribution: $\alpha_{ij} \sim N(0.25, 0.1)$. B) The interaction strengths in Figure 2.4D-1 follow a uniform distribution ($\alpha_{ij} \sim U(-0.5, 0.5)$). C) The interaction strengths in Figure 2.4D-2 follow a unimodal distribution ($\alpha_{ij} \sim N(0, 0.15)$). All interactions were restricted to lie between –0.5 and 0.5, i.e., the normal distributions were truncated at –0.5 and 0.5.



Appendix Figure 2.4 - Network structures used in the different case studies. A) An example of a random network and its corresponding interaction matrix. B) An example of a structured network with interaction modules and its corresponding interaction matrix. The modular networks are based on a cross-feeding structure between producers and consumers (with equal numbers of producers and consumers). Between producers (P_i ; i = 1:5) and consumers (C_j ; j = 1:5), positive interactions (indicated in green) are more likely to occur, because metabolites excreted by the producers are consumed by the consumer species. Among producers or among consumers, the interactions are predominantly negative (indicated in red) as these species are more likely to compete for similar resources. C) An example of a structured network with interaction hubs and its corresponding interaction matrix. The hub-species network contains species (H_i) with unusually high numbers of ecological interactions compared to other species in the network. This can occur when some species perform a central role in the microbial ecosystem, for example when a hub-species produces a metabolite that is required for growth by many other species.



Appendix Figure 2.5 - The effect of a_{ij} on the correlations between the abundances of two bacterial species for different interaction mechanisms.

The two species can co-exist under certain combinations of a_{ij} (Box 2.1). The abundances of the two species N_1 and N_2 at equilibrium are shown as scatterplots and were obtained by running the two-species Lotka-Volterra model with $K_1 = 1.5$, $K_2 = 1.1$, $r_1 = 1$, $r_2 = 2$, and a_{ij} drawn randomly from normal distributions with different combinations of means and standard deviations. In A) the two distributions have different means and standard deviations: $a_{12} \sim N(|0.5|, 0.1)$ and $a_{21} \sim N(|0.7|, 0.2)$. In B) the distributions have identical means, but different standard deviations: $a_{12} \sim N(|0.5|, 0.2)$ and $a_{21} \sim N(|0.5|, 0.1)$. For exploitative interactions we also show the situations that negative correlations can occur when the exploitative benefit displays much more variation than the harm to the other species, i.e., $a_{12} \sim N(-0.5, 0.01)$ and $a_{21} \sim N(0.5, 0.2)$ for exploitative interaction type 1, and $a_{12} \sim N(0.5, 0.2)$ and $a_{21} \sim N(-0.5, 0.01)$ for exploitative interaction type 2. In C) interactions are randomly drawn from distributions with different means and identical standard deviations: $a_{12} \sim N(|0.6|, 0.1)$ and $a_{21} \sim N(|0.3|, 0.1)$. In D) the interactions have identical strengths for the two species, namely $|a_{12}| = |a_{21}|$. The mutualistic interactions are drawn from the distribution $a_{12} = a_{21} \sim U(0, 2.5)$, for competition and exploitative interactions we show two different scenarios, namely $|a_{12}| = |a_{21}| \sim U(|0.4|, |2.5|)$ (upper graph) and $|a_{12}| = |a_{21}| \sim U(0, |0.4|)$ (lower graph). Because the two species have different carrying capacities, the two situations of exploitative interactions are different. The grey polygon indicates the area where coexistence is possible. Note that the ranges of the axes are different in each subplot.

Chapter 3

Wavelet clustering analysis for characterizing community structure

Elisa Benincà[1], Susanne Pinto[2], Bernard Cazelles[3, 4, 5], Susana Fuentes[1], Sudarshan Shetty[1, 6], Johannes A. Bogaards[7, 8]

1 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
2 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
3 UMMISCO, Sorbonne Université, Paris, France
4 INRAE, Université Paris-Saclay, MaIAGE, Jouy-en-Josas, France
5 Eco-Evolution Mathématique, IBENS, UMR 8197, CNRS, Ecole Normale Supérieure, Paris, France
6 Department of Medical Microbiology and Infection prevention, Virology and Immunology research Group, University Medical Center Groningen, the Netherlands
7 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
8 Amsterdam Institute for Infection and Immunity (AI&I), Amsterdam UMC, Amsterdam, the Netherlands

Sci Rep. 2023 May 17;13(1):8042. doi: 10.1038/s41598-023-34713-8.

Wavelet clustering analysis as a tool for characterizing community structure in the human microbiota

Abstract

Human microbiota research is helped by the characterization of microbial networks, as these may reveal key microbes that can be targeted for beneficial health effects. Prevailing methods of microbial network characterization are based on measures of association, often applied to limited sampling points in time. Here, we demonstrate the potential of wavelet clustering, a technique that clusters time series based on similarities in their spectral characteristics. We illustrate this technique with synthetic time series and apply wavelet clustering to densely sampled human gut microbiota time series. We compare our results with hierarchical clustering based on temporal correlations in abundance, within and across individuals, and show that the cluster trees obtained using either method are significantly different in terms of elements clustered together, branching structure, and total branch length. By capitalizing on the dynamic nature of the human microbiota, wavelet clustering reveals community structures that remain obscured in correlation-based methods.

Introduction

The human microbiota is the collective of microbial communities living on the various surfaces of the human body. These communities consist of microorganisms which do not live in isolation but interact with each other and with their human host.[252, 284] In the past decade, thanks to advances in sequencing techniques and data analyses, an increasing number of studies have attempted to gain ecological insights from microbiota abundance data, e.g., by reconstructing networks of interacting species with the nodes representing the microorganisms and the edges representing the dependencies between them.[285]

Most of the studies that aim to reconstruct the network of interacting species are based on measures of co-occurrence, e.g., using correlations between pairs of species as proxies of between-species dependencies.[8, 250, 286] Despite the popularity of such methods in microbiota studies,[230, 250, 251] their usefulness in describing community structure is still a matter of debate.[114, 229, 287] While these co-occurrence studies are often performed on a relatively large number of individuals, they are limited to one or a few sampling points in time, presenting a mere snapshot of the dynamic microbiota. Other methods infer the ecological network by fitting an a priori chosen population-dynamic model to time series data of the microbial community.[100, 254, 288] These methods have the limitation that the inferred community structures strongly rely upon the assumptions that are intrinsic to the chosen model, and require considerable prior knowledge of the community of interest. There are also examples where the ecological interactions are inferred from repeated measurements around steady states.[255] This circumvents the need for a priori specification of a population dynamic model but makes the implicit assumption that the microbial system tends towards a stable equilibrium.

However, many experimental and field studies have shown the presence of complex dynamics in ecological communities, such as alternative stable states,[2, 97, 105] oscillations, and chaos,[5, 6, 281, 282] questioning the steady states assumptions for the human microbiota. These dynamics are driven by a complex interplay between intrinsic factors (e.g., interaction mechanisms between organisms such as competition, mutualism, and parasitism) and external perturbations (e.g., environmental conditions and interventions).[6, 289, 290] Complex dynamics are also likely to occur in the human microbiota, because the bacterial communities living in our body are characterized by a plethora of interactions[291] and are also affected by external perturbations (e.g., diet, use of antibiotics, and travel patterns).[59, 292, 293] A study with a thousand healthy western individuals suggested the existence of tipping elements in the intestinal microbiota,[102] reflecting the presence of alternative attractors and the possibility of more complex microbiota dynamics. The presence of complex dynamics in the human microbiota has not yet been demonstrated, probably due to the paucity of long and dense time series of the human microbiota. However, the study with one of the longest time series of human microbiota measurements available shows strong variability in the abundance of the bacteria over time, indicating that the human microbiota might not be at the presumed steady state.[46]

To advance our ecological understanding of the human microbiota, methodology is needed that can exploit the temporal information in microbiota time series data without a priori knowledge of data generating mechanisms or steady-state assumptions. In the last decade, many methods have been developed to model the abundances of compositionally sampled data with the purpose of either fitting or predicting the temporal dynamics of the microbiota communities.[294-296] Here, we perform wavelet clustering analysis, a technique that clusters time series based on similarities in their periodical patterns.[297] This technique, which is commonly applied in climate and engineering studies,[298] more recently gained popularity in ecological,[290] and epidemiological studies.[299-301] Wavelet clustering analysis has only recently been applied to time series derived from 16S rRNA gene amplicon data to reveal coastal plankton community structure,[302] but, to our knowledge, our study is the first application to human gut microbiota data. The novelty of the wavelet clustering approach, relative to prevailing co-occurrence or time series methodologies in human microbiota research, is that it is able to characterize community structure on the basis of collective temporal behaviour of the microbiota, without directly fitting a dynamic model or reconstructing the network of interacting species.

We illustrate wavelet clustering first with synthetic time series and then with densely sampled time series of human gut microbiota data from a male and female subject.[46] For both examples, we compare our results with clustering obtained on the basis of correlations in bacterial abundances over time. Our results show that correlation-based clustering is significantly different from clustering using wavelets. Wavelet clustering uncovered more diverse community structures and retained more of the differences between the male and the female subject compared to methods using temporal correlation. The results of this work highlight how the choice of method determines the type of communities found in microbiota data analysis. This is particularly important, considering that most of the putative microbiota communities, and their associations with a particular disease state or physical host condition, strongly rely on prevailing correlation-based methods or steady-state assumptions. Our results suggest that wavelet clustering readily capitalizes on the dynamic nature of the human microbiota and reveals more diverse community structures than those based on temporal correlations or associations.

Methods

Wavelet analysis

Wavelet analysis makes use of a periodic function (the mother-wavelet). The relative importance of periodicities (wavelet power) is then plotted in contour plots as a function of time (wavelet power spectra). Here, we use as mother-wavelet the Morlet wavelet, which is particularly suited for detecting periodicities.[298, 303] Significance of the detected periodicities is assessed using a Markov surrogate significance test.[304] Statistical significance is assessed by testing against the null hypothesis that observed periodicities are identical to those generated by a stochastic Markov process, characterized by the same mean, the same variance, the same distribution of values and the same short-term autocorrelation structure. More detailed information on wavelet analysis is provided elsewhere.[6, 305-307]

Wavelet clustering

The wavelet spectra are compared using a procedure based on the maximum covariance analysis.[297] To be more precise, as described in Rouyer, Fromentin et al. (2008), the distance matrix is computed based on leading patterns and singular vectors obtained using matrix decomposition analysis.[297] Matrix decomposition analysis relies on a singular value decomposition performed on the covariance matrix between two wavelet power spectra. This enables construction of a distance matrix based on the wavelet power spectra. Only periodicities with a confidence higher than 90% have been considered in the computation of the dissimilarity matrix. Wavelet analysis and wavelet clustering were performed using wavelet software written in Matlab which is available at Bernard Cazelles' research page (www.biologie.ens.fr/~cazelles/bernard/Research.html).[297]

Comparison among cluster trees

We quantified similarities between cluster trees using the B_k statistic (i.e., Fowlkes-Mallows index).[308] The B_k statistic measures the degree of similarity between two hierarchical clusters. Consider two hierarchical trees C_1 and C_2, each with the same number of elements n and partition each tree to produce $k = 2, \ldots, n-1$ subclusters for each tree. For each value of k we can compute the quantity $m_{i,j}$ which quantifies the number of objects in common between the i^{th} cluster of C_1, and the j^{th} cluster of C_2. The statistic B_k is then defined:

$$B_k = \frac{T_k}{\sqrt{P_k Q_k}} \qquad \textbf{Eq. 3.1}$$

where:

$$T_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{i,j}^2 - n \qquad \textbf{Eq. 3.2}$$

$$P_k = \sum_{i=1}^{k} \left(\sum_{j=1}^{k} m_{i,j}\right)^2 - n \qquad \textbf{Eq. 3.3}$$

$$Q_k = \sum_{j=1}^{k} \left(\sum_{i=1}^{k} m_{i,j}\right)^2 - n \qquad \textbf{Eq. 3.4}$$

B_k is calculated for all the k partitions and B_k takes values between 0 and 1; $B_k = 1$ indicates that k subclusters in each tree correspond completely whereas $B_k = 0$ indicates that the subclusters in each tree don't correspond at all.

Details on the B_k statistic are described in Fowlkes et al. (1983).[308] The B_k statistic has been calculated using the 'dendextend' R package.[309] The computed values of B_k are then plotted as a function of k. The significance of the B_k values is tested against the null hypothesis that the two cluster trees are not related. A one-sided rejection line (with significance level of 5%) is drawn based on the asymptotic distribution of B_k values, for each k, under the null hypothesis of no relation between the clusters.

Calculation of total branch length

The total branch length was calculated by summing the lengths of connecting segments in the tree using the 'treeheight' function of the 'vegan' R package.[310]

Microbiota data

In our analysis, we used previously published time series of the gut microbiota of two healthy subjects, one male and one female, on which fecal samples have been taken for 15 and 6 months, respectively.[46] The V4 variable region of the 16S rRNA gene was amplified by PCR and sequenced on an Illumina Genome Analyzer IIx. In the original paper of Caporaso et al. (2011)[46] the raw sequences were clustered in Operational Taxonomic Units (OTU) using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline. However, recent studies have shown that the use of OTUs is more prone to produce noisy features which are artifacts of sequencing errors.[208] Nowadays, the use of Amplicon Sequence Variants (ASV) data has been shown to be more reliable than OTU's.[208]

Following the same line, here we used the ASV gut microbiota data of Caporaso et al. (2011) which is available at the Earth Microbiota Project (EMP) platform (earthmicrobiome.org).[46] The ASV data provided at the EMP platform have been generated from the raw sequence data with the Deblur pipeline[311] and the detailed protocol is provided in Thompson et al. (2017).[312] The data for human microbiota time series was obtained from 'emp_deblur_150bp.release1.biom' by filtering to keep only samples from the Qiita study ID 5501.

We removed singletons and ASV sequences assigned to mitochondria and chloroplasts. We assembled the taxa at the genus level and this yielded 578 unique genera. For both the male and female subject, we first removed samples with less than 500 reads, then we transformed the time series to relative abundances and then we made a selection of genera, using a bootstrapping method[313] with a prevalence value of 25% and a relative abundance threshold value of 0.005 (i.e., select the genera in which the relative abundance has a value higher than 0.005 in at least 25% of the samples). We disregarded the taxa that were not identified as uniquely defined genera. This yielded a total of 19 genera for the male subject and of 12 genera for the female subject. The aim of our analysis is to compare clusters (and techniques to obtain these clusters) among the two different subjects. Therefore, we considered in our analysis the genera that were present in at least one subject, yielding a total of 19 genera for each subject. Processing of the data from ASV to the core-microbiota taxa was done using the 'phyloseq'[314] and 'microbiota'[313] R packages. Subsequently, we applied a centered log-ratio (CLR) transformation to the relative abundance time series using the 'compositions' R package.[315] The CLR transformed time series of the selected genera are shown in Figure 3.2. Wavelet analysis requires equidistance between subsequent datapoints, therefore we interpolated the time series of both subjects using cubic Hermite interpolation to obtain data with equidistant time intervals of 1.6 days (the mean time interval of the original data of the male subject is 1.6 days and the female subject is 1.5 days), yielding a total of 336 data points for the male subject and of 131 data points for the female subject.

Before performing wavelet analysis to the data, the microbiota CLR transformed time series were rescaled using a Box-Cox transformation to suppress sharp peaks, homogenize the variance and approximate a normal distribution. For each time series the optimal parameter of the Box-Cox transformation has been estimated by optimizing the normal probability plot correlation coefficient using the 'EnvStats' R package (see Appendix Figure 3.1).[316]

Results

Wavelet cluster analysis

Wavelet analysis enables investigation of time series characterized by different periodicities and is particularly suited for time series which are not stationary, as applies to many biological systems. We first illustrate this technique using synthetic time series (Figure 3.1A left hand side). Consider for instance time series 1 and 2: they are stationary and oscillate at the same periodicity of eight days, but in antiphase. They are therefore characterized by the same wavelet spectrum: a significant period of eight days (orange area inside the black dotted line) occurring along the entire time span of 100 days. The average wavelet spectrum, which is an estimation of the classical Fourier spectrum, is also identical among the two time series (see plot at the far most right-hand side). If one considers time series 7 and 8, one may see that they are showing opposite patterns. Time series 7 oscillates fast at a periodicity of about four days in the first 50 days and then slows down and oscillates at a periodicity of about 20 days in the second half of the time series. Time series 8 is doing exactly the opposite, it oscillates slowly with a periodicity of about 20 days in the first half of the time series and then oscillates with a periodicity of about four days in the second half of the time series. While the average wavelet spectrum is identical for both time series, the wavelet spectra are showing opposite patterns and are therefore able to depict the differences between the temporal behaviour in the oscillations of the two time series (Figure 3.1A).

The wavelet spectra are then compared using a procedure based on maximum covariance analysis which enables construction of a distance matrix based on the wavelet power spectra.[297] The constructed distance matrix is used to build a cluster tree based on the WARD agglomeration criterion (Figure 3.1B).[317] For comparison, we also constructed a Spearman dissimilarity matrix calculated as $d = 1 - \rho$ (where ρ is the correlation coefficient), using all data points in the time series pairs. The Spearman dissimilarity matrix is also used to construct a cluster tree based on the WARD agglomeration criterion (Figure 3.1C). We compare the wavelet clustering with a clustering based on Spearman's correlation, because the latter is a common method used in microbiota studies to infer relationships between microorganisms.[253] One may immediately observe substantial differences between the trees obtained with the two different methods (Figure 3.1B and 3.1C). The time series are clustered differently within the trees according to the two methods, but also branching structure and the total length of the branches is noticeably different.

Time series 1 and 2 are close together in the wavelet cluster tree (Figure 3.1B), but they fall apart in the Spearman cluster tree (Figure 3.1C). The first results from the fact that the two time series have identical wavelet spectra, which indicates that the time series oscillate at the same periodicity. However, they are considered dissimilar in correlation analysis, because the time series are in antiphase (i.e., the peaks of one time series coincide with the troughs of the other time series and vice versa).

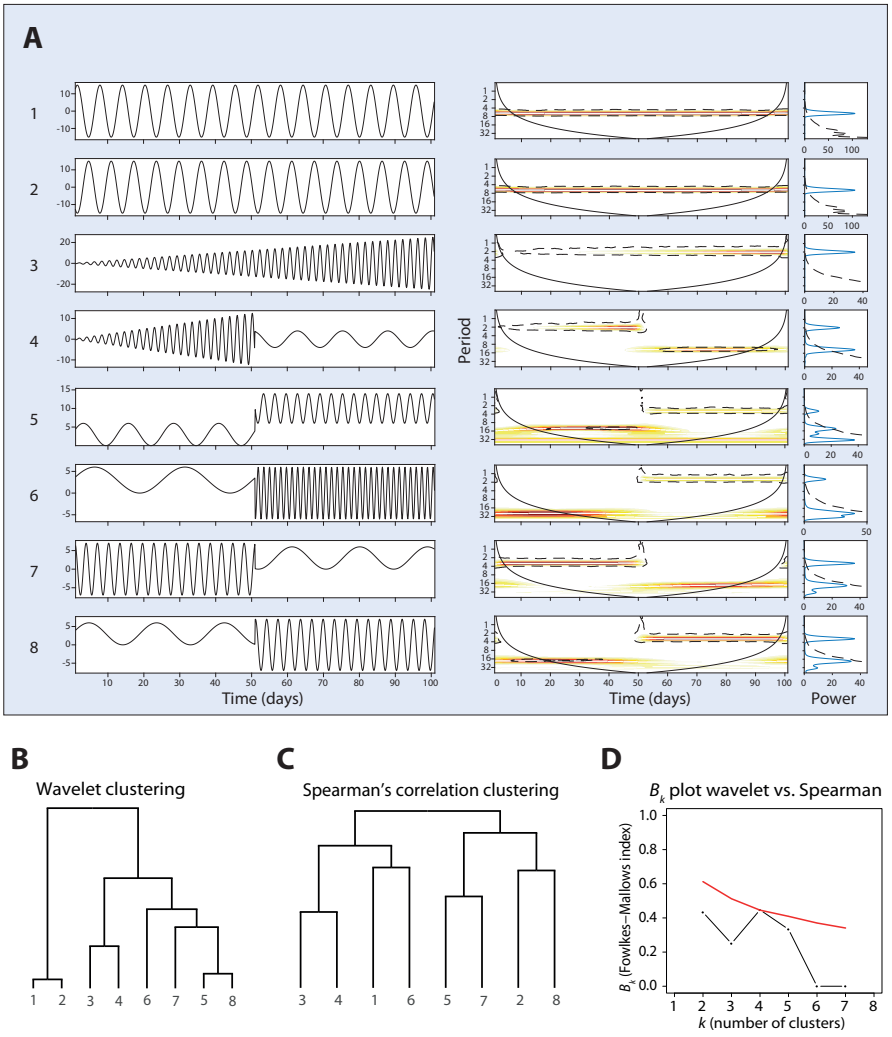


Figure 3.1 - Illustration of wavelet clustering analysis with synthetic time series. A) Wavelet analysis of synthetic time series: synthetic time series (left hand side) characterized by different periodicities; wavelet spectra (right hand side) and average wavelet spectra (far right) of the synthetic time series. Color codes represent wavelet power and range from low (white) to high (red). Black dotted lines enclose the 5% significance areas computed using a Markov surrogate significance test. The solid black line delimits the cone of influence, where edge effects become important. Clustering of the synthetic time series based on two methods. In B), clustering is based on the wavelet spectra. The cluster tree is constructed by grouping the time-frequency patterns of the time series using maximum covariance analysis. In C), clustering is based on Spearman's correlations calculated for each pair of time series. The correlations are used to compute the dissimilarity matrix which is used to cluster the data. For both methods the hierarchical clustering of the time series is performed using the WARD agglomeration criterion. D) Comparison of the hierarchical clusters obtained using the B_k statistic.[308] Black dots represent the B_k values plotted against the k number of clusters in which each tree has been partitioned. Red line represents the one-sided rejection region based on the asymptotic distribution of B_k values, for each k, under the null hypothesis of no relation between the clusters (significance $\alpha = 5\%$).

Similarly, time series 5 and 8 cluster together in the wavelet tree but they fall apart in the Spearman cluster tree. Both time series 5 and 8 oscillate slowly at a periodicity of about 13 and 20 days, respectively, in the first part of the time series but then oscillate faster (at a periodicity of about four days) in the second part of the time series. Therefore, their wavelet spectra are very similar.

If the synthetic time series would represent the dynamical behaviour of microorganisms, one would conclude from the Spearman cluster tree that microorganisms 1 and 2 (or 5 and 8) are not or only weakly related, because when one microorganism is highly abundant then the other one has very low abundance (and the other way around). The wavelet clustering instead shows that these microorganisms are strongly connected because they oscillate with similar periodicities and therefore share the same dynamical properties, which may point to ecological interdependence e.g., through parasitic interactions or neutral niche competition.

In addition to visual inspection, we used the B_k statistic to quantify the similarity in cluster trees constructed with the two methods.[308] The B_k statistic assesses the chance-corrected proportion of items that two cluster trees have in common, as a function of the number of subclusters k that the two trees are partitioned into. Plotting B_k versus k gives a quantitative representation of the similarity between two cluster trees (black dots in Figure 3.1D). The red line represents the 95% rejection region under the null hypothesis of no relation between the trees. For all partitions k, the blacks dots fall below the red line, hence we cannot conclude that the trees calculated with the wavelets and the Spearman's correlations for the synthetic time series are significantly related.

In Box 3.1 we give an additional demonstration of wavelet clustering analysis applied to the outputs of an ecological model of four consumers and four resources. In this case, wavelet clustering accurately captures the competitive coupled dynamics between consumers and resources, whereas clustering based on Spearman's correlation does not (Figure 3.A - panels D and E in Box 3.1).

Application to human microbiota data

We tested our approach, as illustrated for the synthetic time series, on real data of microbiota communities. We used previously published gut microbiota time series of two healthy subjects, one male and one female, from whom fecal samples had been collected for 15 and 6 months, respectively.[46] We considered the data at genus level and we selected the same 19 genera for the male and the female subject. A detailed description of the data and of the selection criterion is provided in the methods.

Time series of the selected genera for the male and the female subject are shown in Figure 3.2. CLR transformed relative abundances over time show remarkable fluctuations. Some genera (e.g., *Lachnospira* and *Roseburia* in the male subject; *Bacteroides* in both subjects) show a clear wax and wane in their dynamical pattern. There are other genera (e.g., *Campylobacter* and *Finegoldia* in the female subject) that show more spiky dynamics, dominated by low CLR transformed relative abundances, but with few very high peaks.

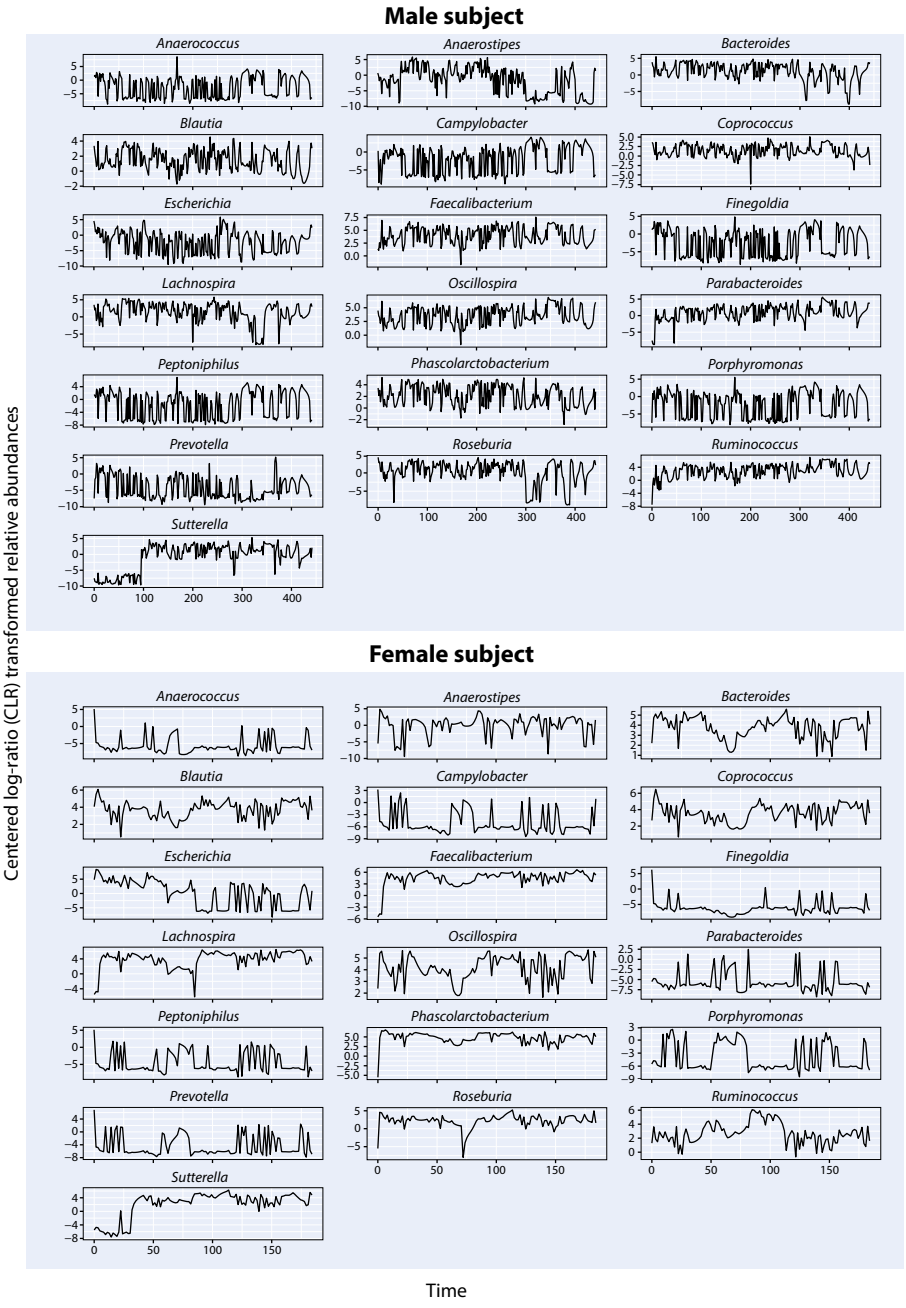


Figure 3.2 - Gut microbiota time series of CLR transformed relative abundances for selected genera. Male (upper graphs) and female (lower graphs) subject. The time series show clear fluctuations. Note the distinct time axes in the male and the female subject.

Box 3.1 - Wavelet clustering applied to the dynamics of four consumers feeding on four resources. In this section we give an extra demonstration of the potential of wavelet clustering by performing the analysis on the outputs of a simplified ecological model describing the dynamics of four consumers and four resources. The model is a modified version of the previously published model of Vandermeer of two species feeding on two resources.[318-320]

The model reads as follows:

$$\frac{dC_i}{dt} = \frac{aR_iC_i}{1+bR_i} - mC_i \qquad \textbf{Eq. 3.A}$$

$$\frac{dR_i}{dt} = r_iR_i\left(\frac{K - R_i - \alpha_{ij}R_j}{K}\right) - aR_i\left(\frac{C_i}{1+bR_i}\right) \qquad \textbf{Eq. 3.B}$$

$$\frac{dC_k}{dt} = \frac{aR_kC_k}{1+bR_k} - mC_k \qquad \textbf{Eq. 3.C}$$

$$\frac{dR_k}{dt} = r_kR_k\left(\frac{K - R_k - \alpha_{kl}R_l}{K}\right) - aR_k\left(\frac{C_k}{1+bR_k}\right) \qquad \textbf{Eq. 3.D}$$

for $i = 1, 2$ and $k = 3, 4$ and $i \neq j$ and $k \neq l$, where C_i and C_k are the abundances or densities of the i^{th} and the k^{th} consumers, respectively, and R_i and R_k denote those of the i^{th} and the k^{th} resources. The parameters r_i and r_k represent the intrinsic growth rates of the i^{th} and the k^{th} resource, respectively. m is the mortality rate of the consumers, α_{ij} is the competition coefficient between resource 1 and 2, α_{kl} is the competition coefficient between resource 3 and 4, a is the resource consumption rate, b is the functional response parameter (with higher values denoting diminished response in consumer growth at a given resource abundance), and K is the carrying capacity of each resource, which we assume for simplicity to be the same for all four resources.

The model consists of two separated food webs of two consumers each feeding on one resource (Figure 3.A - panel A). Consumer C_1 feeds on resource R_1, consumer C_2 feeds on resource R_2 and the two resources R_1 and R_2 negatively interact with a parameter α_{12}. Similarly, consumer C_3 feeds on resource R_3, consumer C_4 feeds on resource R_4, and the two resources R_3 and R_4 negatively interact with a parameter α_{34}. In Figure 3.A (panel B left hand site) are shown the temporal dynamics of the four consumers and the four resources. We applied wavelet analysis to all eight of the time series (Figure 3.A - panel B right hand side) and we used this information to build the cluster tree (Figure 3.A - panel C). Wavelet clustering identifies two big subclusters: subcluster 1 with consumers C_3 and C_4 and resources R_3 and R_4, and subcluster 2 with consumers C_1 and C_2 and resources R_1 and R_2. Wavelet clustering successfully identifies the two separated food webs. In addition, inside each cluster we observe that each consumer is clustered together with its own resource (C_1 with R_1, C_2 with R_2, C_3 with R_3, and C_4 with R_4). For comparison we build a tree based on Spearman's correlation (Figure 3.A - panel D). In contrast to wavelet clustering, clustering based on Spearman's correlation is not able to identify neither the two distinct food webs, neither the pairs of consumers-resources. Clustering based on Spearman's correlation is substantially different from clustering based on wavelets as it is shown by the corresponding B_k plot (Figure 3.A - panel E).

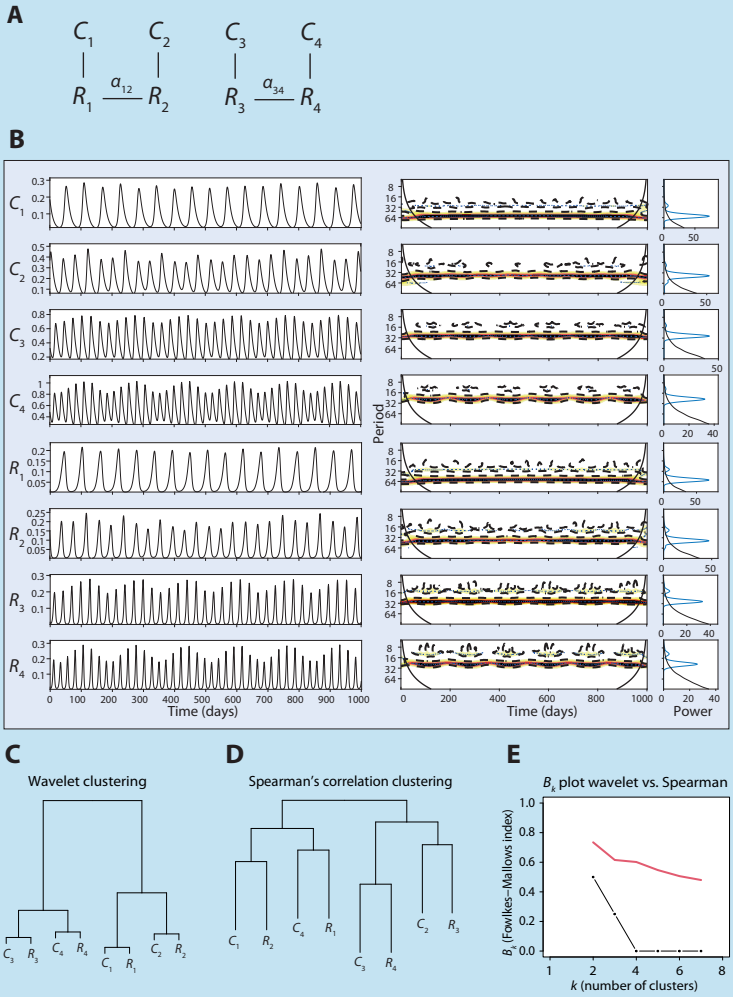


Figure 3.A - Application of wavelet clustering to the outputs of a model with four consumers feeding on four resources. A) The model consists of two separated food webs of two consumers on two resources. The two resources within each food web negatively interact with a competition coefficient a. B) (Left) Outputs of the resources-consumers model. Simulations have been run for 2000 time units. The plots shown here covers the last 1000 time units of the simulation. Parameters: $K = 1$; $a = 2$; $b = 1.3$; $m = 0.1$; $r_1 = 0.2$; $r_2 = 0.4$; $r_3 = 0.8$; $r_4 = 1.2$; $a_{12} = 0.8$; $a_{34} = 0.4$; (Right) Wavelet spectra and average wavelet spectra (far right) of the model outputs. Color codes represent wavelet power and range from low (white) to high (red). Black dotted lines enclose the 5% significance areas computed using a Markov surrogate significance test. The solid black line delimits the cone of influence, where edge effects become important. C) Clustering based on the wavelet spectra. The cluster tree is constructed by grouping the time-frequency patterns of the time series using maximum covariance analysis. D) Clustering based on Spearman's correlations calculated for each pair of time series. The correlations are used to compute the dissimilarity matrix which is used to cluster the data. For both methods the hierarchical clustering of the time series is performed using the WARD agglomeration criterion. E) Comparison of the hierarchical clusters obtained using the B_k statistics. Black dots represent the B_k values plotted against the k number of clusters in which each tree has been partitioned. Red line represents the one-sided rejection region based on the asymptotic distribution of B_k values, for each k, under the null hypothesis of no relation between the clusters (significance $\alpha = 5\%$).

To capture possible similarities in the dynamical patterns of the bacteria, we applied wavelet analysis to each of the bacterial time series in both subjects. Wavelet spectra detected several significant periodicities in the fluctuations of bacteria both for the male (Figure 3.3) and the female subject (Figure 3.4). A first visual inspection of the spectra already reveals similarities between the dynamical patterns of the bacteria. For instance, in the male subject (Figure 3.3), *Porphyromonas*, *Phascolarctobacterium*, and *Peptoniphilus* show common periodicities of about 30–40 days co-occurring for approximately 100 days at the end of the time series. In addition, *Campylobacter* and *Roseburia* clearly show common periodicities of 64 days occurring approximately in the last 150 days of the time series, whereas *Blautia* and *Coprococcus* share this periodicity at the beginning of the time series. Common patterns are less clear in the female subject (Figure 3.4), though some similar periodicities can be identified. For instance, many genera show the same periodicity of about 60 days occurring along the entire length of the time series.

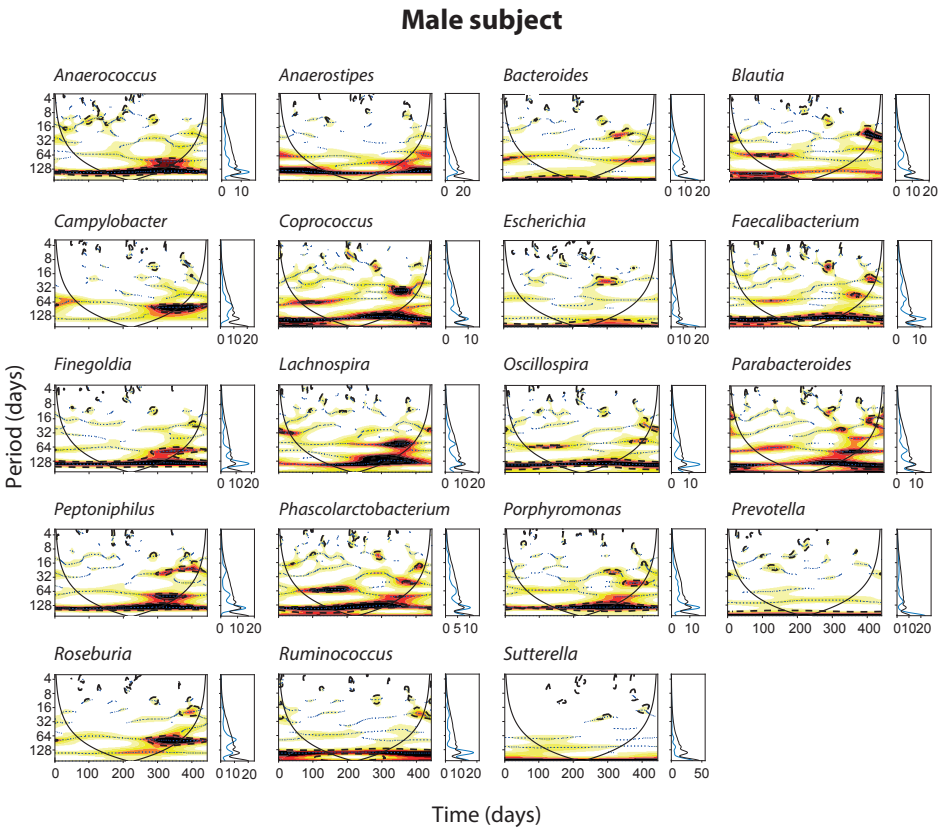


Figure 3.3 - Wavelet analysis of time series for selected genera in the male subject. For each genus the wavelet spectrum (left) and the average wavelet spectrum (right) are computed. Color codes represent wavelet power and range from low (white) to high (red). Black dotted lines enclose the 5% significance areas computed using a Markov surrogate significance test. The solid black line delimits the cone of influence, where edge effects become important.

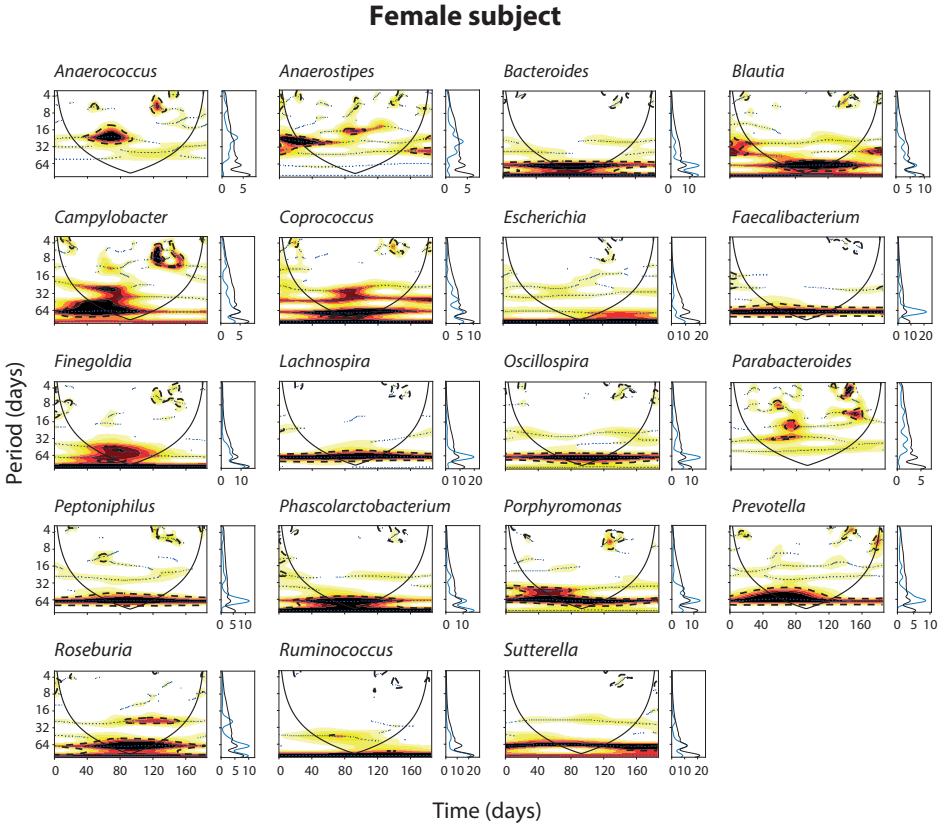


Figure 3.4 - Wavelet analysis of time series for selected genera in the female subject. For each genus the wavelet spectrum (left) and the average wavelet spectrum (right) are computed. Color codes represent wavelet power and range from low (white) to high (red). Black dotted lines enclose the 5% significance areas computed using a Markov surrogate significance test. The solid black line delimits the cone of influence, where edge effects become important.

With the wavelet spectra at hand, we built trees based on the wavelet distance matrix as described for the synthetic time series. Both the clusters based on wavelet spectra for the male and the female subject show a clear partition in two subgroups (Figure 3.5A and 3.5B). The clusters based on Spearman's correlations for the male and the female subjects are also characterized by two main subclusters (Figure 3.5C and 3.5D). Although there are few bacteria that are clustered together with both methods (i.e., *Peptoniphilus*, *Finegoldia*, *Porphyromonas*, and *Anaerococcus* in the male subject), the two methods yield very different clusters. For instance, *Bacteroides* and *Prevotella* are clustered together in the male subject with the wavelet method, but they are in two different clusters in the male subject with the correlation method. The case of *Prevotella* and *Bacteroides* resembles the example of signals 1 and 2 (or 5 and 8) illustrated before: two time series with similar dynamical properties are clustered together based on wavelets but are considered not related by the correlation method.

Also, visual comparison of the clusters obtained using wavelets (Figure 3.5A and 3.5B) with the clusters obtained by pairwise correlations (Figure 3.5C and 3.5D) reveals substantial differences between the two methods in the positioning of branches within the two subclusters and in the total length of the branches.

Of note, total branch length (see 'Methods') is substantially higher in the wavelet cluster tree as compared to the tree based on Spearman's correlations (male subject: 80.9 vs. 27.6; female subject: 70.0 vs. 21.9). Further visual comparison of the trees based on wavelets among the two subjects also reveals that the members of each subcluster are substantially different between the male and the female subject (compare Figure 3.5A with Figure 3.5B). In contrast, comparison of the cluster trees based on correlations shows that many bacteria that are clustered together in the male subject are also clustered together in the female subject (compare Figure 3.5C with Figure 3.5D).

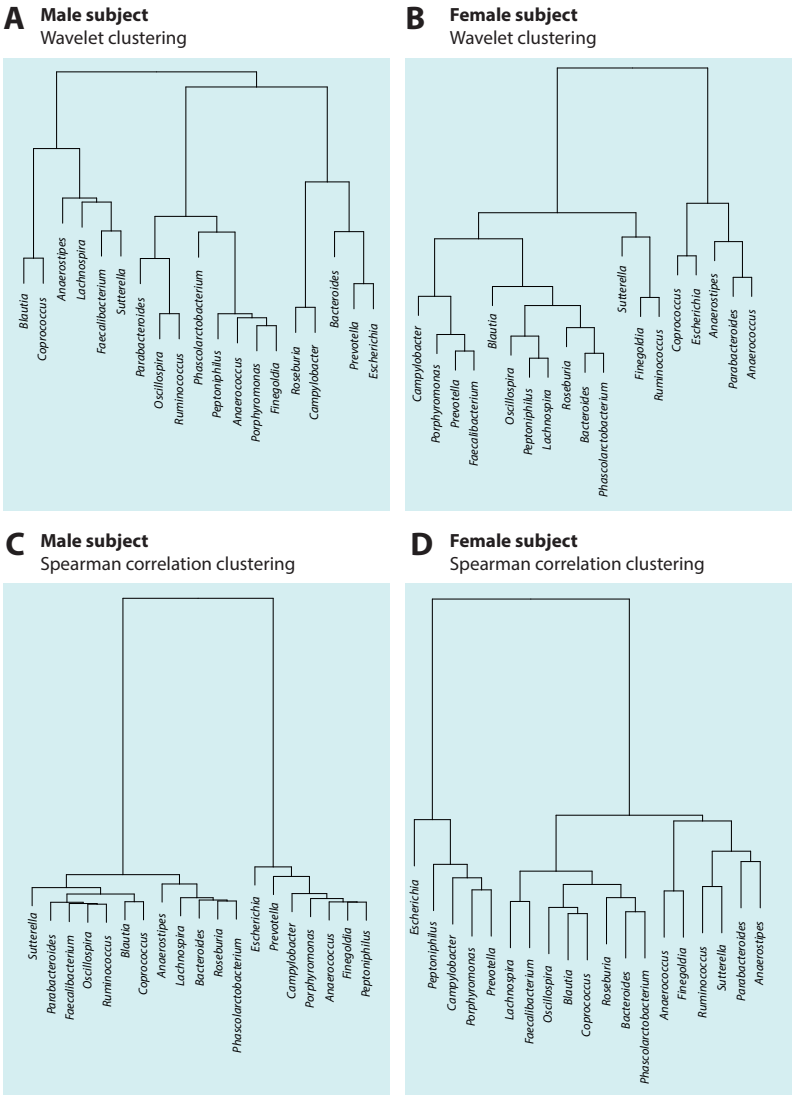


Figure 3.5 - Clustering for the male and female subjects based on different methods. Cluster tree obtained using the dissimilarity matrix obtained from the wavelet clustering analysis for A) the male subject and B) the female subject. Cluster tree obtained using the dissimilarity matrix obtained from the Spearman's correlation matrix for C) the male subject and D) the female subject.

To further quantify the similarities between subjects and methods we calculated the B_k statistic as we did for the synthetic time series. For low values of k, the dots in Figure 3.6A and 3.6C fall below the 95% rejection line. Thus, wavelet clustering and Spearman's clustering are not significantly related when the community is partitioned into a limited number of subclusters, and this holds for both the male and female subject. This is likely because the wavelet clustering method accounts for other features (i.e., the spectral characteristics of the bacterial dynamics and their time evolution) than the correlation-based methods, which only consider quantities averaged over the whole series. For higher values of k, the dots sometimes fall above the rejection line (Figure 3.6A and 3.6C), meaning that wavelet clustering and Spearman's clustering get significantly related at some higher resolution when certain subclusters become apparent. For comparison (Figure 3.6B and 3.6D) we also applied the B_k statistic to correlation-based trees constructed with the Spearman's correlation and with the Pearson's correlation coefficient (trees not shown). For all k partitions (except the maximum partition for the male subject), the trees calculated with these two correlation methods are instead, as it could be expected, significantly related.

Finally, we also assessed the similarity between the two subjects. Interestingly, we found no evidence for related wavelet clusters between the male and female subjects, as all dots fall below the 95% rejection line irrespective the number of k partitions (Figure 3.6E). In contrast, in the B_k plot of the Spearman's correlation-based clustering, the majority of dots fall above the 95% rejection line (Figure 3.6F), indicating significantly related clusters for almost all subpartitions between the male and female subject. This suggests that wavelet clustering not only uncovers more diverse community structures within individuals, but might also be more sensitive towards subtle differences in community structures across individuals.

Discussion

Developments in high-throughput sequencing have improved our ability to track the temporal variability of microbial communities. This has led to an increase in longitudinal data from a variety of different microbiota ranging from wastewater,[321] marine,[322-324] freshwater,[325] and terrestrial[326, 327] environments. These time series offer unprecedented opportunities to gain ecological insights into microbial community dynamics and the mechanisms governing them, and to track the response of the microbial systems to external perturbations.

Ideally, long time series are required to capture the periodic patterns of microbial dynamics and reveal community structures. Unfortunately, only few of such datasets exist in human microbiota studies.[46, 59, 328, 329] This probably reflects the relative difficulty to repeatedly sample the human microbiota in comparison to a natural field habitat (e.g., sampling strongly relies on the consent of the host to provide sampling material at a regular basis). As a result, the majority of studies on human microbial community structures have relied on sparse data and methods based on co-occurrence, which may have produced biased associations, e.g., towards positive correlations.[253, 268, 269] Clearly, there is a need to shift from a static to a dynamical approach, that takes into account the temporal development of bacterial communities and can shed new light on microbial community structure.[50] This also has bearing on the ability to employ microbiota data for clinical practice, as more and more studies move from association to prediction of disease course, e.g., exacerbation of inflammatory bowel disease (IBD),[330] and treatment response in *Clostridioides difficile* infection.[190]

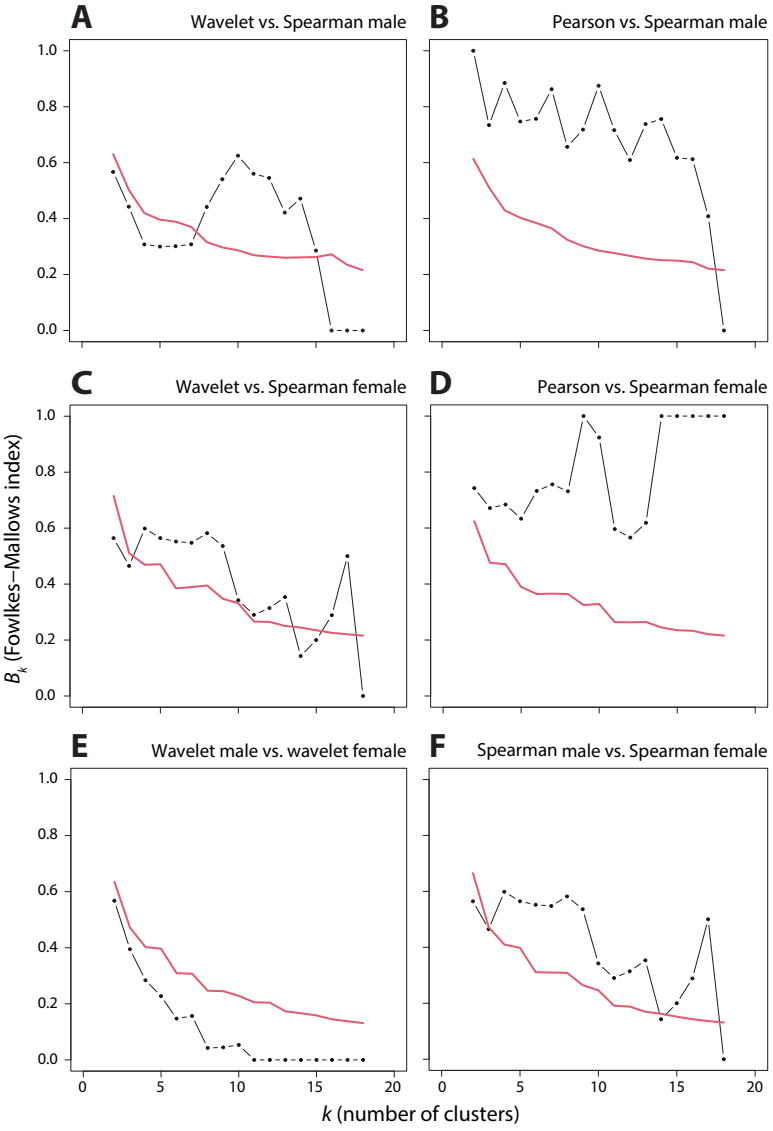


Figure 3.6 - Comparison of hierarchical clusters using the B_k statistic.[308]
Black dots represent the B_k values plotted against the k number of clusters in which the tree has been partitioned. Red line represents the one-sided rejection region based on the asymptotic distribution of B_k values, for each k, under the null hypothesis of no relation between the clusters (significance $\alpha = 5\%$). A) Comparison of the tree based on wavelets and the tree based on Spearman's correlations for the male subject. B) Comparison of the tree based on Pearson's correlations and the tree based on Spearman's correlations for the male subject. C) Comparison of the tree based on wavelets and the tree based on Spearman's correlations for the female subject. D) Comparison of the tree based on Pearson's correlations and the tree based on Spearman's correlations for the female subject. E) Comparison of the trees based on wavelets for the male and female subject. F) Comparison of the trees based on Spearman's correlations for the male and female subject.

Interestingly, our reanalysis of the widely used Caporaso et al. (2011) data reveals some novel important patterns.[46] The trees obtained with the two different methods show significant differences in the way microbial genera are clustered together. For instance, there are cases where pairs of bacteria are clustered together in the male and female subject when using correlations, but not when using wavelets. For example, according to wavelet analysis *Blautia* and *Coprococcus* only cluster together in the male subject, and *Phascolarctobacterium*, *Roseburia*, and *Bacteroides* only in the female subject, whereas these genera are clustered together in both subjects with the correlation-based method. In general, similarity of the cluster trees between subjects seems to be stronger with the correlation-based method than with wavelet clustering, for which we found no evidence for significant relations between the male and female trees. Tree correspondence according to clustering method within subjects was more ambiguous, as similarity also depends on tree resolution. This emphasizes how sensitive the clustering is to the type of method chosen.

In addition, we also note differences in the pattern of branching and in the total branch length of the cluster trees. Studies have shown that the total length of the branches in a traits tree is indicative of the functional diversity in ecosystems.[331] Analogously, total branch length can here be considered as an indicator of the diversity of community structure. While we are not considering functional traits here, we could speculate that the higher total length observed in the wavelet clustering of the microbiota time series is indicative of a higher diversity in community structure as compared to the correlation-based method. A likely explanation is that wavelet analysis is able to detect dependencies that are not apparent in correlations, whereas the reverse is not the case: highly correlated time series are still detectable in wavelet spectra. Thus, wavelet clustering can extract more information on the dependencies within microbial communities than is reflected in mere correlations.

Looking at the clusters identified by the wavelet method one can speculate about possible interaction mechanisms between the bacteria. For instance, in the male subject, two genera are observed together, *Blautia* and *Coprococcus*. Members of genus *Blautia* are known to produce acetate and lactate which is shown to support improved growth of *Coprococcus in vitro*.[332] *Coprococcus* bacteria can convert lactate and acetate to butyrate, a short chain fatty acid that is associated with a healthy microbiota.[333] This mutualistic mechanism could potentially lead to similar dynamical patterns and explain why these bacteria co-occur in the same cluster. Although these 'potential' interaction mechanisms are based on associative dynamical patterns of 16S rRNA gene sequence data they may provide ground for further investigation of these interactions *in vitro* and *in vivo*. In addition wavelet cluster analysis can be used as a starting point for investigation for time series causality inference methods such as Granger causality[334, 335] or convergence-cross mapping.[336, 337] For instance, there are methods that are able to estimate Granger's causality from wavelet spectra of time series data.[338, 339] Application to a complex system such as the microbiota has not yet been done and can be subject of investigation in future studies.

In ecological and epidemiological studies, wavelet analysis is often used to evaluate the effect of external factors, such as climatic or meteorological variables, on species or disease dynamics. Examples include studies which evaluate the effect of external factors on the spread of dengue fever,[340] malaria,[341] and cholera,[342] or on the dynamics of communities of benthic organisms,[6] marine[343] and freshwater plankton,[344] or fish.[290, 345]

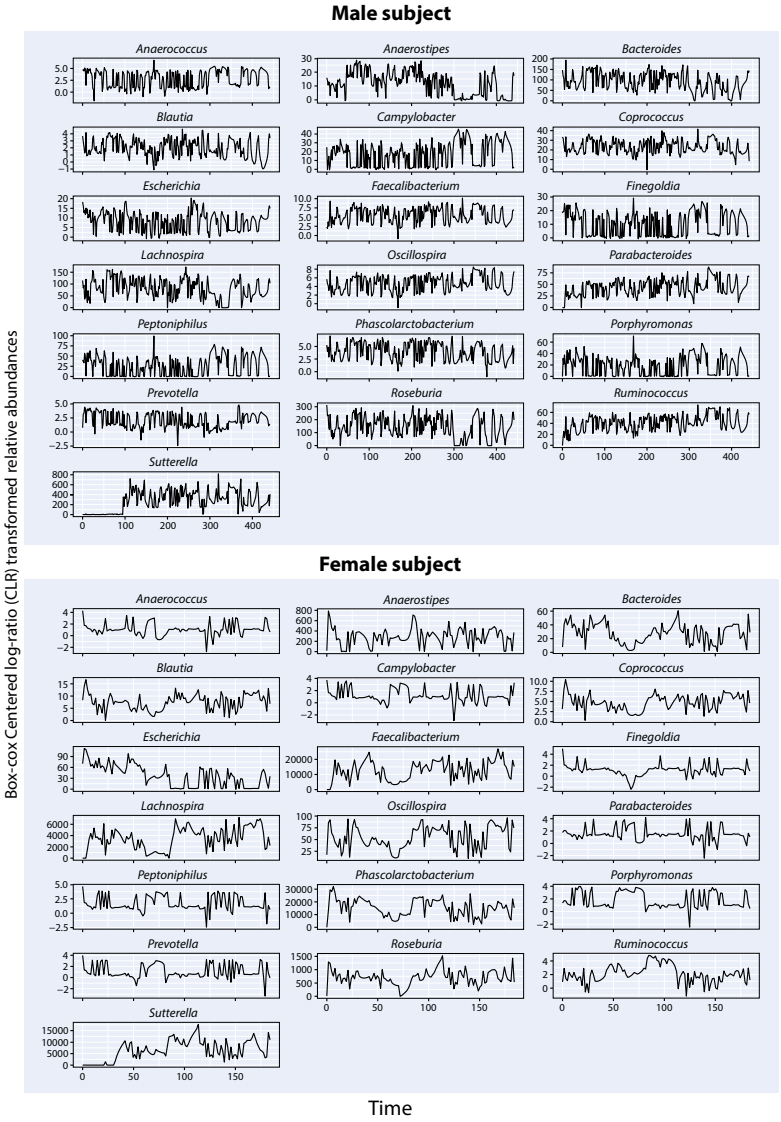
In an analogous way, when longitudinal studies on human microbiota dynamics become more widely available, metadata can be exploited using wavelet analysis to evaluate the effect of interventions, as for instance vaccination, the use of antimicrobials or probiotics, fecal microbiota transplantation, and cancer treatment.

The reader interested in using the wavelet clustering approach might wonder how many points are needed for applying such an analysis. The limits in the number of data points for wavelet analysis are similar to those of Fourier analysis and depend on the periodic components that one wants to highlight. For instance, Murdoch et al. (2002)[346] suggest that with a minimum time series length of 25 time units one can identify periodicities between two time units (the Nyquist frequency) and 8–10 time units. Cazelles et al. (2012)[300] are more conservative and they suggest time series with a minimum length of 30–40 time units which allows detection of a maximum periodicity equal to 20–25% of the total length of the time series. Another practical aspect is that wavelet analysis requires equidistant data. Although this might appear as a limiting factor, this requirement can easily be addressed. For instance, when possible, an experiment or a sampling strategy could be designed in such a way to obtain equidistant sampling points. If this is not possible, there are interpolation methods that can be used to obtain equidistant data. Different interpolation methods should be tested, and the interpolated data should be checked against the original data to see if the general dynamical behaviour is unaffected by the interpolation. This is the approach taken in this study. In addition, as for Fourier analysis, there are extensions of wavelet analysis that can be applied to non-equidistant data.[347-351]

In our study we analysed the time series of two individuals, and we compared the wavelet dendrograms of the two subjects using a pairwise metric. Ideally, new longitudinal human microbiota studies will track the joint dynamics of much more than two individuals. When time series of multiple subjects become available, one might want to compare dendrograms among classes of individuals (e.g., individuals of the same gender or patients versus healthy controls). Instead of a pairwise metric between individuals, our analysis could then be applied to consensus dendrograms between classes of individuals to assess how communities differ with respect to the condition of interest.[352]

To summarize, wavelet cluster analysis has the big advantage of accounting for non-stationary dynamics which are often preponderant in biological systems. In addition, we show that it appears to be a sensitive method for recovering microbial community structure from densely sampled microbiota time series. By taking into account the spectral features of bacterial abundance and their time evolution that are ignored in methods focusing on co-occurrence at any one time point, wavelet clustering analysis is able to extract more information on the dependencies within microbial communities, and to uncover more diverse communities within and across individuals than conventional methods. The results show that interpretation of microbial networks and communities, inferred on the basis of only a few sampling points in time, should be done with care, and be compared to alternatives.

Appendix of Chapter 3



Appendix Figure 3.1. Box-cox transformed CLR time series of selected genera in the male (upper graphs) and the female (lower graphs) subject. The relative abundance time series of both subjects have been interpolated using cubic Hermite interpolation to obtain data with equidistant time intervals of 1.6 days (the mean time interval of the original data of the male subject is 1.6 days and the female subject is 1.5 days), yielding a total of 336 data points for the male subject and of 131 data points for the female subject. Subsequently, we applied a CLR transformation to the relative abundance time series using the 'compositions' R package.[315] Before performing wavelet analysis on the data, the microbiota CLR transformed time series were rescaled using a Box-Cox transformation to suppress sharp peaks, homogenize the variance and approximate a normal distribution. For each time series the optimal parameter of the Box-Cox transformation has been estimated by optimizing the normal probability plot correlation coefficient using the 'EnvStats' R package.[316]

Part II

Gut microbiota and inflammatory bowel disease

Chapter 4

Heterogeneous associations of gut microbiota with Crohn's disease activity

Susanne Pinto[1], Elisa Benincà[2], Gianluca Galazzo[3, 4], Daisy M. A. E. Jonkers[3, 5], John Penders[3, 4], Johannes A. Bogaards[6, 7]

1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
2 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
3 School for Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University, Maastricht, the Netherlands
4 Department of Medical Microbiology, Infectious Diseases and Infection Prevention, Maastricht UMC, Maastricht, the Netherlands
5 Department of Gastroenterology-Hepatology, Maastricht UMC, Maastricht, the Netherlands
6 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
7 Amsterdam Institute for Infection and Immunity (AI&I), Amsterdam UMC, Amsterdam, the Netherlands

Gut Microbes. 2024 Jan-Dec;16(1):2292239. doi: 10.1080/19490976.2023.2292239.

Heterogeneous associations of gut microbiota with Crohn's disease activity

Abstract

The multifactorial involvement of gut microbiota in Crohn's disease (CD) necessitates robust analysis to uncover potential associations with specific microbes. CD has been linked to certain bacteria, but reported associations vary widely across studies. This inconsistency may result from heterogeneous associations across individual patients, resulting in no apparent or only weak relationships with the means of bacterial abundances. We investigated the relationship between bacterial relative abundances and disease activity in a longitudinal cohort of CD patients (n = 57) and healthy controls (n = 15). We applied quantile regression, a statistical technique that allows investigation of possible relationships outside the mean response. We found several significant and mostly negative associations with CD, especially in lower quantiles of relative abundance on family or genus level. Associations found by quantile regression deviated from the mean response in relative abundances of Coriobacteriaceae, Pasteurellaceae, Peptostreptococcaceae, Prevotellaceae, and Ruminococcaceae. For the family Streptococcaceae we found a significant elevation in relative abundance for patients experiencing an exacerbation relative to those who remained without self-reported symptoms or measurable inflammation. Our analysis suggests that specific bacterial families are related to CD and exacerbation, but associations vary between patients due to heterogeneity in disease course, medication history, therapy response, gut microbiota composition, and historical contingency. Our study underscores that microbial diversity is reduced in the gut of CD patients, but suggests that the process of diversity loss is rather irregular with respect to specific taxonomic groups. This novel insight may advance our ecological understanding of this complex disease.

Introduction

Crohn's disease (CD) is a chronic inflammatory disorder that can affect any part of the digestive tract, but mostly involves the ileum and colon.[353] The disease is characterized by periods of inflammation (exacerbation) interspersed by periods without symptoms (remission). During exacerbation, the patients are suffering from a range of different symptoms, including diarrhea, abdominal pain, bloody stool, fatigue, and weight loss. Prolonged inflammation can lead to severe complications, such as damage to the gastrointestinal tract and malnutrition.[353] While the exact cause of CD is unknown, an inappropriate immune response against commensal gut bacteria, host genetics, and environmental factors are all thought to be involved in disease pathophysiology.[354] The gut microbiota in CD patients is characterized by a reduced diversity and lower long-term stability as compared to healthy individuals.[355] Also, shifts in abundance of specific bacterial genera or families have been associated with CD,[356] its disease course,[177] and disease activity.[181]

Several studies have investigated relations between specific microbial groups and CD. *Faecalibacterium prausnitzii* (Ruminococcaceae), *Clostridium leptum* (Clostridiaceae), and

Clostridium coccoides (Clostridiaceae) were found to be negatively associated with CD as well as disease activity.[181, 356, 357] Conversely, the family Enterobacteriaceae was found to be positively associated with CD and with disease activity.[171, 358] However, the patterns of association with specific microbes are not always consistent among studies. Within the Bacteroidaceae family conflicting results were found. For example, CD patients showed both lower relative abundances,[356, 358] as well as higher relative abundances[357] in *Bacteroides* (Bacteroidaceae) compared to healthy individuals.

The inconsistency in findings might be partly due to technical artifacts, such as differences between studies in sequencing methods to quantify gut microbiota composition and the compositional nature of data obtained by most next generation sequencing (NGS) techniques. Another explanation is that the heterogeneous responses among patients may derive from multifactorial dependencies, between microbial elements themselves and between gut microbiota and host factors, such as treatment with immunomodulatory drugs, lifestyle, and diet, but also in underlying disease characteristics such as disease location, severity, and epigenetic immune regulation.[22, 359] This heterogeneity among patients is reflected by a strong variation in disease course, the response to medication, and the need for surgery among subgroups of patients.[360] The involvement of specific bacterial groups in CD will likewise depend on multiple factors. Some of these factors can be accounted for when relating CD to gut microbiota composition, although correction relies on adequate model specification which is difficult in multi-factorial systems. Moreover, many factors which may strongly determine the observed relationships between bacterial abundance and CD activity have not been identified or are not routinely measured. One such factor is the order in which specific bacteria have been acquired throughout life. Rapid colonization by maternal and environmental bacteria occurs within days of birth and is unique per person. The temporal development of the microbiota is directed, implying that the growth of certain species precedes the growth of others, leading to the unique microbiomes in adult life. This historical contingency of gut microbiota might also influence how microbes react to future perturbations in that gut community.[67]

The multi-factorial involvement of specific microbial groups with CD necessitates robust analysis to uncover possible associations, as there may be no apparent or only weak relationships with the means of bacterial abundances. Here, we apply quantile regression, an extension of the general linear model that allows for investigation of relationships across different quantiles of the distribution of a response variable.[361, 362] Quantile regression extends regression of the mean to the analysis of the entire conditional distribution of the response variable.[362] Examining quantile regression functions across the entire range of quantiles provides a more complete view of the response variable distribution than achieved by standard regression analysis.[361] The idea behind this method is that not all individuals are equally responsive to changes in abundance of specific bacterial groups, due to hidden bias and complex dependencies in ecological datasets.[361] Quantile regression is less sensitive to outliers than conventional regression and is not dependent on homoscedastic errors.[363] In particular, we tested whether associations between relative abundances of specific families with CD can be found with a clinical diagnosis (i.e., remission vs. exacerbation), but also with specific markers (i.e., fecal calprotectin (FC), serum C-reactive protein (CRP), and the Harvey Bradshaw index (HBI)) of disease activity in repeatedly sampled CD patients and healthy controls.[181]

Methods

Data and procedures

The study population has previously been described in Galazzo et al. (2019).[181] A total of 57 CD patients were included in this study. Demographic variables and subject characteristics are provided in Appendix Table 4.1 and medication use between visits is provided in Appendix Table 4.2. The CD patients formed a subset of the Inflammatory Bowel Disease South Limburg Cohort.[169] As a reference group, 15 healthy cohort (HC) subjects, all without any gastrointestinal disease, gastrointestinal symptoms, or comorbidities, were recruited among the controls who participated in the Maastricht Irritable Bowel Syndrome (IBS) Cohort.[364] Clinical data, blood, and feces were collected at two time points. The CD group comprised 22 remission-exacerbation (RE) patients with baseline sampling at time of remission and subsequent sampling during an exacerbation, and 35 remission-remission (RR) patients, with two subsequent samples while maintaining remission, i.e., without any flares in between subsequent samples. The median time between baseline and follow up samples was 14 (IQR 11–21), 20 (8–36), and 13 (12–16) weeks for RR patients, RE patients, and HCs, respectively (Appendix Table 4.2). All study subjects gave written informed consent prior to participation. Both studies have been approved by the Medical Ethics Committee of Maastricht University Medical Center and have been registered in the US National Library of Medicine (www.clinicaltrials.gov; NCT02130349 and NCT00775060, respectively).

Fecal samples were collected at home, kept at room temperature, and brought to the hospital within 12 hours after defecation. Part of the fecal sample of the CD patients was sent to the laboratory of Clinical Chemistry for routine analysis of FC. The remaining part was aliquoted and frozen at −80°C for microbiota analysis. Disease activity was defined by FC, serum CRP, and HBI. Patients were included in the study when patients were in remission at baseline, i.e., FC < 100 µg/g and CRP < 5 mg/L or FC < 100 µg/g, CRP < 10 mg/L, and HBI ≤ 4. Exacerbation at the second time point was defined by FC > 250 µg/g or FC > 100 µg/g with at least a 5-fold increase from baseline (Appendix Figure 4.1). The fecal microbiota composition was assessed by Illumina Miseq sequencing of the V4-region of the 16S rRNA gene. A detailed description of metagenomic DNA isolation, sequencing, and quality control is provided in the supplementary information of Galazzo et al. 2019.[181] The 16S rRNA gene sequencing data are released in the European Nucleotide Archive. The accession number is: PRJEB62578 (ERP147674). Information on microbial profiling and the selection of bacterial families for quantile regression analysis can be found in Box 4.1.

Box 4.1 - Data procedures and family selection. Data demultiplexing, length and quality filtering, and clustering of reads into Operational Taxonomic Units (OTUs) at 97% sequence identity was done using the online Integrated Microbial Next Generation Sequencing (IMNGS) platform using default settings except for minimum and maximum length for amplicons, which were set at 100 and 500 bp, respectively.[365] After quality filtering, binning, and removing unassigned reads, sequences were clustered in 640 OTUs. Normalization was performed by dividing OTU counts per sample for their total count (sample depth) and then multiplying the obtained relative abundances by the lowest sample depth. OTU sequences assigned to chloroplasts were removed prior to the statistical analyses. Then, the 18 different families used in the main text were selected by removing rare reads (not seen more than three times in at least 20% of the samples). We also performed quantile regression on the remaining families (which were not selected by the base case threshold), these results are placed in the Appendices.

Linear quantile regression mixed models (LQMM)

The quantile regression model takes the form $Q_{Y|X}(\tau) = X\beta_\tau$, where $Q_{Y|X}(\tau)$ denotes the τ^{th} quantile of the response variable Y, which is predicted from a vector X of explanatory variables with quantile specific parameters β_τ. The τ^{th} quantile is the inverse of the cumulative distribution function of Y, i.e., $q_Y(\tau) = F_y^{-1}$ or reciprocally $F_Y(q_\tau) = P(Y \leq y) = \tau$; where $\tau \in [0,1]$. It denotes the smallest value where the probability of finding an even smaller value is less than or equal to τ, whereas the probability of finding a larger value is less than or equal to $1 - \tau$.[361] A parametric distribution is assumed for the deterministic part of the model, but the random error part does not assume any distributional form. Further information on inclusion of covariates and model building strategy is supplied in Box 4.2.

Box 4.2 - Model building strategy. The analysis was divided in three parts. First, we investigated whether the relative abundances of the bacterial families could be explained by the group to which each individual belongs (i.e., healthy control (HC), remission-remission (RR), or remission-exacerbation (RE)). We added the interaction with visit number, to allow for different temporal changes in bacterial relative abundance over time between healthy controls, CD patients who experienced an exacerbation at the second visit, and those who remained in remission. Secondly, we excluded the healthy individuals from the model and investigated whether the relative abundances of the bacteria could be explained by disease activity (i.e., remission vs. exacerbation) in the CD patient group. Thirdly, we additionally investigated if the relative abundances of the bacterial families could be related to a quantitative disease indicator (i.e., HBI, CRP, or FC) other than the clinical definition of disease activity (remission and exacerbation). The models contain two time points per individual. Therefore, we used a random intercept per patient as well as a random effect for the variable 'visit number', because temporal changes in bacterial family's relative abundances may differ within patients, even when accounting for the fixed effect of disease trajectory (e.g., experiencing an exacerbation at the second visit).

Prior to the analyses, relative abundances were multiplied with 1000 and log-transformed with the natural-log function assuming a lower detection limit of 100 reads (which is ¼th of the lowest measurable value in the data). Prior to variable selection, all models contained the variables sex (male vs. female), smoking (current, ex, or never), and age (centered around mean age of 39.6 years) (Appendix Table 4.1). The models for CD patient cohort data also contained the disease-specific variables disease location (colonic (C), ileal (I), or ileocolonic (IC)), age at diagnosis (younger than 40 years (0) or older than 40 years (1)), surgery (no (0) vs. yes (1)), disease phenotype (non-stricturing/non-penetrating vs. stricturing/penetrating), and current treatment (mesalazines (no (0) vs. yes (1)), thiopurines (no (0) vs. yes (1)), biologicals (no (0) vs. yes (1)), induction (no (0) vs. yes (1)), and proton pump inhibitors (PPI) (no (0) vs. yes (1))) (Appendix Table 4.1). Variable selection was performed by running all possible models and then selecting the model with the lowest Bayesian Information Criterion (BIC) in the 50% quantile. For the model with the disease indicators (HBI, CRP, and FC), variable selection was performed on a model that contained all three indicators. For the sake of comparison, the selected variables were also used in the separate models for HBI, CRP, and FC. FC was divided by 1000 to improve numerical precision in quantile regression. Moreover, HBI, CRP, and FC are measured on different scales, therefore the data was normalized beforehand to make the quantile regression model estimates comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles.

We used the 'lqmm' R package (version 1.5.5)[366] of the R statistical analysis software (www.R-project.org) to perform the quantile regression analysis. Analyses were performed separately for each bacterial family. Although genus level might be preferred, this would have resulted in too many models. Therefore, we only looked at certain genus levels, when significant results were found at family level (within the base case selection as described in Box 4.1). To accommodate repeated sampling on the individual level, we employed a linear quantile mixed model (LQMM) framework. A mixed model contains both fixed effects and random effects, and can then account for correlation in repeated measurements from the same individual as these are likely to be more similar than observations from different individuals.[367] We estimated the series of quantile regression functions from the 10th to the 90th percent quantile. We used the Benjamini-Hochberg (BH) procedure per quantile to control for the expected proportion of 'false discoveries' across microbial families.[266] However, the BH procedure assumes independency in multiple testing, which is likely not the case in the gut microbiota. Therefore, the BH correction might provide too conservative estimates, and we choose to also report the unadjusted results.

Results

Differences in abundance between healthy individuals and CD patients

We found several associations between the relative abundances of bacterial families with CD, and more specific with remission or disease exacerbation (Figure 4.1). The quantiles that were significantly associated with CD are different per bacterial family (Appendix Figure 4.2). For example, patients with baseline sampling at time of remission and subsequent sampling during an exacerbation (RE) displayed significantly distinct distributions in relative abundance in the family Coriobacteriaceae (Figure 4.2A), both at baseline (visit 1) and at the second visit, compared to the healthy control subjects (HC). At baseline, there was a positive association in the higher quantiles and over time (at time of exacerbation) there was a negative association in the lower quantiles. This means that the distribution of Coriobacteriaceae abundance among RE patients is skewed to higher values at baseline, but to lower values at the follow-up visit, as compared to healthy controls (see also Figure 4.1D). However, these effects were no longer significant after BH correction (Appendix Figure 4.3). For Coriobacteriaceae we also found a significant relation in the higher quantiles of patients in the RE group compared to the patients with two subsequent samples while maintaining remission (RR) (Appendix Figure 4.4). Thus, a significant fraction of patients in the RE group had higher Coriobacteriaceae abundance than healthy individuals and RR patients at baseline.

The family Erysipelotrichiaceae (Figure 4.2B) displayed negative associations in relative abundance over almost all quantiles (except the most upper quantiles) for both patient groups compared to the healthy controls. We still found significant differences after BH correction, but these were only present in the lowest quantiles (Appendix Figure 4.3). Looking at genus level, the genera *Holdemania* and *Turicibacter* both displayed a similar pattern of significant results (Appendix Figures 4.5 to 4.7). We did not find a significant difference among the patients in the RE and RR group (Appendix Figure 4.4). This implies that the relative abundance of Erysipelotrichiaceae is severely skewed to lower values in CD patients (see also Figure 4.1G). The same kind of relationship was also found for Ruminococcaceae (Figure 4.2D). Similar to Erysipelotrichiaceae, the associations were only found at baseline, suggesting that this is a characteristic of CD, but not related to disease activity.

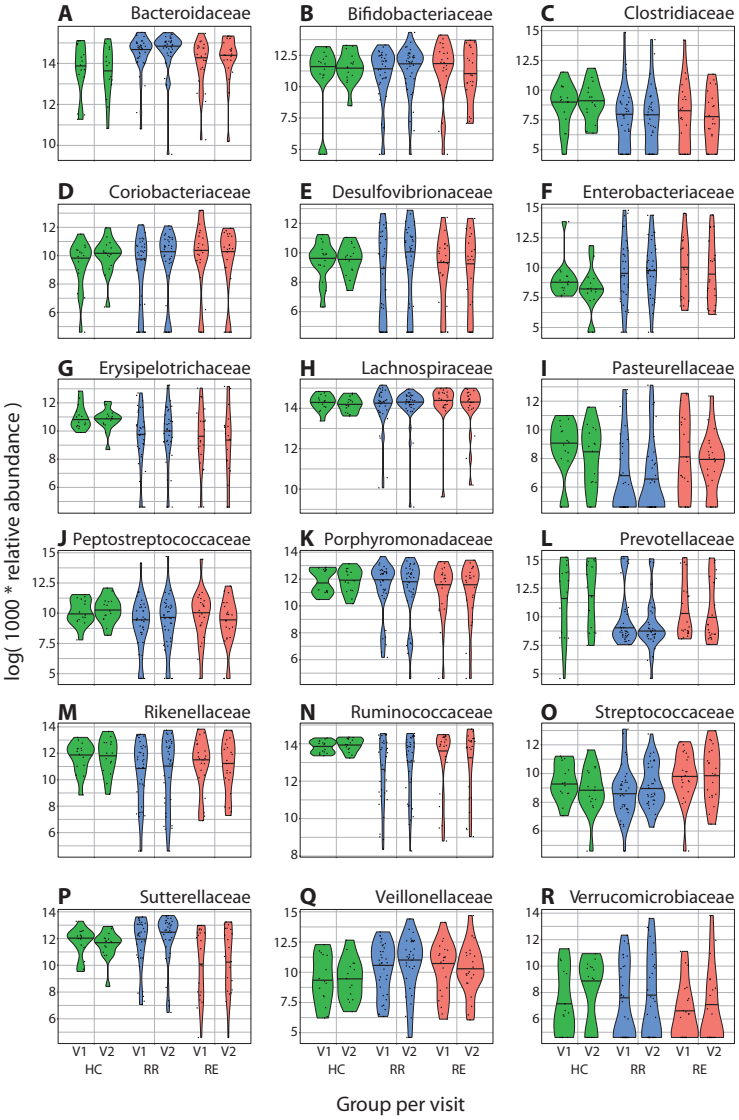


Figure 4.1 - Violin plots of transformed relative abundances of base case bacterial families by group and time point. In green the healthy controls (HC), in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50% quantile is shown with a black horizontal line. Genera are given in Appendix Figure 4.5 and families outside the base case criterion in Appendix Figure 4.8.

As another example, the relative abundances of Sutterellaceae among patients in the RE group were significantly skewed to lower values compared to both the healthy controls and the RR group patients (Figures 4.1P, 4.2F, and Appendix Figure 4.4). We also found a significant negative relation between the abundance of the family Pasteurellaceae in the RR group at baseline compared to the healthy controls (Figure 4.2C). For the family Streptococcaceae, we did not find many significant associations at baseline (except for one quantile), but the association for the RR x visit variable was significant for almost all quantiles (Figure 4.2E). This means that patients from the RR group experienced stronger increases in relative abundance of Streptococcaceae over time as compared to the healthy controls.

We also found a significant difference between the RR and RE patient groups for the family Streptococcaceae, with the RE patients having elevated abundances across the entire quantile range (Appendix Figure 4.4, see also Figure 4.1 - panel O). In our data, Pasteurellaceae and Streptococcaceae both only consisted of one classified genus. When refining the analyses for these classified genera, we did not find significant results within *Mannheimia* (Pasteurellaceae) or *Streptococcus* (Streptococcaceae) (Appendix Figures 4.5 to 4.7).

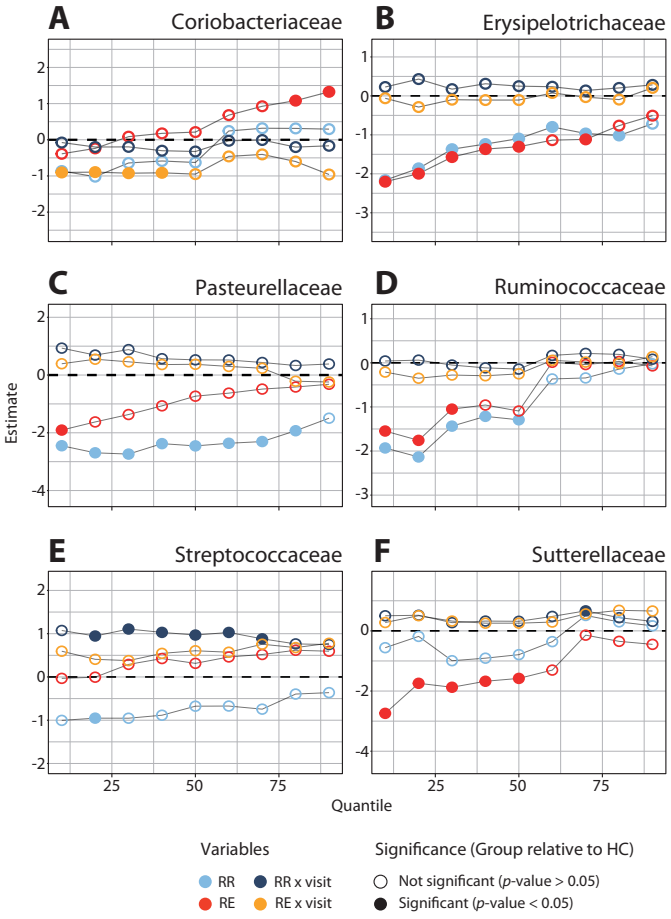


Figure 4.2 - Examples of quantile regression profile plots for some of the base case families. A) Coriobacteriaceae, B) Erysipelotrichaceae, C) Pasteurellaceae, D) Ruminococcaceae, E) Streptococcaceae, and F) Sutterellaceae. The various dots represent sample estimates (*y*-axis) of differences in relative abundance compared to healthy controls across the 10th to the 90th percent quantile (*x*-axis). Differences at baseline (visit 1) are visualized in light blue (RR) and red (RE), while the interaction with visit number (in dark blue and orange) displays the difference in changes over time. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance at that particular quantile. Significant variables (*p*-value < 0.05) are indicated with a closed circle. The other base case families are in Appendix Figure 4.2.

On top of the examples given in Figure 4.2, we also found that the relative abundances of Clostridiaceae, Desulfovibrionaceae, Peptostreptococcaceae, Prevotellaceae, and Rikenellaceae in CD patients were different from the relative abundances in the microbiota of healthy controls (Figure 4.3, Appendix Figure 4.2). These results, except for the families Prevotellaceae and Rikenellaceae, remained significant after BH correction (Appendix Figure 4.3). Most significant relations were negative and were found in the lower quantiles (Figure 4.3), meaning that CD patients more often displayed negatively than positively skewed abundance distributions (see also Figure 4.1). Besides, only a few associations between bacterial family abundance and covariates were found, with sex being the only significant covariate (males having higher abundances than females) (Figure 4.3). We also identified some significant associations in the families which fall under the sensitivity analyses of the families outside the base case selection criterion (Appendix Figure 4.8 to 4.10). We found that the relative abundances of Victivallaceae and Clostridiales fam. *i.s.* XI were different from the healthy controls for both the patients that stayed in remission and the patients that experienced an exacerbation. We also found significant results for the family Enterococcaceae for the patients that stayed in remission compared to the healthy controls and the families Actinomycetaceae and Lactobacillaceae for the patients that experienced an exacerbation at the second visit (Appendix Figure 4.9). However, these results disappeared after applying BH correction for multiple testing (Appendix Figure 4.10).

We compared our results from the LQMM models with the results obtained from an ordinary linear mixed-effects model (with similar variables as used in the LQMM models) using the 'lme' function from the 'nlme' R package (version 3.1).[368] Example code of the LQMM and LME models can be found on the GitHub repository (susannepinto/Quantile-Regression-CD). Most associations found by quantile regression could also be found with ordinary regression, as the mean response in the linear mixed-effects model provides somewhat of an average response over all quantiles. Nevertheless, some differences were also noticeable (Appendix Figure 4.11). For example, the family Coriobacteriaceae has a positive estimate in the higher quantiles for patients in the RE group relative to the healthy control group, which is not visible in the mean response (Appendix Figure 4.11D). Likewise, patients in the RR group displayed significant reductions in abundance in the lower to middle quantiles of Prevotellaceae and Streptococcaceae abundance, that were apparent in a reduced mean response, but without statistical significance. Conversely, the reduced mean responses regarding Ruminococcaceae in both RR and RE patients hide the fact that reductions only apply to lower and middle quantiles of abundance (Figure 4.1N and Appendix Figure 4.11N). Comparable findings were obtained for Pasteurellaceae and Peptostreptococcaceae. In some instances, linear mixed-effects regression yielded imprecise (cf. Lachnospiraceae) or biased (cf. group × visit in Prevotellaceae) estimates as compared to quantile regression (Appendix Figure 4.11).

Gut microbiota changes in relation to Crohn's disease activity

The relation between bacterial family abundance and disease activity (exacerbation) was mainly negative across the quantile range, indicating reduced abundance among RE patients as compared to RR patients at the second visit. However, this association was only statistically significant for upper quantiles of Coriobacteriaceae after adjustment for covariates (Figure 4.4). Instead, significant associations were revealed with several clinical variables (e.g., phenotype, surgery, proton pump inhibitors (PPI), and biologicals), suggesting that differences between RR and RE patients might have been confounded by disease-specific variables (Figure 4.4).

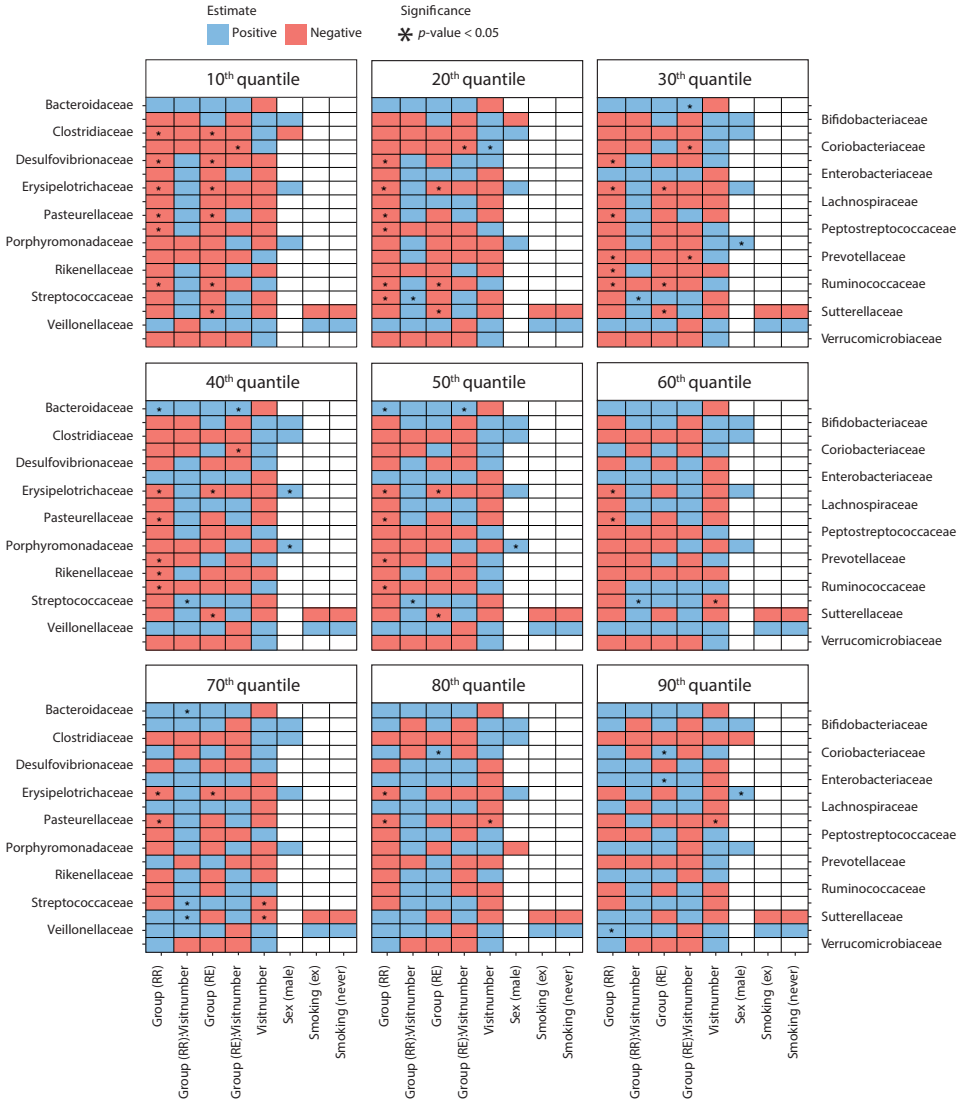


Figure 4.3 - Heatmap of quantile regression estimates across quantiles of relative abundance for base case families and common variables.

The model included all groups, i.e., healthy control (HC), remission-remission (RR), and remission-exacerbation (RE), with healthy controls as reference group. The red boxes indicate negative regression estimates, the blue boxes indicate positive regression estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*'), results adjusted with the BH procedure are given in Appendix Figure 4.3. Other families outside the base case criterion are given in Appendix Figure 4.9 and Appendix Figure 4.10.

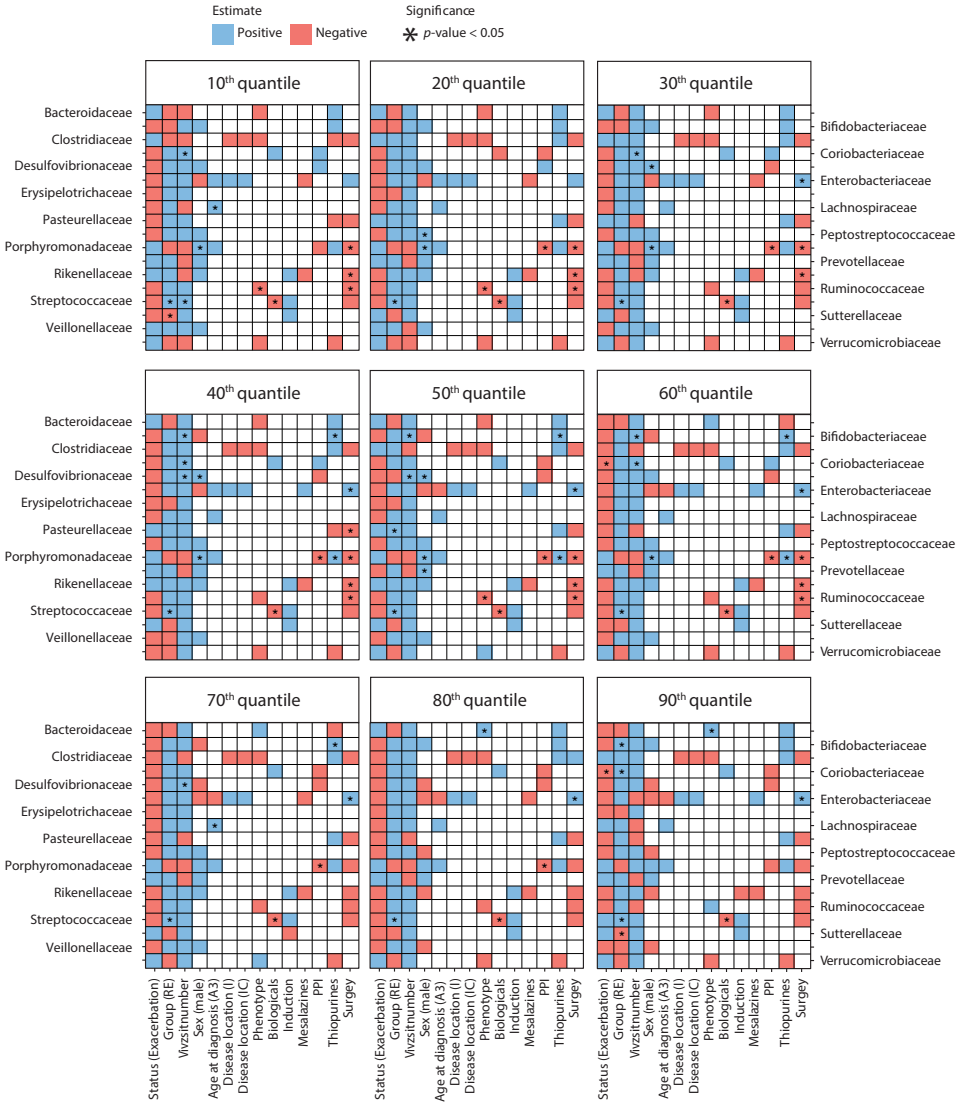


Figure 4.4 - Heatmap of quantile regression estimates across quantiles of relative abundance for base case families and clinical variables in CD patients only. The red boxes indicate negative regression estimates, the blue boxes indicate positive regression estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*'), results adjusted with the BH procedure are given in Appendix Figure 4.12.

Of note, many associations with disease activity also disappeared after BH correction for multiple testing (Appendix Figure 4.12), but the finding that RE patients had elevated Streptococcaceae abundances across the entire quantile range (irrespective of visit number) remained significant; likewise, treatment with biologicals remained significantly associated with lower Streptococcaceae abundance (Figure 4.4 and Appendix Figure 4.12). Further results on genera and other families can be found in Appendix Figures 4.13 to 4.16.

Bacterial relative abundances in relation to different disease activity indicators

When comparing regression on clinically defined exacerbation with different indicators of disease activity, we found that especially FC levels gave distinct results compared to the other indicators (Figure 4.5 and Appendix Figure 4.17). For almost all bacterial families, we did not observe a signal (estimates around zero) for clinical status (remission or exacerbation), CRP, and HBI after correction for clinical variables. In contrast, the normalized estimates of FC were much stronger (Appendix Figure 4.17), and significantly negative across the entire quantile range for Porphyromonadaceae and Verrucomicrobiaceae (Figure 4.5). The results of Porphyromonadaceae still hold after BH correction (Appendix Figure 4.18D). Further results on genera and other families can be found in Appendix Figures 4.19 to 4.22.

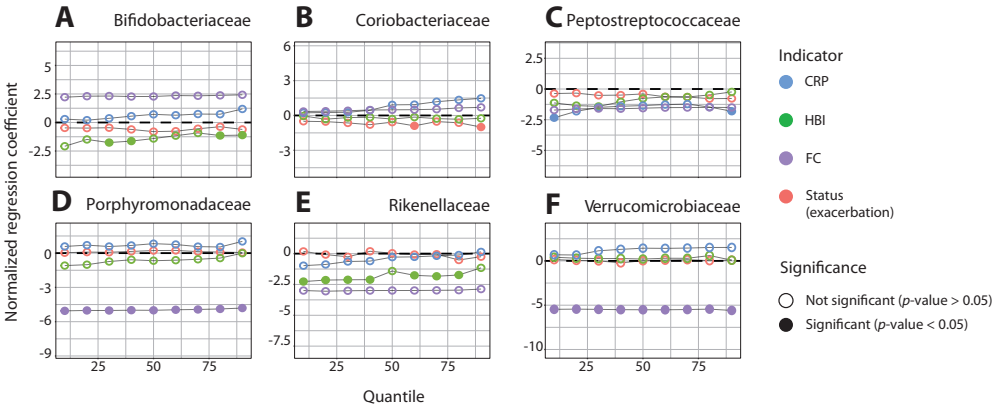


Figure 4.5 - Quantile regression profile plots for different disease activity indicators and clinical variables in CD patients only. For the families: A) Bifidobacteriaceae, B) Coriobacteriaceae, C) Peptostreptococcaceae, D) Porphyromonadaceae, E) Rikenellaceae, and F) Verrucomicrobiaceae. Only families with significant results are given, the other families are in Appendix Figure 4.17. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance for that particular quantile. The regression estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data was normalized beforehand to make the regression coefficients comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05) are indicated with a closed circle, results adjusted with the BH procedure are given in Appendix Figure 4.18.

Discussion

In this study, we investigated the possible associations between the relative abundance of specific bacterial families with CD and disease activity. We relied on quantile regression to uncover relationships that are not restricted to the mean response across CD patients. We found mainly negative associations with CD at family or genus level, especially in lower quantiles of relative bacterial abundance. These results are consistent with the frequently cited reduced microbial diversity in the gut of CD patients compared to healthy controls, but they also highlight that reductions for specific microbes are usually limited to a minority of patients. Thus, while CD coincides with a loss of microbial diversity in the gut, the process of diversity loss seems rather irregular with respect to specific taxonomic groups.

Associations outside of the mean response can be a result of heterogeneity among CD patients and may arise from the complex interactions among members of the microbiota. Microbes alter the environment through metabolic by-products, creating new ecological niches that promote diversification. However, some metabolites may adversely affect the growth of other microorganisms.[8] Whether systemic changes, such as those induced by CD, lead to niche reduction or expansion for a particular microbe probably depends as much on the characteristics of that particular microbe as on the microbial ecosystem of the individual host. If specific bacterial groups respond to disease or disease activity in some of the patients but not in others, their associations with CD or disease activity are more likely to be found in upper or lower quantiles than in the mean or median response across CD patients.

Interestingly, almost all significant associations we found were negative and applicable to the lower quantiles of bacterial abundance. While positive associations in upper quantiles have been attributed to unmeasured factors that limit the potential response to a positive stimulus,[361] this opposite pattern is reminiscent of an ecosystem response to stress: the ability to maintain healthy bacterial abundances is gradually lost once the system gets close to a tipping point.[97] The loss of some species in the microbial network can still be compensated for by other species with similar ecosystem functions (functional redundancy), but the loss of too many may lead to a loss of resilience and critical transition to an alternative stable state.[104, 369] Although the existence of tipping points in the onset or exacerbation of CD has not been demonstrated, a large-scale study by Lahti et al. (2014) showed distinct bimodal abundance patterns of certain bacterial species (i.e., tipping elements) among healthy human hosts.[102] These species were present in either a high or low abundance state, supporting the idea of alternative stable states in the human gut microbiota. Taken together, concepts of ecosystem resilience and critical transitions in the gut microbiota may explain why some individuals respond strongly to systemic changes, as induced by CD, while others display a more robust microbiota composition.[22]

Associations between the relative abundance of bacterial families and CD across the entire quantile range can also be identified with methods that focus on the mean response, such as ordinary linear regression. Families that exhibit such uniform responses could be considered to represent keystone bacterial groups, as their response to CD or disease activity is less dependent on other microbes or host factors as compared to families that are only responsive in some patients. This feature of robustness would be preferred for clinical diagnostics, prioritization for treatment, or monitoring of disease course, because guidelines can then be developed and used for all CD patients. However, it is also important to understand the less generic differences in the microbiota of CD patients because less robust associations may

also shed light on etiology and progression of CD and may provide leads for personalized treatment strategies. This is especially important considering the heterogeneous disease course, therapy response, and potentially contributing factors to microbiota perturbations. Moreover, the results of our analysis might help to reconcile inconsistencies in previously published findings as regards involvement of specific bacterial families in CD.

Our results confirm previously identified associations of CD with the families Erysipelotrichaceae, Peptostreptococcaceae, Prevotellaceae, Rikenellaceae, Ruminococcaceae (e.g., *Faecalibacterium prausnitzii*), and Veillonellaceae.[171, 181, 358] In addition, we also identified previously unreported associations between CD and the families Coriobacteriaceae, Desulfovibrionaceae, Streptococcaceae, and Sutterellaceae. Mixed results (both negative and positive associations) have been reported for the family Bacteroidaceae (e.g., *Bacteroides fragilis*).[171, 357, 358] Our results suggest that these mixed results can be explained by a change in association (from positive to negative) across the quantile range. Likewise, we found a negative association with the family Pasteurellaceae when the patients were compared to healthy individuals, especially with regard to patients who remained in remission. This is in contrast with previous results showing a positive association between relative abundances of Pasteurellaceae and CD.[171]

Previous studies did not make use of quantile regression to identify possible associations between the microbiota and factors related to inflammation. Most studies compared samples based on their means or medians (by Student's *t*-test, Mann-Whitney U test, or the analysis of variance), without taking into account the confounding effect of covariates, such as medication use or the age of the patient.[173, 356-358] Other studies used methods that can take covariates into account, such as generalized linear regression models, but these still only consider the mean count or relative abundance and do not consider distinct associations across patients.[171, 173, 175, 357] Lastly, supervised classifiers (e.g., Random Forest) and clustering algorithms (e.g., agglomerative hierarchical clustering) are used to predict the presence or activity of disease by the pattern in relative abundance of many families at once.[181] While these methods are not constrained by the strict assumptions of regression models, they have difficulties in dealing with repeated measurements and covariates. In addition, these methods require many patients, which are often not available in longitudinal clinical cohorts.

A practical advantage of quantile regression is its usefulness in situations when assumptions of other methods are violated. For example, quantile regression does not require homoscedastic and normally distributed data. On the contrary, the method enables detection and description of changes in the conditional distribution of the response variable when there is heteroscedasticity, skewness, or kurtosis in the data.[361] Another limitation of quantile regression is that it is hard to use for the purpose of prediction. Nevertheless, quantile regression is powerful when heterogeneous response distributions should be expected, e.g., if many interdependencies and potentially limiting factors play a role. If those co-factors are differently distributed among patients and not included in the model, they lead to (hidden) bias in conventional regression but can be dealt with in quantile regression.[370]

We found several significant associations of bacterial abundance with the presence of CD. However, associations with disease activity were less evident in our data. Although we found some differences between the two groups of CD patients, we only found significantly elevated abundances for Streptococcaceae (at baseline) and Coriobacteriaceae (during active disease) in patients experiencing an exacerbation relative to patients remaining in remission, and the latter effect disappeared after correction for multiple testing.

Multiple explanations are possible for the lack of significant associations between exacerbation and remission. Firstly, the microbiota of CD patients might not be responsive to exacerbation as compared to remission. Multiple studies underline our finding of no clear significant differences between remission and active disease.[177, 181] In other words, the observed differences in bacterial abundances are disease-related community differences that even persist in the absence of active inflammation, and therefore this pattern is not significantly reflected between the disease states.[371, 372]

However, most studies, including ours, also likely lacked the statistical power to find such potential differences. A second possibility is that potential associations are confounded with other factors that are likely to play a role in shaping the microbiota, such as disease severity,[175] disease duration,[177] disease location,[173] treatments,[373] and host characteristics (such as smoking).[172] Treatments such as PPIs have been shown to change gut microbiota and individuals undergoing surgery will have been given antibiotics to prevent infection.[373] Lastly, we might not have looked at the right taxonomic level, while the resolution of taxonomic profiling could impact the accuracy and specificity of our findings. Most differences are possibly present only at the species or strain levels, or may even require metabolic or functional analysis.[21, 22, 172] Moreover, a change in relative abundance at taxonomic level might not reflect a change in ecosystem functioning, as expansion in certain species can compensate for the loss of another (functionally similar) species.[22] Nonetheless, we did find stronger associations with fecal calprotectin than with the clinical definition of CD activity. This suggests that a quantitative measure of inflammation carries information about the microbial involvement in disease activity. As the level of fecal calprotectin is only a proxy of inflammation in the gut, the associations might become even clearer when specific immunological markers, or even hormones, would be used.

It is important to acknowledge several limitations that may impact the generalizability and interpretation of our findings. Firstly, our study was conducted within a relatively small cohort of CD patients (n = 57) and healthy controls (n = 15). While this cohort size allowed us to perform longitudinal analyses, it may limit the generalizability of our results to broader CD populations. Also, quantile regression is not insensitive to outliers, especially in the highest and lowest quantiles when there is not much data left for estimation, potentially affecting the robustness of our statistical analyses. With a larger dataset, one might consider a finer quantile division to obtain a smoother quantile regression profile, while with a smaller dataset, a coarser division would be more appropriate. Furthermore, as with any observational study, causation cannot be inferred from our results, and further mechanistic investigations are needed to elucidate whether there is a role of the highlighted bacterial families in CD pathogenesis. Lastly, the dynamic nature of the gut microbiota and potential temporal variations were not extensively explored in this study, which might have limited our ability to capture the full spectrum of microbial changes associated with CD over time. In light of these limitations, caution should be exercised when interpreting our results, and future research hopefully addresses these constraints through larger, more diverse cohorts and a finer taxonomic resolution to provide a more comprehensive understanding of the gut microbiota's role in CD.

In this study, we showed that associations between CD with relative bacterial abundances can be different for subsets of individuals. Our findings revealed significant negative associations with CD for several bacterial families, such as Pasteurellaceae, Peptostreptococcaceae, Prevotellaceae, and Ruminococcaceae, highlighting their potential roles in CD pathogenesis.

Furthermore, the significant differences in the relative abundance of Sutterellaceae and Streptococcaceae among CD patients who experienced exacerbations, relative to those who maintained remission, had not been seen before and underscore the dynamic nature of microbial associations in relation to disease activity. The subtle variations observed in the family Coriobacteriaceae, which could not be seen in the mean response, further emphasize the complexity of these relationships.

Importantly, our study underscores the heterogeneity of CD and its impact on gut microbiota, suggesting that associations may only become evident when considering patients' diverse disease courses, medication histories, therapy responses, and gut microbiota compositions. Associations with specific bacterial families may only be detectable in a minority of patients, hence they cannot generally be considered to identify CD or disease activity. The novelty of our study lies in its rigorous approach to exploring associations in subsets of patients, acknowledging the heterogeneity between them. In such situations, quantile regression is a useful tool for distilling potential relationships that may remain unidentified by commonly used methods. We recommend its use in even larger cohorts, to gain a better understanding of CD in relation to the gut microbiota.

Appendices of Chapter 4

Appendix Table 4.1 - Clinical and demographic information of healthy controls and CD patients (RR and RE).
Note that two samples per individual (*n* = 72 individuals) were collected (*n* = 144 samples).

	HC (*n* = 30)	RR (*n* = 70)	RE (*n* = 44)	Overall (*n* = 144)
Sex				
Female	14 (46.7%)	50 (71.4%)	24 (54.5%)	88 (61.1%)
Male	16 (53.3%)	20 (28.6%)	20 (45.5%)	56 (38.9%)
Smoking				
Ex	4 (13.3%)	34 (48.6%)	20 (45.5%)	58 (40.3%)
Never	26 (86.7%)	20 (28.6%)	20 (45.5%)	66 (45.8%)
Current	0 (0%)	16 (22.9%)	4 (9.1%)	20 (13.9%)
Age				
Mean (SD)	26.7 (5.93)	42.6 (12.8)	43.6 (16.8)	39.6 (14.7)
Median (min, max)	25.0 (20.0, 45.0)	43.0 (17.0, 67.0)	42.5 (19.0, 68.0)	38.5 (17.0, 68.0)
Disease location[a]				
C	NA	16 (22.9%)	14 (31.8%)	30 (20.8%)
I	NA	24 (34.3%)	14 (31.8%)	38 (26.4%)
IC	NA	30 (42.9%)	16 (36.4%)	46 (31.9%)
Age at diagnosis[b]				
A2	NA	62 (88.6%)	28 (63.6%)	90 (62.5%)
A3	NA	8 (11.4%)	16 (36.4%)	24 (16.7%)

Table continued on next page

Surgery				
0	30 (100%)	54 (77.1%)	36 (81.8%)	120 (83.3%)
1	0 (0%)	16 (22.9%)	8 (18.2%)	24 (16.7%)
Phenotype				
0	NA	52 (74.3%)	24 (54.5%)	76 (52.8%)
1	NA	18 (25.7%)	20 (45.5%)	38 (26.4%)
Mesalazines				
0	30 (100%)	60 (85.7%)	35 (79.5%)	125 (86.8%)
1	0 (0%)	10 (14.3%)	9 (20.5%)	19 (13.2%)
Thiopurines				
0	30 (100%)	46 (65.7%)	28 (63.6%)	104 (72.2%)
1	0 (0%)	24 (34.3%)	16 (36.4%)	40 (27.8%)
Biologicals[c]				
0	30 (100%)	32 (45.7%)	15 (34.1%)	77 (53.5%)
1	0 (0%)	38 (54.3%)	29 (65.9%)	67 (46.5%)
Induction				
0	30 (100%)	57 (81.4%)	36 (81.8%)	123 (85.4%)
1	0 (0%)	13 (18.6%)	8 (18.2%)	21 (14.6%)
PPI[d]				
0	30 (100%)	56 (80.0%)	28 (63.6%)	114 (79.2%)
1	0 (0%)	14 (20.0%)	16 (36.4%)	30 (20.8%)
HBI[e]				
Mean (SD)	NA	2.41 (2.95)	3.00 (3.39)	2.08 (2.97)
Median (min, max)	NA	1.00 (0, 11.0)	2.00 (0, 13.0)	1.00 (0, 13.0)
Missing	NA	0 (0%)	2 (4.5%)	2 (1.4%)
CRP[f]				
Mean (SD)	NA	2.57 (2.05)	3.76 (3.28)	2.31 (2.61)
Median (min, max)	NA	2.00 (0.9, 11.0)	2.80 (0.9, 13.0)	1.40 (0, 13.0)
Missing	NA	5 (7.1%)	8 (18.2%)	13 (9.0%)
FC[g]				
Mean (SD)	NA	28.6 (20.9)	290 (742)	102 (426)
Median (min, max)	NA	14.0 (14.0, 98.0)	110 (14.0, 4900)	15.0 (0, 4900)

[a] Disease location: colonic (C), ileal (I), or ileocolonic (IC)

[b] Age at diagnosis: younger than 40 years (A2) or older than 40 years (A3)

[c] All biological treatments concerned anti-TNF therapy

[d] PPI: proton pump inhibitors

[e] HBI: Harvey Bradshaw index (Appendix Figure 4.1)

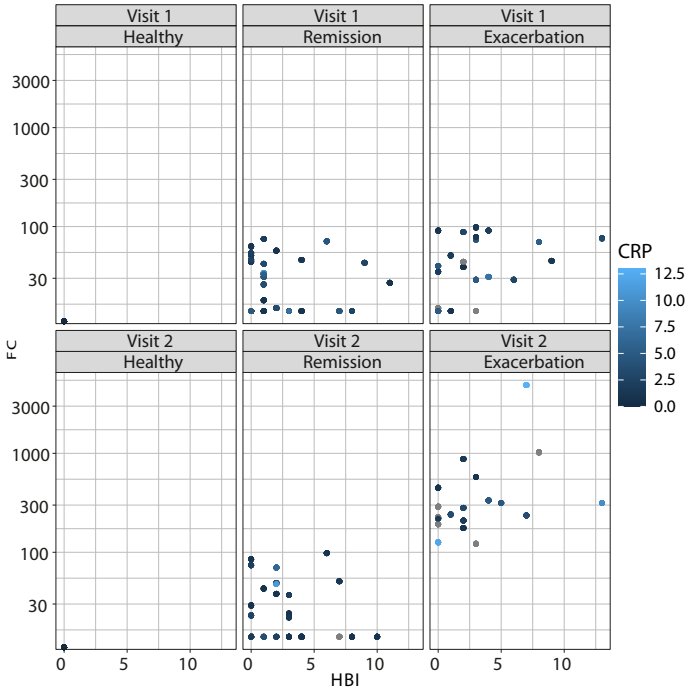
[f] CRP: serum C-reactive protein (Appendix Figure 4.1)

[g] FC: fecal calprotectin (Appendix Figure 4.1)

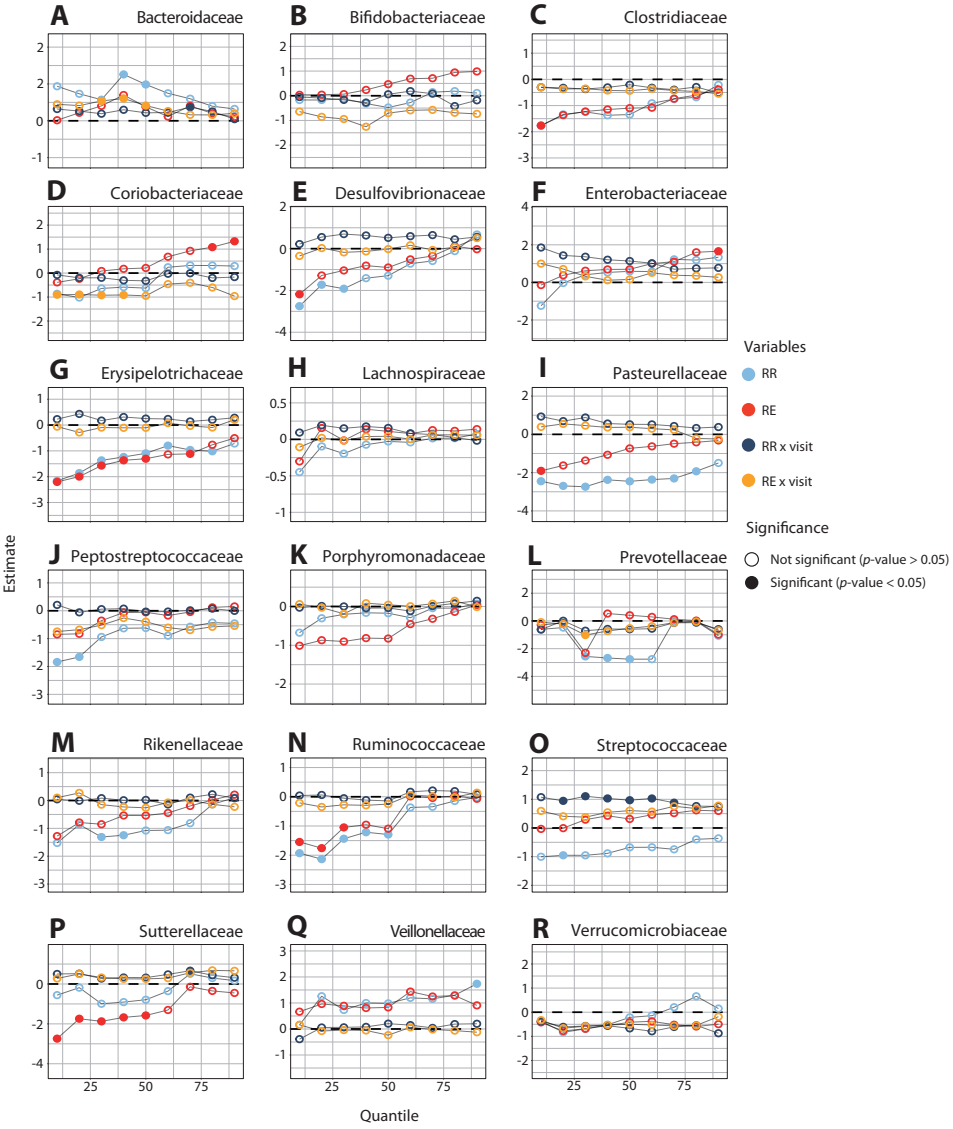
Appendix Table 4.2 - Medication use and time between sampling moments for remission and exacerbation samples.				
	RR (*n* = 35)		RE (*n* = 22)	
	Remission	Remission	Remission	Exacerbation
Medication[a]				
Mesalazine	5 (14.3%)	5 (14.3%)	4 (18.2%)	5 (22.7%)
Thiopurines	11 (31.4%)	11 (31.4%)	9 (40.9%)	7 (31.8%)
Biologicals	18 (51.4%)	19 (54.3%)	13 (59.1%)	15 (68.2%)
Corticosteroids	1 (2.9%)	0 (0%)	1 (4.5%)	1 (4.5%)
PPI	7 (20%)	7 (20%)	8 (36.4%)	8 (36.4%)
Antibiotics[b]	1 (2.9%)	0 (0%)	1 (4.5%)	0 (0%)
Time between sampling moments (week, median, IQR)	14 (11–21)		20 (11–21)	

[a] Six RR and five RE patients had a medication change between consecutive samples during the study period. In the RR group, mesalazine was stopped by one patient, prednisone by one patient and biologicals by two patients, while one patient started mesalazine and one patient started with biologicals. In the RE group, two patients started with biologicals, two patients stopped with thiopurines, and one patient started with mesalazine.

[b] Ciprofloxacin and cotrimoxazole were used two and one month prior to sample collection, respectively.

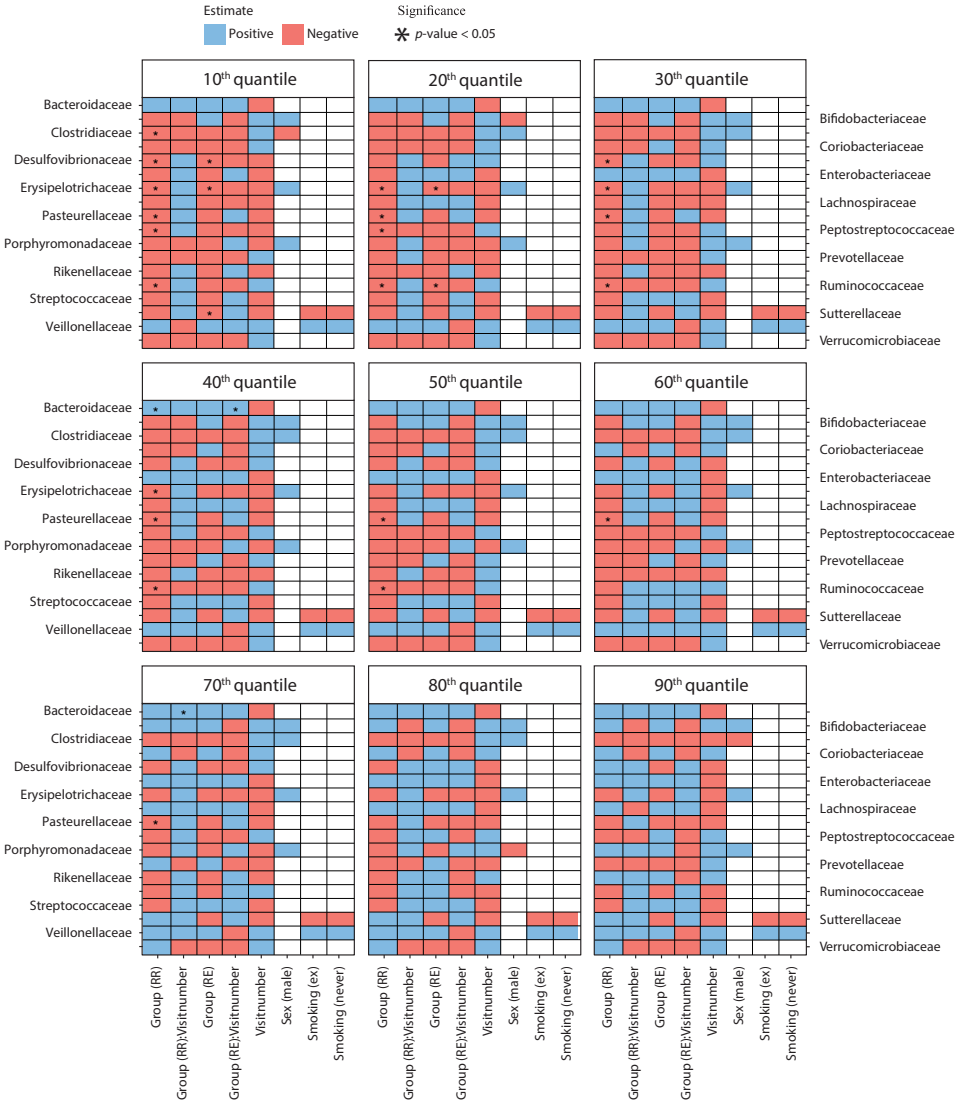


Appendix Figure 4.1 - Disease indicators (FC, serum CRP, and HBI). Remission at baseline was defined by FC < 100 µg/g and CRP < 5 mg/L or FC < 100 µg/g, CRP < 10 mg/L, and HBI ≤ 4. Disease activity at the second visit was defined by FC, serum CRP, and HBI, i.e., FC > 250 µg/g or FC > 100 µg/g with at least a 5-fold increase from baseline.

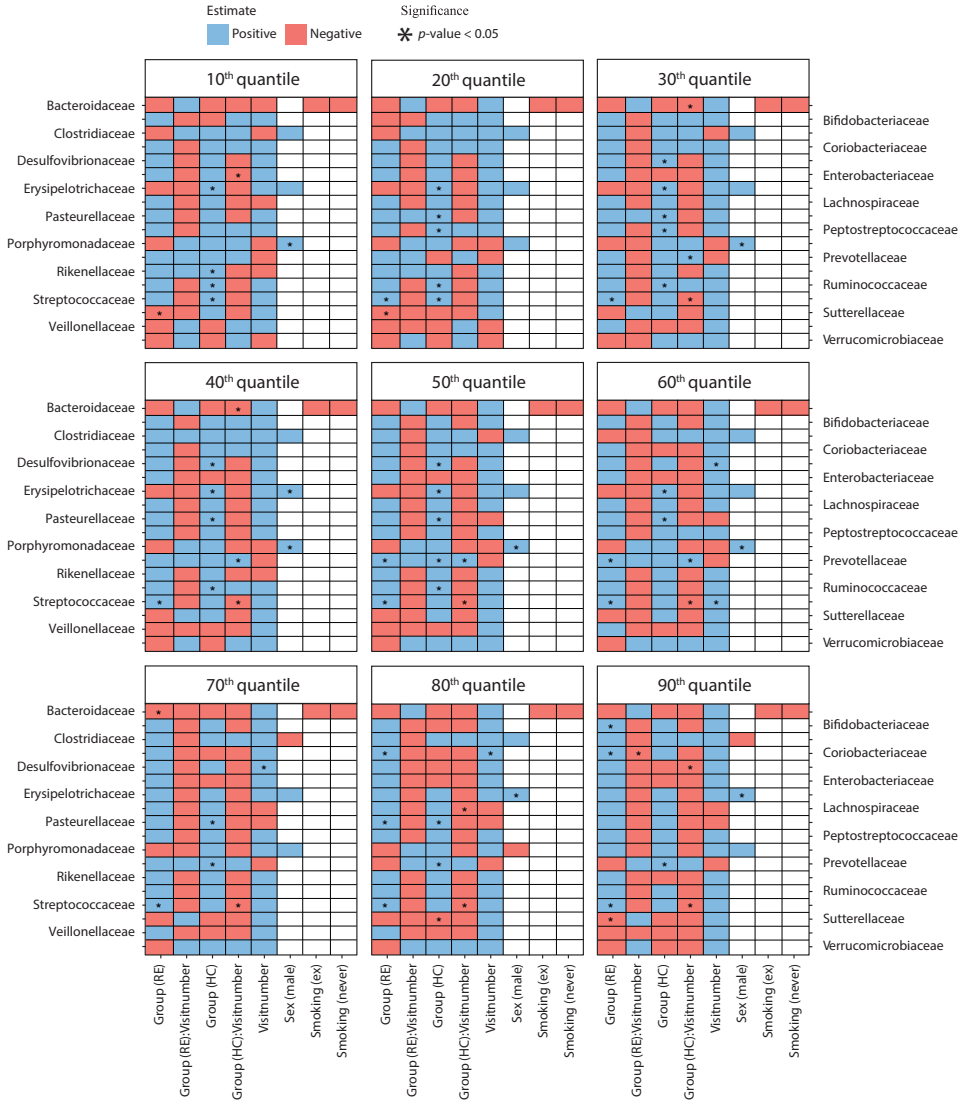


Appendix Figure 4.2 - Quantile regression profile plots for all base case families. The various dots represent sample estimates (*y*-axis) of differences in relative abundance compared to healthy controls across the 10th to the 90th quantile (*x*-axis). Differences at baseline (visit 1) are visualized in light blue (RR) and red (RE), while the interaction with visit number (in dark blue and orange) displays the difference in changes over time. The dotted line at zero indicates no difference compared to healthy controls. When the points are above the dotted line there is a positive effect of disease group on relative abundance, whereas points below the dotted line imply a negative effect of disease group on relative abundance at that particular quantile. Significant variables (*p*-value < 0.05) are indicated with a closed circle.

Sensitivity analyses: Differences in abundance between healthy individuals and CD patients

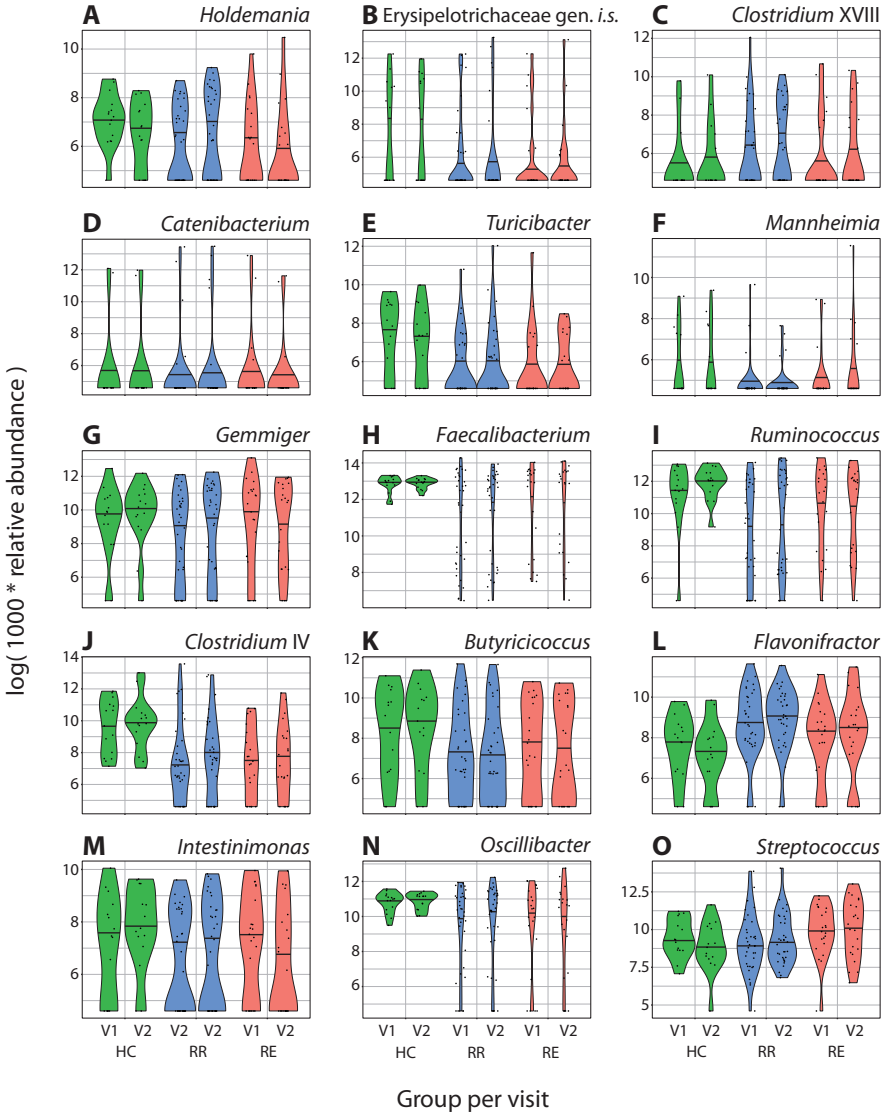


Appendix Figure 4.3 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and common variables, with *p*-values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').

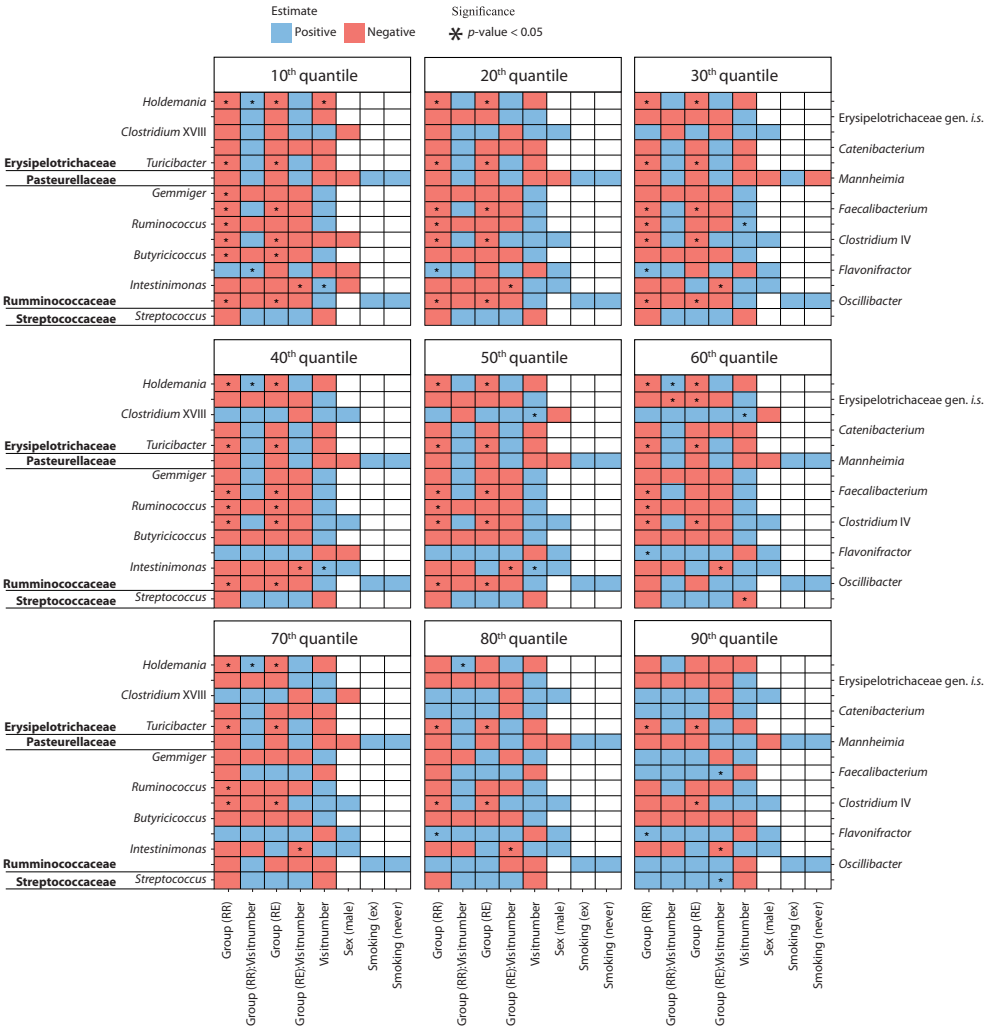


Appendix Figure 4.4 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and common variables, using the RR group as reference instead of the HC group.

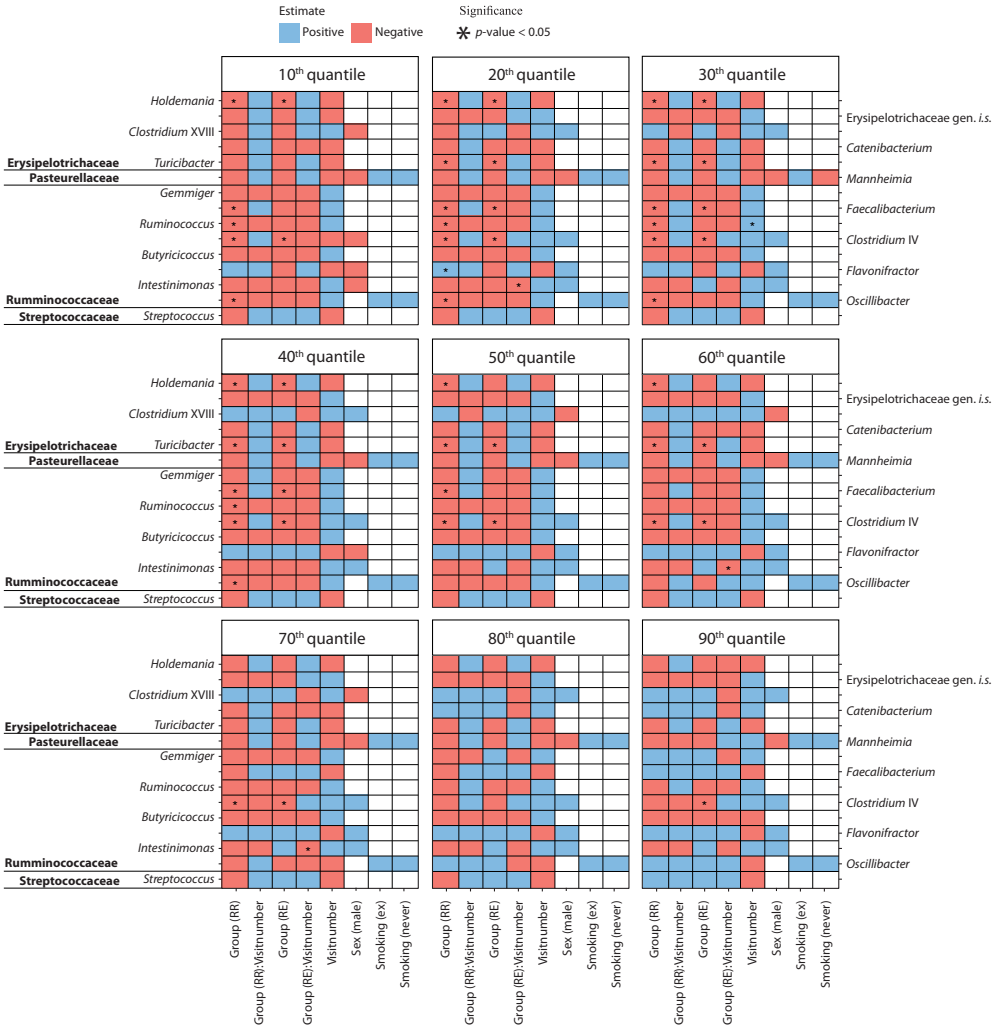
The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*').



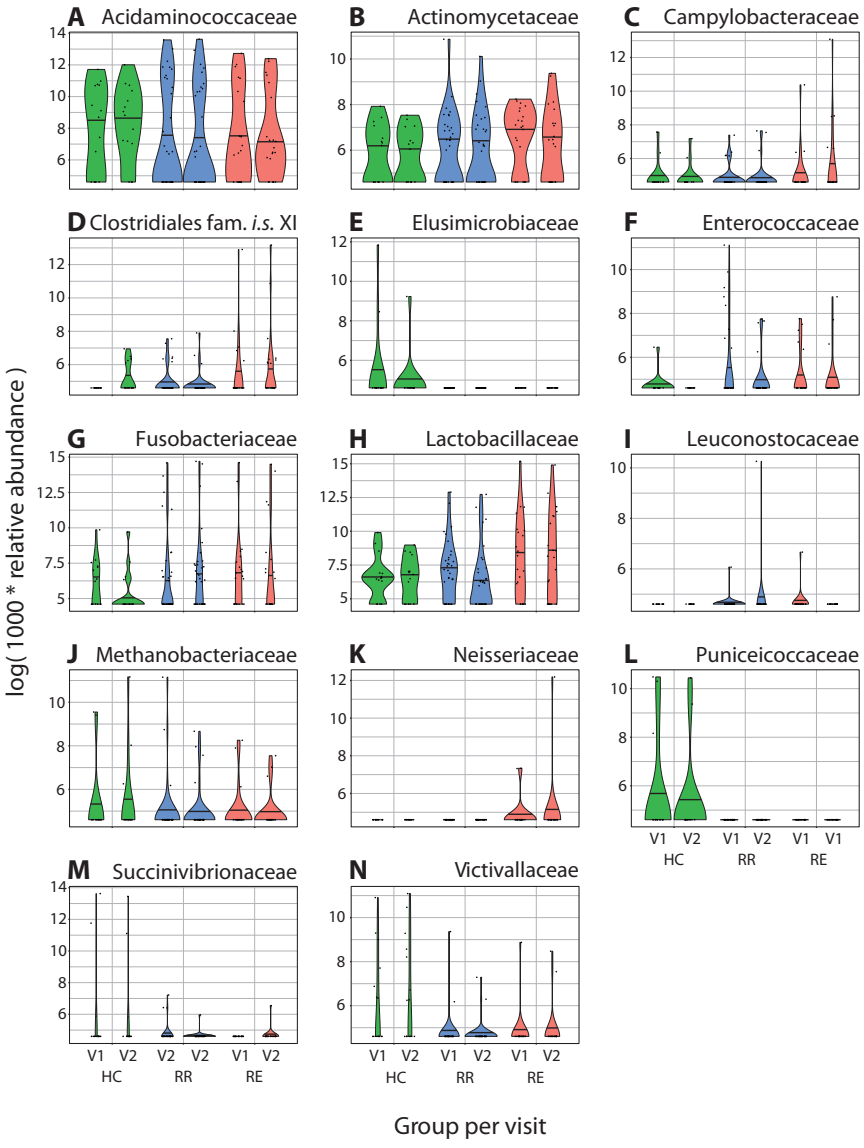
Appendix Figure 4.5 - Violin plots of transformed relative abundances for selected genera by group and time point. In green the healthy controls, in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50th quantile is shown with a black horizontal line.



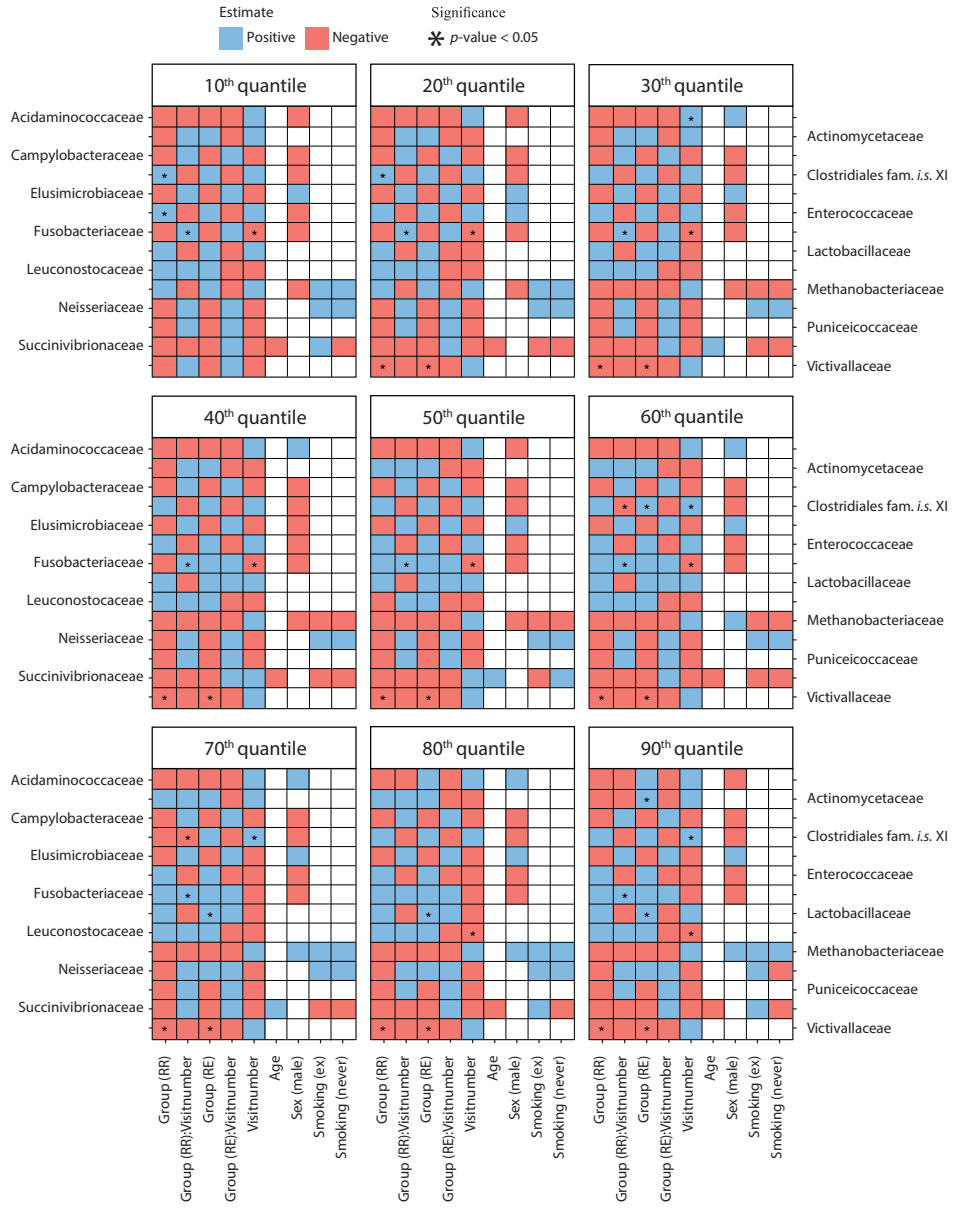
Appendix Figure 4.6 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and common variables. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*').



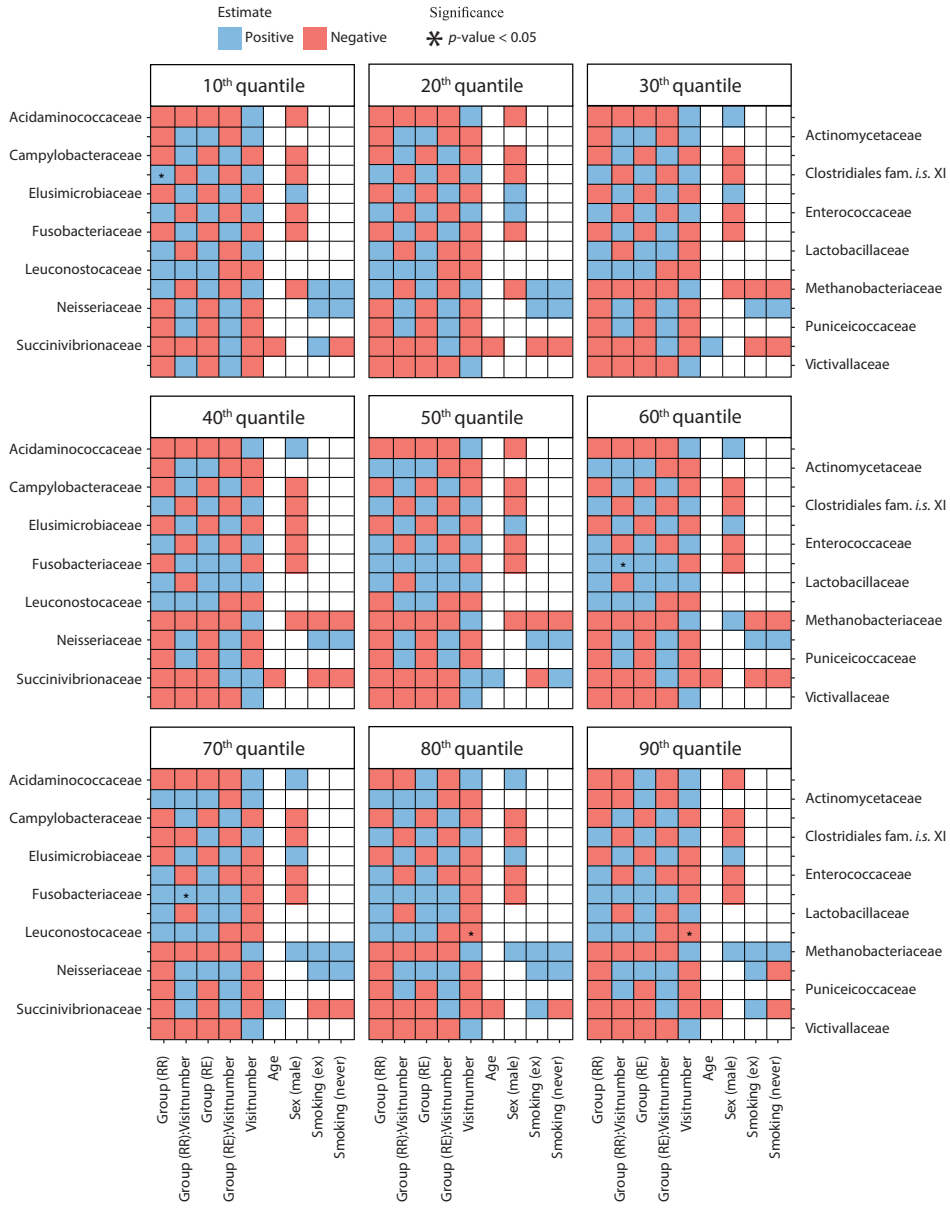
Appendix Figure 4.7 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and common variables, with *p*-values adjusted using the BH procedure. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').



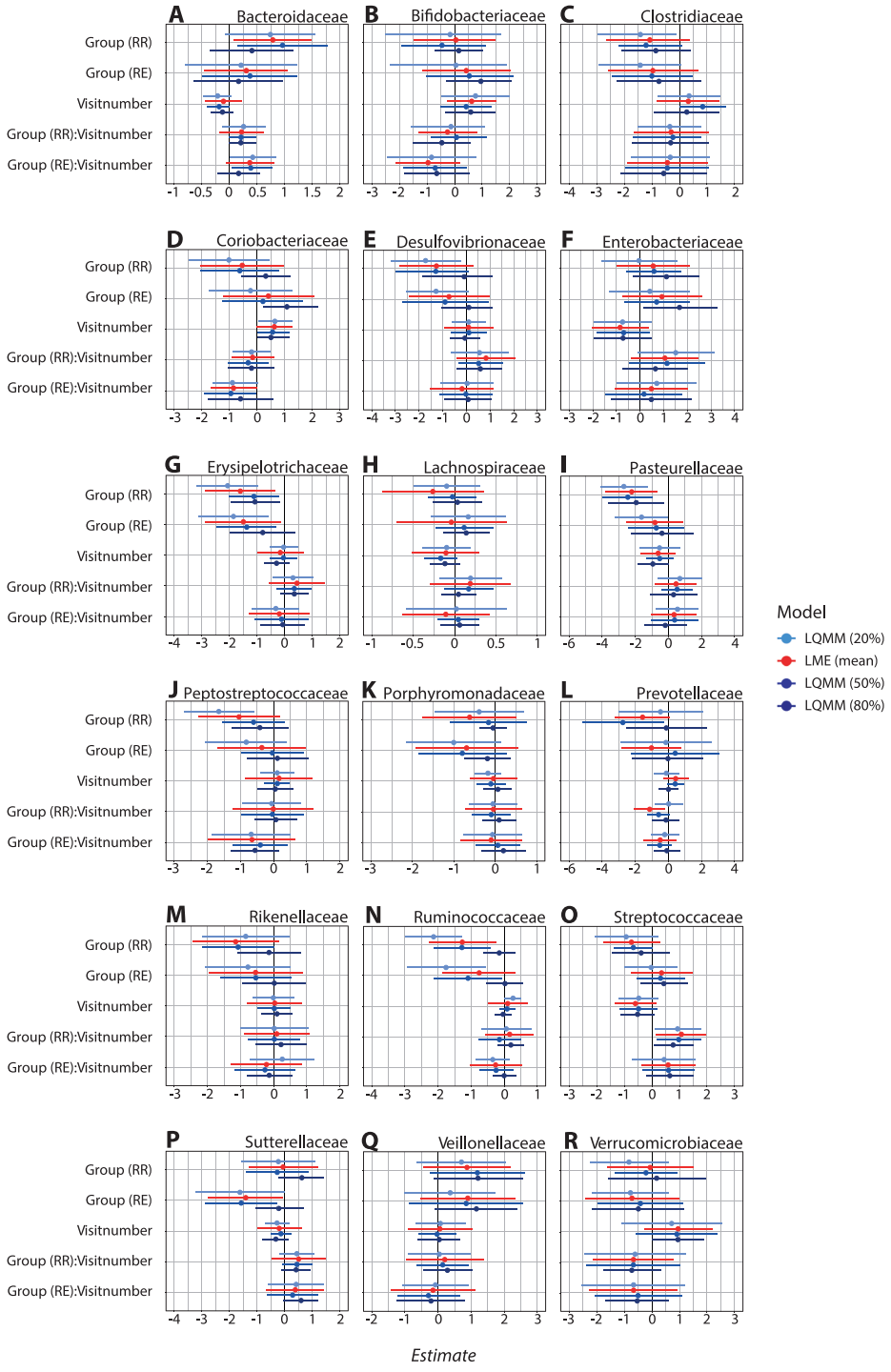
Appendix Figure 4.8 - Violin plots of transformed relative abundances for families outside the base case selection criterion by group and time point. In green the healthy controls, in blue the RR group, and in red the RE group, all visualized per time point (V1 = visit 1 and V2 = visit 2). Patients in the RE group are in remission during the first visit and experience an exacerbation during the second visit. The 50th quantile is shown with a black horizontal line.



Appendix Figure 4.9 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and common variables. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*').

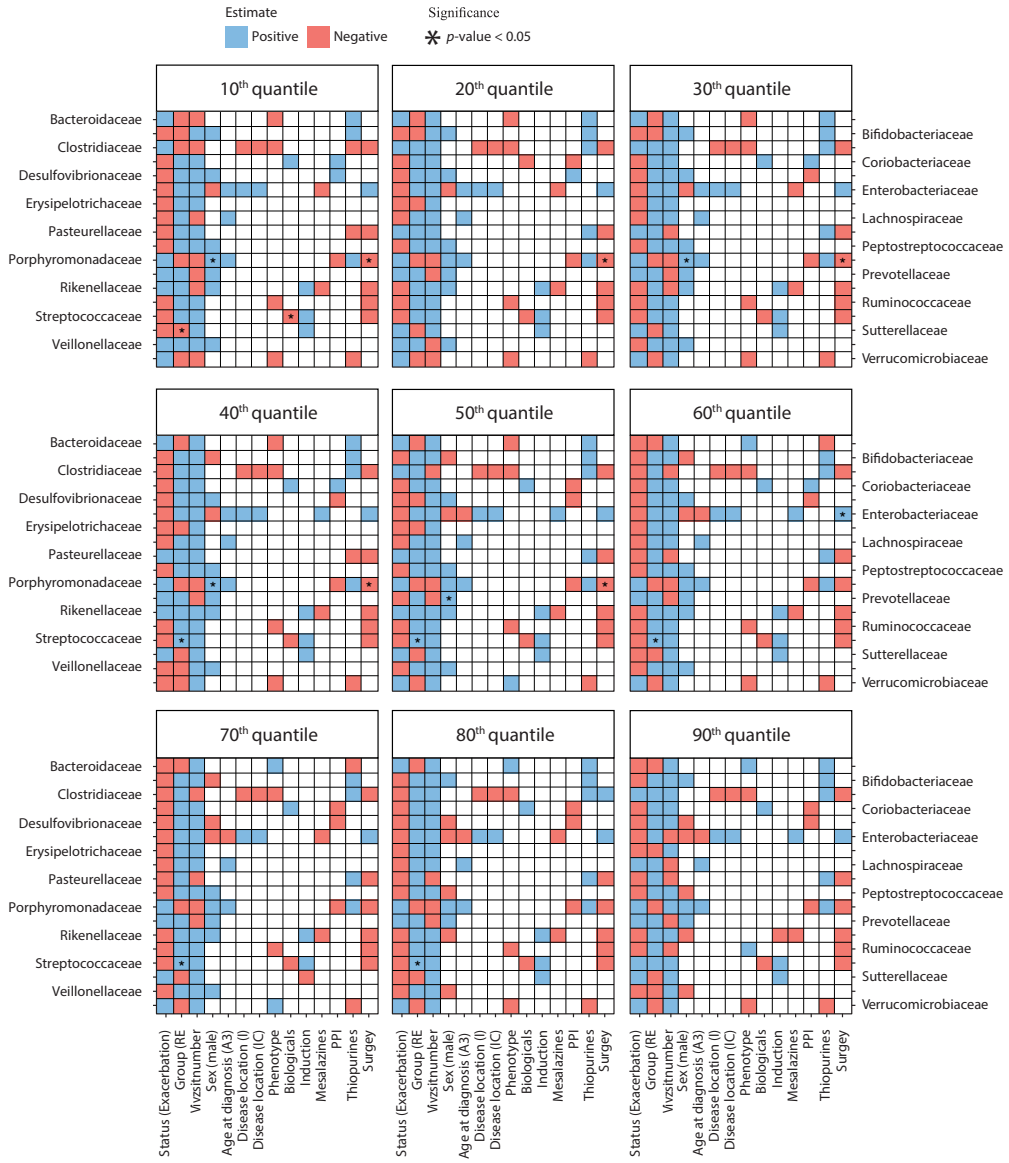


Appendix Figure 4.10 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and common variables, with *p*-values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').

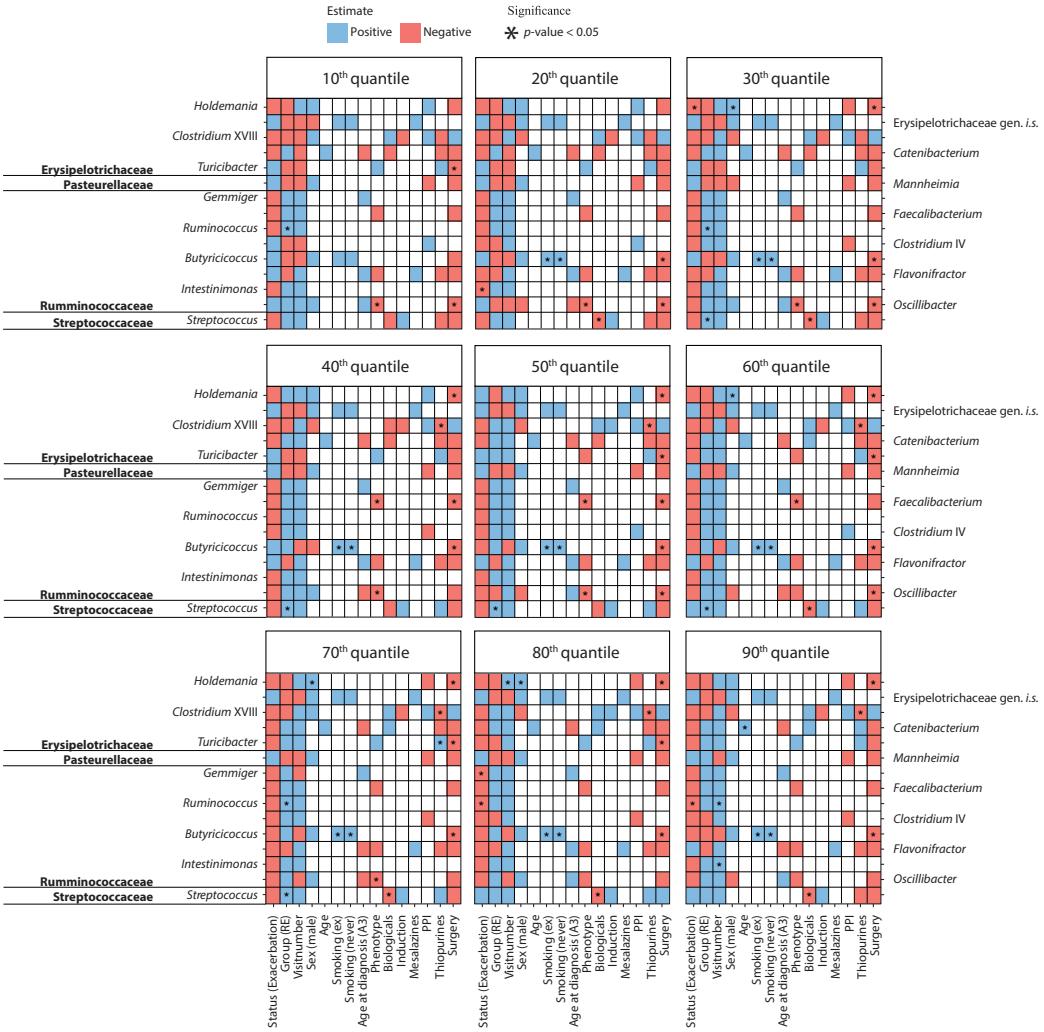


Appendix Figure 4.11 - Comparison of the LQMM analysis results (20%, 50%, and 80% quantiles) with linear mixed-effects model results. The point estimates, 95% confidence intervals, and a reference line at 0 (in black) are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different (*p*-value < 0.05) from 0.

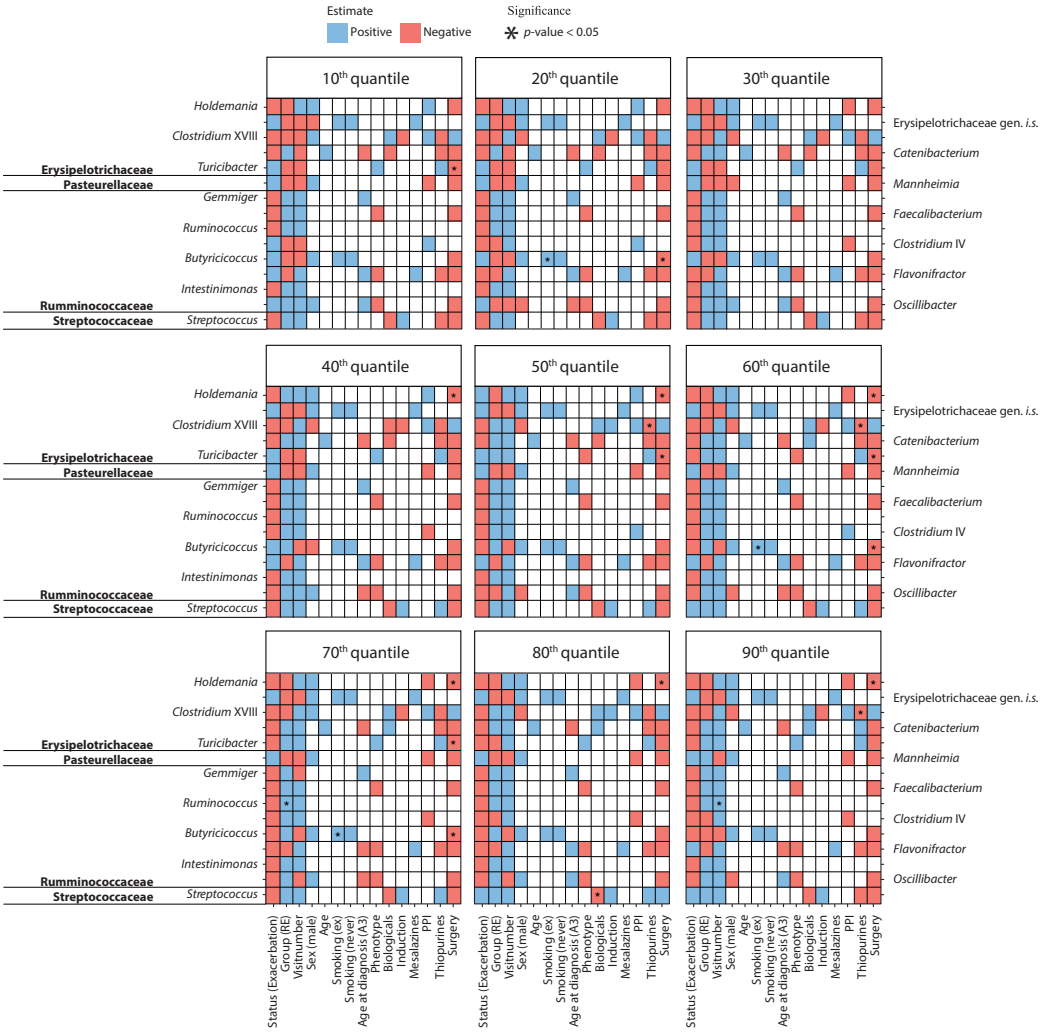
Sensitivity analyses: Gut microbiota changes in relation to Crohn's disease activity



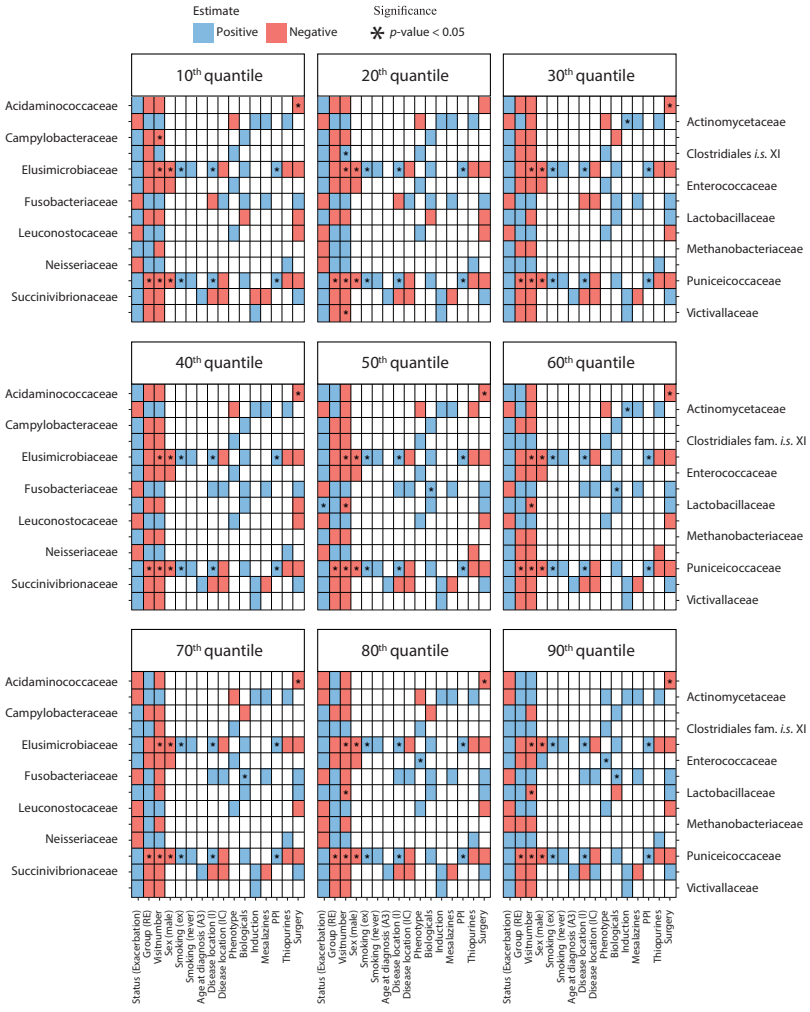
Appendix Figure 4.12 - Heatmap of quantile regression estimates across quantiles of the relative abundance for base case families and clinical variables in CD patients only, with *p*-values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').



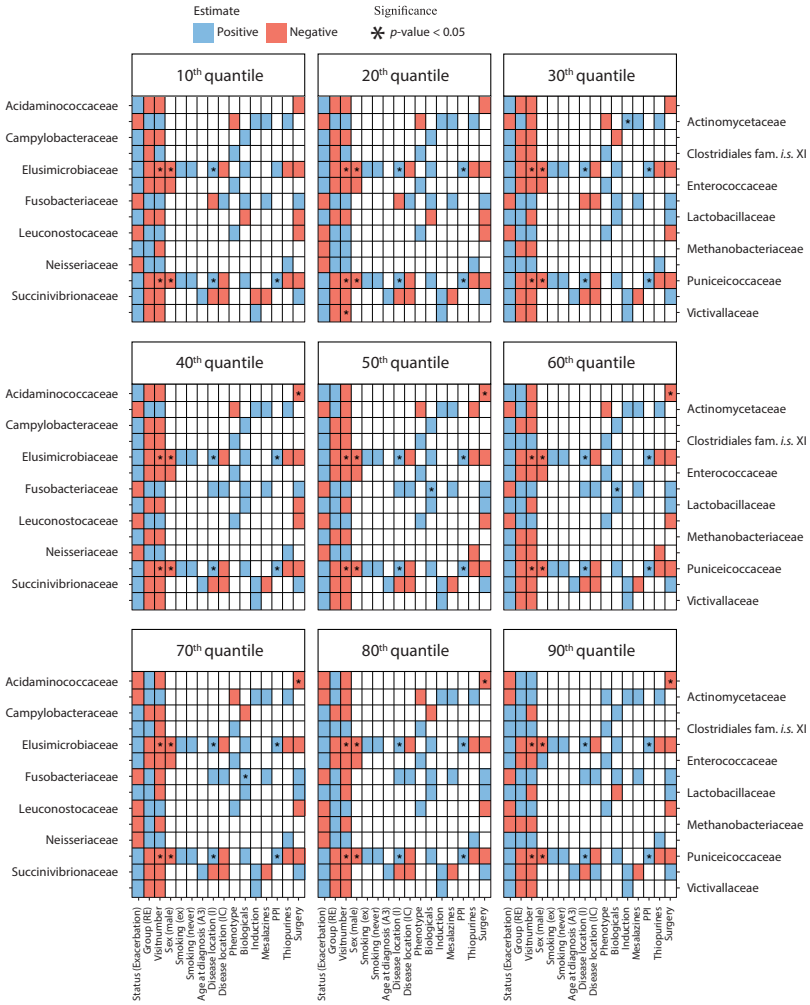
Appendix Figure 4.13 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and clinical variables in CD patients only. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05) are indicated with an asterisk ('*').



Appendix Figure 4.14 - Heatmap of quantile regression estimates across quantiles of the relative abundance for selected genera and clinical variables in CD patients only, with *p*-values adjusted using the BH procedure. The corresponding family names are placed in bold on the left. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').

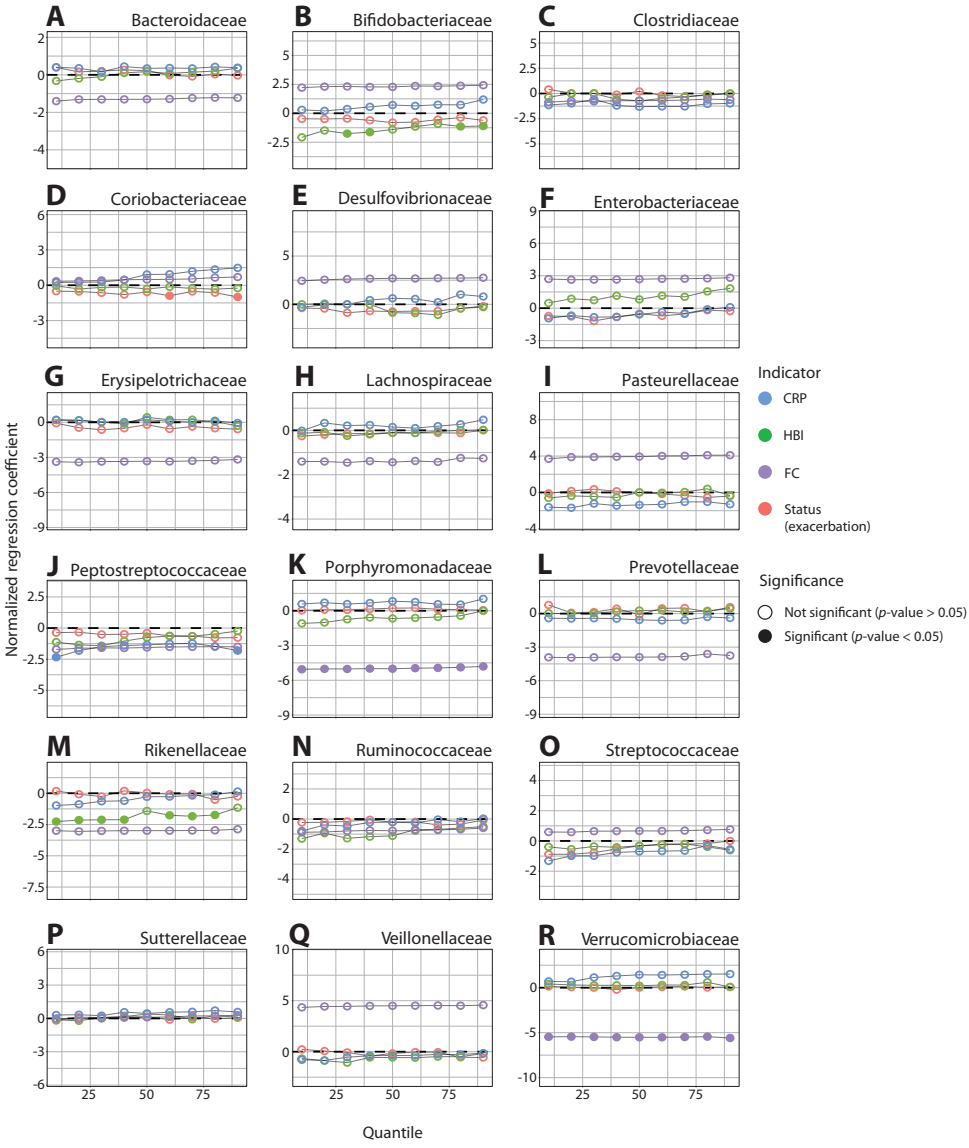


Appendix Figure 4.15 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and clinical variables in CD patients only. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (p-value < 0.05) are indicated with an asterisk ('*').

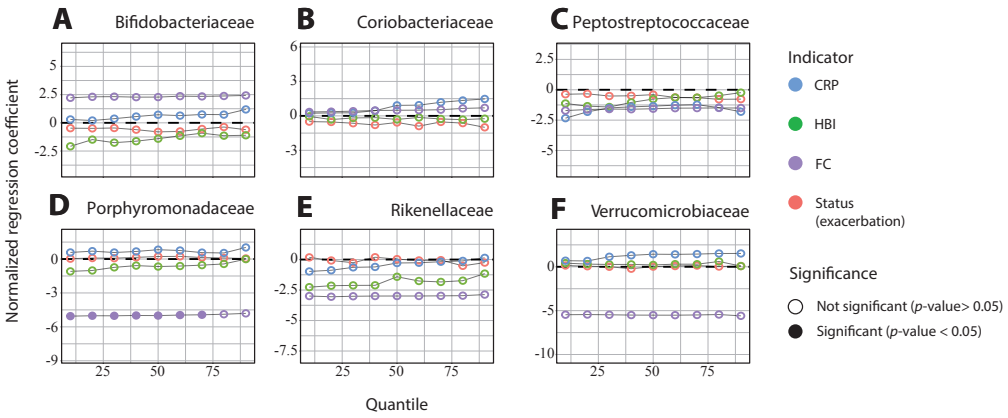


Appendix Figure 4.16 - Heatmap of quantile regression estimates across quantiles of the relative abundance for families outside the base case selection criterion and clinical variables in CD patients only, with *p*-values adjusted using the BH procedure. The red boxes are negative estimates, the blue boxes are positive estimates, and the empty boxes are the variables that were not selected during variable selection. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with an asterisk ('*').

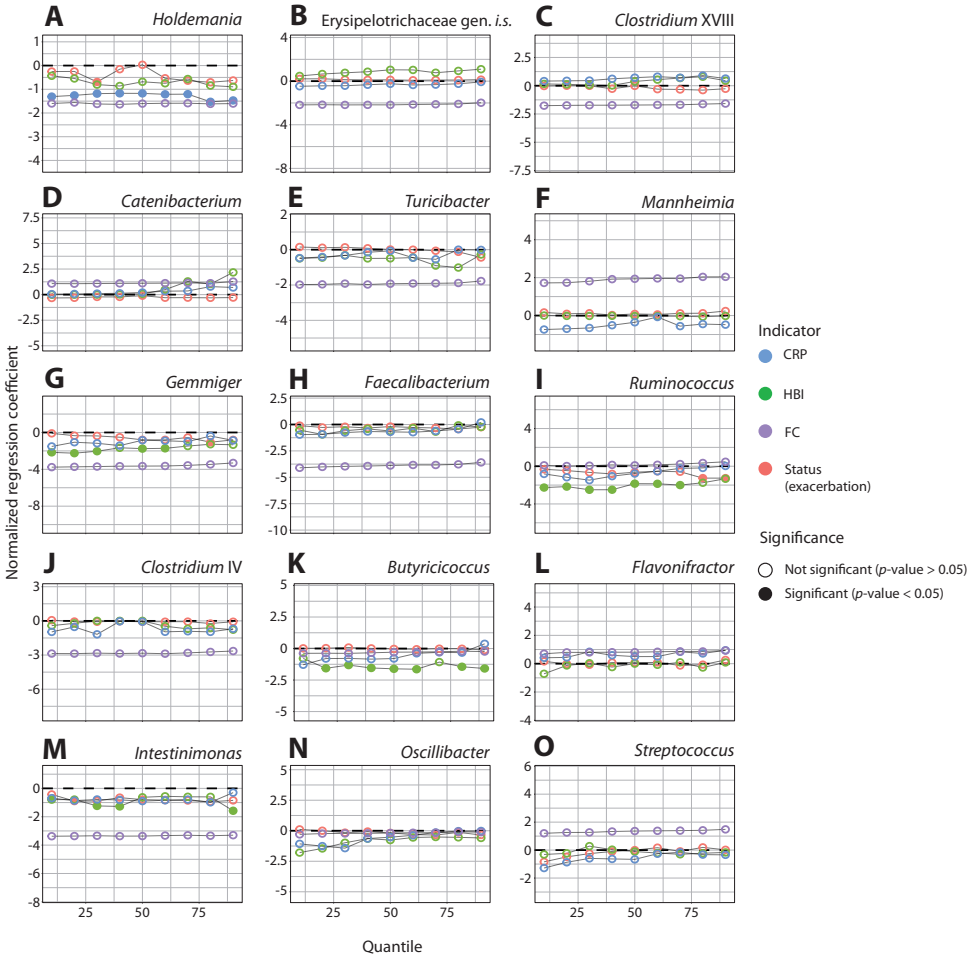
Sensitivity analyses: relative abundances of bacterial families in relation to different disease activity indicators



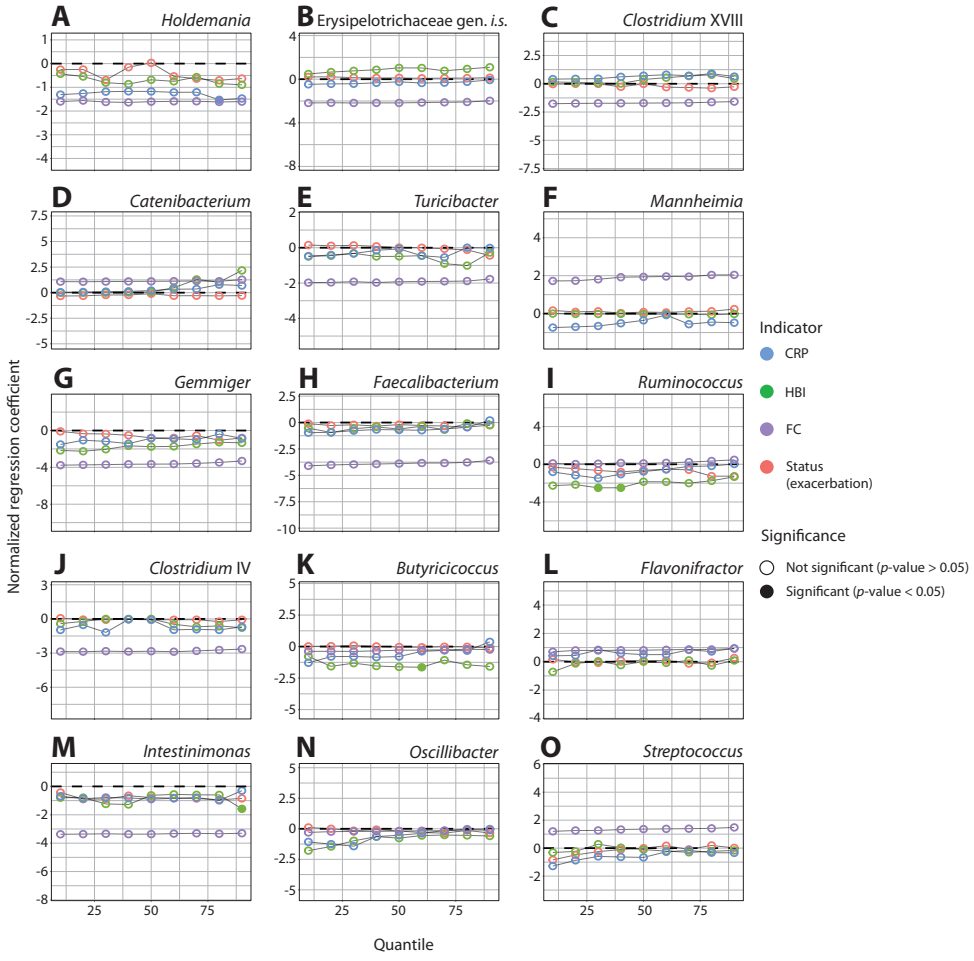
Appendix Figure 4.17 - Quantile regression profile plots for different disease activity indicators and clinical variables for all base case families in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05) are indicated with a closed circle.



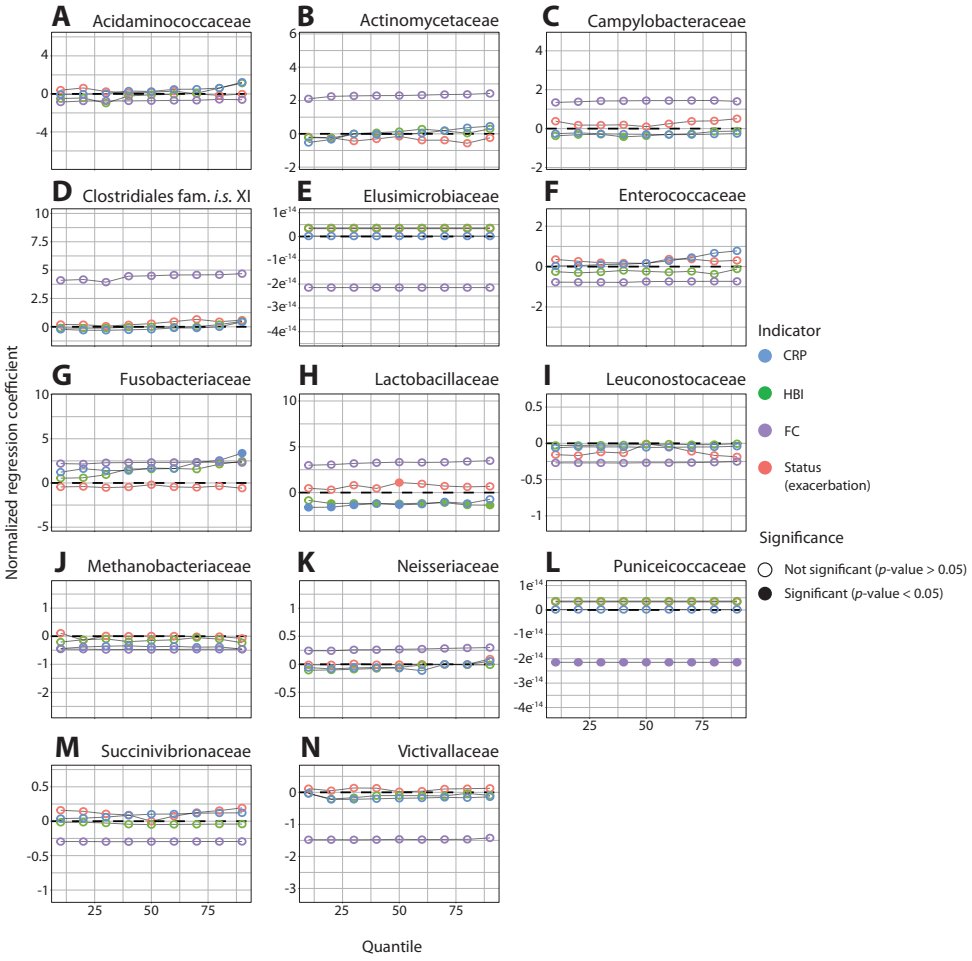
Appendix Figure 4.18 - Quantile regression profile plots for different disease activity indicators and clinical variables for base case families in CD patients only, with p-values adjusted using the BH procedure. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05 after BH adjustment) are indicated with a closed circle.



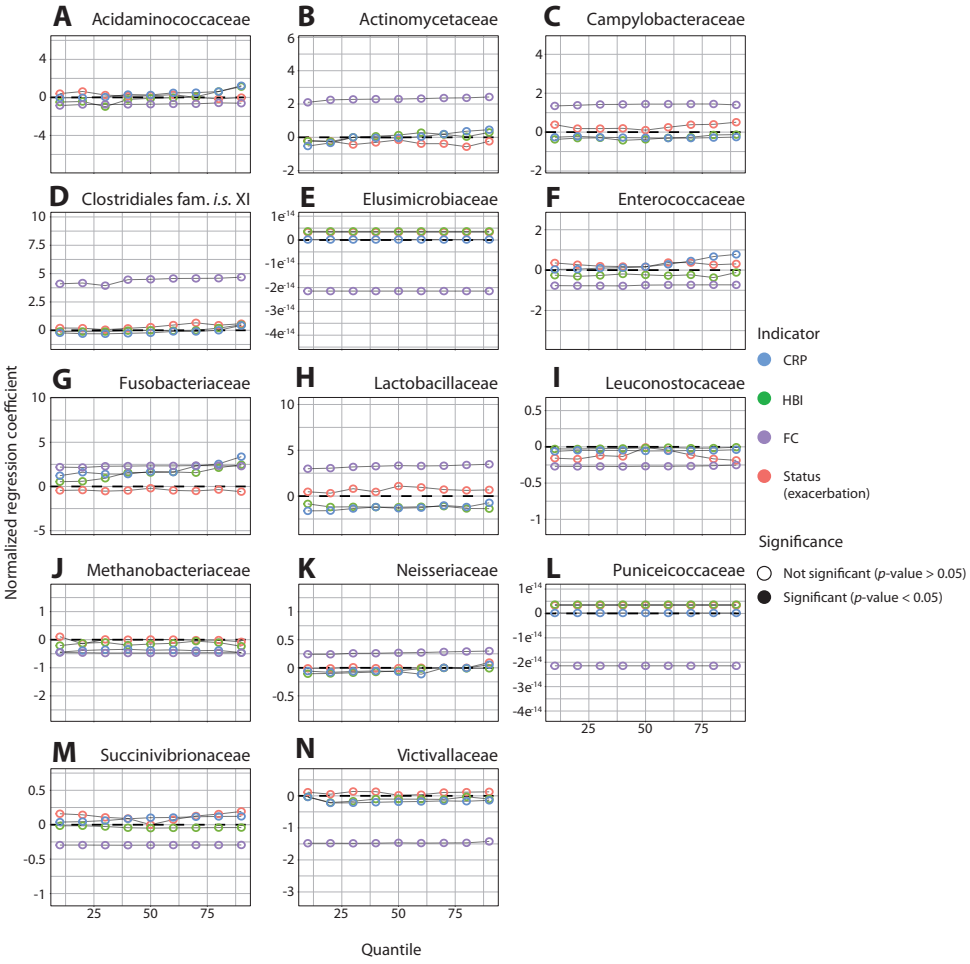
Appendix Figure 4.19 - Quantile regression profile plots for different disease activity indicators and clinical variables for selected genera in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05) are indicated with a closed circle.



Appendix Figure 4.20 - Quantile regression profile plots for different disease activity indicators and clinical variables for selected genera in CD patients only, with p-values adjusted using the BH procedure. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (*p*-value < 0.05 after BH adjustment) are indicated with a closed circle.



Appendix Figure 4.21 - Quantile regression profile plots for different disease activity indicators and clinical variables for families outside the base case selection criterion in CD patients only. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05) are indicated with a closed circle.



Appendix Figure 4.22 - Quantile regression profile plots for different disease activity indicators and clinical variables for families outside the base case selection criterion in CD patients only, with p-values adjusted using the BH procedure. The estimates for clinical status, HBI, CRP, and FC were estimated in different models, therefore the data were normalized beforehand to make the models comparable. For this purpose, the values for HBI, CRP, and FC were divided by the difference between the 5th and 95th percentiles. Significant variables (p-value < 0.05 after BH adjustment) are indicated with a closed circle.

Part III

Ecological determinants of FMT treatment success

Chapter 5

Dynamics of gut microbiota after FMT

Susanne Pinto[1], Dominika Šajbenová[1], Elisa Benincà[2], Sam Nooij[3], Elisabeth M. Terveer[3, 4], Josbert J. Keller[4, 5, 6], Andrea E. van der Meulen-de Jong[5], Johannes A. Bogaards[7, 8*], Ewout W. Steyerberg[1*]

1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
2 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
3 Leiden University Center for Infectious Diseases (LUCID) Research, Leiden University Medical Center, Leiden, the Netherlands
4 Netherlands donor Feces bank, Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands
5 Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, the Netherlands
6 Department of Gastroenterology, Haaglanden Medisch Centrum, The Hague, the Netherlands
7 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
8 Amsterdam Institute for Infection and Immunity (AI&I), Amsterdam UMC, Amsterdam, the Netherlands

* These authors contributed equally

J Crohns Colitis. 2025 Feb 4;19(2):jjae137. doi: 10.1093/ecco-jcc/jjae137.

Dynamics of gut microbiota after fecal microbiota transplantation in ulcerative colitis: success linked to control of Prevotellaceae

Abstract

Fecal microbiota transplantation (FMT) is an experimental treatment for ulcerative colitis (UC). We aimed to study microbial families associated with FMT treatment success. We analysed stools from 24 UC patients treated with four weekly FMTs after randomization for pretreatment during three weeks with budesonide (*n* = 12) or placebo (*n* = 12). Stool samples were collected nine times pre-, during, and post-FMT. Clinical and endoscopic response was assessed 14 weeks after initiation of the study using the full Mayo score. Early withdrawal due to worsening of UC symptoms was classified as non-response. Nine patients (38%) reached remission at week 14, and 15 patients had a partial response or non-response at or before week 14. With a Dirichlet multinomial mixture model we identified five distinct clusters based on the microbiota composition of 180 longitudinally collected patient samples and 27 donor samples. A Prevotellaceae-dominant cluster was associated with poor response to FMT treatment. Conversely, the families Ruminococcaceae and Lachnospiraceae were associated with a successful clinical response. These associations were already visible at the start of the treatment for a subgroup of patients and were retained in repeated measures analyses of family-specific abundance over time. Responders were also characterized by a significantly lower Simpson dominance compared to non-responders. The success of FMT treatment of UC patients appears to be associated with specific gut microbiota families, such as control of Prevotellaceae. Monitoring the dynamics of these microbial families could potentially be used to inform treatment success early during FMT.

Introduction

Ulcerative colitis (UC) is a chronic inflammatory disorder affecting the colon. Symptoms experienced by patients during disease exacerbation include bloody stools, diarrhea, and abdominal pain.[353] The etiology of UC is multifactorial, involving complex interplay between the host immune system, gut microbiota, and genetic and environmental factors.[354, 355, 374] UC patients exhibit reduced microbial diversity and alterations in the composition of their gut microbiota compared to healthy individuals.[355, 375] Notably, a decrease in Bacillota (formerly Firmicutes), especially Clostridia (such as *Clostridium*, *Roseburia*, and *Faecalibacterium*), and Verrucomicrobia, along with an overgrowth of species from the Enterobacteriaceae family (such as *Escherichia coli* or *Klebsiella* spp.), have been observed.[91, 196, 376, 377] Studies investigating associations with common Bacteroidota in the human gut, such as the Bacteroidaceae and Prevotellaceae families, have yielded conflicting results.[196, 200, 356, 376-379]

The current approach to treat UC focuses on attenuating the hyperactive immune response using pharmaceutical drugs, such as local immune suppression with 5-aminosalicylates (5-ASA) or systemic immune suppression with prednisolone, thiopurines, biologics, or small molecules.[196] However, many patients do not derive lasting benefits from these interventions and may even experience severe side effects.[380] Fecal microbiota transplantation (FMT) has emerged as a promising alternative treatment for microbiota-associated disorders, particularly in the treatment of recurrent *Clostridioides difficile* infection.[189, 190, 381] FMT involves transferring fecal matter from a healthy donor to a patient with the aim of modulating the microbiota composition towards a more favourable state. The effectiveness of FMT in UC is limited, with a lower response rate observed as compared to FMT treatment of *Clostridioides difficile* infection.[193] A recent meta-analysis comprising six randomized controlled trials (RCT) reported a short-term clinical response in only half of the patients with active UC following FMT administration.[193] The specific host factors influencing successful FMT response in UC are still unclear, and the donor characteristics that influence patient response to clinical success after FMT remain uncertain.[23, 199]

A small pilot study in patients with Crohn's disease suggests an additional value of FMT in maintaining remission after successful induction therapy with corticosteroids.[380, 382] Achieving or maintaining remission after FMT may be associated with engraftment of donor bacteria.[382, 383] We hypothesized that reducing inflammation promotes engraftment of the healthy donor microbiota, which in turn may result in clinical improvement in inflammatory bowel disease (IBD). To further explore the effects of corticosteroids on engraftment and clinical response, we performed a randomized study investigating the effects of three weeks budesonide pretreatment prior to FMT in patients with UC. The primary analysis showed no association between pretreatment or overall engraftment with clinical response. This may be because the anti-inflammatory potential of budesonide is limited after three weeks. However, there was a significant donor-dependent effect on engraftment, although the study was not powered to detect differences regarding clinical endpoints.[384] In the current study we aimed to further identify longitudinal associations between the microbiota composition and clinical response to FMT treatment. We explored differences in gut microbiota dynamics between patients with clinical remission and non-responders following FMT treatment.

Methods

The study population

For the current study we used the stool samples collected from 24 UC patients included in our previously described FMT trial (Appendix Table 5.1).[384] Patients were randomly assigned to be pretreated daily for three weeks with oral budesonide (9 mg) or with a placebo, and for treatment with FMT suspensions from donor D07 or D08 (block randomization). Inclusion criteria included being at least 18 years old and having a confirmed diagnosis of mild to moderate UC, defined as a full Mayo score ranging from 4 to 9 (including a partial Mayo score and endoscopic sub score of 1 or 2). Exclusion criteria included, among others, proctitis, antibiotic use, surgery within the last 6 weeks, or received other treatments within 12 weeks prior to study entry.

The following clinical and demographic information was collected for each patient in the study (Appendix Table 5.1): sex, age at baseline (years), donor ID (D07 or D08), pretreatment (placebo or budesonide), and clinical outcome at week 14.

Patients who did not complete the study because of progressive symptoms or disease were considered treatment failures and classified as non-responders. At week 14, nine patients were in clinical and endoscopic remission (hereafter called responders), 14 patients were non-responders, and one patient was a partial responder. We included this last patient in the non-responder group.

Clinical and laboratory procedures

Patients received a weekly FMT for four times (at the end of weeks 3, 4, 5, and 6 after randomization) from the Netherlands Donor Feces Bank (NDFB), either from donor D07 or donor D08 following standard protocols for donor screening, sample collection, sample preparation, sample storage, and FMT infusion.[385] The samples used for the different FMTs came from different donations. Before every FMT the patients fasted for at least six hours. A bowel lavage with two liters of macrogol solution (Kleanprep) was performed one day before the first FMT to cleanse the intestine. No changes in diet or medication were reported by the physician who monitored the patients during the study.

Stool samples of the patients were collected once at baseline, once after the pretreatment phase (but still before the FMT treatment), one week after every FMT (four times; designated Post-1 to Post-4), and three times as a follow-up, at 8, 10, and 14 weeks after randomization.[384] In total we collected 81 stool samples in the responder group (n = 9 patients) and 99 stool samples in the non-responder group (n = 15 patients). Stool samples of donors D07 and D08 were collected regularly, and a total of 27 samples (n = 13 samples for donor D07 and n = 14 samples for donor D08) were used for analysis.

Microbiota composition

DNA was extracted from the collected stool samples (both from the donors and recipients) and sequenced by Diversigen (New Brighton, MN, USA) with the Illumina NovaSeq platform (100 bp single-end reads to a median depth of 2.9 million reads). Raw reads mapping to the human genome were removed using bowtie2 (version 2.4.2)[386] and the GRCh37 reference genome, and reads were quality-trimmed using fastp (version 0.20.1),[387] both part of an in-house workflow (git.lumc.nl/snooij/metagenomics-preprocessing). The mOTUs3 workflow (version 3.0.1) was used to generate taxonomic profiles.[388, 389] Unassigned, human-derived, Archaeal, and low-quality reads were removed from the data, which resulted in 93 different families (i.e., 1552 unique mOTUs). The mOTUs3 database includes taxa based on metagenomic bins that have not been formally classified, which are listed as '*incertae sedis*' (*i.s.*). Due to the sparsity of the data and the relatively small number of patients, the analyses performed at taxonomic genus rank lacked the statistical power needed to provide robust and reliable results. For this reason, the data were aggregated to family level prior to the statistical analysis. All analyses were performed using R software (R version 4.2.2) and R code is available on the GitHub repository (susannepinto/FECBUD_microbiome).

Differences in relative abundances of specific microbial families among responders and non-responders were tested for statistical significance in repeated measures analyses, as described in the 'longitudinal models of bacterial relative abundances' section. The average relative abundances of the same bacterial families were calculated for each donor from multiple samples, considering the donor samples were not collected at the same time points as the patient samples. Differences between donor D07 and donor D08 were tested with Pearson's χ^2 test and p-values were corrected for multiple hypothesis testing with the Bonferroni method.

Principal component analysis

We performed principal component analysis (PCA) on the Aitchison distances calculated between each pair of patient microbiota profiles. The Aitchison distance is often used in microbiota data because it takes into account the compositionality of the data.[219, 390] The Aitchison distance involved each patient sample undergoing the centered log-ratio (CLR) transformation and then obtaining the Euclidean distance between each pair of samples, as implemented in the 'microViz' R package.[391]

Dirichlet multinomial mixture models

We used the Dirichlet multinomial mixture (DMM) clustering algorithm to identify distinct clusters of samples based on their microbial abundance profiles. DMM assumes that the microbial abundances in each sample follow one of a given number of multinomial distributions, the number of which is determined by the assumed number of clusters in the data. We used the 'dmn' function from the 'DirichletMultinomial' R package to cluster patient and donor samples.[392] The parameters of the different clusters are estimated by maximizing the likelihood of the observed data given the assumed model, with a Dirichlet prior for relative abundances of the bacterial families to facilitate parameter estimation and prevent overfitting. The prior consisted of a mixture of Dirichlets with $k = 1, ..., K$ to represent the K clusters, with hyperparameters denoting cluster-specific weights and relative abundances. Next, the bacterial families in each cluster were ranked based on the posterior difference between the cluster in a multi-cluster solution versus a one-cluster model. A more detailed description of DMM models is presented elsewhere.[393] Considering that the DMM clustering algorithm uses stochastic likelihood optimization with random initial parameter values, we performed the clustering algorithm 1000 times and chose the model with the lowest Laplace value, indicating a better parsimonious fit of the model to the data.

Data were clustered according to a combination of patient and donor samples. As a sensitivity analysis, we also applied the algorithm in the following situations: patient samples only; patient samples excluding a patient who was placed in a distinct cluster relative to all other patients (patient 102); patient samples excluding patients who both had only two samples available (patients 109 and 117).

Longitudinal models of bacterial relative abundances

Mixed models were used to model the changes over time in relative abundance for each of the 15 most abundant bacterial families in the patient samples. Regarding the distribution of relative abundance, many families had a high proportion of zeros, resulting in right-skewed distributions. All abundances, except for Ruminococcaceae, were therefore transformed with an arcsine square root transformation to approximate normally distributed data. We modelled the relative abundances of the 15 selected bacterial families separately in 15 different longitudinal models with a linear mixed-effects model (LMM), possibly augmented with a zero-inflation component (ZILMM). The 'lme4' R package was used for constructing LMMs and the 'glmmTMB' R package was used for constructing ZILMMs.[394, 395] To account for the correlation of repeated observations within each patient, both random slopes and random intercepts were considered as potential models for each bacterial family. Note that the dataset was too small for the specification of predictors in the zero-inflation component. To incorporate possible non-linearity in relative abundance trajectories over time into the model, a natural cubic spline (with the 'ns' function from the 'splines' R package) with a knot at week 8 (the beginning of the follow-up phase) was considered for all models.

Model preference was based on the lowest Akaike Information Criterion (AIC) and model diagnostics, judged by a QQ-plot and a plot of residuals against predicted values. All choices per family are given in Appendix Table 5.2.

The longitudinal models further included the variables: clinical outcome (non-responder vs. responder), time (possibly with a cubic spline), and an interaction with time and clinical outcome (non-responder vs. responder). The interaction determined whether there was a divergence in the relative abundance of a particular family between non-responders and responders, with statistical significance assessed by Wald tests.[396] The inclusion of the patient-specific variables donor (donor D07 vs. D08), pretreatment (budesonide vs. placebo), age, and sex (female vs. male) in the model was dependent upon testing their role as confounders or contribution to the model fit. This assessment involved examining whether their inclusion led to a greater than 15% change in the primary coefficients (notable influence on the model's outcome) or a significant Likelihood Ratio Test (contribution of the variable to the model fit); with flexibility allowed for a variable to meet one of these criteria during the evaluation process.

Simpson dominance

Simpson dominance was used to summarize microbiota diversity of each sample. We calculated this measure (the sum of the squared relative abundances) with the 'dominance' function from the 'microbiome' R package.[397] The Simpson dominance estimates the probability that two random entities taken from a sample represent the same bacterial family within a patient's microbiota. Hence, a higher Simpson dominance means a higher concentration of species from the same family in the sample, which corresponds with a less diverse microbiota. To account for the correlation of repeated observations within each patient, the Simpson dominance was modelled with a random-intercepts LMM (with the 'lme' function from the 'nlme' R package).[398] A log transformation was applied to the Simpson dominance measure to correct for non-normality. The regression parameter of primary interest was the relationship between Simpson dominance and clinical response, either as a main effect (denoting baseline differences in diversity) or in interaction with time (denoting divergence in diversity between responders and non-responders over time). Additional parameters included the effects of sex and time. Similar to the longitudinal LMM of bacterial families, time was modelled as a continuous variable with a natural cubic spline (knot at week 8). The effects of pretreatment, donor, and age were negligible and therefore not included in the model. Wald tests were performed to test for statistical significance of the clinical response variables jointly in the model.

Results

Microbiota community composition of donors, responders, and non-responders

The fecal microbiota composition between the two donors was distinctly different (Figure 5.1 and Appendix Figure 5.1). Donor D07 had a significantly higher relative abundance of the families Clostridiaceae, Clostridiales fam. *i.s.* (i.e., an unclassified family within the order Clostridiales), Ruminococcaceae, and Veillonellaceae compared with donor D08, whereas donor D08 had a significantly higher relative abundance of Bacillota fam. *i.s.* and Lachnospiraceae (Figure 5.1, Appendix Figure 5.1, and Appendix Table 5.3).

Overall, the most abundant bacterial family in the patients was Ruminococcaceae. However, from the second time point onwards, the relative abundance of Prevotellaceae continued to increase in the microbiota of the non-responders. Prevotellaceae overtook Ruminococcaceae as the most abundant family for non-responders at Post-1 and remained the most abundant for the remaining time points (Figure 5.1, Appendix Figure 5.1, and Appendix Figure 5.2). Compared to the non-responders, Lachnospiraceae and Oscillospiraceae seemed to become more abundant in the responder group over time (Figure 5.1, Appendix Figure 5.1, and Appendix Figure 5.2).

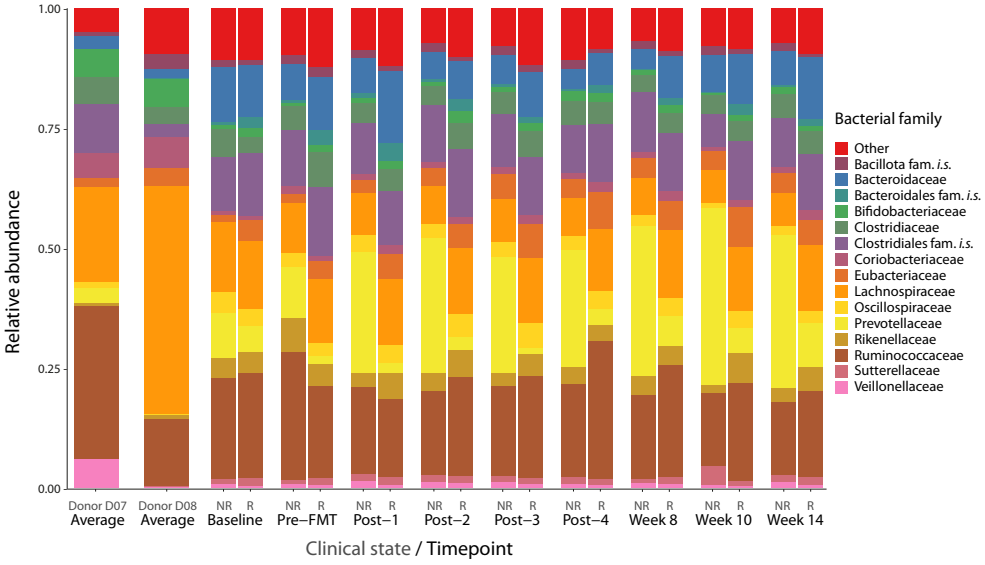


Figure 5.1. Average microbiota composition of the 15 most abundant bacterial families. Abundances were followed over time for the two donors, non-responders (NR), and responders (R). Here, the 'other' category includes all remaining bacterial families.

PCA results for donors and patients

The first two components in PCA of patient and donor samples, based on the Aitchison distance, explained 24% of the total variation in the data (Figure 5.2). The samples of donor D08 clustered away from the patients' samples, driven by a difference in the relative abundance of Lachnospiraceae (Figure 5.2). Patients treated with an FMT from donor D08 showed a higher responder rate than those from donor D07 (Appendix Table 5.1). The difference in distance between non-responders and responders seemed to be explained by the relative abundance of Prevotellaceae (Figure 5.2). This applied particularly to the patients who received an FMT from donor D08 (Appendix Figure 5.3). Only a few patient samples seemed to traverse considerable Aitchison distance over time. Notably, the patients whose microbiota became more donor-like over time were more often non-responders (e.g., patients 110 and 111) (Appendix Figure 5.4).

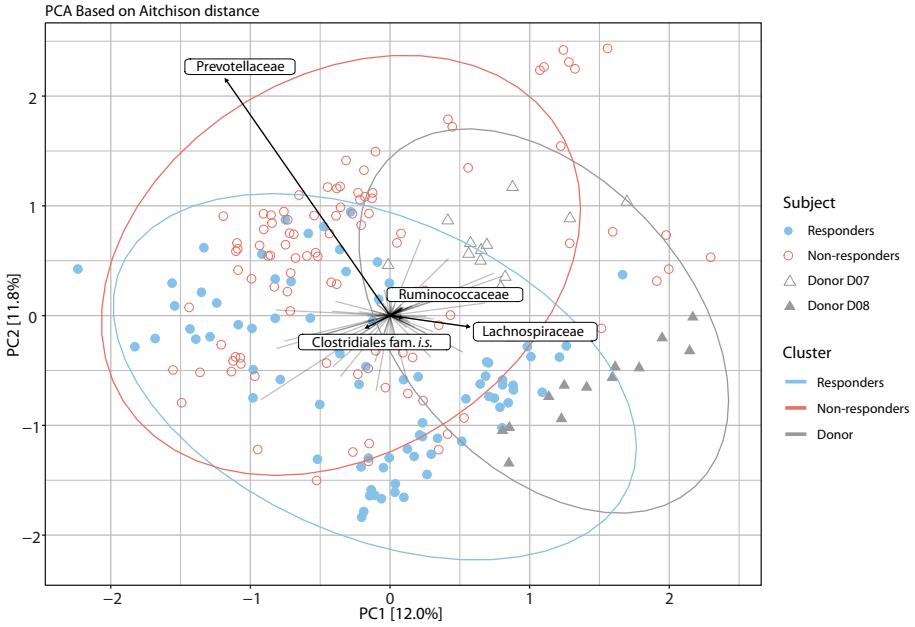


Figure 5.2. PCA plot with Aitchison distances in microbiota profiles, showing the distance between sample types. The PCA plots include data ellipses around the different groups and loading vectors of families to obtain an initial visualization about the extent of separation between non-responder, responder, and donor samples. The different symbols, closed circles, open circles, open triangles, and closed triangles, represent responders, non-responders, donor D07, and donor D08, respectively, while the different colors indicate the various groups (responders, non-responders, and donors).

Sample clustering with Dirichlet multinomial mixture models

Over 1000 iterations, a five-clusters model was selected as the best-fitting model (i.e., having the lowest Laplace value). Figure 5.3 and Appendix Figure 5.5 show that Ruminococcaceae was present in all clusters whereas Lachnospiraceae, Bacteroidaceae, and Clostridiales fam. *i.s.* were present in four of the five clusters. The relative abundances of those families in each cluster differed: clusters 1 and 4 were dominated by Ruminococcaceae and Lachnospiraceae, whereas clusters 2 and 5 were dominated by Ruminococcaceae and Clostridiales fam. *i.s.* Prevotellaceae was the only family almost defining an entire cluster (cluster 3).

Cluster 1 appeared to be associated with a successful clinical response, while cluster 3 appeared to be associated with non-response (Figure 5.4). For the patient samples, 56% of responder samples were classified into cluster 1, and 38% into cluster 2, whereas 42% of non-responder samples were classified into cluster 3 (Figure 5.4B). All donor samples, except for one, were assigned to cluster 4 (Appendix Figure 5.6). Five non-responder patient samples were also assigned to cluster 4 (Figure 5.4A). This donor-dominated cluster disappeared in sensitivity analysis on patient samples only (Appendix Figure 5.7A), resulting in the reassignment of the corresponding patient samples to cluster 2. Patient 102 was responsible for the existence of a separate cluster (cluster 5), with all its measurements belonging to that cluster. Removal of this patient in a sensitivity analysis resulted in the disappearance of that cluster, with re-assignment of the other corresponding samples to cluster 2 (Appendix Figure 5.7B). Removing patients with only two measurements (patients 107 and 119) had a minor impact on the results (Appendix Figure 5.7C).

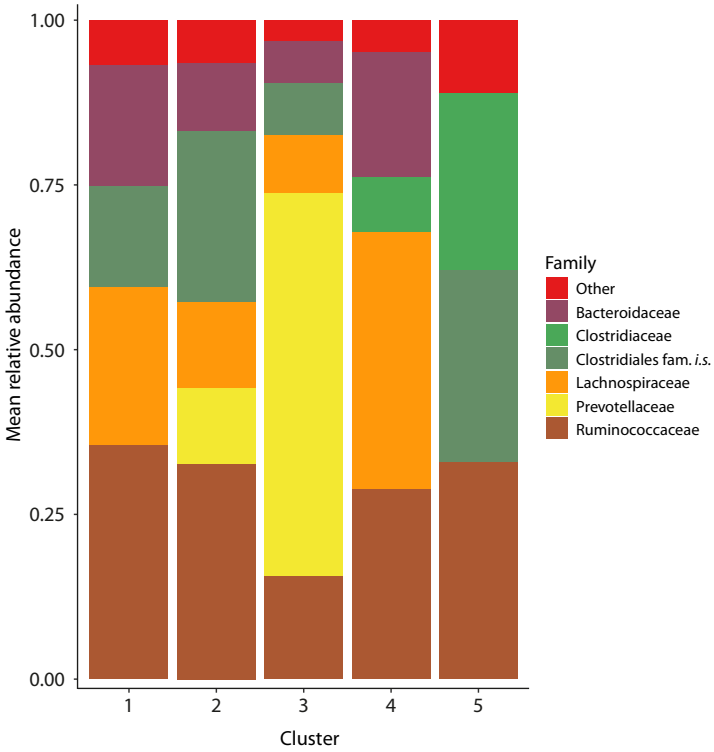


Figure 5.3. Mean relative abundance of bacterial families in the five clusters. Clusters are detected by the Dirichlet multinomial mixture model.

Out of 24 patients, nine patients (38%) remained in the same cluster for all of their provided samples (Figure 5.4A). An alluvial plot of patient samples showed the substantial changes in sample membership and cluster size throughout the clinical trial (Figure 5.4C). There was a mixture of non-responder and responder samples in cluster 1 at the beginning, with most samples at baseline being classified into cluster 1. There was then a shift toward more responder samples in cluster 1 from Pre-FMT onwards. Samples in cluster 1 were exclusively composed of responder samples at time points Post-4, Week 10, and Week 14. Cluster 3 was fully composed of non-responder samples after pretreatment and after every FMT treatment (Figure 5.4C).

Coloring samples by their cluster membership in the PCA plot of Aitchison distances showed separation among clusters 1, 2, and 3, with cluster 2 being the intermediate cluster (Appendix Figure 5.6). The Prevotellaceae vector was pointed in the direction of cluster 3, corresponding to a potential association between this cluster and non-response (Appendix Figure 5.6), possibly driven by the donor (Figure 5.2 and Appendix Figure 5.3). There appeared to be some separation between donor samples, a majority of which were in cluster 4, and patient samples. Donor D08 samples were close to cluster 1 samples. Meanwhile, donor D07 samples were positioned near cluster 2 samples (Appendix Figure 5.6). Finally, samples from cluster 5 were tightly grouped together, likely because they all originated from the same patient.

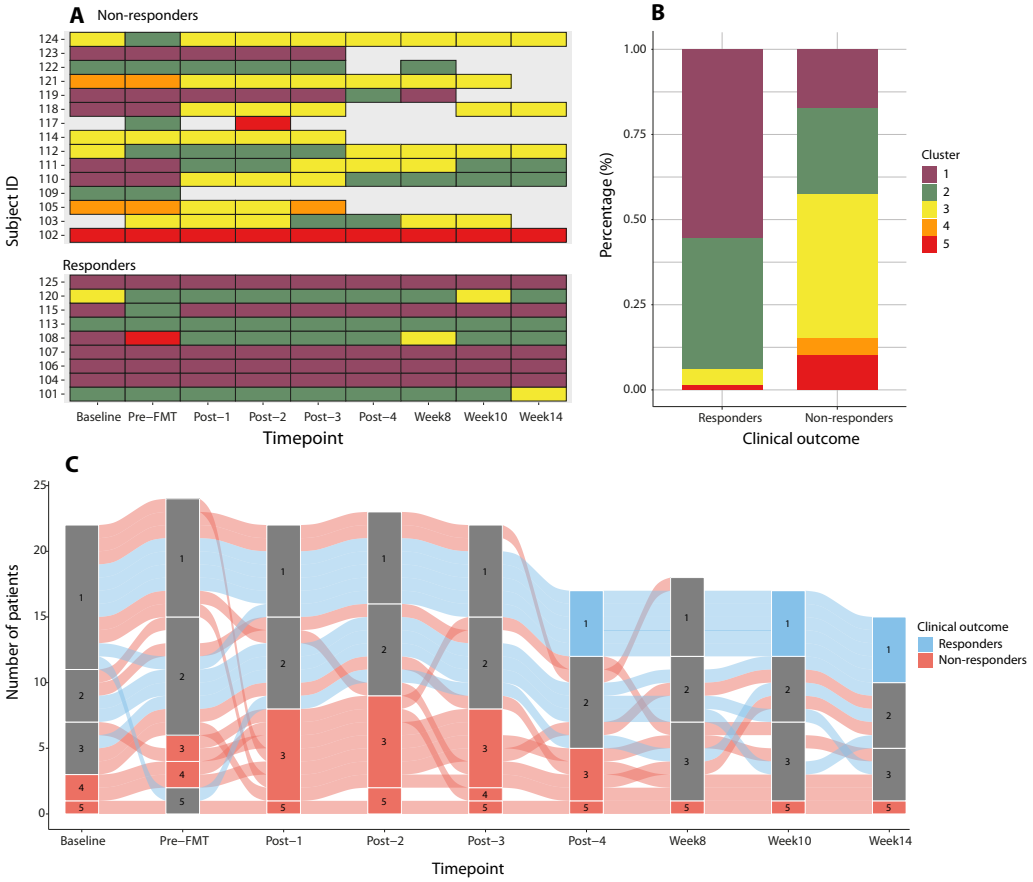


Figure 5.4. Clustering of donor and patient samples. A) Cluster membership over time per patient for the non-responders (upper facet) and responders (lower facet). Lack of colored bar indicates that no stool sample was collected at that time point. B) Percentage of each cluster for non-responders and responders. C) Alluvial plot of patients distributed over the different clusters over time. This plot displays the distribution of clusters per time point and whether each cluster is comprised of only one clinical group (e.g., only non-responders) for every time point. A grey box means that the cluster at that time point contains both samples from responder and non-responder patients, a colored box only contain responder samples or only non-responder samples.

Mixed models of bacterial families

Responders and non-responders showed significantly different trajectories in relative abundance over time for the families Prevotellaceae, Lachnospiraceae, Ruminococcaceae, Oscillospiraceae, and Sutterellaceae (Figure 5.5, Appendix Table 5.2, and Appendix Figure 5.2). Prevotellaceae showed the greatest difference in trajectory between responders and non-responders over time. Note that the preferred model for Prevotellaceae had a straightforward linear trajectory and used only the original time variable instead of splines. The family Prevotellaceae consisted of four named genera, of which *Prevotella* (especially *Prevotella copri*) was the most abundant (Appendix Figure 5.8).

There were four families with a significant donor effect, namely Veillonellaceae, Rikenellaceae, Sutterellaceae, and Bifidobacteriaceae (Appendix Table 5.2). Notably, removal of the donor variable from the model for Sutterellaceae diminished the significance of the main effect

related to clinical response. This observation underscores the role of the donor variable in influencing the association between Sutterellaceae and clinical response. Rikenellaceae and Bacillota fam. *i.s.* had a significant sex effect, Veillonellaceae had a significant pretreatment effect (Appendix Table 5.2). None of these other significant covariates altered the statistical significance of clinical response. This observation suggests that the estimated associations were not confounded by these covariates.

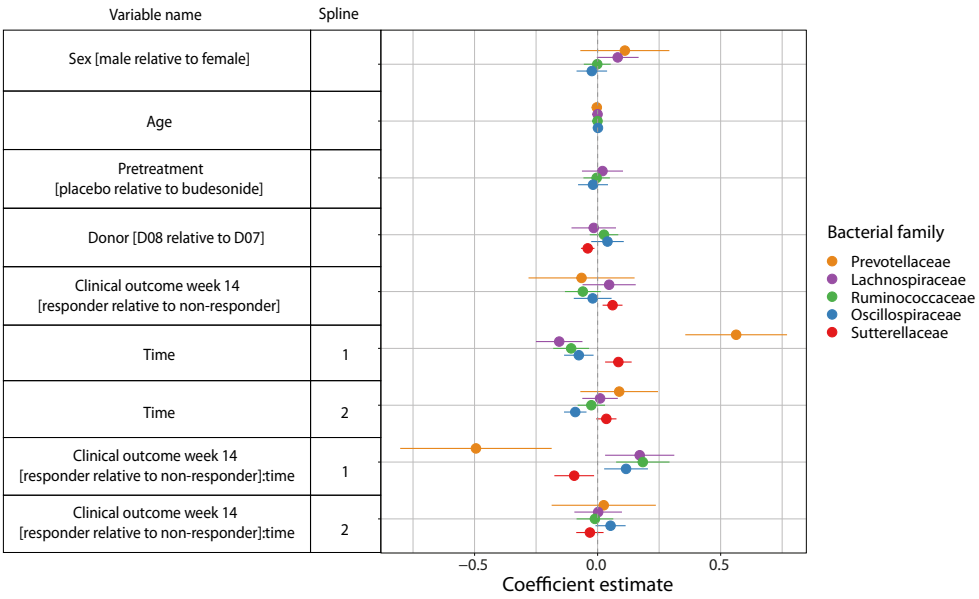


Figure 5.5. Results of the mixed models. Only the families among the 15 most abundant families (Prevotellaceae, Lachnospiraceae, Ruminococcaceae, and Oscillospiraceae) for whom we found a significant effect in relation to clinical response with the Wald test are shown. The point estimates, 95% confidence intervals, and a reference line at zero are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different from 0. All *p*-values are given in Appendix Table 5.2.

Simpson dominance

The steadily increasing relative abundance of Prevotellaceae in non-responders found before was reflected in the Simpson dominance. Simpson dominance was higher for non-responders compared to responders, especially throughout the follow-up period (Figure 5.6). There was a significant difference between the Simpson dominance in responder and non-responder patients (Wald test: *p*-value = 0.004). Our study was too small to determine whether this difference already existed at baseline or developed over time (Appendix Table 5.4). The LMM random-intercept model suggested that there was also a significant sex effect (Appendix Table 5.4). However, sex did not alter the statistical significance of clinical response. This observation suggests that the estimated associations were not confounded by the sex of the patients.

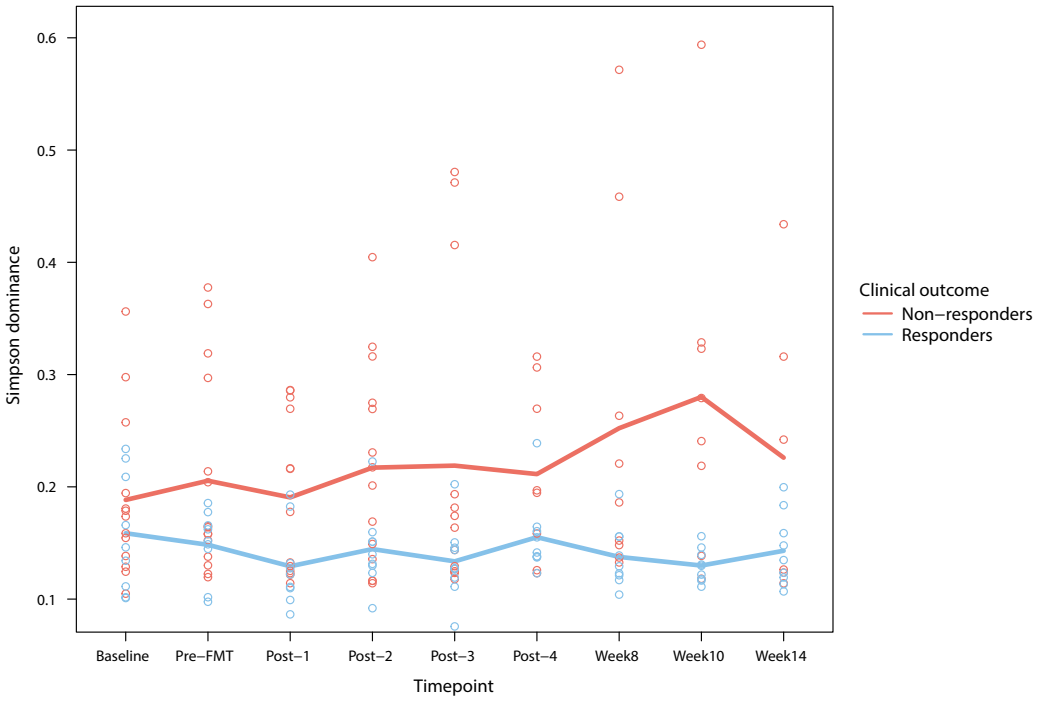


Figure 5.6. Change in Simpson dominance calculated for non-responders and responders. The points indicate the individual measurements of the patients. The lines are the mean Simpson dominance per group.

Discussion

Inflammatory bowel diseases (IBD) such as ulcerative colitis (UC) have been linked to alterations in both the composition and metagenomic function of the gut microbiota.[355, 375] In this study, we employed a wide range of analytical techniques to investigate potential associations between microbiota and clinical outcomes following FMT in UC patients. A subgroup of the cohort (9 of 24 patients) reached a successful combined clinical and endoscopic remission after the FMT treatment, and our results suggest that this response may be related to certain gut microbiota families. Specifically, longitudinal models and cluster analysis of repeatedly measured compositional data indicated that the success of FMT treatment of UC patients appears to be associated with control of Prevotellaceae. Conversely, our analyses also highlighted a potentially beneficial role of Lachnospiraceae and Ruminococcaceae in FMT treatment response. Furthermore, we identified several other bacterial families, including Oscillospiraceae and Sutterellaceae, that exhibited associations with clinical remission. The clustering results indicated that differences in the gut microbiota of responders versus non-responders might already be apparent early during the treatment. If this result can be confirmed by larger studies, clinical success may be predicted from early microbiota analysis after the first FMT treatment and mitigating actions, for example, stopping, personalizing, or changing the treatment, might be envisioned.

Donor-related microbiota characteristics may potentially impact the clinical efficacy of FMT.[198] Intriguingly, we observed marked differences between the donors' and the patients' microbiota. Amongst patients who responded well to FMT, gut microbiota composition did not transition fully to resemble that of the donors at the end of follow-up.

This contrasts with earlier studies that suggested that a donor-like microbiota is preferred after FMT treatment,[23, 198, 199, 383] and suggests that some complementarity in microbiota compositions between donors and recipients is required for a successful clinical response.[199, 200] In other words, the complementarity of the donor-patient pairing seems more important to achieve clinical remission than attaining a donor-like microbiota. The samples of donor D08 clustered closer to cluster 1 (associated with a successful clinical response), and the samples of donor D07 were closer to cluster 3 (indicating non-response). Note that an FMT from donor D08 resulted in relatively more treatment success in the patients than donor D07. Also, donor D08 seemed to have a more diverse microbiota than donor D07, although not statistically significant. Donor gut microbiome diversity has been associated with a higher clinical response before.[399] In addition, higher post-FMT diversity has been associated with remission, suggesting that the variety of introduced organisms may promote recovery.[23] It was already noted that donor D08 was the more successful donor; however, intriguingly, this was the donor with the least engraftment.[384] This observation suggests that the persistent transfer of microbes may not be the prime reason for clinical success. Possibly, the transient exposure to an external microbial community might still induce a beneficial change in the recipient's gut environment. It is also possible that patients who received FMT from donor D08 had a more favourable starting state, while those who received FMT from donor D07 required stronger microbiota changes to move to a more favourable state. Further investigations are warranted to unravel the intricate dynamics underlying the observed outcomes.

This study provides novel evidence for a potential association between control of Prevotellaceae at a moderate abundance and favourable clinical outcomes following FMT in UC patients. In addition, the Simpson dominance measure suggests that Prevotellaceae constituted a sizable proportion of the microbiota in non-responsive FMT patients throughout the course of the clinical trial. Screening the patients (and donors) for Prevotellaceae before and during treatment, and matching donors to patients accordingly might improve the response rate. However, a previous study suggested that higher levels of *Prevotella* (a genus level within Prevotellaceae) may confer health benefits in UC patients after treatment. For instance, studies on UC patients who underwent drug and surgical treatments, excluding FMT, demonstrated that responders had higher baseline levels of *Prevotella* compared to non-responders.[378] Notably, a previous FMT trial on IBD patients did not report any detrimental effects of increased *Prevotella* abundance, despite observing a substantial increase in this bacterium in their patients after FMT treatment.[200] They classified *Prevotella* as a colonizing bacterium, as its abundance in patients reached a level comparable to that in the donors. Of note, in our study, responders also maintained levels of Prevotellaceae comparable to donors, but in non-responders there was a clear overgrowth. The conflicting role of *Prevotella* in human health has been attributed to the high diversity within the *Prevotella* genus. While the majority of *Prevotella* species are commonly found in healthy individuals, certain strains may be implicated in disease pathogenesis.[400, 401] For instance, *Prevotella intestinalis* has been shown to induce intestinal inflammation upon colonization in mice.[379] *Prevotella melaninogenica* and *Prevotella oralis* have been characterized as tipping elements.[402] This means that *Prevotella* stands out as a bimodal group, with either a high or low abundance state, and can be a pivotal driver in the context of microbial ecosystem stability. This finding was reiterated in a recent investigation into the involvement of gut microbiota families with Crohn's disease activity, where we found that associations with Prevotellaceae were among the most heterogeneous across individual patients (see Chapter 4).[403]

In contrast to Prevotellaceae, other bacterial families have shown associations with positive clinical outcomes. Specifically, the families Lachnospiraceae, Ruminococcaceae, and Oscillospiraceae have also been found to increase following FMT in patients with UC in other studies.[195] Lachnospiraceae and Ruminococcaceae may play a role in modulating the immune response and inflammatory pathways in the colon.[195] Earlier attempts to cluster the gut microbiota of healthy and unhealthy individuals showed clusters dominated by *Bacteroides*, *Prevotella*, and *Ruminococcus*.[404-406] While our study identified clusters dominated by Prevotellaceae and Ruminococcaceae, we did not find clusters dominated by *Bacteroides* (i.e., Bacteroidaceae). This discrepancy could be due to differences in the study populations, or the specific methodologies used for microbiota analysis. Interestingly, contrary to previous literature, the expected increase in Clostridiaceae among responders was not observed in the present study. This discrepancy in Clostridiaceae abundance may be attributed to variations in FMT protocols employed across different clinical trials or the low number of patients in this study.[407] In addition, in contrast to the present study, previous research has reported an increased abundance of Enterobacteriaceae in UC patients who did not respond to drug and surgical interventions, with higher levels being associated with mucosal inflammation.[378] Discrepancies in Enterobacteriaceae abundance may stem from differences in the types of UC treatments employed, for example, when FMT was not involved as a treatment modality.[378] In the context of FMT, a study involving IBD patients who underwent FMT revealed the presence of a dysbiotic *Bacteroides* cluster, as well as an Enterobacteriaceae cluster. Donors were subjected to cluster analysis and categorized into *Prevotella* or *Bacteroides* clusters. Interestingly, the clinical outcome of FMT varied depending on the cluster of both the patients and their respective donors.[200]

The longitudinal study design of our trial, with protocolized data collection across all stages of FMT, enabled a uniquely fine-grained view of gut microbiota dynamics during and after FMT in UC patients. Our study allowed us to assess changes on an almost weekly basis. RCTs with a strong longitudinal component often involve a smaller number of patients with more frequent repeated measures, as compared to RCTs that focus on clinical outcomes. For example, in a recent clinical trial 42 patients provided a single stool sample for microbiota analysis before FMT, followed by another single sample after FMT.[200] Another clinical trial included 12 patients who submitted stool samples weekly throughout their 12-week FMT treatment and at the 18-week follow-up.[408] A limitation of our study is that the results of statistical analyses should be interpreted with caution due to multiple tests in a small number of patients. Yet, most associations found in cluster analysis were retained in repeated measures analyses where we also accounted for the correlation of repeated observations within each patient. Moreover, despite the relatively small number of patients (n = 24) and donors (n = 2), both DMM and PCA clustering utilize all 180 patient samples and 27 donor samples available, rather than considering observations per patient.

Microbiota data are compositional, high-dimensional, and often zero-inflated.[217, 219] Moreover, the intestinal microbiota exhibits complex interactions, including competition and cooperation, that form intricate networks.[8, 252] These characteristics pose challenges to analytical methods, such as mixed models, which are commonly employed to investigate temporal variation and potential differences in bacterial abundance trajectories among clinical groups. Our analysis was limited by the individual modeling of each bacterial family, neglecting the interplay and interactions between families within the microbiota network. However, results obtained by supervised models of family-specific abundance over time were in line with results obtained by unsupervised methods (PCA and DMM clustering) that use

community characteristics. Cluster analysis has been widely employed to explore the relationship between gut microbiota and conditions such as child gut development, depression, obesity, and IBD.[200, 409-411] Conventionally, unsupervised methods are suitable for exploratory analyses.[393] If the distinct clusters that we identified are confirmed in further larger-scale longitudinal analyses, this may lead to tailored diagnosis and treatment approaches based on specific cluster characteristics.[412] In our study, this would, for example, mean that the FMT treatment is stopped or changed to another donor when patients are found to be in the Prevotellaceae-dominated cluster during the treatment. While clustering techniques provide valuable insights, it is important to recognize that they depend on various choices by the modeler, including cutoffs and priors, which may lead to different clustering results.

Our study is admittedly rather exploratory in nature, but consistently revealed indications of a potential association between controlled abundances of Prevotellaceae with successful clinical and endoscopic remission following FMT treatment in UC patients. Moreover, we also highlighted a potential beneficial role of Lachnospiraceae and Ruminococcaceae. This provides a basis for new hypotheses regarding the role of gut microbiota in UC. Therapeutic interventions may be refined in the future, with early prediction of clinical outcomes and more personalized FMT treatments.

Appendices of Chapter 5

Appendix Table 5.1 - Clinical and demographic information of responders and non-responders.

	Responders[a]	Non-responders[b]
	Number (Percentage)	**Number (Percentage)**
Patients	9 (38%)	15 (63%)
Samples	81 (45%)	99 (55%)
Missing	0	36
Sex		
% Female[c]	6 (67%)	6 (40%)
Pretreatment		
% Budesonide[c]	5 (56%)	8 (53%)
Donor		
% D07[c]	2 (22%)	10 (67%)
	Mean (SD)	**Mean (SD)**
Age	45 (17)	48 (16)

[a] Remission (i.e., response) was defined at week 14 as no symptoms (partial Mayo score of 2 with no individual sub score of > 2) and an endoscopic Mayo score 0–1.

[b] All other patients, including those with a partial response (a decrease of at least 3 points in the partial Mayo score and at least 1 point at the endoscopic Mayo score) at week 14 and patients who left the study early, were classified as non-responders.

[c] Percentages calculated separately for responders and non-responders.

Appendix Table 5.2 - Model choice and mixed models results for the 15 most abundant families.

Significant results are obtained via a χ^2 statistic (Wald test). Significant results are highlighted in bold and blue. Absence of a *p*-value means that the variable was not included in the model.

Families	Model choice	Sex (*p*-value)	Age (*p*-value)	Pretreatment (*p*-value)	Donor (*p*-value)	Clinical outcome[a] (*p*-value)
Bacillota fam. *i.s.*	LMM (random intercepts)	**0.004**	0.760	0.877	0.535	0.428
Bacteroidaceae	LMM (random intercepts)	0.243	0.377	-	0.794	0.052
Bacteroidales fam. *i.s.*	ZILMM (random intercepts)	0.182	-	-	-	0.546
Bifidobacteriaceae	ZILMM (random intercepts)	0.230	-	-	**0.023**	0.104
Clostridiaceae	ZILMM (random intercepts)	0.377	0.694	-	-	0.439
Clostridiales fam. *i.s.*	ZILMM (random slopes)	0.197	0.280	0.821	0.629	0.909
Coriobacteriaceae	ZILMM (random intercepts)	0.618	0.027	-	0.825	0.146
Eubacteriaceae	LMM (random slopes)	0.509	0.701	0.499	0.337	0.661
Lachnospiraceae	LMM (random intercepts)	0.059	0.904	0.640	0.734	**0.014**
Oscillospiraceae	LMM (random intercepts)	0.459	0.135	0.550	0.233	**0.020**
Prevotellaceae	LMM (random intercepts)	0.230	0.251	-	-	**< 0.001**
Rikenellaceae	ZILMM (random intercepts)	**< 0.001**	0.061	-	**0.038**	0.181
Ruminococcaceae[b]	LMM (random intercepts)	0.963	0.708	0.891	0.381	**0.011**
Sutterellaceae	ZILMM (random intercepts)	-	-	-	**0.004**	**0.010**
Veillonellaceae	ZILMM (random slopes)	0.589	0.503	**< 0.001**	**0.046**	0.435

[a] Wald test on multiple parameters: Responders, Responders x time point (first and second spline)

[b] No transformation

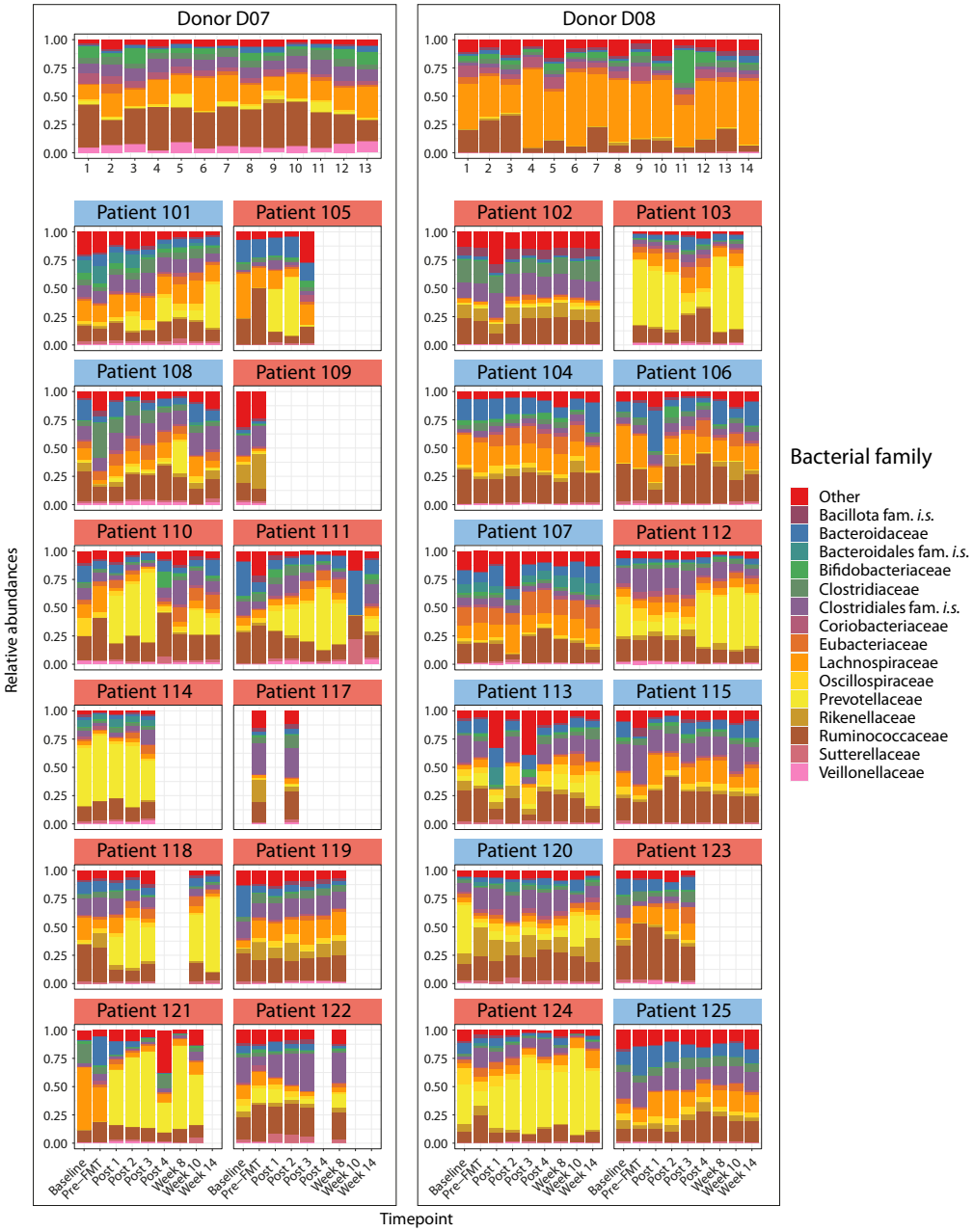
Appendix Table 5.3 - Significant differences in bacterial abundances between the two donors (for donor D07 *n* = 13 and for donor D08 *n* = 14 samples).
The results are obtained with the independence test. Significant results are highlighted in bold and blue.

Family	Mean relative abundance		*p*-value[a]
	Donor D07	Donor D08	
Bacillota fam. *i.s.*	0.0072	0.0327	**< 0.001**
Bacteroidaceae	0.0265	0.0199	0.311
Bacteroidaceae fam. *i.s.*	0.0003	0.0002	0.722
Bifidobacteriaceae	0.0575	0.0575	0.100
Clostridiaceae	0.0572	0.0365	**0.003**
Clostridiales fam. *i.s.*	0.1033	0.0265	**< 0.001**
Coriobacteriaceae	0.0501	0.0660	0.144
Eubacteriaceae	0.0204	0.0367	0.060
Lachnospiraceae	0.1971	0.4755	**< 0.001**
Oscillospiraceae	0.0126	0.0024	0.006
Prevotellaceae	0.0314	0.0000	**0.004**
Ruminococcaceae	0.3183	0.1400	**< 0.001**
Sutterellaceae	0.0018	0.0027	0.371
Veillonellaceae	0.0582	0.0000	**< 0.001**

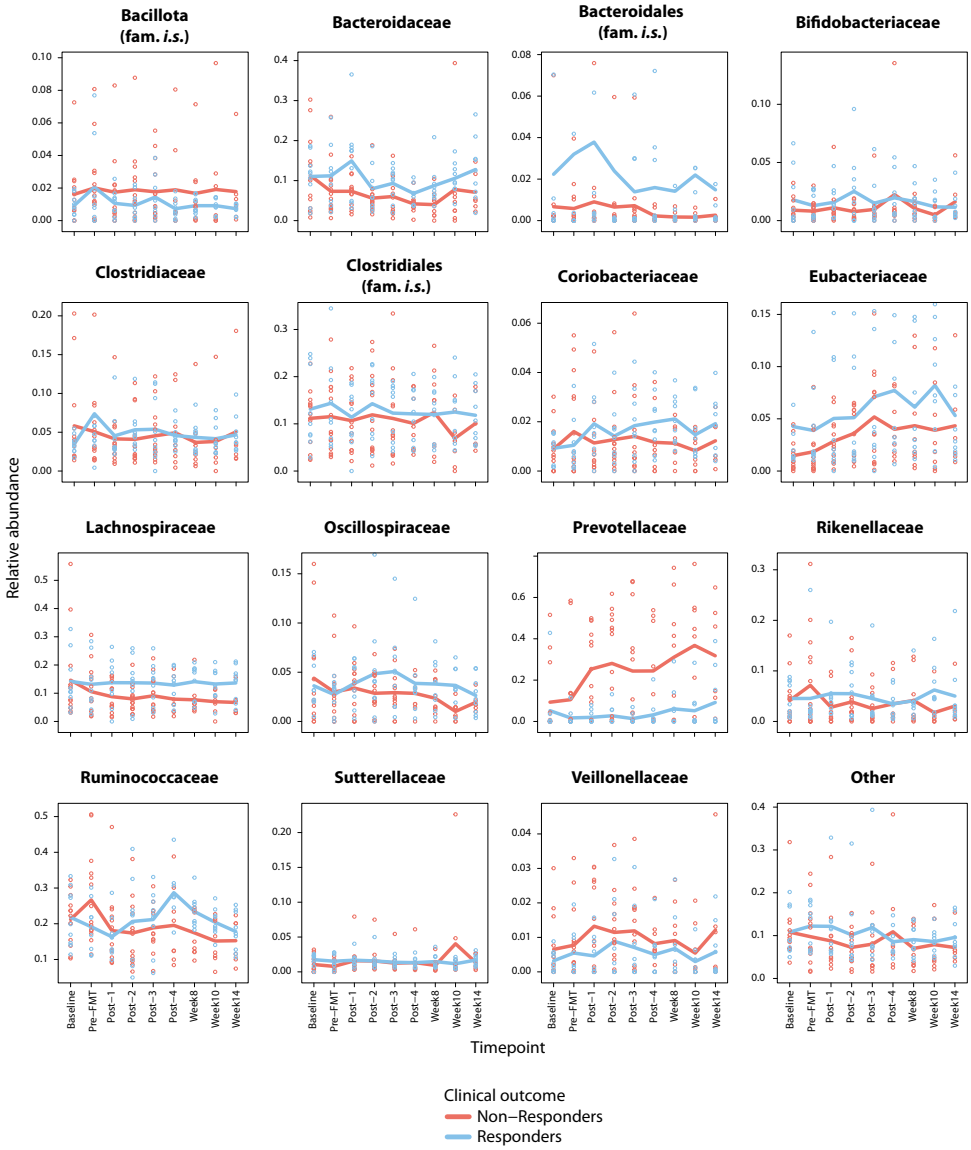
[a] After a Bonferroni correction in which the adjusted *p*-value threshold was 0.004

Appendix Table 5.4 - Regression coefficients and *p*-values of the Simpson dominance random-intercepts LMM.
Significant results are highlighted in bold and blue.

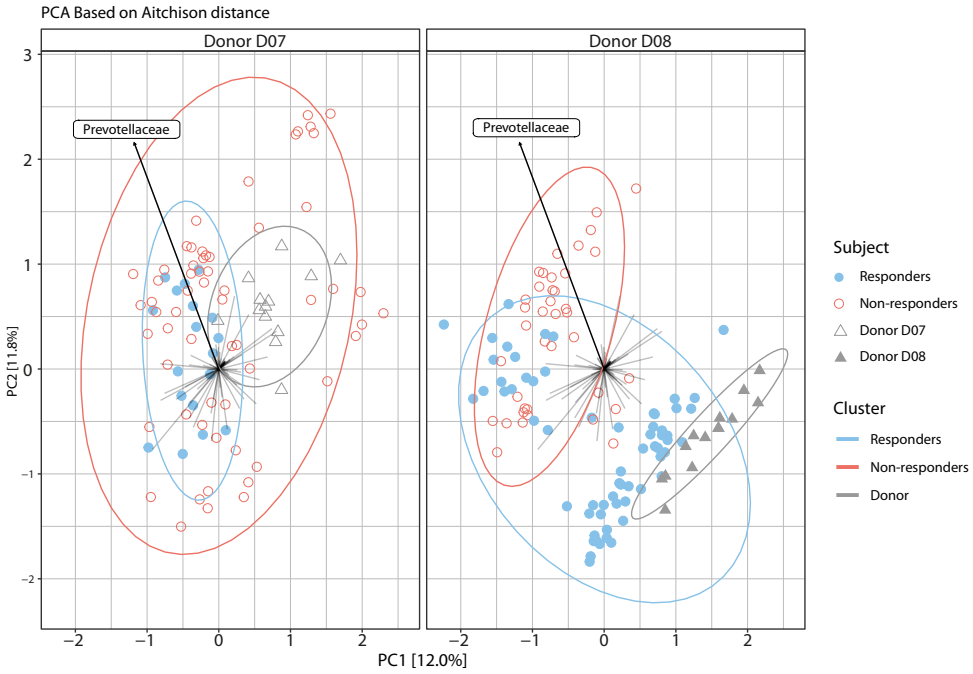
Predictors	Estimates	Standard error	*p*-value
(Intercept)	–1.88	0.1	**< 0.001**
Sex (male relative to female)	0.27	0.09	**0.01**
Clinical outcome (responder relative to non-responder)	–0.14	0.13	0.30
Time (1st spline)	0.30	0.16	0.06
Time (2nd spline)	0.23	0.12	0.06
Clinical outcome (responder relative to non-responder) * Time (1st spline)	–0.2	0.24	0.07
Clinical outcome (responder relative to non-responder) * Time (2nd spline)	–0.22	0.16	0.17



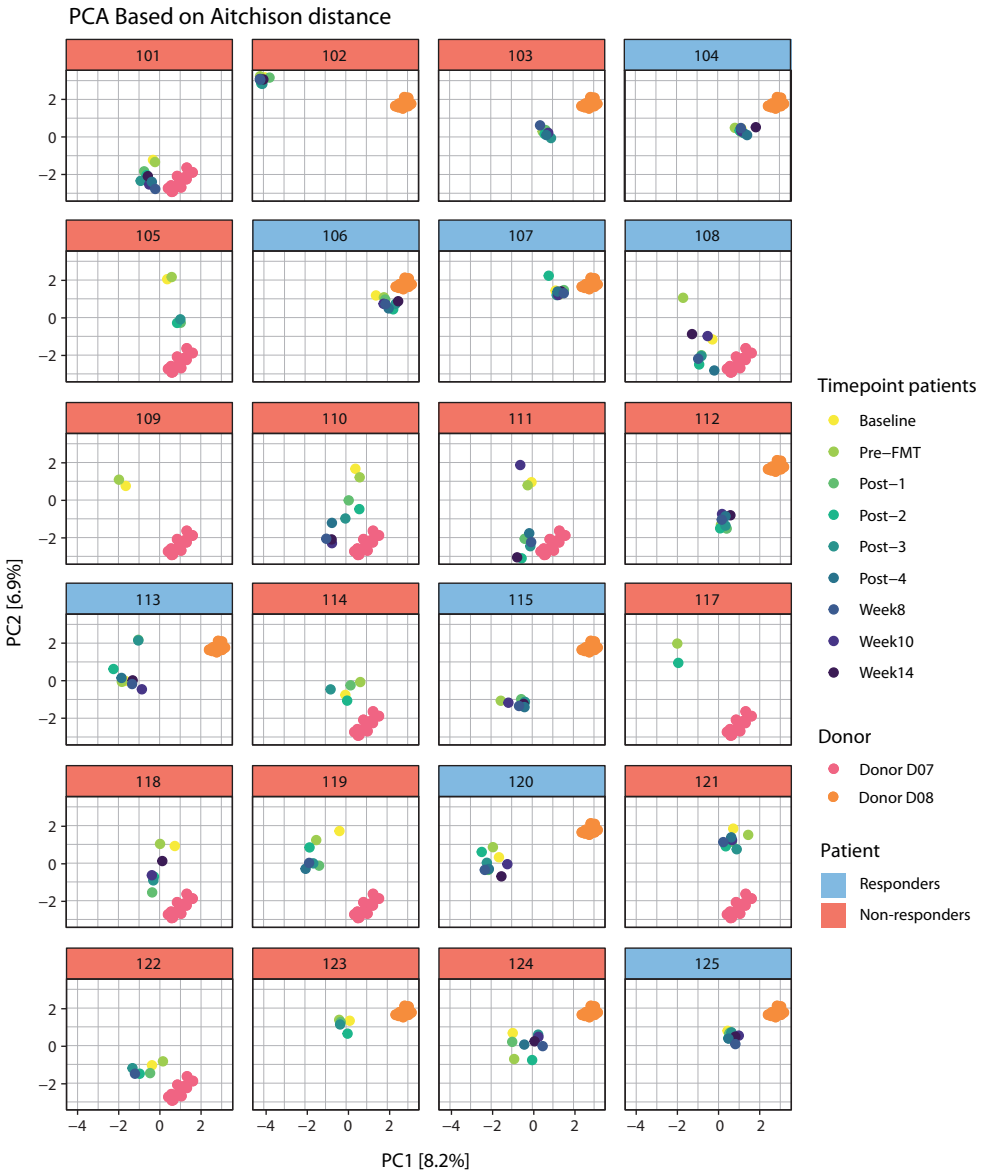
Appendix Figure 5.1 - Composition of the 15 most abundant families in the donors and the patients' microbiota over time. The 12 patients at the left-hand side of the plot (under the plot of donor D07) were treated with feces from donor D07. The 12 patients at the right-hand side of the plot (under the plot of donor D08) were treated with samples of donor D08. Patients with a blue title are responders, patients with a red title are non-responders.



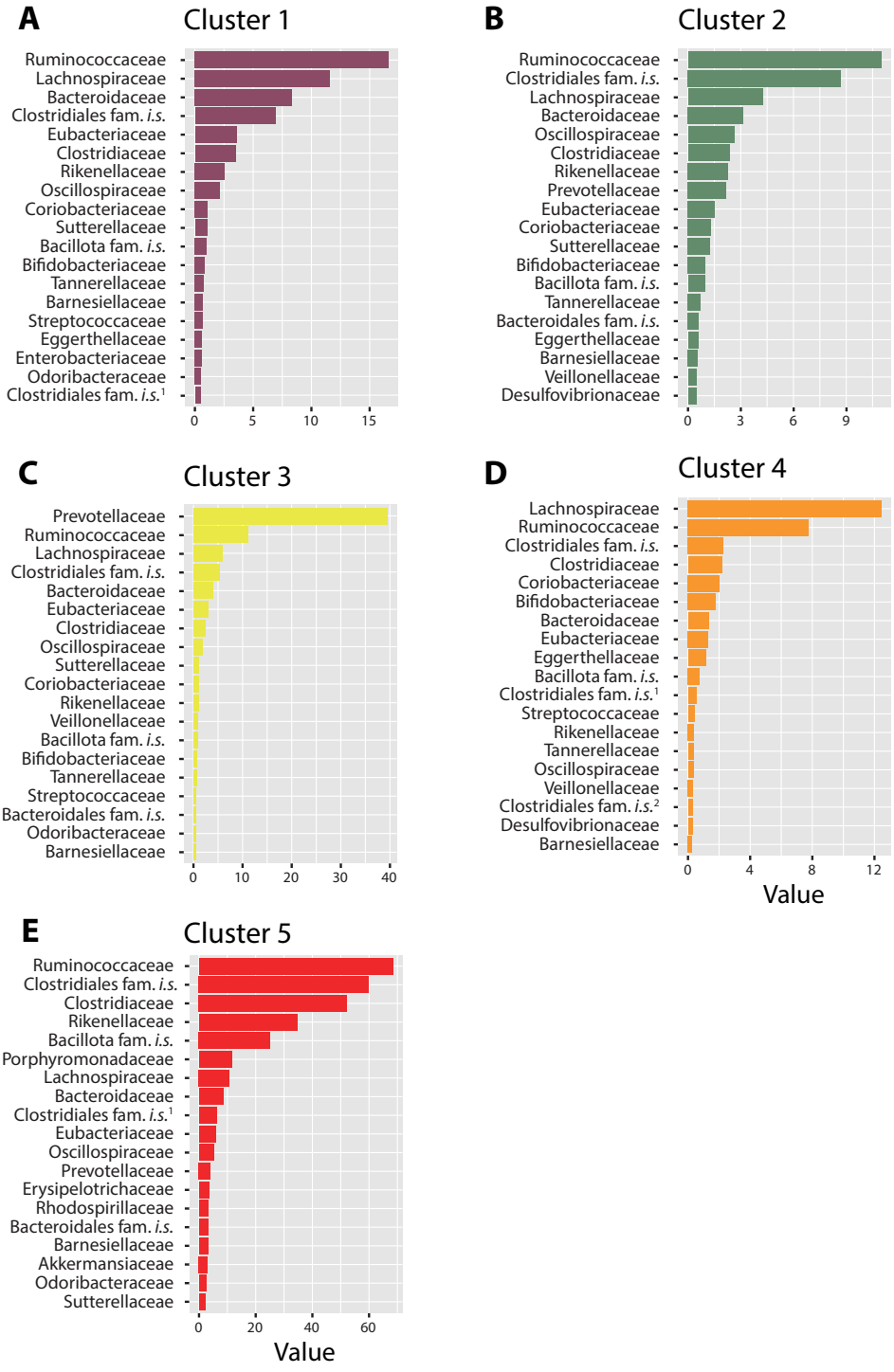
Appendix Figure 5.2 - Relative abundances over time of the 15 most abundant bacterial families. The points indicate the individual measurements of the patients. The lines are the mean relative abundances per group (responders in blue and non-responders in red).



Appendix Figure 5.3 - PCA plot with Aitchison distances in microbiota profiles differentiated per donor. The PCA plots include data ellipses around the different groups (e.g., blue for the responders, red for the non-responders, and grey for the donors) and a loading vector of Prevotellaceae to obtain an initial visualization about the extent of separation between responders, non-responders, and donor samples. The different symbols, closed circles, open circles, open triangles, and closed triangles, indicate responders, non-responders, donor D07, and donor D08, respectively.

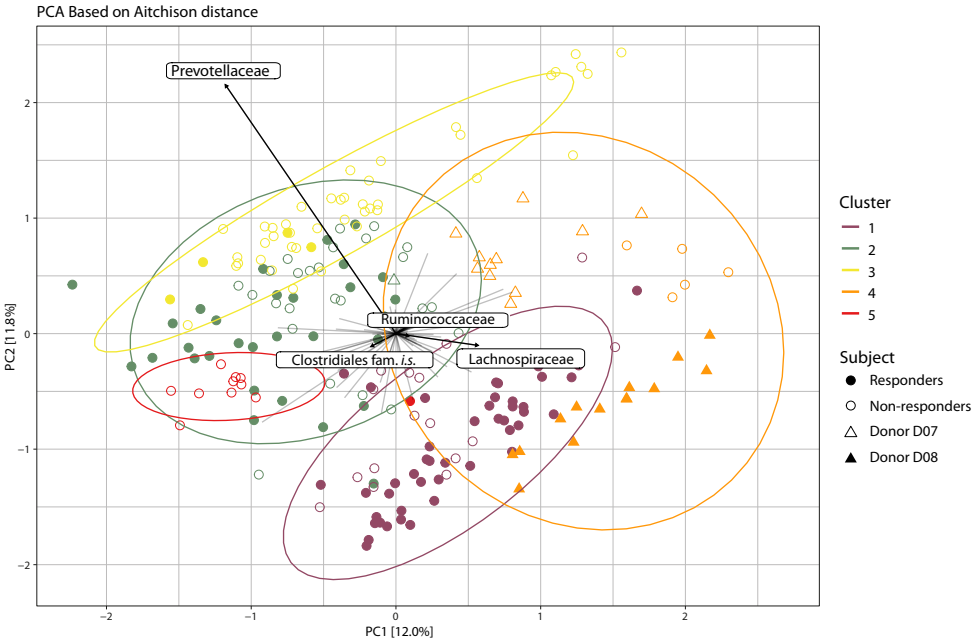


Appendix Figure 5.4 - Plot with Aitchison distances in microbiota profiles differentiated per patient and corresponding donor. Patients with a blue title are responders, patients with a red title are non-responders.

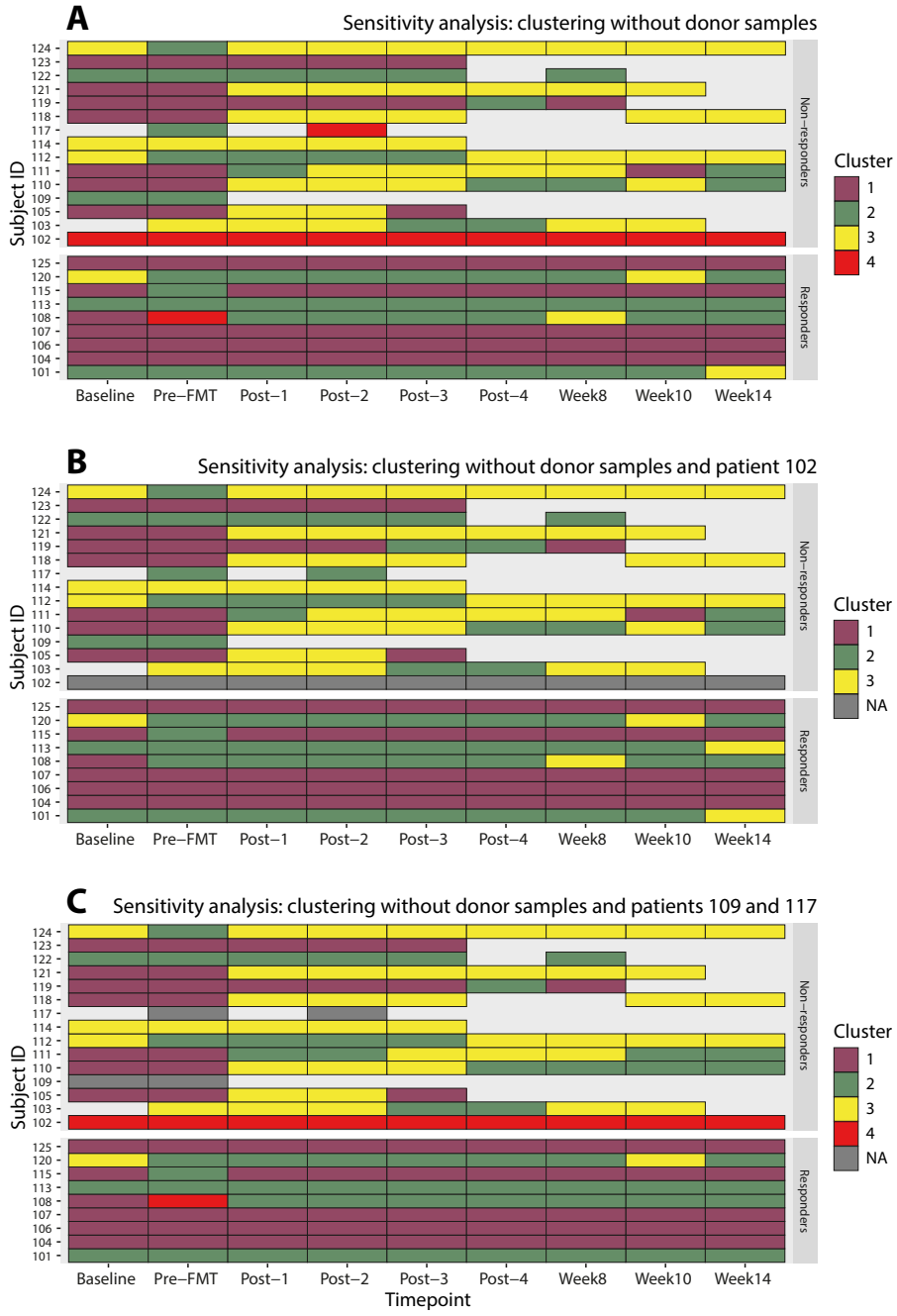


[1] Family is Lachnospiraceae or Clostridiaceae
[2] Family is Eubacteriaceae or Ruminococcaceae

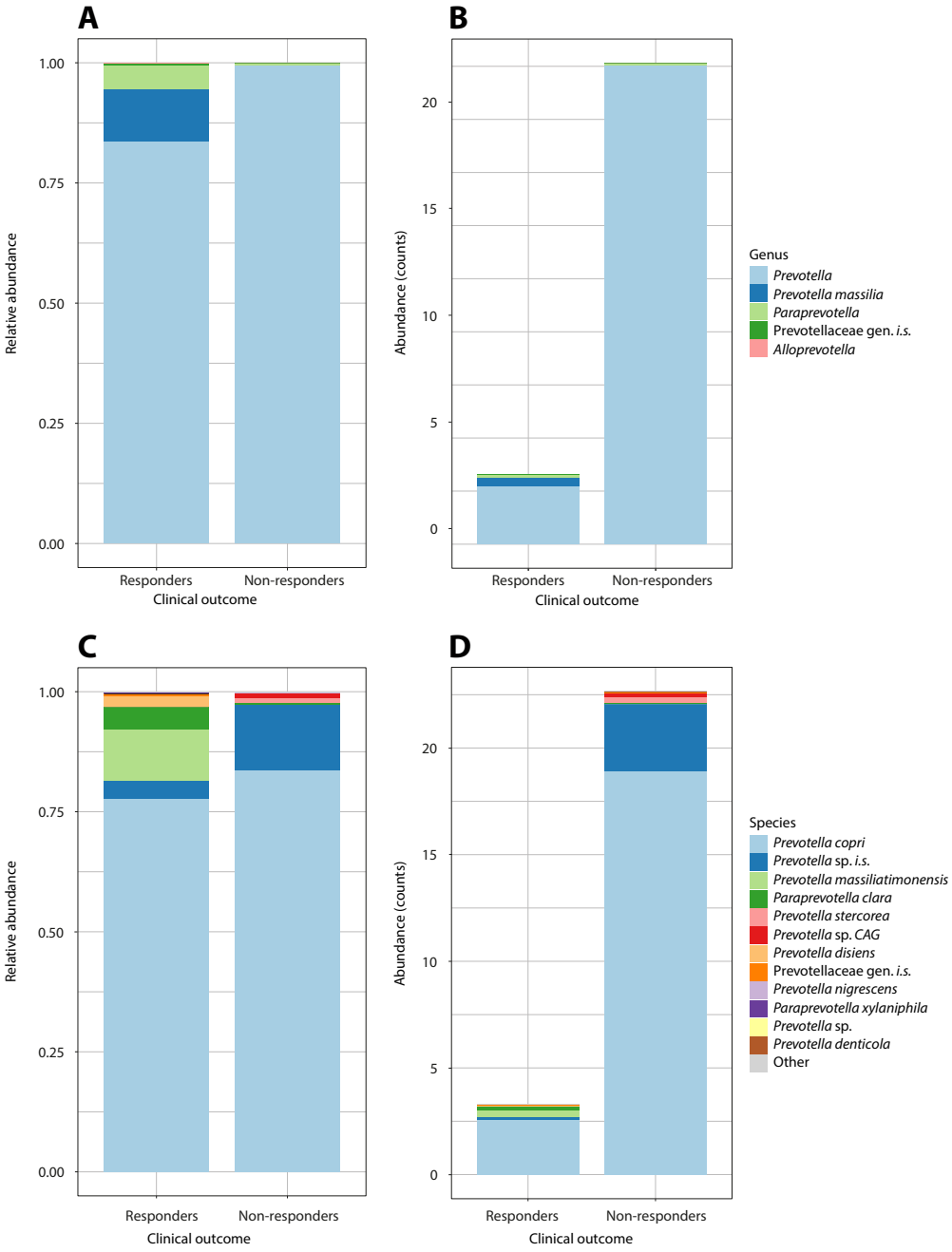
Appendix Figure 5.5 - Importance of the contribution of different families to each cluster. A) Cluster 1, B) Cluster 2, C) Cluster 3, D) Cluster 4, and E) Cluster 5.



Appendix Figure 5.6 - PCA plot with Aitchison distances in microbiota profiles for different clusters, showing the taxa that generally differ across the samples. The PCA plots include data ellipses around the different Dirichlet clusters and loading vectors of families to obtain an initial visualization about the extent of separation between patient (responders and non-responders) and donor samples. The different symbols, closed circles, open circles, open triangles, and closed triangles, indicate responders, non-responders, donor D07, and donor D08, respectively.



Appendix Figure 5.7 - Sensitivity analyses of DMM models. A) patient samples only, B) patient samples excluding patient 102 (with a distinct microbiota from all other patients), and C) patient samples excluding patients 109 and 117 (only two samples available for those patients).



Appendix Figure 5.8 - Genera (panels A and B) and species (panels C and D) within the Prevotellaceae family. Relative abundances (panels A and C) and counts (panels B and D) are given.

Chapter 6

Ecological dynamics of donor and host microbial species following FMT

Susanne Pinto[1], Elisa Benincà[2], Sam Nooij[3], Elisabeth M. Terveer[3, 4], Josbert J. Keller[4, 5, 6], Andrea E. van der Meulen-de Jong[5], Ewout W. Steyerberg[1], Johannes A. Bogaards[7,8]

1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
2 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
3 Leiden University Center for Infectious Diseases (LUCID) Research, Leiden University Medical Center, Leiden, the Netherlands
4 Netherlands Donor Feces Bank, LUCID Medical Microbiology & Infection Control, Leiden University Medical Center, Leiden, the Netherlands
5 Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, the Netherlands
6 Department of Gastroenterology, Haaglanden Medisch Centrum, The Hague, the Netherlands
7 Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
8 Amsterdam Institute for Infection and Immunity (AI&I), Amsterdam UMC, Amsterdam, the Netherlands

This manuscript has been published in *ISME Communications* with a new title:
Ecological resilience in ulcerative colitis: microbial dynamics of donor and resident species in a longitudinal fecal microbiota transplantation study

ISME Commun. 2025 Jul 16;5(1):ycaf119. doi: 10.1093/ismeco/ycaf119.

Ecological dynamics of donor and host microbial species in the treatment of ulcerative colitis with fecal microbiota transplantation

Abstract

Fecal microbiota transplantation (FMT) has emerged as a promising treatment for the chronic immune-mediated disease ulcerative colitis (UC). However, the ecological dynamics underlying clinical remission remain poorly understood. To investigate these dynamics, we analysed data from 24 UC patients treated with four rounds of FMT donated by two healthy donors. Microbiota samples from patients were collected at nine standardized time points before, during, and after treatment, covering a period of 14 weeks. Additionally, 27 donor samples were analysed. Species detected in the recipients' gut microbiota were categorized into ecological categories based on their origin and temporal dynamics: species already present in the host pre-FMT, species derived from the donor, or novel species, i.e., absent before FMT in both host and donor samples but detected later. Overdispersed Poisson regression models with random effects were employed to model the number of species within each category over time. Furthermore, we investigated the change in relative abundance for species present in the host pre-FMT. The results revealed that host species with higher relative abundances prior to FMT were more likely to persist following FMT. Notably, patients who achieved combined clinical and endoscopic remission at week 14 retained a significantly higher number of host species compared to non-responders. In contrast, non-responders initially exhibited a higher colonization of donor species than responders, but colonization rate decreased significantly over time in non-responders. These findings suggest that clinical remission following FMT is associated with a resilient patient gut community, capable of controlled incorporation of donor species, without replacing resident species.

Introduction

Fecal microbiota transplantation (FMT) is the transfer of fecal matter, including gut microorganisms, from the intestine of a healthy donor to a diseased recipient with the goal of modulating the recipient's disturbed microbiota.[189, 190, 381] FMT has been demonstrated to be effective in recurrent *Clostridioides difficile* infection,[189, 381] but the success rate is lower for more complex diseases, such as inflammatory bowel disease (IBD).[193, 413] A possible cause for the lower success rate of FMT in complex diseases is the tendency of the recipient's microbiota to revert to its original pre-FMT adverse state.[23] Transition to a healthier state is likely helped by the successful colonization (engraftment) of donor-derived microorganisms. Therefore, it has been suggested that the success of FMT depends on the donor's gut microbial diversity and composition.[210, 399] The extent to which shifts in the patient's microbiota towards the donor's microbiota are beneficial for resolving gut disturbances remains unclear.[23, 195, 200, 414] This donor-centric view has been challenged, and the importance of the recipient and procedural factors to determine FMT outcomes has been highlighted.[199, 415-417]

In previous analyses of the FMT trial for ulcerative colitis (UC) we examined the engraftment of specific microbial species following FMT, and their associations with clinical remission (see also Chapter 5).[384, 414] For this, we analysed the data from a randomized controlled trial (RCT) involving 24 UC patients treated with four rounds of FMT donated by two healthy donors. Interestingly, we observed that the rate of microbial engraftment did not correlate with successful clinical remission,[384] a paradox also noted in a meta-analysis conducted by Schmidt et al. (2022) involving 316 FMT procedures.[199] In their study, clinical success was not correlated with donor strain colonization or replacement of recipient species. Instead, recipient factors seemed to play a more important role in determining FMT outcomes than donor-related factors.[199] The seemingly limited role of engraftment in predicting clinical outcome of FMT defies the super-donor hypothesis and necessitates deeper investigation into the ecological changes underlying clinical remission.

In this study, the role of donor and host microbial species in determining clinical outcome of FMT is investigated further by applying the conceptual framework introduced by Schmidt et al. (2022)[199] to a longitudinal setting. We capitalize on a randomized controlled trial[384] with dense repeated sampling to map the succession dynamics in the recipient's gut microbiota of UC patients following FMT treatment in relation to clinical remission. Our analysis focuses on ecological dynamics on a species level, categorizing all taxa based on their origin and temporal presence: already present in the host before FMT, derived from the donor, or detected during or after the FMT therapies while absent in both the pre-FMT host and the donor.

Methods

The study population

A total of 24 adult patients experiencing mild to moderate exacerbations of UC were included in a double-blind randomized controlled trial conducted at LUMC.[384] Written informed consent was obtained from all study participants prior to their participation. Demographic variables and subject characteristics are provided in Appendix Table 6.1, with further details on the study population and clinical characteristics provided by van Lingen et al. (2024) and in Box 6.1.[384]

Following pretreatment with either budesonide (n = 12) or placebo (n = 12), patients received four fecal transplants at weekly intervals. Donors (D07 and D08) were randomly assigned. FMTs were infused in the duodenum via a nasoduodenal tube or gastroscope.[385] Stool samples were obtained before and after the pretreatment phase, before every FMT (four times), and 1 week, 4 weeks, and 8 weeks after treatment. At the end of the study, at week 14, a sigmoidoscopy was performed to assess the endoscopic Mayo score. Remission (i.e., response) was defined at week 14 as no symptoms (partial Mayo score of 2 with no individual sub score of > 2) and an endoscopic Mayo score 0–1. Partial remission was defined as a decrease of at least 3 points at the partial Mayo score and at least 1 point at the endoscopic Mayo score. A total of nine patients achieved remission, and one patient achieved partial remission. Of the 14 non-responders, 10 patients left the study early (in total 2 patients did not finish all four FMT treatments) because their symptoms worsened.[384, 414]

For this study, we defined a responder as a patient in remission after FMT (n = 9). Non-responders were defined as having activity despite FMT (non-responders and partial responders, n = 15).

Box 6.1 - Patient inclusion criteria, treatment protocols, and study design. The patients were included in the study if they had a full Mayo score of 4–9 and a colonoscopy with a Mayo endoscopic sub score of 1–2 within four weeks before study entry. Patients were excluded from this study if they had used antibiotics (< 6 weeks), used oral corticosteroids (< 8 weeks), surgical treatment (< 12 weeks), treatment with any investigational drug in another trial (< 12 weeks), significant signs of active infectious gastro-enteritis or enterocolitis, or any other significant medical illnesses. During the study, the medication and diet of the patients was not changed. Patients randomly received daily treatment for three weeks with either 9 mg budesonide or a placebo drug (Appendix Table 6.1). One day before the first FMT a bowel lavage with two liters of Kleanprep (macrogol solution) was performed to cleanse the intestine. Before every fecal transplantation the patients did not eat for at least six hours. The fecal donor suspensions were provided by the Netherlands Donor Feces Bank (NDFB). Collected fecal samples were stored and prepared at the LUMC following standard protocols.[385] Further details on the study population and clinical characteristics are provided by van Lingen et al. (2024).[384]

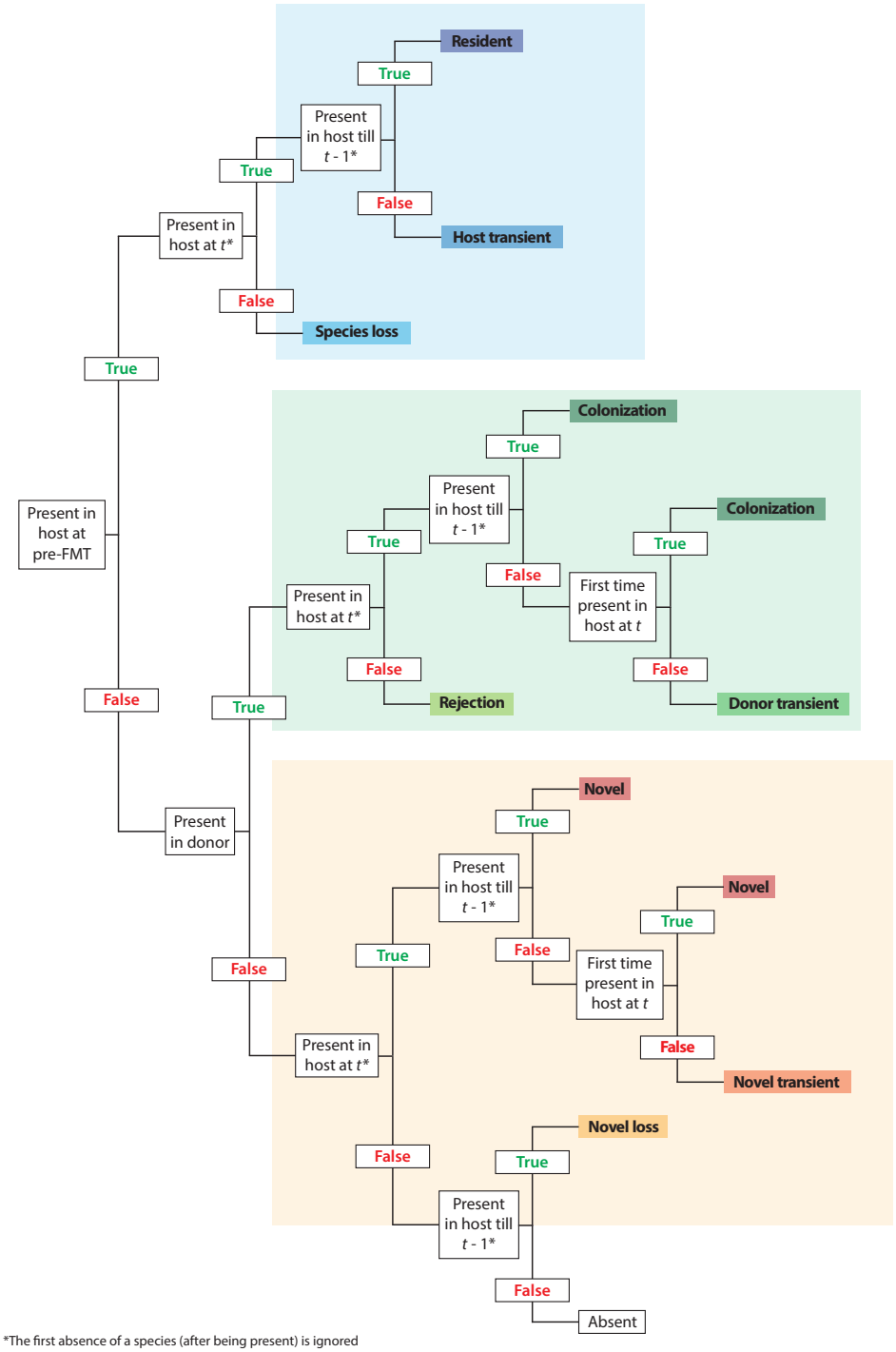
Microbiota data

DNA was extracted from the donor and recipient stool samples and shotgun sequenced with 100 bp single-end reads to a median depth of 2.9 million reads by Diversigen (New Brighton, Minneapolis, USA) using the Illumina NovaSeq platform. Raw reads mapping to the human genome were removed using bowtie2 (version 2.4.2)[386] and the GRCh37 reference genome and reads were quality-trimmed using fastp (version 0.20.1),[387] both part of an in-house workflow (git.lumc.nl/snooij/metagenomics-preprocessing). The mOTUs3 workflow (version 3.0.1) was used to generate taxonomic profiles.[388, 389] Unassigned, human-derived, archaeal, and low-quality reads were removed from the data, which resulted in 1552 unique mOTUs. For the sake of simplicity, we use the term 'species' to refer to unique mOTUs throughout. The results table was then imported into R (version 4.2.2) for analysing the data, visualizing the results and performing the statistical tests. R code is available on the GitHub repository (susannepinto/FECBUD_microbiome).

Mapping ecological categories

Respectively 13 and 14 samples were available for donor D07 and donor D08. Note that every recipient received FMT material from only one of the donors. We could not match every recipient sample to a specific donor sample used for the FMT, because not every donor sample used for FMT was sequenced. Therefore, we created a dataset with the core microbiota for each donor. The core donor microbiota was defined as having its relative abundance higher than 0.1% in at least one sample. The core donor microbiota yielded 120 and 84 unique species for donors D07 and D08, respectively.

Subsequently, we created a presence or absence dataset of all species per recipient and per time point, and every species was assigned to an ecological category per recipient and per time point based on its origin and presence over time, according to the decision tree presented in Figure 6.1 (detailed explanation Box 6.2).



*The first absence of a species (after being present) is ignored

Figure 6.1 - Decision tree used to assign species to ecological categories. The categories are based on the origin and presence of a species over time. First, the species was compared to the pre-FMT host samples, then to the core donor microbiota. Next, the presence or absence at all previous time points was considered to assign the species to an ecological category. Note that we ignored the first absence of a species when categorizing species as lost or as transient upon re-detection. In Sensitivity 1 we evaluated whether this choice had an impact on the results (Box 6.2).

Per recipient, for every species ever present at any time point in the recipient, or present in the microbiota of the associated donor, a comparison was made with the recipient's pre-FMT sample and with the microbiota of the corresponding donor. All species present in the recipient's pre-FMT sample were placed into a host category (Resident, Host transient, or Species loss), depending on the pattern of presence over time. If species were unique for the donor relative to the recipient's pre-FMT samples, species were placed into a donor category (Colonization, Donor transient, or Rejection). If species were not present in the host pre-FMT or in the microbiota of the donor, they were classified as a novel species (Novel, Novel transient, or Novel loss). Within these broad categories, a species was further categorized as a stable (Resident, Colonization, or Novel), intermittent (Host transient, Donor transient, or Novel transient), or previous occupant (Species loss, Rejection, or Novel loss) in the microbiota, depending on the presence at that moment and at the previous time points. Because absence in microbiota data can also mean that the abundance was under the detection limit, in the base case we allowed, for each species, the occurrence of one single absence without direct consequences for categorization in the rest of the time series. Due to the way the categories are defined, some categories cannot occur at the first time points. For example, a donor-derived species first had to colonize the gut (colonization), then be absent for at least two time points (absence ignored (NA) and Rejection), and then be detected again to be categorized as a Donor transient species (Box 6.2).

In sensitivity analyses, we tested some variations to the base case criteria regarding the temporal information used for categorizing the species. In Sensitivity 1 we did not allow the occurrence of any absence when categorizing species into either of the host, donor, or novel categories (Figure 6.1). In Sensitivity 2 we only considered the presence or absence at the previous time point instead of all the previous time points (Appendix Figure 6.3). In contrast, in Sensitivity 3 the presence of species at all time points is considered in the categorization of species at a particular time point (Appendix Figure 6.4). Sensitivity 4 is the same as Sensitivity 3 but with the added criterion of not allowing the occurrence of any absence (Appendix Figure 6.4). In Box 6.2 examples on categorization of species and the differences between the sensitivity analyses are illustrated.

Box 6.2 - Examples illustrating the categorization of the species in the base case and in the four sensitivity analyses.

Sensitivity analyses

In Sensitivity 1 we did not allow the occurrence of any single absence when categorizing species as lost or as transient upon re-detection (in either the host, donor, or novel categories). Secondly, in Sensitivity 2 we only considered the previous time point instead of all previous time points (Appendix Figure 6.3). Therefore, the species can switch more frequently between ecological categories. In Sensitivity 3 and in Sensitivity 4, we considered the full time series (also future points) before assigning them to a category with and without considering a single absence, respectively (Appendix Figure 6.4).

Species present in the host pre-FMT

In the base case scenario, a host species was present in one of the pre-FMT samples of the host (Example 6.1). The resident species has been present up to a specific time point, however, we have ignored a single absence of the species. If the species was absent for two or more time points up to the current one, the species was categorized as a host transient species. The third possible category for a host species is based on the absence of the species at a specific time point and is called 'Species loss'.

Example 6.1 - A species present in the host pre-FMT can be categorized as Resident (Res), Host transient (HT), or Species loss (SL).

			Time point						
	Donor	Host pre-FMT	1	2	3	4	8	10	14
	Absent	Present	Present	Present	Absent	Present	Present	Present	Present
A - Base case			Res	Res	NA	Res	Res	Res	Res
B - Sensitivity 1			Res	Res	SL	HT	HT	HT	HT
C - Sensitivity 2			Res	Res	SL	HT	Res	Res	Res
D - Sensitivity 3			Res	Res	NA	Res	Res	Res	Res
E - Sensitivity 4			HT	HT	SL	HT	HT	HT	HT

Species identified in both the host pre-FMT and the donor are categorized as host species into the groups: Resident (Res), Host transient (HT), and Species loss (SL) (Example 6.2).

Example 6.2 - A species both present in the host pre-FMT and in the donor will be categorized as a host species into: Resident (Res), Host transient (HT), or Species loss (SL).

			Time point						
	Donor	Host pre-FMT	1	2	3	4	8	10	14
	Present	Present	Absent	Absent	Present	Absent	Absent	Present	Present
A - Base case			NA	SL	HT	SL	SL	HT	HT
B - Sensitivity 1			SL	SL	HT	SL	SL	HT	HT
C - Sensitivity 2			SL	SL	HT	SL	SL	HT	Res
D - Sensitivity 3			NA	SL	HT	SL	SL	HT	HT
E - Sensitivity 4			SL	SL	HT	SL	SL	HT	HT

Donor-derived species

A donor species is a species that was not detected in the host pre-FMT, and that was present in the core donor microbiota (Example 6.3). Again, there are three possible categories: Colonization, Donor transient, and Rejection. Species are categorized according to rules similar to how the host species are categorized (Colonization similar to Resident, Donor transient similar to Host transient, and Rejection similar to Species loss). However, a species can still be placed in the Colonization category after being absent for some time points, as it is possible that a species does not colonize directly after the first FMT, but that it needs time to establish in the gut. Note that also in this category a species is allowed and ignored if it is absent once, but only after being present.

Example 6.3 - A species not present in the host pre-FMT, but present in the donor can be categorized as Colonization (C), Donor transient (DT), or Rejection (Rej).

			Time point						
	Donor	Host pre-FMT	1	2	3	4	8	10	14
	Present	Absent	Present	Present	Absent	Present	Present	Absent	Present
A - Base case			C	C	NA	C	C	Rej	DT
B - Sensitivity 1			C	C	Rej	DT	DT	Rej	DT
C - Sensitivity 2			C	C	Rej	DT	C	Rej	DT
D - Sensitivity 3			DT	DT	NA	DT	DT	Rej	DT
E - Sensitivity 4			DT	DT	Rej	DT	DT	Rej	DT

Novel species

A novel species has not been present or was under the detection limit in the pre-FMT host samples, as well as in the core donor microbiota (Example 6.4). Similar to colonizing species, novel species can also enter the microbiota of the host later. However, where a donor species is in that case categorized as 'Rejected', the novel species is not categorized as 'Novel loss', but as 'Absent' and not taken into account in the analyses, until the species has been present once.

Example 6.4 - A species not present in either the host pre-FMT or the donor can be categorized as Novel (N), Novel transient (NT), or Novel lost (NL), from the moment the species appeared in the patient samples.

			Time point						
	Donor	Host pre-FMT	1	2	3	4	8	10	14
	Absent	Present	Present	Present	Absent	Present	Present	Present	Present
A - Base case			-	N	N	N	NA	N	NL
B - Sensitivity 1			-	N	N	N	NL	NT	NL
C - Sensitivity 2			-	N	N	N	NL	NT	NL
D - Sensitivity 3			-	NT	NT	NT	NA	NT	NL
E - Sensitivity 4			-	NT	NT	NT	NL	NT	NL

Modeling the number of species across ecological categories

We modelled the number of species across ecological categories by means of overdispersed Poisson regression models with random effects to accommodate correlation between repeated measurements per recipient. For this, we employed a generalized linear mixed-effects model (GLMM) with a negative binomial family and a log-link using the 'glmer.nb' function from the 'lme4' R package.[395] The temporal evolution of the expected log-number of species in each category was modelled with a spline transformation of the original time variable (in weeks since start of FMT treatment). Estimates from the spline model were compared to those from a linear model in a sensitivity analysis, by modeling the expected log-number of species as a simple linear function of time. Possible differences in succession dynamics between responders and non-responders were investigated by adding the treatment response variable as a covariate to the model, and through specification of interaction terms with time and ecological category. Patient-specific variables, namely, donor (donor D07 vs. D08), pretreatment (budesonide vs. placebo), age, and sex (female vs. male), were included based upon their role as possible confounders.

Change in population abundances of host-derived species

To explore the dynamics of host-derived species in response to FMT in more detail, we investigated the relative abundance over time for the species that were already present

in the host pre-FMT. Results reveal the distribution of abundance differences at particular time points across subjects per ecological category for the species that were already present pre-FMT. In addition, we compared the baseline distributions among species that were later categorized as resident, host transient, and species lost among both responders and non-responders. Finally, we also calculated the differences in microbial abundance before and after FMT for all species that were present in the recipients' pre-FMT samples. Because several non-responder patients quitted early during the study, we only included patients who completed all four rounds of FMT and had at least one post-FMT sample (n = 18 patients, of whom 9 were defined as responders) and used the last available post-FMT measurement when calculating the difference in relative abundance before and after FMT. Because the abundance distributions were right-skewed, we used a natural-log transformation of the abundances. Consequently, the abundance differences on the log scale can be interpreted as proportional differences on the original scale (in percentage differences). To assess the significance of these differences between responders and non-responders, linear mixed-effects models (LMM) were applied, accounting for the correlation of repeated observations within each patient (using the 'lmer' function from the 'lme4' R package).[395]

Results

Succession of host-derived, donor-derived, and novel species following FMT

To study the succession dynamics of species during and after FMT in our UC cohort, we modelled the number of species across ecological categories and investigated differences between responders and non-responders (Figure 6.2). In these models, donor and sex were included as covariates, while pretreatment and age were not relevant as confounders. Appendix Figure 6.1 shows the specific parameter estimates of the model depicted in Figure 6.2.

At the start of the study, we observed a significantly higher number of host species in the resident categories (species that were present in the patient's gut pre-FMT) among responders compared to non-responders, and this difference persisted over time (Figure 6.2A). Although the number of resident species declined over time in both responders and non-responders this decrease was not statistically significant. In contrast, the number of host transient species increased significantly over time in both patient groups (Figure 6.2B). Of note, this increase may be partly attributable to the definition of host-derived species being transient upon re-detection after temporary absence. Non-responder patients exhibited a significantly greater loss of host species over time compared to responders, in whom the number of host species lost decreased significantly over time (Figure 6.2C).

Conversely, non-responders were initially colonized by a significantly higher number of donor species compared to responders. However, the number of colonizing species in non-responders significantly declined over time, whereas it remained constant in responders (Figure 6.2D). The number of donor transient species was similar between the two patient groups at the start of the study and showed a significant increase over time, especially in non-responders. However, this category remained relatively small and differences according to treatment response were not significant (Figure 6.2E). The number of rejected donor species was higher at baseline and over time for non-responders compared to responders, however this difference also did not reach statistical significance (Figure 6.2F).

The number of novel species detected post-FMT was similar for both responders and non-responders and remained constant in time (Figure 6.2G). The number of novel transient species increased significantly over time; this increase was more or less similar for both the responders and non-responders (Figure 6.2H). Initially, the responders lost significantly more novel species than the non-responders, but over time the latter group lost significantly more novel species than the responders (Figure 6.2, panel I).

We also found significant differences between responders and non-responders in the host transient and novel transient categories when applying a linear model instead of splines for the temporal evolution of the number of species in each category (Appendix Figure 6.2). It should be noted that these categories contained relatively few species, and the lack of statistical significance when using splines is likely explained by a reduced statistical power. Importantly, all differences between responders and non-responders identified by the spline model were retained in the linear model for category size (Appendix Figure 6.2).

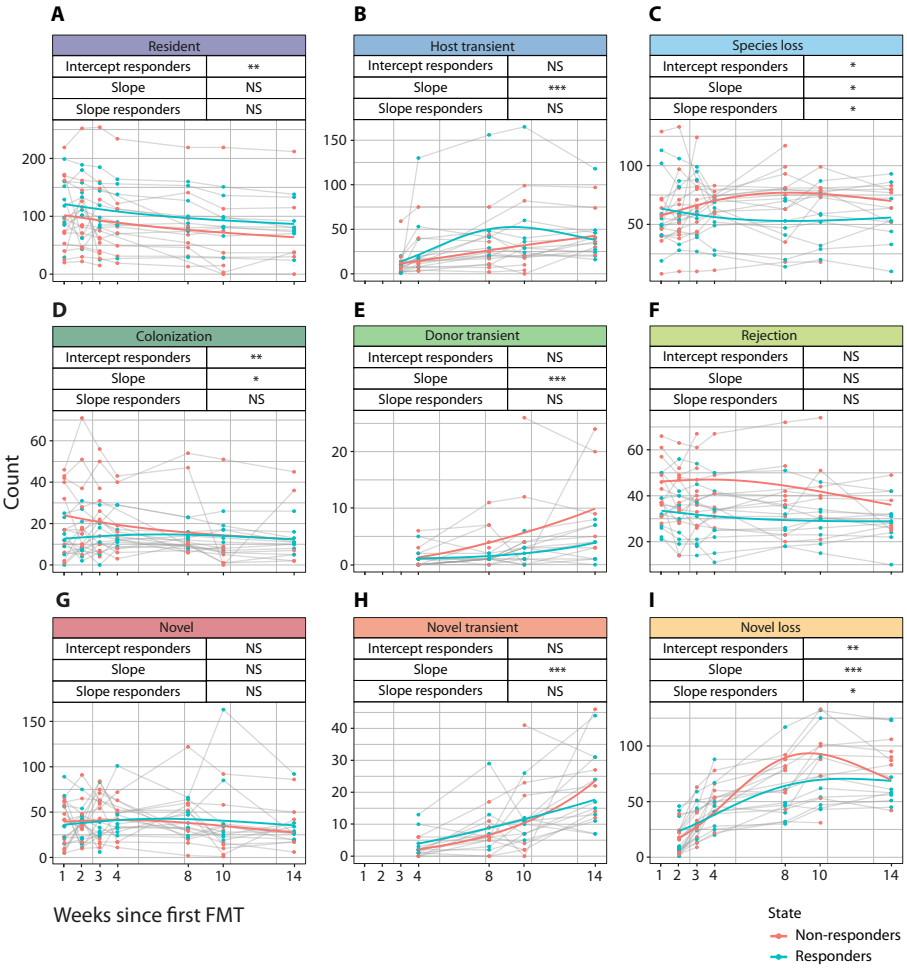


Figure 6.2 – Temporal changes in the number of species per ecological category. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the *y*-axis per category. The model contained a random intercept per patient to account for repeated measurements.

Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05; NS = not significant).

Sensitivity analyses

We conducted four different sensitivity analyses concerning the categorization of the species. To illustrate the effect of categorization on the rates of change over time, we generated a plot of the average slope estimates according to each sensitivity analysis (Appendix Figures 6.5 to 6.10). Sensitivity analysis 1 resulted in a slightly stronger decline in the number of species for the resident, colonization, and novel categories (Appendix Figures 6.5, 6.9, and 6.10). This outcome is a logical consequence of the criterion that a species can no longer be absent for a single time point. Consequently, the likelihood of a species moving to a different category (transient or loss) increased, since it was by definition not possible to return to the categories denoting stable presence over time. This resulted in transient categories having higher intercepts, but the average slopes remained unchanged for all other categories (Appendix Figures 6.5, 6.9, and 6.10). Similarly, for Sensitivity analysis 2, no substantial differences from the base case were found (Appendix Figures 6.6, 6.9, and 6.10). The most profound differences were noted in the slopes of the resident and transient categories. The slopes of the transient categories were smaller, especially for the host-derived species among non-responders (Appendix Figures 6.6, 6.9, and 6.10). Sensitivity analyses 3 and 4 led to more stable patterns over time, especially for the resident category, as compared to both the base case scenario and the other sensitivity analyses (Appendix Figures 6.7 to 6.10). This stability can be attributed to the modifications in the category assignment criteria in Sensitivity analyses 3 and 4, where stable presence is defined at all time points. Consequently, fewer species were assigned to the resident, colonization, and novel categories and more to the transient categories (Box 6.2).

Relative abundances of host resident species pre- and post-FMT

We further assessed changes in the relative abundance of species present in the gut prior to treatment to investigate whether the relative abundance pre-FMT is indicative of the category that a species will reach post-FMT. Host transient species displayed significantly lower relative abundances at all time points compared to resident species (Figure 6.3A and Appendix Table 6.2). In both responders and non-responders, recipient species with higher pre-FMT relative abundances were more likely to remain in the recipient's gut and become resident species, compared to recipient species that were transient or lost (Figure 6.3B, Appendix Figure 6.11, and Appendix Table 6.2). Therefore, our findings show that initial microbiota composition is associated with post-FMT composition. The differences in relative abundance of host resident species between the pre-FMT measurement and the last available post-FMT measurement were centered around zero (Figure 6.3C). A positive difference indicates an increase in the relative abundance of resident species following FMT, while a negative difference denotes a decrease. Thus, approximately equal numbers of resident species showed either a positive or negative response to FMT. No significant differences were found between responders and non-responders regarding relative abundances of resident species in response to FMT (Figure 6.3C, Appendix Figure 6.12, and Appendix Table 6.2).

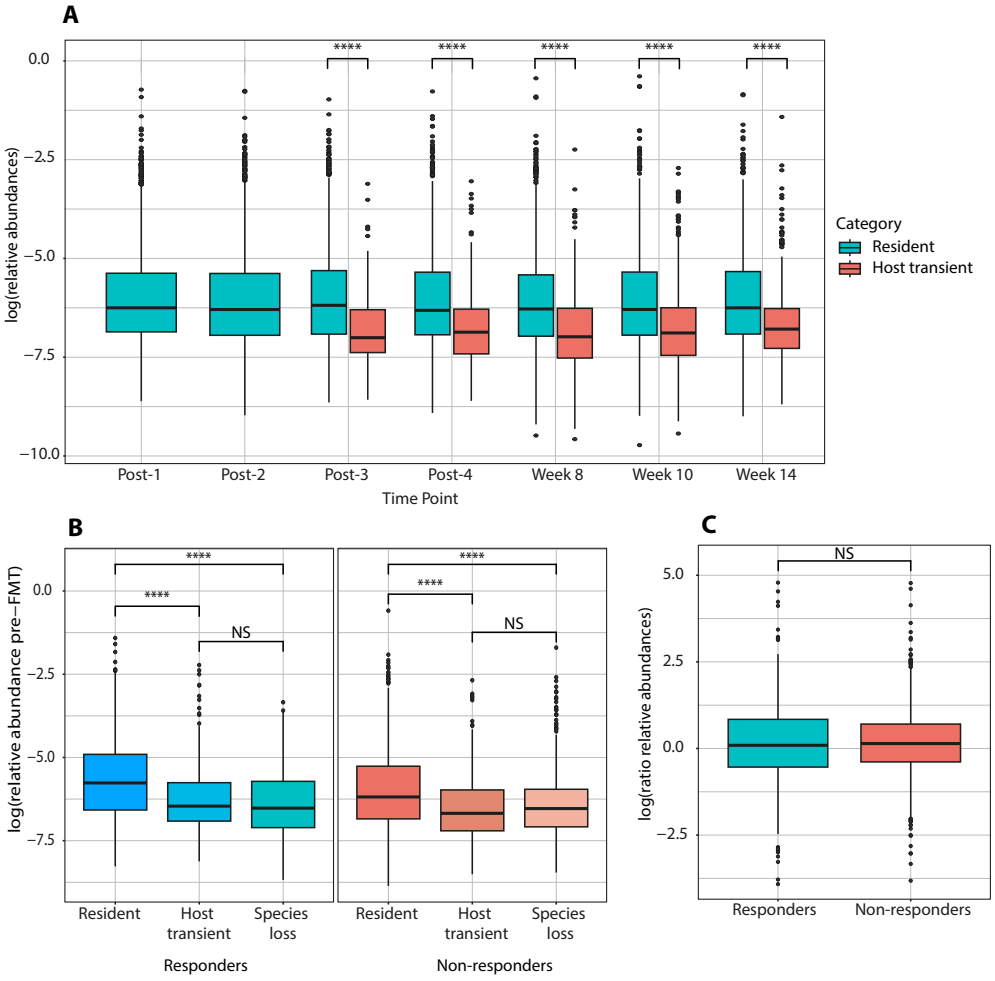


Figure 6.3 - Comparison of relative abundances of species in different categories.
A) Relative abundances of Resident (blue) and Host transient (red) species over time. Here, no distinction has been made between responders and non-responders. B) Relative abundance of host species at pre-FMT measurement. The relative abundances in species categorized as Resident, Host transient, and Species loss species between responders (blue) and non-responders (red) are not significant (Appendix Table 6.2). C) Difference in relative abundance in resident species between pre-FMT and last available post-FMT measurement for responders (blue) and non-responders (red). Significance was tested with linear mixed-models and shown in the plots (**** = p-value < 0.0001, *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05; NS = not significant).

Discussion

The success of FMT for UC is ultimately determined by whether the patient achieves clinical and endoscopic remission after treatment. It has been suggested that treatment success is related to the extent to which the recipient's microbiota composition shifts towards that of the donor.[399, 418] However, we found no evidence supporting this link, in line with several other studies.[199, 200, 384, 414, 415]

We used an ecological framework of succession to investigate microbiota dynamics associated with clinical success of FMT. Microbial species were categorized as pre-existing in the host before FMT, donor-derived, or newly detected. We found that responders retained a higher number of host species compared to non-responders. Although non-responders initially exhibited colonization by more donor species than responders, this colonization in non-responders declined over time and eventually became equal to the levels observed in responders. These findings suggest that a successful clinical response to FMT may be facilitated by a microbiota receptive to colonization without compromising the resident microbiota. Additionally, non-responders lost substantially more novel species over time compared to responders, indicating that newly detected species failed to establish stably within the non-responder gut microbiota. This finding suggests less robust alterations in gut microbiota composition among non-responders. A successful FMT may induce a shift in which the recipient's microbiota integrates donor and novel species, achieving a balanced coexistence to restore the gut microbial ecosystem. This observation aligns with earlier research.[188, 199] Our study expands upon previous analyses using longitudinal analysis of UC patients, thereby providing a fine-grained view of the ecological dynamics over time of donor and host species following FMT.

FMT can be seen as a perturbation experiment on the gut microbiota, creating a dynamic interplay between donor and recipient communities, which may open ecological niches for other microorganisms.[95, 199] The balance between the engraftment of beneficial microorganisms and competition with deleterious microorganisms in the recipient gut, combined with systemic host processes, such as the modulation of immune responses and the interaction with (external) environmental factors and genetic characteristics, could initiate clinical remission.[413] The process of microbial invasion involves various challenges that incoming microorganisms need to overcome to establish colonization and influence the existing microbial community. It is important for the invading species to achieve sufficient metabolic activity in the gut to interact with the resident community. This interaction may also be achieved by transient species, indicating that permanent colonization is not always necessary.[95] Analogous to nurturing an ecosystem such as a crop field through biological control, FMT necessitates the introduction of donor species with healthy functional properties to modify the recipient's system rather than inducing wholesale changes that might lead to the extinction and replacement of existing microbial inhabitants. Therefore, the recipient microbiota must exhibit a degree of resilience, allowing it to integrate donor species without completely altering its composition. FMTs may also strengthen recipient species by introducing beneficial spores or metabolites, thereby enhancing the stability and functionality of the recipient's own microbiota.[103] The stability of the microbiota is maintained through controlled species loss, ensuring that introduced organisms integrate harmoniously with the pre-existing ecosystem.

The outcome of FMT is influenced by a range of ecological processes, spanning from neutral or stochastic factors (e.g., donor propagule pressure) to adaptive or selective factors (e.g., niche competition and differentiation).[199, 419] This indicates a complex mechanism of action of FMT in patients with UC, necessitating the establishment of a novel homeostasis between the donor and recipient microbiota. This complexity may also explain why prolonged FMT treatment with multiple donor infusions appears necessary in UC, as repeated exposure may be required to achieve an optimal balance between recipient and donor microbiota. This approach contrasts with the FMT treatment of recurrent *Clostridioides difficile* infections (rCDI), which is characterized by a depleted microbiota that can be effectively restored with a single infusion, with a cure rate of about 80%.[190]

The success of FMT may not be reliant on resembling the donor's microbiota, but rather on establishing a complementary relationship, emphasizing the importance of selecting donors whose microbiota optimally aligns with the recipient's specific needs.[399] Unlike the developmental stages of a child's microbiota, the gut microbiota of a UC patient is already an established, independent microbial community. This pre-existing microbiota makes the introduction of new species and the induction of change considerably more challenging.[67, 68] Tailoring the selection of FMT donors to those enriched in taxa capable of restoring disturbed metabolic pathways in the recipient might enhance the effectiveness of FMT, particularly in metabolic dysfunction associated diseases.[23, 200, 399] For example, incoming species that are metabolically complementary to the recipient's community, by introducing novel functions or by occupying previously unfilled niches, may be more likely to colonize the resident community.[87, 420] In addition, a high gut microbial diversity in the donor and low diversity in the recipient may further influence the success of colonization.[1, 200]

From an ecological perspective, our findings suggest that donor and recipient species can coexist. We might hypothesize that they occupy distinct metabolic niches. Moreover, we observed that species with a higher abundance prior to FMT (the main 'founders') are more likely to persist during the FMT than species with a lower abundance. This implies that the competitive strength of the resident species is related to their abundance, indicating that within each metabolic niche, communities are built by random winners, driven by stochastic colonization.[84] This is in line with ecological studies showing that functional differences create opportunities for coexistence (niche theory). However, within each niche functionally similar species can coexist, and communities are structured to random stochastic rules (neutral theory).[421] Within the gut microbiota, species often have overlapping functions, allowing them to replace each other and take over specific functional traits if one species is perturbed or removed.[22]

This study has several limitations. The first concerns the classification of patients into responders and non-responders. Patients who dropped out early due to worsening symptoms were classified as non-responders. Microbiota data were not collected for these patients, which potentially introduces bias into the results for the non-responder group. Moreover, the study concerns only 24 UC patients and the time series up to 14 weeks represents only a snapshot of the dynamic process of microbial succession. This sample size is too small to draw definite conclusions and further investigation into longer-term outcomes is necessary to gain a more comprehensive understanding.[422] A third limitation is the sequencing depth (2.9 million 100 bp single-end Illumina reads), which does not allow for definitive determination of whether an absent species was actually absent in the host or donor, or simply undetected.[87]

Also, the low sequencing resolution makes it impossible to determine whether the same strain present in the donor sample successfully colonized the recipient's gut microbiota or whether the donor and host strains coexisted or were replaced following FMT. Lastly, we did not have data to directly link the unique donor sample used for FMT to the corresponding recipient samples. Therefore, we used the combined microbiota data, which may have led to the misclassification of some low-abundance colonizing species from the donor as novel species.

By applying an ecological perspective to FMT, our study sheds new light on the importance of ecological principles, such as succession of microorganisms and the resilience of the recipient's system, in shaping therapeutic outcomes. Our study reveals the ecological dynamics of the gut microbiota during and after FMT in patients with UC, with a particular focus on the dynamics of recipient, donor, and novel species. Contrary to some previous studies, the overall engraftment of the donor microbiota did not emerge as the most important factor for FMT success in this study.[399, 415] The key factor influencing the response may not be the overall engraftment of donor species, but rather the recipient's ability to retain resident species while simultaneously enriching with novel and donor species. Thus, successful FMT hinges on fostering a microbiota shift that complements rather than compromises the existing ecosystem. This ecological interpretation aids in understanding the mechanism through which FMT may induce clinical remission and also underscores the nuanced interplay between donor and recipient microbiota essential for therapeutic efficacy.

Appendices of Chapter 6

Appendix Table 6.1 - Clinical and demographic information of responders and non-responders.		
	Responders[a]	**Non-responders**[b]
	Number (Percentage)	**Number (Percentage)**
Patients	9 (38%)	15 (63%)
Samples	81 (45%)	99 (55%)
Missing	0	36
Sex		
% Female[c]	6 (67%)	6 (40%)
Pretreatment		
% Budesonide[c]	5 (56%)	8 (53%)
Donor		
% D07[c]	2 (22%)	10 (67%)
	Mean (SD)	**Mean (SD)**
Age	45 (17)	48 (16)

[a] Remission (i.e., response) was defined at week 14 as no symptoms (partial Mayo score of 2 with no individual sub score of > 2) and an endoscopic Mayo score 0–1.

[b] All other patients, including those with a partial response (a decrease of at least 3 points in the partial Mayo score and at least 1 point at the endoscopic Mayo score) at week 14 and patients who left the study early, were classified as non-responders.

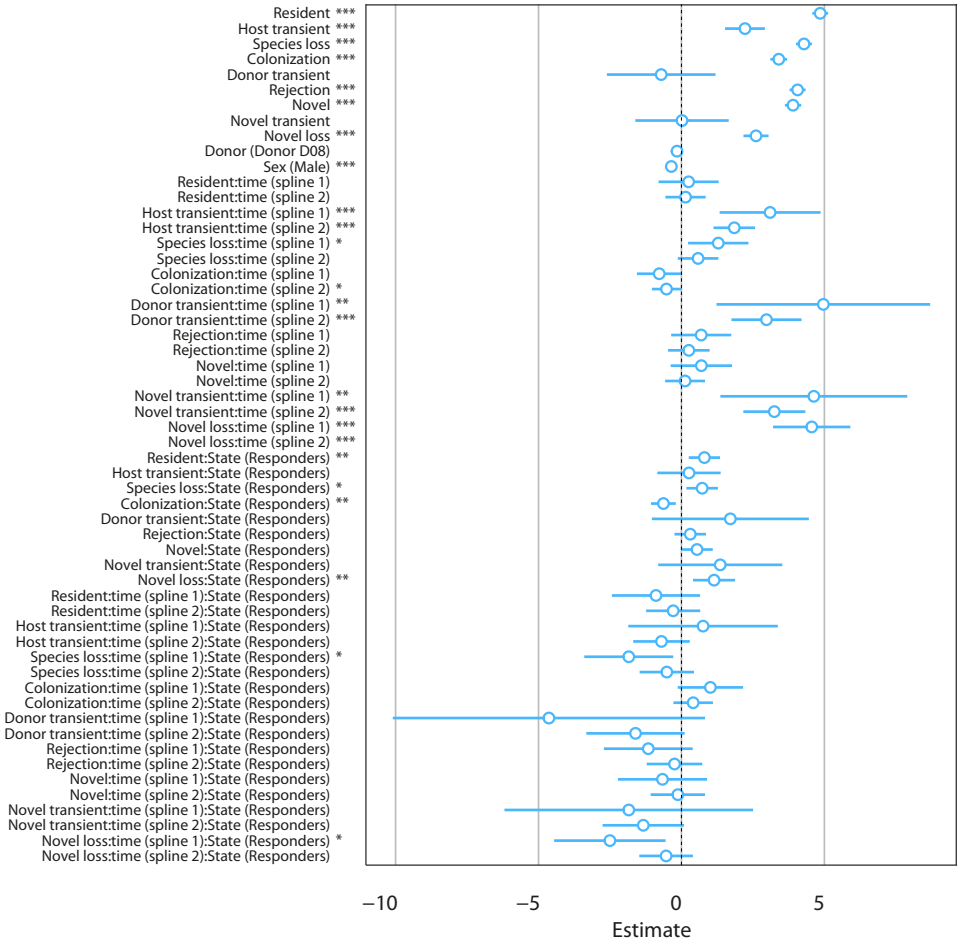
[c] Percentages calculated separately for responders and non-responders.

Appendix Table 6.2 - Model estimates and *p*-values for the differences in relative abundances.

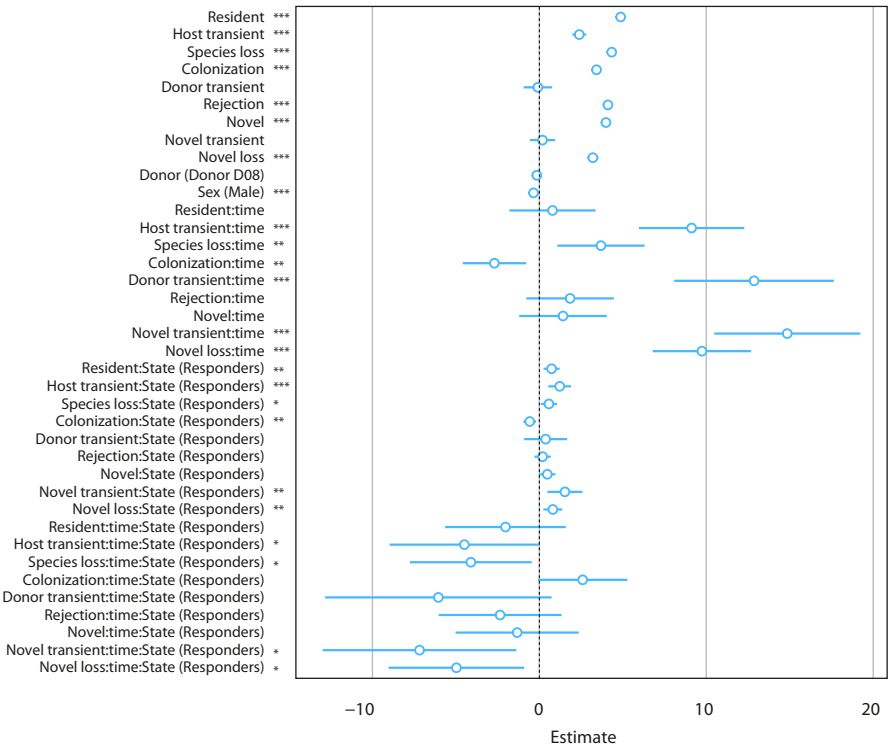
Results are visualized in Figure 6.3. Multiple models were used to test the differences. Significant results are highlighted in bold and blue.

A) Abundance differences per time point				
Post-3				
Intercept	–5.72735	0.15259	–37.535	**< 2e-16**
Host transient	–0.68310	0.08039	–8.497	**< 2e-16**
Post-4				
Intercept	–5.91299	0.10204	–57.95	**< 2e-16**
Host transient	–0.83215	0.06461	–12.88	**< 2e-16**
Week 8				
Intercept	–5.95367	0.10251	–58.08	**< 2e-16**
Host transient	–0.83115	0.05676	–14.64	**< 2e-16**
Week 10				
Intercept	–5.75390	0.18769	–30.66	**3.73e-10**
Host transient	–0.84484	0.06146	–13.75	**< 2e-16**
Week 14				
Intercept	–5.90331	0.09321	–63.33	**< 2e-16**
Host transient	–0.78173	0.05636	–13.87	**< 2e-16**
B) Relative abundance pre-FMT				
Categories within responders				
Intercept	–5.48590	0.19461	–28.189	**2.17e-09**
Host transient[a]	–0.63418	0.07692	–8.245	**3.91e-16**
Species loss[a]	–0.71556	0.06999	–10.224	**< 2e-16**
Categories within non-responders				
Intercept	–5.60069	0.18367	–30.49	**5.3e-10**
Host transient[a]	–0.72155	0.07792	–9.26	**< 2e-16**
Species loss[a]	–0.80577	0.06514	–12.37	**< 2e-16**
Differences in Resident species between responders and non-responders				
Intercept	–5.6414	0.1732	–32.57	**2.69e-16**
State (Responders)	0.1314	0.2432	0.54	0.597
Differences in Host transient species between responders and non-responders				
Intercept	–6.3584	0.1989	–31.973	**5.92e-16**
State (Responders)	0.2003	0.2797	0.716	0.484
Differences in Species loss species between responders and non-responders				
Intercept	–6.3788	0.2133	–29.902	**1.34e-14**
State (Responders)	0.1883	0.3053	0.617	0.546
C) Ratio relative host species abundances between responders and non-responders				
Intercept	0.184309	0.152030	1.212	0.246
State (Responders)	0.006944	0.213633	0.033	0.975

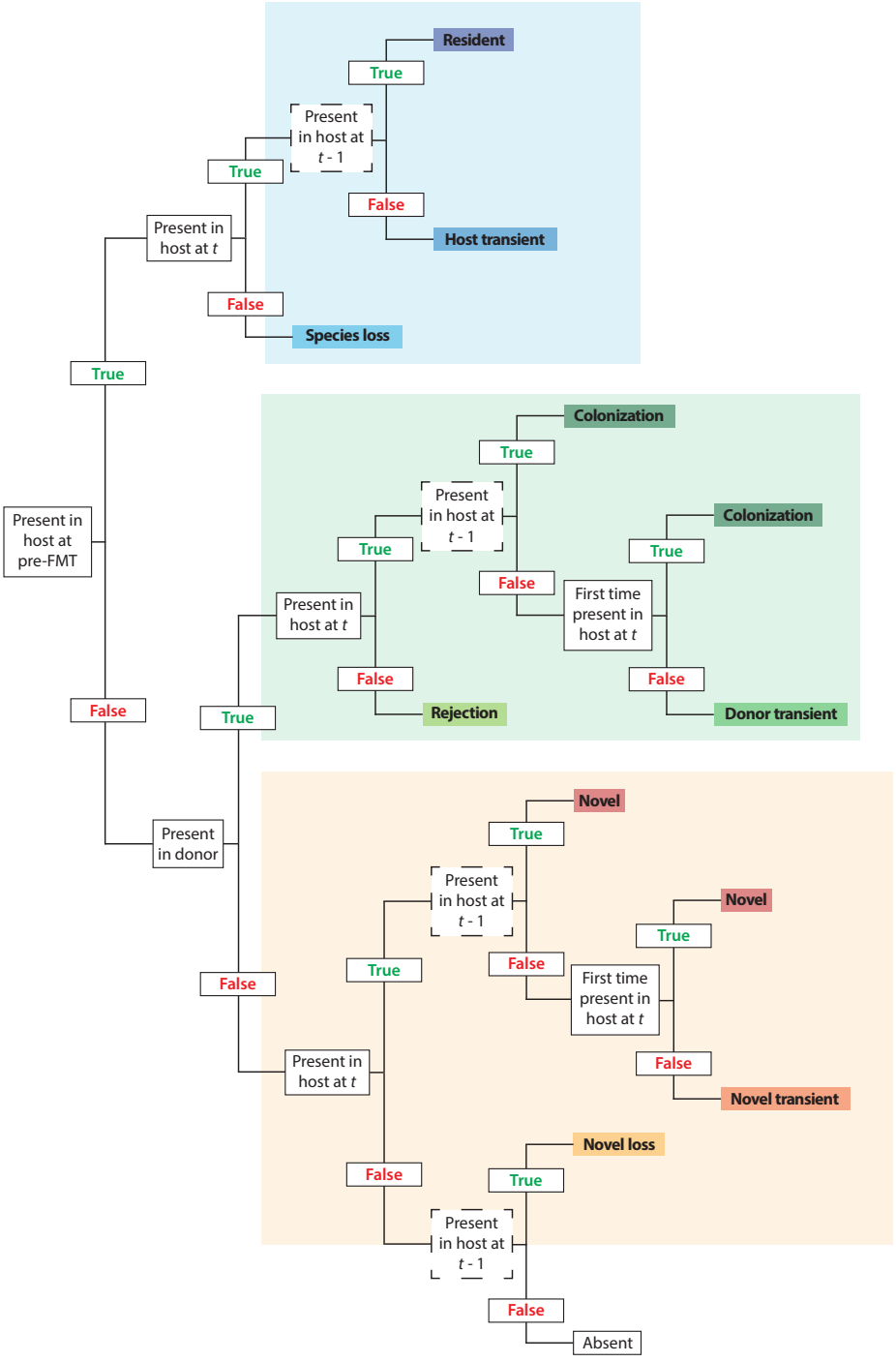
[a] The difference between the host transient and species loss categories for responders and non-responders was tested in a separate model and was not significant (*p*-values were 0.303 and 0.343 for responders and non-responders, respectively).



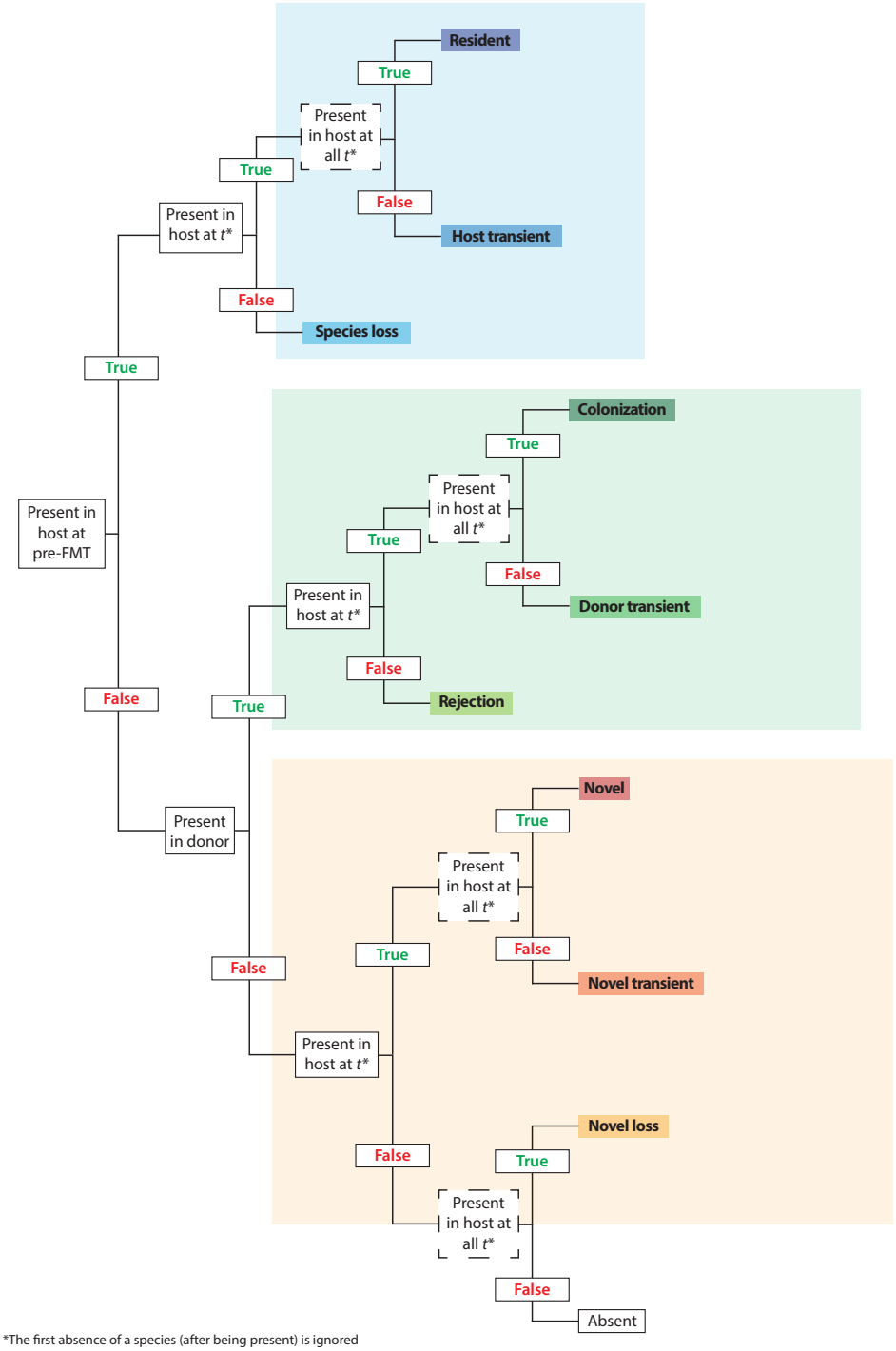
Appendix Figure 6.1. Results of modeling (with a spline) the number of species per ecological category in the base case. The point estimates, 95% confidence intervals, and a reference line at 0 are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different from 0. The original time variable was modelled with a spline rescaled to denote time in weeks since the start of FMT. The model contained a random intercept per patient to account for repeated measurements.



Appendix Figure 6.2 - Results of modeling (without a spline) the number of species per ecological category in the base case. The point estimates, 95% confidence intervals, and a reference line at 0 are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different from 0. Contrary to the base case, the original time variable was not modelled with a spline. Time was rescaled to denote time in weeks since the start of FMT. The model contained a random intercept per patient to account for repeated measurements.

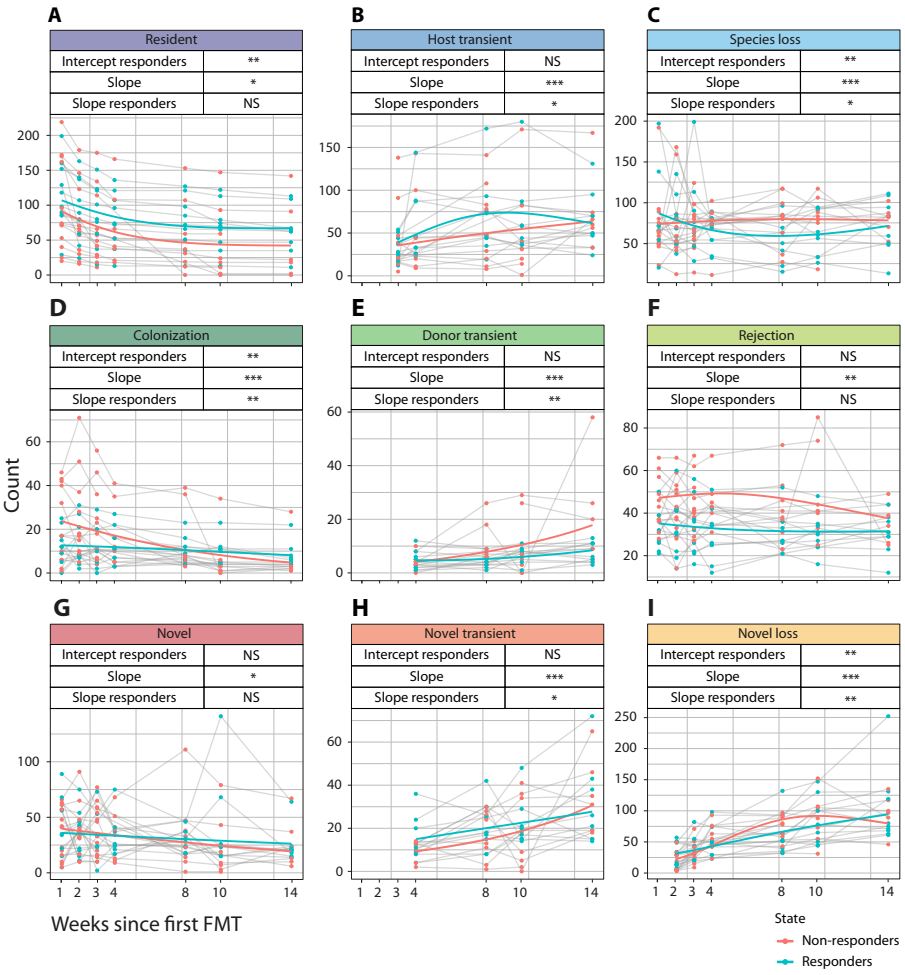


Appendix Figure 6.3 - Decision tree for Sensitivity 2 analysis to assign species to ecological categories according to different inclusion criteria as in the base case analysis. The categories are based on the origin and presence of a species over time. First, the species was compared to the pre-FMT host samples, then to the core donor microbiota. Next, the presence or absence at only the previous time point was considered to assign the species to an ecological category. Differences with the base case scenario, where all previous time points were considered, are indicated with a dotted line around the box (see also Box 6.2).

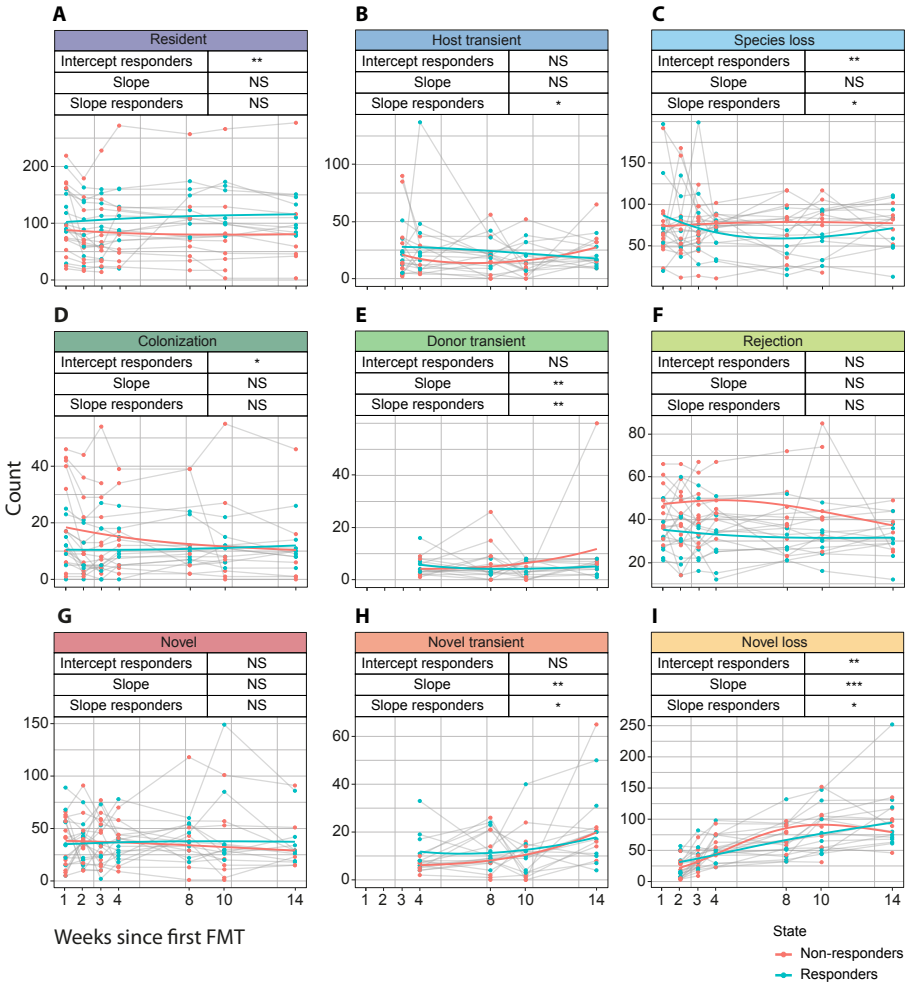


*The first absence of a species (after being present) is ignored

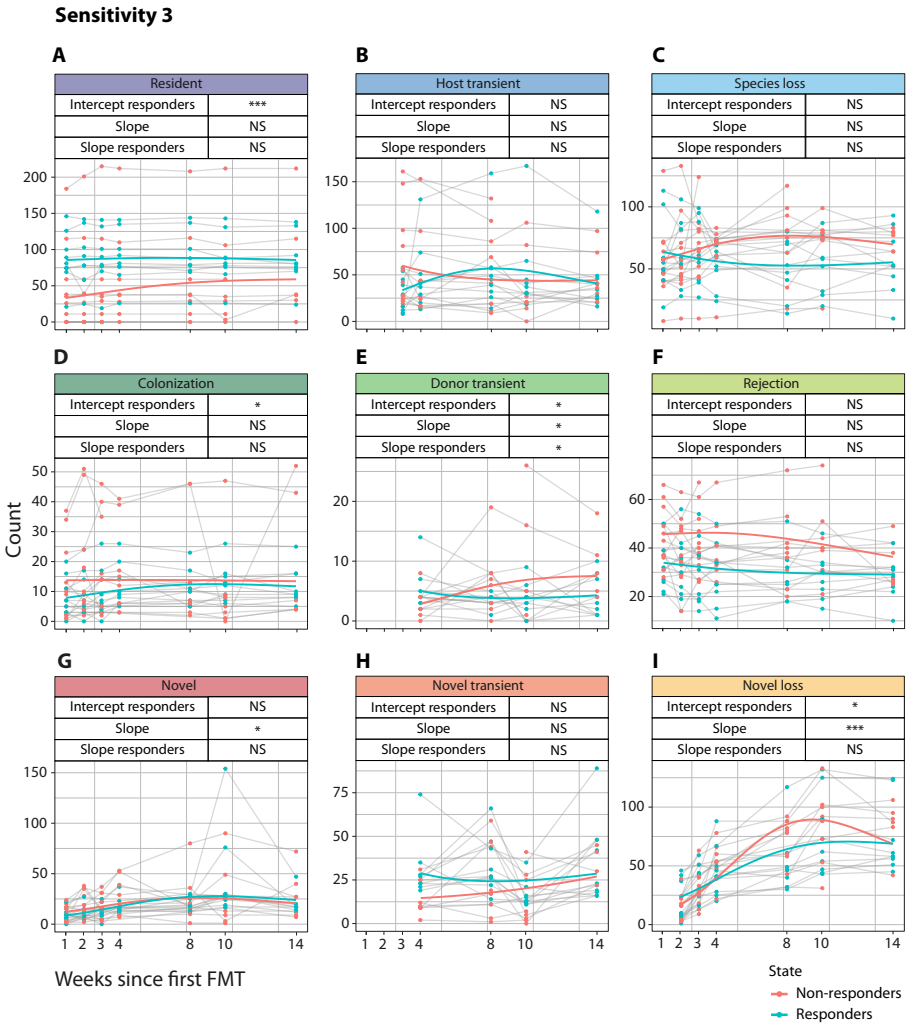
Appendix Figure 6.4 - Decision tree for Sensitivity 3 and 4 analyses to assign species to ecological categories according to different criteria as in the base case analysis. The categories are based on the origin and presence of a species over time. First, the species was compared to the pre-FMT host samples, then to the core donor microbiota. Next, the presence or absence at all time points was considered to assign the species to an ecological category. Differences with the base case scenario, where only previous time points were considered, are indicated with a dotted line around the box (see also Box 6.2).



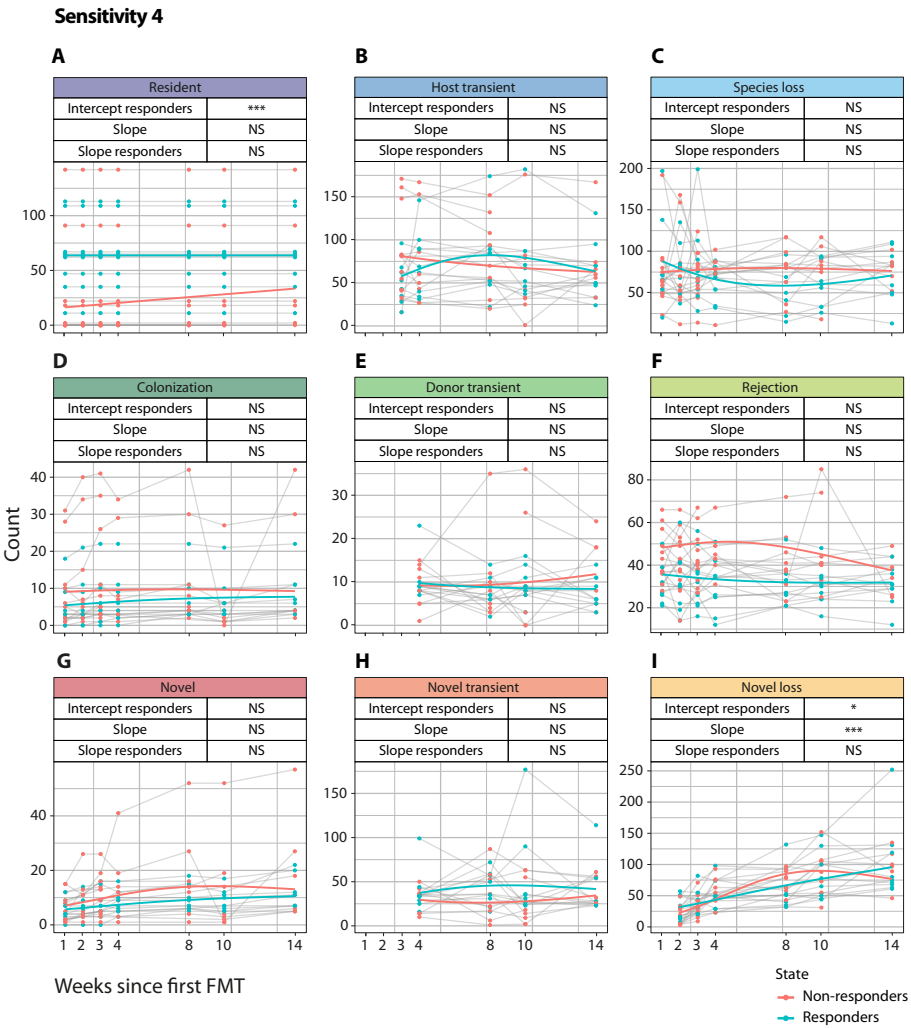
Appendix Figure 6.5 – Temporal changes in the number of species per ecological category for Sensitivity 1. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the *y*-axes. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = *p*-value < 0.001; ** = *p*-value < 0.01; * = *p*-value < 0.05; NS = not significant).



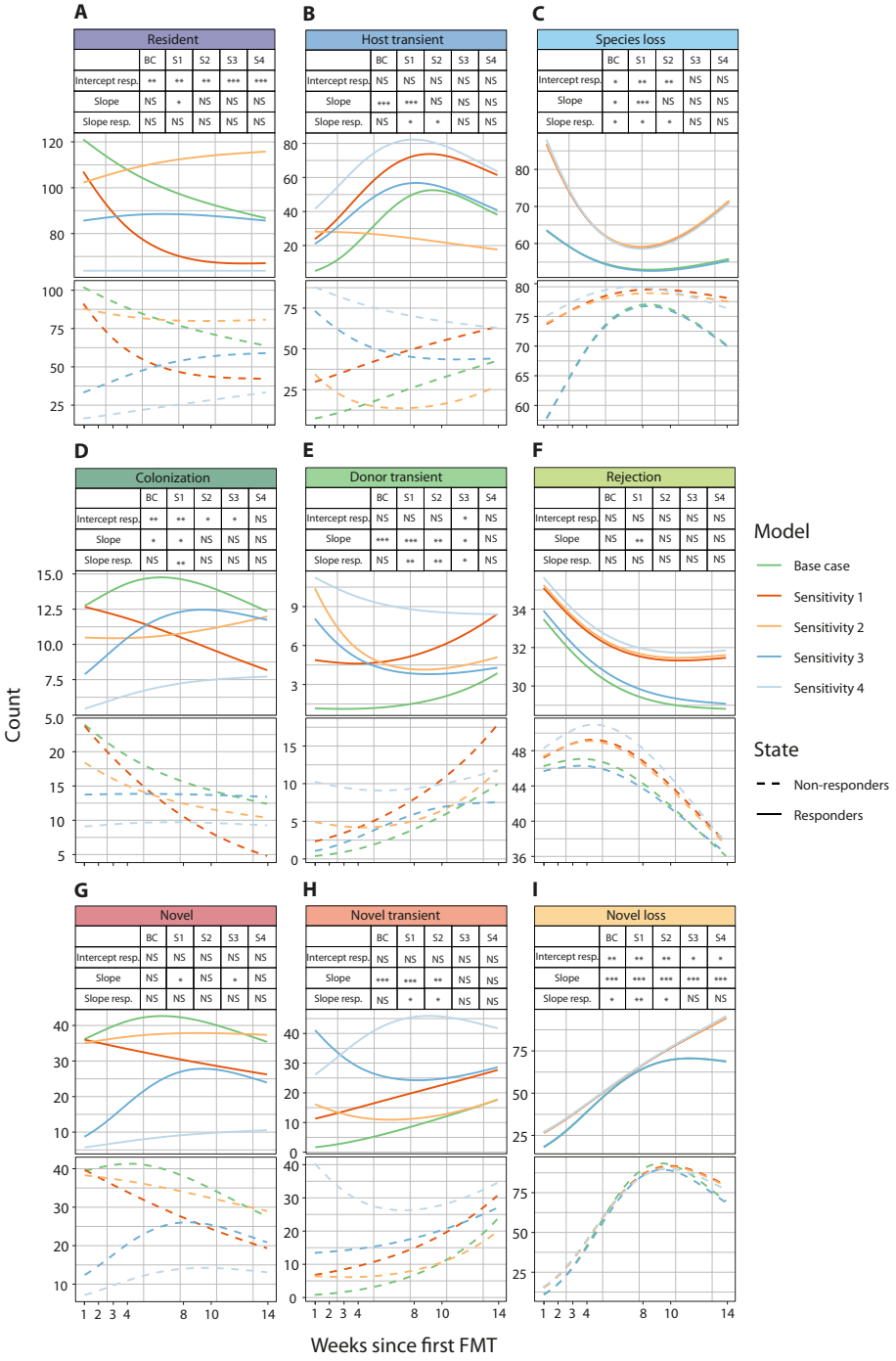
Appendix Figure 6.6 - Temporal changes in the number of species per ecological category for Sensitivity 2. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the *y*-axes. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = *p*-value < 0.001; ** = *p*-value < 0.01; * = *p*-value < 0.05; NS = not significant).



Appendix Figure 6.7 - Temporal changes in the number of species per ecological category for Sensitivity 3. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the *y*-axes. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = *p*-value < 0.001; ** = *p*-value < 0.01; * = *p*-value < 0.05; NS = not significant).



Appendix Figure 6.8 - Temporal changes in the number of species per ecological category for Sensitivity 4. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the *y*-axes. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = *p*-value < 0.001; ** = *p*-value < 0.01; * = *p*-value < 0.05; NS = not significant).



Appendix Figure 6.9 - Average temporal changes in the number of species per ecological category for the base case (BC) and all Sensitivity analyses (S1, S2, S3, and S4). Upper plots are for responders (solid lines) and lower plots for non-responders (dashed lines) to the treatment. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (*** = *p*-value < 0.001; ** = *p*-value < 0.01; * = *p*-value < 0.05; NS = not significant).

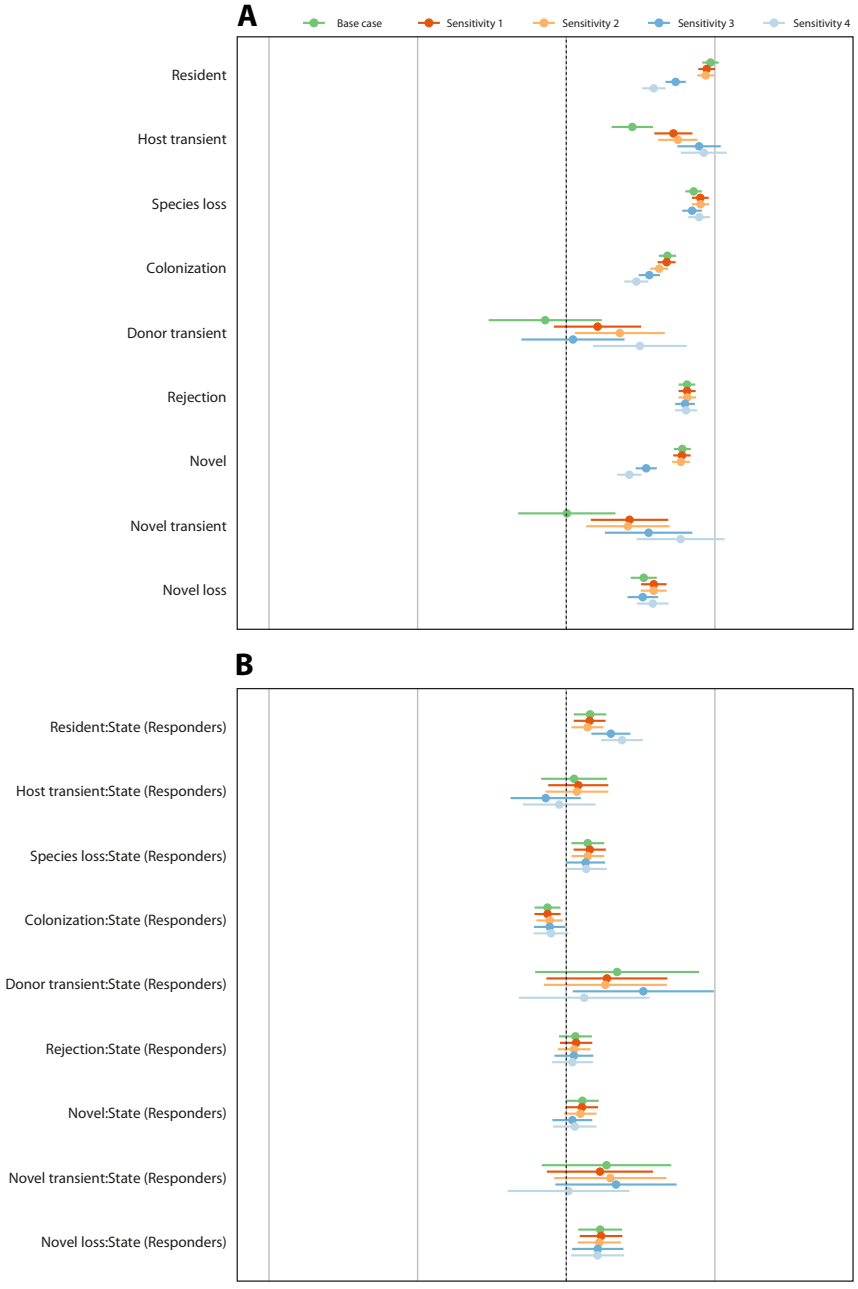
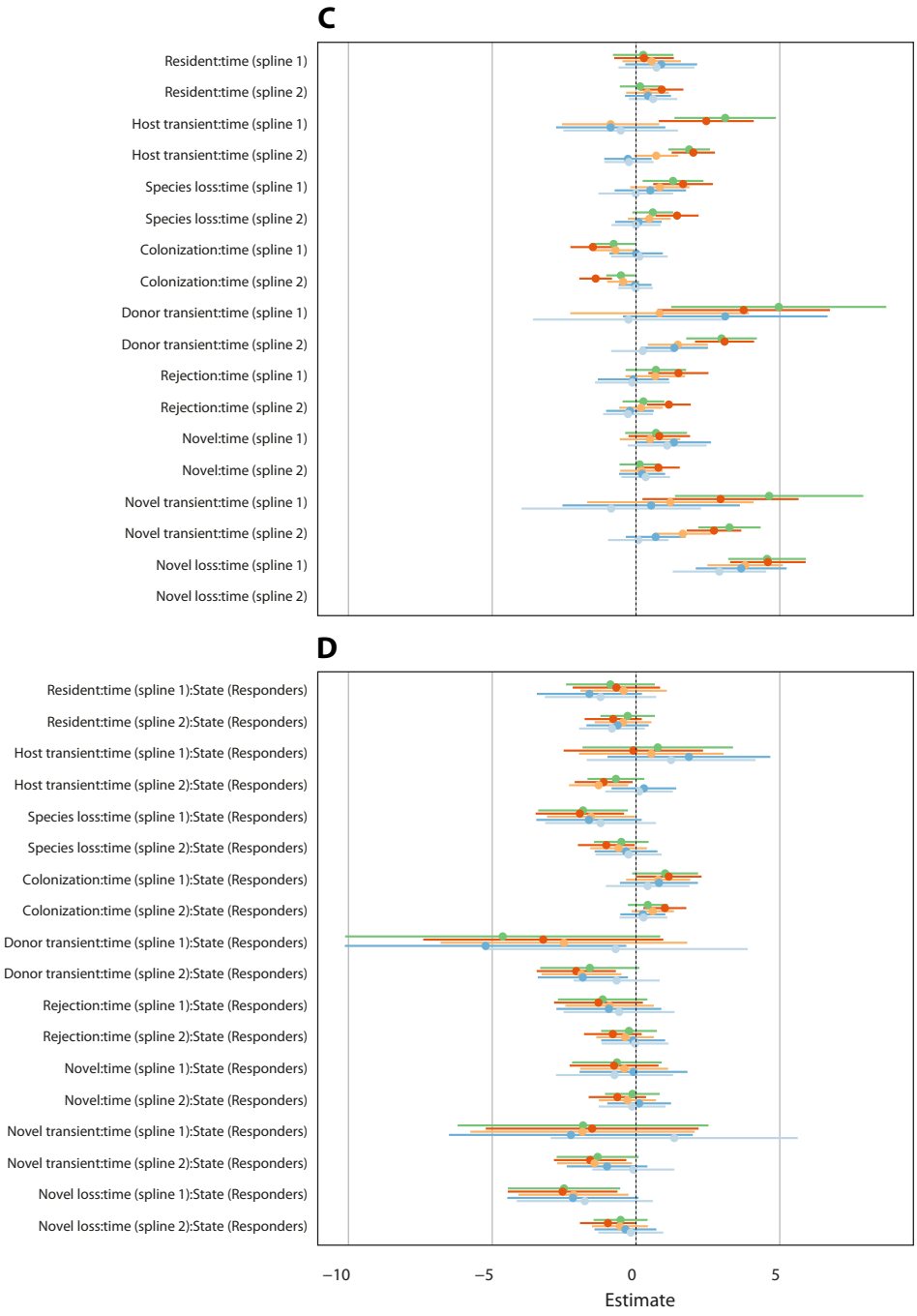
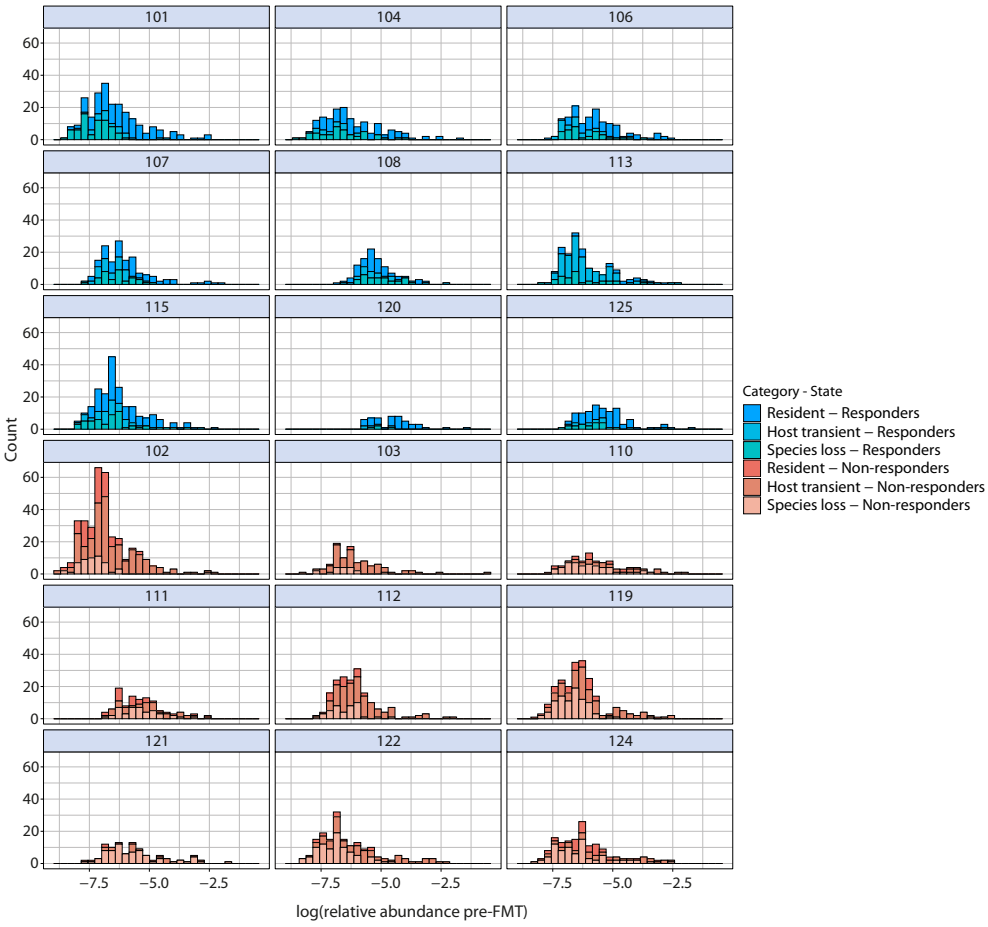


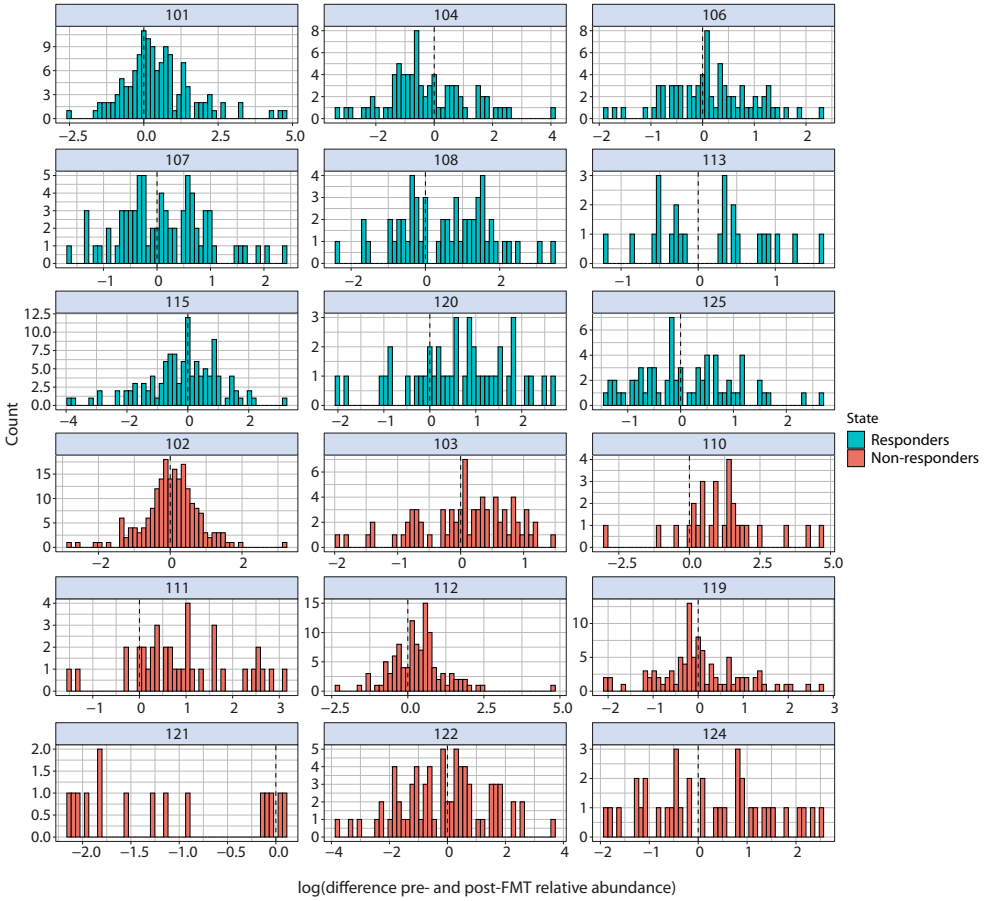
Figure continued on next page



Appendix Figure 6.10 - Distribution of the number of species per ecological category for the base case and all sensitivity analyses, estimated by overdispersed Poisson regression models with random effects and splines.
The models contain random intercepts per patient to account for repeated measurements. The point estimates, 95% confidence intervals, and a reference line at 0 are shown. When the horizontal lines do not cross the vertical reference line, the coefficients are significantly different from 0. A - D) Model output is presented for variables grouped into four categories for clarity.



Appendix Figure 6.11 - Histograms showing the relative abundances of host species (Resident, Host transient, and Species loss) pre-FMT. Only patients that completed the treatment and had at least one post-FMT sample are included in this plot. Because the data had skewed distributions, we used a natural-log transformation of the abundances to normalize the data and homogenize the variance.



Appendix Figure 6.12 - Histograms showing the distribution of the differences in relative abundances (between pre- and post-FMT) of resident species.

Only patients that completed the treatment and had at least one post-FMT sample are included in this plot. The striped vertical line indicates no change in abundance between pre- and post-FMT. Because the data had skewed distributions, we used a natural-log transformation of the abundances to normalize the data and homogenize the variance.

Chapter 7

General discussion and future perspectives

General discussion and future perspectives

Main findings

In this thesis, we applied a theoretical framework and used methodologies derived from the field of ecology to investigate the dynamic properties and characteristics of the human gut microbiota. In this way, we aimed to contribute to a better understanding of the complex microbial ecosystem of the human gut and its association with inflammatory bowel disease (IBD) course (i.e., exacerbation or remission). Additionally, we examined microbial changes following an intervention with fecal microbiota transplantation (FMT). Addressing these aims requires a thorough examination of the human gut microbiome, its dynamics, and the key factors influencing the functioning of this microbial ecosystem. This dissertation contributes to these goals in several ways.

First, we studied the correspondence between correlation-based networks and the underlying network of ecological interactions. Our results demonstrated that correlations could indicate the presence of bacterial interactions, at least in a simulation setting. Interactions were recovered with precision exceeding recall, indicating that the likelihood of missing interactions was higher than the likelihood of finding false positive interactions when using correlations in cross-sectional abundance as their proxy. However, we also showed that asymmetric interaction types cannot be detected and that there are many factors that may worsen these results, such as measurement noise. Unfortunately, biomedical data are always subject to measurement errors, particularly in microbiota studies where data are obtained through sequencing processes.[118] Furthermore, microbiota data are also influenced by host-specific variation in process parameters (process noise) and sampling under various (non-equilibrium) conditions, all of which will influence the inference, though not necessarily in an adverse way.[423] Therefore, while correlations may hint at interactions, independent validation is needed to confirm their presence and to ensure that these correlations represent genuine biological interactions with meaningful implications. Until then, we should continue to refer to these correlations as associations rather than interactions. Moreover, in our second study we showed that wavelet clustering uncovers more diverse community structures compared to analyses based on temporal correlations. We revealed significant differences between these methods and suggested that the correlation-based approaches might overlook certain dynamical aspects of microbial communities. This comparison highlights the potential of wavelet clustering to use the temporal fluctuations and complexity inherent in the human microbiota for characterizing community structure, offering a more nuanced understanding than correlation-based methods alone.

Second, our objective was to describe specific associations between microbial abundances and Crohn's disease (CD), in particular with exacerbation of disease. In doing so, we made the analogy between the gut microbiota in an unhealthy host with an ecosystem under stress. We found that microbial diversity is reduced in the gut of CD patients, and that the process of diversity loss is irregular with respect to specific taxonomic groups. If this process of loss of species continues for an extended period, it may eventually lead to an unhealthy and possibly irreversible state. Moreover, in this study we showed that associations of relative bacterial abundances with CD can be different for subsets of individuals. A practical, though

undesirable implication of this finding is that it seems very difficult to pinpoint specific gut microbes as biomarkers or therapeutic targets for CD patients.

Third, we studied bacterial associations with clinical treatment success of FMT in ulcerative colitis (UC) and investigated the succession of the microbiota during and after the treatment. By means of several analytical techniques, such as longitudinal modeling and cluster analysis, we identified potential associations between specific gut microbiota families and clinical outcomes. Our findings suggest that the success of FMT in UC patients may be linked to the control of Prevotellaceae, with potentially beneficial roles attributed to Lachnospiraceae and Ruminococcaceae. Notably, clustering analysis indicated that differences in the gut microbiota between responders and non-responders may manifest early during treatment. Moreover, successful FMT seems to be associated with a resilient gut community that is open to colonization by donor species, while maintaining the original community to some degree. This suggests that a balanced coexistence of host and donor species can induce a shift in which the recipient's microbiota evolves towards a healthier community.

Stability and variability in microbiota dynamics

Over the past 15 years, microbiological research has flourished, driven by technological advancements that have significantly expanded our knowledge concerning the ecology of gut microbiota and its relation to health and disease.[424] The beneficial functions provided by our microbiomes offer potential for improving human health. Therefore, efforts have been made to understand the temporal variations in our microbiota to define 'stable' and '(un)healthy' dynamics.[21, 45, 62] Early attempts to classify the gut microbiota introduced the concept of 'enterotypes', distinct clusters characterized by an enrichment of *Bacteroides*, *Prevotella*, or *Ruminococcus*.[405] However, this early classification was only based on metagenomics from 39 individuals, and much larger studies have challenged the distinctness of these enterotypes, suggesting a more gradient-like distribution with varying levels of *Prevotella* and *Bacteroides*.[233, 412, 425]

The microbiota is acknowledged to be highly specific to individuals, displaying relative stability in adults, with regular fluctuations in the composition over time.[45, 46, 51] These fluctuations suggest that long-term stability of human gut microbial communities is influenced by the tendency of the intestinal ecosystem to maintain internal stability (homeostasis), owing to the coordinated response to any stimulus that disturbs its normal condition.[62] This prompts inquiries into whether fundamental ecosystem 'rules' governing microbiota (group)dynamics can be distilled from a collection of individual microbiota, and to what extent each represents a unique ecosystem with its own host-specific microbial dynamics (Figure 7.1).[426] If microbiota dynamics were independent from the host, then the presence of the same species should result in the same relative proportions of those species, and interventions could be devised to regulate microbial states across different individuals.[284] On the other hand, if the dynamics are strongly host-specific, personalized interventions should be designed, considering not only the unique microbial state of an individual but also the specific host factors of the microbial ecosystem.[46, 55, 59] However, studying this is very difficult due to the presence of latent or unknown parameters (related to lifestyle or diet for example) influencing microbiota composition.[155] The factors contributing to microbiota variation are still not fully understood.[57]

Consequently, comparative analyses between patient and healthy cohorts yield many different dysbiotic states or sets of microbial biomarkers that are dependent on a specific comparison, and the definition of a normal healthy microbiota remains unsatisfactorily answered. Moreover, it is still unclear whether the structure of the gut microbial community shifts gradually within individuals or transitions between distinct community states, and whether such states are consistent among different individuals.[233, 405, 427]

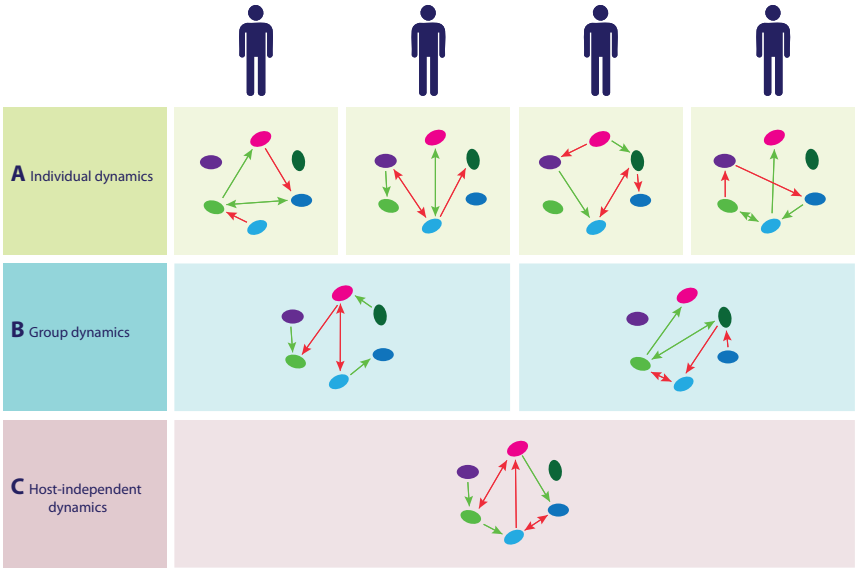


Figure 7.1 - Illustration of microbial dynamics through ecological networks. Microbial dynamics are illustrated through an ecological network, wherein nodes symbolize species and edges depict interspecies interactions (green and red arrows denote positive and negative interactions, respectively). A) The underlying dynamics or networks are unique to each subject. B) Subjects within the same group exhibit shared dynamics or networks, which markedly differ from those of other groups. C) Different subjects display identical underlying dynamics or networks. Note that subjects can also differ in species composition or in the relative abundances of each species. This figure is based on Bashan et al. (2016).[426]

Broader insights from the literature

Part I - Ecological structure in the human gut microbiota

Microbial interactions can yield diverse outcomes, ranging from positive impacts such as mutualism, where species exchange metabolic products to benefit each other, to negative impacts on participating species. These interactions shape community patterns and inhibit the outgrowth of certain species. In Chapter 2, we assessed the reliability of correlation-based methods for inferring microbial interaction networks. Unraveling the network of interactions within ecological systems, particularly in studies of the human microbiome, is challenging. Technical issues in constructing networks from sequencing data, such as compositionality and the predominance of zeros, combined with the influence of often unmeasured environmental factors, make the networks difficult to interpret and susceptible to potential biases.[118] Additionally, data generated from assays may be censored by detection limits, causing species to remain undetected.[203]

Moreover, the presence of a third variable or species (e.g., bacteriophage) can influence the observed correlations, especially if the researcher fails to measure this linked species (Figure 7.2). Correlation-based network analysis typically results in too many spurious edges.[118] Addressing these challenges has led to the development of various co-occurrence methods, such as CoNet, SparCC, and SPIEC-EASI.[237, 253, 428] Interestingly, in evaluations, classical correlation measures often perform just as well as the more sophisticated algorithms.[118, 429]

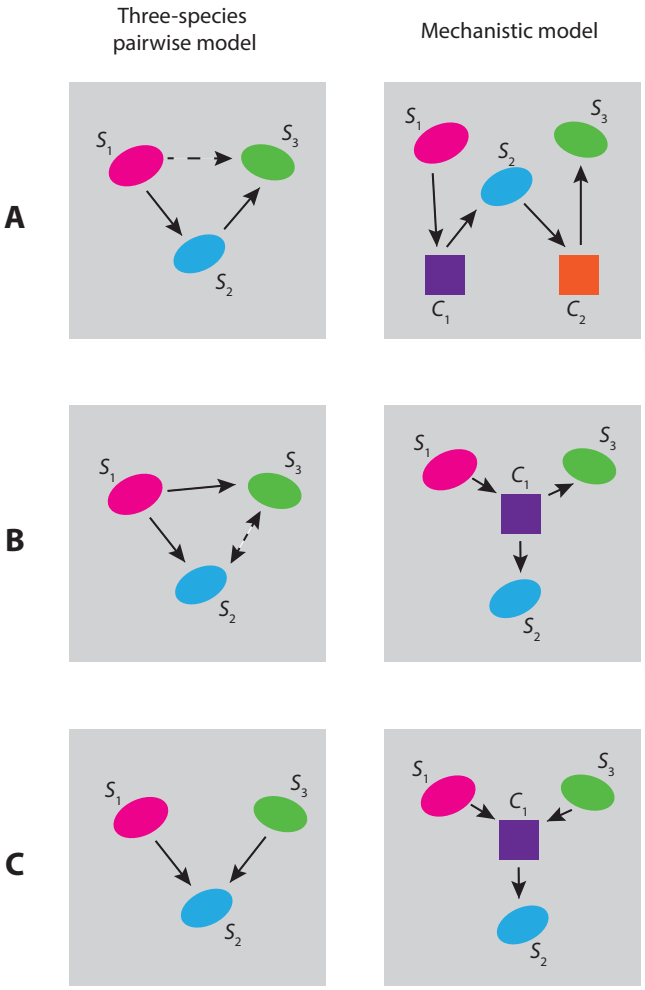


Figure 7.2 - Interaction networks between three species.
Direct interactions are indicated by a solid arrow, indirect interactions are given by a dashed arrow. A) The interactions utilize separate compounds, C_1 and C_2, as mediators. Interaction chain: Species S_1 influences S_2, which in turn affects S_3. B) In this scenario, S_1 initiates a change where S_2 and S_3 interact only when S_1 is present. Modified interaction: Species S_1 influences both S_2 and S_3. Species S_3 consumes mediator C_1, altering the interaction between S_2 and S_3. C) Modified interaction: Both S_1 and S_3 contribute compound C_1, which stimulates S_2. S_1 and S_3 do not directly interact regardless of S_2. This figure is based on Momeni et al. (2017).[240]

To address potential confounding in pairwise interactions, we employed partial correlations in Chapter 2 to infer the correlation network. See Figure 7.3 for a comparison between plain and partial correlations in a real dataset. For most microbes, ecological interactions are poorly understood, necessitating the *de novo* construction of ecological interaction networks without guiding assumptions or a gold-standard set of interactions for validation.[100, 111, 238, 251, 430, 431] Therefore, we used the generalized Lotka-Volterra (gLV) model with simulated interactions to study the correspondence between correlations and interactions.[232] gLV models are widely employed in ecological studies to simulate the dynamics within bacterial communities.[100, 111, 232, 254, 423, 432] This approach enabled us to define the species-species interaction terms and incorporate variations in model parameters to reflect the variability among hosts. The gLV model, while versatile, has drawbacks: it only describes pairwise interactions, disregards immigration and environmental effects, and maintains constant and additive interaction strengths.[49, 57, 100, 232, 240, 433-435] In Chapter 3 we also used an ecological model. Here, we simulated the dynamics of four consumers and four resources to provide an additional dataset to evaluate the accuracy of wavelet clustering in contrast with clustering based on Spearman's correlation.[318-320]

Some scientists tend to approach mathematical models, also the ones used in Chapter 2 and Chapter 3, with skepticism, wary that simplification might sacrifice realism. However, while models may simplify complex systems, they can also serve as invaluable tools for understanding phenomena that are otherwise difficult to grasp.[1] For example, in Chapter 2, we would not have been able to judge the correctness of the correlation matrix without a simulated network that could be used as a ground truth. Models allow scientists to explore hypothetical scenarios, test theories, perform virtual experiments that are impossible or unethical in humans, make predictions, explain complex phenomena, thereby ultimately advance our understanding of the natural world. However, it is imperative to ensure that models are built upon correct assumptions as these can significantly impact model outcomes.

Notably, many studies on microbial communities and their associations with specific disease courses or host conditions heavily rely on a steady-state assumption and the failure to account for non-steady-state dynamics could introduce biases in the findings, leading to an overemphasis on certain taxa while neglecting others that may be important in a non-steady-state context. The microbial interaction network is also likely dynamic, shaped by both negative and positive feedback loops. These feedbacks occur as an organism's metabolic activity alters its environment, influencing its own fitness, and the fitness of competing species, creating ecological niches that drive diversification.[22] Therefore, the niches in the gut might be more comparable to a dynamic river ecosystem than to a more static ecosystem on land, as nutrient flows through the bowel, providing constant resources but also causing constant disturbances and reassembly of microbial communities and interactions.[436]

Future microbiome studies will benefit from larger cohorts, more frequent sampling, and longer follow-up periods to unravel the short- and long-term dynamics of gut microbial communities in real datasets. Longitudinal studies allow for investigating the consistency, or changes, of microbiota patterns over time. Following this, in Chapter 3, we applied a methodology unknown to the microbiome field, namely wavelet clustering analysis. This method clusters time series based on the similarity in their temporal dynamics of microbial communities.

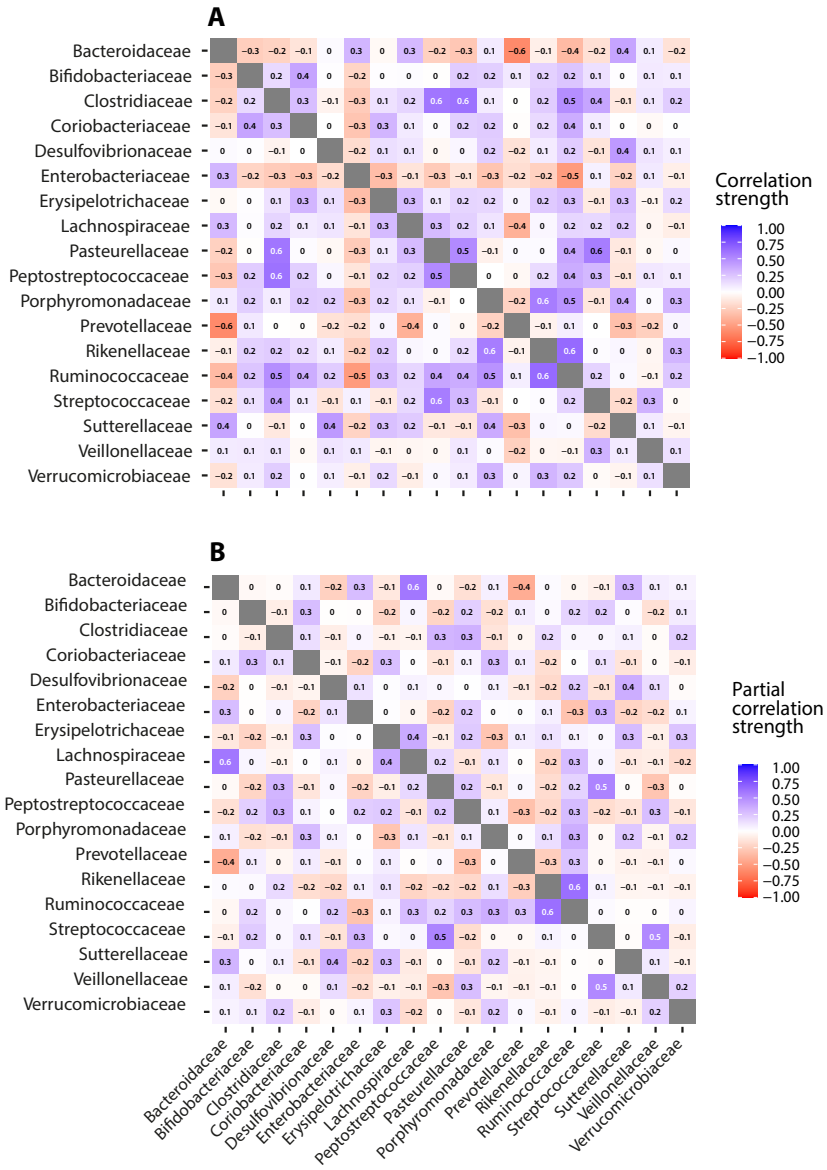


Figure 7.3 - Correlation matrices. Matrices are derived from the dataset presented in Chapter 4 of this thesis. A) Spearman's correlation matrix displaying the pairwise correlations between bacterial families. The correlation matrix provides insights into the linear abundance relationships among variables. B) Spearman's partial correlation matrix illustrating the partial correlations between the bacterial families. Partial correlations help to assess the unique association between bacteria, independent of the interrelated influence of other bacteria. Each cell represents the (partial) correlation coefficient between two variables, with color intensity indicating the strength and direction (e.g., blue is positive and red is negative) of the correlation.

Unlike prevailing co-occurrence methodologies, the novelty of wavelet clustering lies in its ability to characterize community structure based on the collective temporal behaviour of the microbiota, without directly fitting a dynamical model or reconstructing the network of interacting species. While traditional correlation-based methods may offer some, but limited or biased, insights, wavelet clustering enables the extraction of more information on dependencies within microbial communities and can reveal community structures that remain obscured in correlation-based methods.[437] These findings underscore the critical role of longitudinal data and methodological choices in shaping the outcomes of microbiota data analysis.

Mapping ecological networks to predict (temporal) behaviours and discern assembly rules is motivated by the goal of gaining insights into the underlying dynamics that drive microbial ecosystems. Ultimately, this knowledge may be used to establish early warning signals, develop clinical prognostic models, and even engineer stable microbiomes with desired properties.[438-440] The topology of the network often provides insights into the potential explanatory nodes for specific functional properties within the network, allowing for the identification of tightly interrelated modules of variables, such as communities.[3] Additionally, knowledge of the interaction network not only aids in identifying key players within the network (i.e., keystone species) but also facilitates predictions on how microbial communities might respond to diverse stimuli or disturbances, such as alterations in diet or exposure to antibiotics.

Previous research has indicated that correlation-based networks likely capture only a fraction of the interactions occurring in microbiota, with strong symmetric interactions being more readily detected compared to weaker or asymmetric interactions.[216, 235, 273] Correlation-based networks from cross-sectional data are commonly interpreted as representing interspecific interactions.[227] Each significant link in a correlation network suggests a shared process affecting connected nodes; however, we should acknowledge that correlations do not always imply causation or biological meaning.[3, 216] Densities may also vary as a result of an external factor that is not of biological interest.[118] The presence of two species together in one sample, while absent in another, may not necessarily indicate an interaction between them. Instead, they could simply coexist because one sample was taken during a nutrient-rich period that supports the growth of both species independently, whereas the other sample may have been taken at a less favourable time, limiting the growth of both species. Therefore, incorporating additional information about influencing factors can provide a richer, more nuanced picture of the underlying dynamics within the microbiome. Moreover, as most microorganisms form biofilms, i.e., genetically diverse, surface-associated communities embedded in an extracellular polymeric matrix, bacteria primarily interact with others in their immediate neighborhood, with the strength of these interactions diminishing as distance increases.[440, 441] Therefore, the spatial relationships between individual organisms should ideally also be considered in the network, including the nature and strength of their interactions based on their positions within the community.[440] However, before delving into more complex network structures including extensive metadata, it is essential to first gain a thorough understanding of the 'simpler' networks to lay a solid foundation for future analyses.

Part II - Gut microbiota and inflammatory bowel disease

Given the involvement of the microbiome in numerous essential functions, it is not surprising that disturbances in microbiota composition (known as dysbiosis) have been linked to the onset and course of various diseases. Many associations found may not always be disease-specific but rather part of a non-specific, shared response to health or disease.[200, 442] Chapter 4 and Chapter 5 of this thesis address the relationships between bacterial dysbiosis and the disease course of CD, which, along with UC, comprises the pathology of IBD. While CD can occur anywhere in the digestive system, UC is limited to the colon. Both diseases exhibit significant distinctions in microbiota compositions from one another, although less strongly than they differ from healthy subjects.[372] However, the findings regarding disease exacerbation among CD or UC patients are often inconsistent and occasionally even contradictory. For example, previous studies have reported both lower and higher relative abundances of *Bacteroides* (Bacteroidaceae) in CD patients compared to healthy individuals.[356-358] This discrepancy can be attributed in part to technical variations between studies such as differences in DNA extraction methods and sequencing depth, but they may also arise from variations in disease assessment or study populations, as well as potential confounding factors, such as medication use or lifestyle factors that remained unidentified.[56, 171, 225, 443] Coupled with the interindividual variability of the microbiome in gastrointestinal disorders, the pursuit of shared biological signals proves challenging. Moreover, while many studies adopt a cross-sectional study design, longitudinal studies are needed for comparing active and inactive disease.[231, 444] The knowledge gap with regards to consistent and specific dysbiosis signatures poses a challenge to reveal the role of gut microbiota in human diseases.

In Chapter 4 we investigated the multifactorial involvement of specific microbial groups with CD compared to healthy individuals. Additionally, we also investigated associations between the relative abundances of specific bacterial families with disease course (remission vs. exacerbation) and disease activity markers (e.g., fecal calprotectin (FC), serum C-reactive protein (CRP), and Harvey Bradshaw index (HBI)) in repeatedly sampled CD patients.[181] Given the variability among CD patients and the complex microbial interactions, associations with disease may only be weak when considering mean responses. Therefore, it requires robust analysis to uncover these associations, and quantile regression is a promising method given that potential relationships may only be apparent in lower or upper quantiles of relative abundances.[361, 363]

We identified several significant associations between bacterial family abundances and CD, particularly when compared to healthy controls. CD patients exhibited distinct microbial profiles, with several families showing predominantly negative associations. While our results confirmed previously identified associations, including Erysipelotrichaceae, Peptostreptococcaceae, Prevotellaceae, Clostridiaceae, and Ruminococcaceae, we also uncovered novel associations with Coriobacteriaceae, Desulfovibrionaceae, Pasteurellaceae, Sutterellaceae, and Streptococcaceae.[171, 177, 181, 356-358] Notably, Coriobacteriaceae displayed a shift in relative abundance across the disease course, with higher values at baseline in patients who later experienced exacerbation. Additionally, Streptococcaceae demonstrated increased abundance over time in patients with exacerbation, compared to both healthy controls and patients in remission. Conversely, Sutterellaceae was consistently lower in patients with exacerbation as well as those in remission compared to healthy controls. Interestingly, associations with disease activity were generally weaker. We also found that FC levels were negatively correlated with the abundance of Porpyromonadaceae and Verrucomicrobiaceae.

Prevotellaceae were among the most heterogeneous across individual patients. The genus *Prevotella*, which belongs to this family, is involved in saccharolytic fermentation and short-chain fatty acid production. *Prevotella* is generally more prevalent in individuals from rural areas compared to urban populations, potentially due to the higher abundance of *Prevotella* phages and a diet lower in plant-derived complex carbohydrates in urban populations.[445, 446] Additionally, *Prevotella* has been linked to inflammation in other diseases; for instance, *Prevotella bivia* is strongly associated with inflammation in bacterial vaginosis and an increased risk of HIV.[447, 448] In Chapter 5, we also observed associations with Prevotellaceae in UC patients undergoing FMT treatment. Non-responders to FMT showed an increase in Prevotellaceae abundance compared to patients who achieved clinical remission after FMT (i.e., responders). However, our data from Chapter 4 and Chapter 5 do not clarify whether these differences are driven by the disease or factors, such as dietary habits, environmental variables, or other unknown factors that could contribute to the outgrowth of Prevotellaceae in these patients.

Interestingly, nearly all significant associations found with quantile regression in Chapter 4 were negative and primarily observed in the lower quantiles of the bacterial abundances. While positive associations in upper quantiles have been linked to unmeasured factors constraining the potential response to positive stimuli,[361] this contrasting trend resembles an ecosystem responding to stress: as the system nears a tipping point, the ability to sustain healthy bacterial abundances gradually diminishes.[97] However, the loss of certain species within the microbial network can be compensated for by others with similar ecosystem functions (functional redundancy). This redundancy enhances resilience, ensuring the continuity of essential functions important to the host, such as butyrate production.[62, 449] Consequently, when solely studying the compositional profile, the actual functional output of a system presumed to be in 'dysbiosis' might be normal, and vice versa; lack of significant differences in abundance doesn't necessarily indicate a healthy state as the species may lack essential functional genes.[450] However, an excessive loss of species may reduce resilience and cause a critical transition to an alternative stable state.[104, 369] A study setup including proteins secreted by the microbiome would provide insights into how dysbiosis is expressed on the functional level. For instance, in a CD case-control study, a lack of species capable of consuming hydrogen sulfide was identified as a key distinguishing microbiome feature of the disease.[451] Other studies showed the role of butyrate, secreted by pathobionts such as *Fusobacterium*. While butyrate is typically beneficial, it may negatively affect the viability of the intestinal epithelium and potentially contribute to IBD pathogenesis.[452, 453]

Note that, from a statistical point of view, investigating numerous bacterial species across multiple patients poses a significant challenge regarding the multiple hypothesis testing problem. To construct a correlation network or investigate significant differences in microbiota composition, adjustments might be necessary to control for false discoveries. The choice between correction methods depends on the research goal; stricter corrections, such as the Bonferroni approach, may be preferred to demonstrate specific associations, while more general impressions may be sought with less stringent corrections, such as the Benjamini-Hochberg (BH) method. However, even the BH approach might still be too strict when applied to microbiota data, because these methods assume independence among bacterial abundances, which is not valid due to (biological) relationships between species (Figure 7.2 and Figure 7.3) and the compositional nature of the data. Ideally, correction methods should account for correlated species to provide more accurate results. However, there is no solution yet available; therefore, conclusions should be based on a comprehensive review of existing literature in addition to study findings and not on *p*-values alone.

Part III - Ecological determinants of FMT treatment success

Concerning the treatment of dysbiosis, since a groundbreaking study in 2013, FMT has emerged as a treatment option for recurrent *Clostridioides difficile* infection (rCDI).[190] However, rCDI remains the only condition for which FMT is widely accepted as a treatment. In all other indications where FMT has shown promise, its use remains experimental or is considered a last-resort option.[454] One of the challenges with FMT is its inconsistency in (microbiota composition) outcomes.[455] This means that every person will react differently to certain bacteria and that diverse immune responses are activated across patients with different diseases.[456] This variability raises significant safety concerns, because the microbiota could also be altered to an even more undesirable state in the recipient's gut.[426, 457] Similarly, other therapies designed to modulate the microbiome, such as probiotics, have also been associated with adverse outcomes. The PROPATRIA study, a Dutch clinical trial conducted from 2003 to 2007, revealed that patients with acute pancreatitis who received probiotics had a higher mortality rate compared to the control group.[458] However, it remains unclear whether the probiotics themselves or other factors contributed to this increased mortality. Therefore, a 'one-size-fits-all' treatment approach does not ensure safety and efficacy against multifaceted diseases, as evidenced by the inconsistent results of FMT trials for IBD and irritable bowel syndrome (IBS).[197, 459-461] The interaction between two microbial consortia (donor and recipient) during FMT can be likened to a complex pulse perturbation. Possibly, the perturbation caused by bacterial components, metabolites, or bacteriophages may also mediate the effects of FMT. Especially as investigations into auto-FMT have also shown promising results in restoring gut microbiome composition.[462, 463] Clearly, there is a need for a deeper understanding of the dynamics underlying the interaction between donor and recipient microbiota during FMT.[464] This could ultimately lead to a safe and controlled modification from disturbed to desired phenotypes in the recipient.[23]

In the studies detailed in Chapter 5 and Chapter 6, we examined stool samples from 24 patients with mild to moderate UC undergoing FMT. Stool samples were collected at nine time points across the study period, allowing for a comprehensive assessment of gut microbiota dynamics during and post-FMT. Our longitudinal approach provided insights into weekly changes, a perspective often lacking in randomized controlled trials (RCTs) that focus primarily on clinical outcomes. Our results in Chapter 5 suggested that there is a potential for predicting clinical success of FMT treatment based on early microbiota analysis in the early phase of treatment, which would make it possible to adapt treatment strategies accordingly. However, developing a reliable predictive model for this purpose will require substantial additional effort.

It is plausible that differences in microbiota related to clinical success become apparent early during FMT treatment. The order in which species arrive can influence community succession (the predictable change in community composition over time), as early-arriving species can modify resources and environmental conditions, thereby affecting the establishment of later-arriving species. These priority effects can lead to varying successional pathways within the gut ecosystem.[83, 93] This concept is akin to plant ecosystems, where pioneer species prepare the environment for subsequent arrivals. For example, while a particular patch may not always host the same grass species, the presence of any grass helps create conditions that are conducive to the establishment of shrubs. Similarly, the growth of taller plants (regardless of specific species) facilitates the establishment of shade-tolerant species.[96] Therefore, to understand how microbial species interactions shape community dynamics during succession after FMT, we need to focus not just on which species are present, but also on the role each species plays within the community.

Several hypothetical outcomes of the FMT treatment are possible (Figure 7.4). First, the host communities may revert to their initial dysbiotic state if the perturbation is too weak and the dysbiotic state too strong. Therefore, the transferred microorganisms fail to change the microbiome or to establish themselves permanently. Second, due to intrinsic host or environmental factors, an alternative dysbiotic state may emerge, wherein the microbial community, although different in composition, possibly continues to perform detrimental ecosystem services. Third, an alternative healthy state may emerge, characterized by a novel microbiota composition with beneficial properties. Fourth, the microbiota changes to resemble the donor state, ideally incorporating the donor's healthy functions.[23]

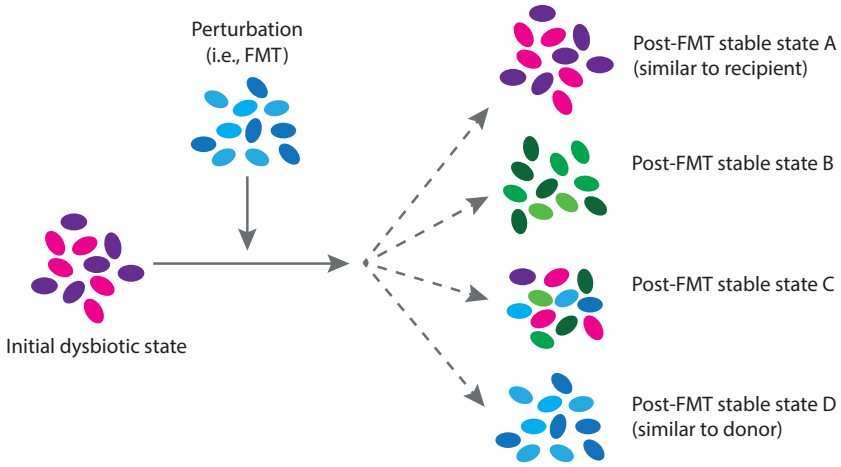


Figure 7.4 - Hypothetical outcomes of FMT treatment on microbiota composition. The interaction between two microbial consortia during FMT treatment may be likened to a complex pulse perturbation intended to transfer the functional properties of a donor microbiota to a recipient. Several potential outcomes can arise. First, one possibility is that the host microbiota returns to its original dysbiotic state (referred to as stable state A), as the introduced microorganisms fail to establish themselves permanently due to an insufficient perturbation. Second, the interaction may lead to the establishment of a completely new microbial community (referred to as stable state B), comprising species neither from the donor nor the original community. This novel community may arise due to a combination of factors such as niche availability, competitive exclusion, and environmental influences. Importantly, this new community could exhibit either beneficial or dysbiotic properties, depending on the specific composition and functional attributes of the newly established species. Third, due to intrinsic host or environmental factors, an alternative state is selected as the outcome (referred to as stable state C), comprising a mix of donor, patient, and new species. Fourth, resilience of the donor community (referred to as stable state D) in the new habitat could lead to a new interaction with long-term transfer of potential beneficial properties. This figure is adapted from Sommer et al. (2017).[23]

In Chapter 6, we applied a methodology inspired by Schmidt et al. 2022 to the same dataset as the one used in Chapter 5 to investigate the extent to which a shift in the patient's microbiota towards the donor microbiota is beneficial for resolving dysbiosis in the patient's gut.[199] Engraftment has long been considered a key mechanism underlying the success of fecal microbiota transplantation.[399] However, insights from earlier studies have raised questions about what happens to all the species involved during the succession phase of the treatment (during and after FMT).[188, 199] Therefore, we categorized species within the recipient's gut microbiota into ecological groups based on their origin and presence over time: those either already present in the host before FMT, derived from the donor, or introduced as novel species (absent in both host pre-FMT and donor samples). Our findings revealed that responders retained more resident species and maintained a more constant level of colonization over time compared to non-responders. This suggests that a favourable response to FMT is facilitated by a microbiota receptive to colonization, without compromising the resident community.

Restoring the microbiota with an FMT treatment is a complex process, as different taxa recover or colonize to varying extents, with some failing to (re)establish entirely.[58, 86, 226, 465] This variability can be influenced by suppression and resource competition between invaders and resident species.[87, 466-468] To mitigate the pressure from the resident species, a bowel lavage was performed prior to the first treatment, allowing for a more conducive environment for donor species to colonize. However, it is likely that the species that successfully colonize the gut after FMT are those closely related to the original inhabitants, as the gut environment provides a suitable niche for their growth.[87, 188, 469] Even if the original species are replaced or supplemented by similar ones, the new microbes may potentially introduce new traits that alter the ecosystem's functionality and metabolic output.[95] Moreover, if donor species may fail to establish, they might still be able to impact the recipient community's functioning and induce autonomous changes through interactions with resident members, for example by horizontal gene transfer or local metabolic activities while passing through.[88, 95, 466, 467] It has been shown that in a fluctuating environment, rapid evolution can destabilize the long-term stability of interactions, potentially enhancing adaptability and resilience or disrupting microbial balance and health.[470] Our study could not determine whether the species that reappear are leftover residents that regrew post-lavage, whether they gained additional functions, or whether they originated from the transplanted donor material if they are identical to the recipient species pre-FMT.

Future directions for microbiome research in health and disease

The recognition of the microbiome's critical role in our health marks a significant shift from traditional clinical perspectives, which often view the body as a battleground between human cells and microbes (i.e., pathogens) to an understanding that embraces the complex ecological community context of the microbiome. A dysbiotic human gut microbiome can be likened to plant or animal communities in a highly disturbed environment, e.g., impacted by overfishing, (abrupt) climate change, habitat loss, ocean acidification, pollution, or an invasive species. Human interventions, such as generic antibiotic use, have demonstrated detrimental effects on both the microbiome and human health, mirroring the irreversible changes observed in disrupted ecosystems where pesticides are used instead of ecological management measures. Therefore, to overcome dysbiosis in complex chronic diseases, we can draw inspiration from strategies such as habitat restoration and targeted removal of invasive species, which have been successfully applied in large-scale biodiversity management. For chronic diseases, an ecological maintenance approach may be more effective than the traditional battlefield strategy.[22, 226]

The limitations of the traditional 'one-size-fits-all' treatment approach, based on broad population averages, have also become increasingly apparent due to the heterogeneity in genotypes and phenotypes of gastrointestinal diseases among human populations. For example, matching donors and recipients by lifestyle and diet could enhance the likelihood of transplanting species that are effective colonizers or providing the resident species with the necessary metabolites that support their growth and function, thereby potentially improving the recipient's microbiome more successfully. Potentially, a better FMT success can also be achieved through the administration of specific prebiotics alongside the microbes. By providing targeted substrates exclusively metabolized by preferred species, prebiotics could create an advantage for them.[471, 472] Such an approach potentially strengthens the recipient's own microbiota and enriches it with species that naturally belong to the same community, leading to a more harmonious and effective community. However, the specific species that are most beneficial and those that are naturally suited to the community still need to be identified before this strategy can become a reality. As our understanding of the microbiome ecosystem advances, doctors will hopefully be equipped with precise disease prevention strategies and more effective treatments in the future.

A recurring theme in microbiome research is the need for large, densely sampled temporal datasets encompassing individuals from diverse backgrounds and lifestyles. Such datasets would be instrumental in unraveling fundamental mechanisms governing ecosystem dynamics in health and disease. Furthermore, studying microbiomes from various geographical regions (e.g., Africa) is important for capturing the global diversity in microbiological research, as most studies to date have focused on the United States, Europe, and Asia.[473] The unique environmental factors, dietary habits, and cultural practices in different regions in the world can significantly influence microbial composition and function.[474] By incorporating diverse microbiomes into our research, we can enhance our understanding of microbial dynamics that could inform health strategies and identify important confounding variables that may influence health outcomes.

Another way forward is to paint a more comprehensive picture of the microbial ecosystem with an integrative ecosystem biology approach that combines multiple omics technologies with host physiological data, and in depth knowledge of bacterial species behaviour and their (chemical) environment.[3, 60, 152, 244, 440, 475] By examining fecal matter in more detail alongside dietary questionnaires or food diaries, we might be able to extract valuable information about the host's diet, offering more insights than what is typically available. Note that the presence of a nutrient in a fecal sample is often assumed to indicate its importance for the microbiome. However, it could also be present because the species have not utilized it, leaving it to be excreted in the feces. Additionally, simultaneous assessment of mediators of reciprocal host-microbe interactions, such as microbial metabolites and immunological parameters, holds promise for identifying causality, discerning what changes first and who or what influences whom at various points in time.[121] At present, a significant question remains unanswered: whether the microbiota differs in various disease states because it causes these states, whether the microbiota differs as a consequence of the patients' disease state, or whether both are caused by the same external factors (for example altered diet or lifestyle). Mixing up association with causality can lead to an overestimation of the clinical relevance and impact of the microbiome on diseases.[214] For example, bacteria associated with unhealthy microbiomes may not necessarily be those directly related to the disease; instead, they could merely be among the few species capable of thriving in a gut environment with reduced diversity (possibly due to chance as described by the neutral theory); or they may play a beneficial role by supporting the host in the restoration of the healthy microbial community.[476]

Future research could also aim to identify not only bacteria, but also other microbes such as Archaea, fungi, and viruses, while exploring their interactions with each other and with bacteria, as well as their potential roles in health and disease. This includes investigating phage therapy as a strategy to target specific bacteria or pathogens, as bacteriophages may regulate intestinal microbiota diversity through mechanisms such as the kill-the-winner principle (which targets the most abundant bacterial species) or by specifically eliminating a species of interest, thereby preventing, for example, the outgrowth of Prevotellaceae in UC patients.[67, 477-479] Cross-domain networks may be important in understanding microbiome dynamics and ecosystem resilience, as there are many correlations with the bacterial microbiome and other domains.[428]

It is important to find a balance between collecting extensive data and maintaining clarity and interpretability. Merely increasing sequencing efforts is insufficient; the analysis pipelines must also continuously evolve to accommodate the influx of new data types and quantities. Moreover, focusing on excessively granular data might lead to a loss of statistical power due to the large number of species or functions relative to the number of patients and the prevalence of rare taxa. Additionally, the fact that different bacterial species can perform the same functional role in different patients may require a much larger sample size or functional assay than is (currently) possible in microbiome studies.[118, 168] On top of that, even the 'simple' networks with only bacteria generated from currently available data are challenging to grasp. Before introducing further complexity, we must step back to reflect on our research designs and develop strategies to effectively comprehend the influx of new information.

Finally, increasing the database of cultured microorganisms and annotated genes is needed for a comprehensive understanding of microbial function and for creating benchmark data to improve the evaluation of tool performance.[118, 480] Without the information about, for example, functional redundancy, dormancy, and phenotypic plasticity, taxonomic data alone offers limited insights into ecosystem processes across space and time.[425, 480, 481] Fundamental research on gut microbiota, including culturing of isolates, remains important for understanding the interspecies interactions and bacterial behaviours and dynamics, as it provides species-specific knowledge.[482] Mechanistic research in wet-lab and (animal) models is also imperative to validate the hypothesized mechanisms of species behaviours, not only for the most abundant ones, as they may not be the most important from an ecological point of view.[483-488] However, replicating complex human gut microbiome interactions (in artificial gut models) poses significant challenges, despite all the current advances in the field, and warrants further improvements.[118, 489-491] Ideally, establishing gold standards for microbiota data analysis and comprehensive reporting of (meta)data would enable more meaningful comparisons across studies, a call made over a decade ago but still largely unaddressed.[492, 493]

Concluding remarks

The journey of microbiome research reveals both the complexities and the promises for enhancing human health. As technology advances, so does our understanding of the microbiome. More fine-grained studies on the (gut) microbiome and its role in human health are needed to provide interpretation and meaning on the differences already found. Sophisticated technologies, such as Artificial Intelligence (AI), machine learning, and network analyses hold potential for identifying patterns within microbiota community data. However, those results should still be considered in light of past discoveries, established methods and models, and longstanding theories from multiple fields. When we combine (mathematical) modeling, theoretical knowledge, and experimental approaches, we gain a more comprehensive understanding of complex biological systems allowing us to validate results, do predictions, uncover underlying mechanisms, and refine our models for more accurate insights, as demonstrated throughout this thesis.

Collaborations across multidisciplinary groups, comprising, among others, (microbial) ecologists, healthcare professionals, complexity scientists, and bioinformaticians will further enrich our research field. Complex systems exist on a spectrum between order and randomness. Although one can get lost in the hairball of a complex network, knowledge from several fields can help. Understanding how systems respond to changes and return to stability enhances our grasp of the complex dynamics within the human gut microbiome. This knowledge can ultimately improve microbiome-modulating strategies and drive innovation of therapeutic strategies. Improved data sharing practices, including publishing raw data in a standardized fashion and statistical code will facilitate higher-quality meta-analyses and the establishment of more robust microbial signatures for diseases.[439, 494] Unfortunately, data accessibility still poses a significant challenge in microbiota research, with researchers frequently withholding study-related data. While concerns about privacy and efforts required for data collection are understandable, limited data sharing impedes scientific advancement.[495-500]

By recognizing that each step brings us closer to harnessing the microbiome's potential to improve human health, we ensure continuous progress and discovery. To truly understand microbial dynamics, it is important to acknowledge that human time is vastly different from bacterial time. Bacteria perceive their environment, resources, and interactions on a much smaller spatial scale. They constantly adapt to their immediate surroundings and rapidly shifting communities. An (microbial) ecological perspective grounded in theory is essential to interpret the impact of the microbiome on our health and disease.

Appendices

References

References

1 Lewin R. Complexity: life at the edge of chaos. 2nd ed: The university of Chicago Press; 1999.
2 Scheffer M. Critical transitions in nature and society: Princeton University Press; 2009.
3 Thurner S, Klimek P, Hanel R. Introduction to the theory of complex systems. United Kingdom: Oxford University Press; 2018.
4 Vermeer B, Zinsmeister S. Grip on complexity: how manageable are complex systems? Directions for future complexity research. The Hage: Netherlands Organisation for Scientific Research; 2014.
5 Benincà E, Huisman J, Heerkloss R, Johnk KD, Branco P, Van Nes EH, et al. Chaos in a long-term experiment with a plankton community. Nature. 2008;451(7180):822-825.
6 Benincà E, Ballantine B, Ellner SP, Huisman J. Species fluctuations sustained by a cyclic succession at the edge of chaos. Proc Natl Acad Sci U S A. 2015;112(20):6389-6394.
7 May RM. Stability and complexity in model ecosystems: Princeton University Press; 2019.
8 Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012;10(8):538-550.
9 Gallagher R, Appenzeller T. Beyond reductionism. Science. 1999;284(5411):79-79.
10 Parrott L, Lange H. An introduction to complexity science. In: Messier C, Puettmann KJ, Coates KD, editors. Managing forests as complex adaptive systems. London: Routledge; 2013.
11 Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006;124(4):837-848.
12 Gupta VK, Paul S, Dutta C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. Front Microbiol. 2017;8:1162.
13 Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560-564.
14 Leviatan S, Shoer S, Rothschild D, Gorodetski M, Segal E. An expanded reference map of the human gut microbiome reveals hundreds of previously unknown species. Nat Commun. 2022;13(1):3863.
15 Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.
16 Deo PN, Deshmukh R. Oral microbiome: unveiling the fundamentals. J Oral Maxillofac Pathol. 2019;23(1):122-128.
17 Lee YH, Chung SW, Auh QS, Hong SJ, Lee YA, Jung J, et al. Progress in oral microbiome related to oral and systemic diseases: an update. Diagnostics (Basel). 2021;11(7).
18 Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. 2015;3:31.
19 Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G, et al. The gut microbiota and host health: a new clinical frontier. Gut. 2016;65(2):330-339.
20 Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. N Engl J Med. 2016;375(24):2369-2379.
21 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-214.
22 Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415):220-230.
23 Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630-638.
24 Yadav M, Verma MK, Chauhan NS. A review of metabolic potential of human gut microbiome in human nutrition. Arch Microbiol. 2018;200(2):203-217.
25 Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science. 2016;352(6285):539-544.
26 Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J. 2017;474(11):1823-1836.
27 Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492-506.
28 Rastelli M, Cani PD, Knauf C. The gut microbiome influences host endocrine functions. Endocr Rev. 2019;40(5):1271-1284.
29 Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R, Goodman AL. Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature. 2019;570(7762):462-467.
30 Quinn RA, Melnik AV, Vrbanac A, Fu T, Patras KA, Christy MP, et al. Global chemical effects of the microbiome include new bile-acid conjugations. Nature. 2020;579(7797):123-129.
31 Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444(7122):1022-1023.
32 Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55(2):205-211.
33 Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology. 2010;139(6):1844-1854.e1841.
34 Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased diversity of the fecal Microbiome in recurrent *Clostridium difficile*-associated diarrhea. J Infect Dis. 2008;197(3):435-438.
35 Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655-662.

36 Safari Z, Gerard P. The links between the gut microbiome and non-alcoholic fatty liver disease (NAFLD). Cell Mol Life Sci. 2019;76(8):1541-1558.
37 Scher JU, Nayak RR, Ubeda C, Turnbaugh PJ, Abramson SB. Pharmacomicrobiomics in inflammatory arthritis: gut microbiome as modulator of therapeutic response. Nat Rev Rheumatol. 2020;16(5):282-292.
38 Levan SR, Stamnes KA, Lin DL, Panzer AR, Fukui E, McCauley K, et al. Elevated faecal 12,13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance. Nat Microbiol. 2019;4(11):1851-1861.
39 Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, et al. Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. Nat Med. 2019;25(11):1772-1782.
40 van Munster KN, Bergquist A, Ponsioen CY. Inflammatory bowel disease and primary sclerosing cholangitis: One disease or two? J Hepatol. 2024;80(1):155-168.
41 Kahharova D, Pappalardo VY, Buijs MJ, de Menezes RX, Peters M, Jackson R, et al. Microbial indicators of dental health, dysbiosis, and early childhood caries. J Dent Res. 2023;102(7):759-766.
42 Othman M, Aguero R, Lin HC. Alterations in intestinal microbial flora and human disease. Curr Opin Gastroenterol. 2008;24(1):11-16.
43 Tap J, Storsrud S, Le Neve B, Cotillard A, Pons N, Dore J, et al. Diet and gut microbiome interactions of relevance for symptoms in irritable bowel syndrome. Microbiome. 2021;9(1):74.
44 Dong TS, Guan M, Mayer EA, Stains J, Liu C, Vora P, et al. Obesity is associated with a distinct brain-gut microbiome signature that connects *Prevotella* and *Bacteroides* to the brain's reward center. Gut Microbes. 2022;14(1):2051999.
45 Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. Science. 2013;341(6141):1237439.
46 Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50.
47 Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science. 2009;326(5960):1694-1697.
48 Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM. Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies. FEMS Microbiol Rev. 2017;41(2):182-199.
49 Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. Science. 2015;350(6261):663-666.
50 Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Microbiol. 2015;25:56-66.
51 Rajilic-Stojanovic M, Heilig HG, Tims S, Zoetendal EG, de Vos WM. Long-term monitoring of the human intestinal microbiota composition. Environ Microbiol. 2012.
52 Olsson LM, Boulund F, Nilsson S, Khan MT, Gummesson A, Fagerberg L, et al. Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. Cell Host Microbe. 2022;30(5):726-739.e723.
53 Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. Nat Genet. 2016;48(11):1407-1412.
54 Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic determinants of the gut microbiome in UK twins. Cell Host Microbe. 2016;19(5):731-743.
55 Gerber GK. The dynamic microbiome. FEBS Lett. 2014;588(22):4131-4139.
56 Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. mBio. 2016;7(4).
57 Gibbons SM, Kearney SM, Smillie CS, Alm EJ. Two dynamic regimes in the human gut microbiome. PLOS Comput Biol. 2017;13(2):e1005364.
58 Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci USA. 2011;108 Suppl 1:4554-4561.
59 David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014;15(7):R89.
60 Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, Troost FJ, et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J. 2012;6(7):1415-1426.
61 Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, et al. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J. 2014;8(4):881-893.
62 Relman DA. The human microbiome: ecosystem resilience and health. Nutr Rev. 2012;70 Suppl 1(Suppl 1):S2-9.
63 Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009;1(6):6ra14.
64 Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLOS Biol. 2007;5(7):e177.
65 Kolenbrander PE, Andersen RN, Blehert DS, Egland PG, Foster JS, Palmer RJ. Communication among oral bacteria. Microbiol Mol Biol Rev. 2002;66(3):486-505.
66 Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci USA. 2011;108 Suppl 1:4578-4585.
67 van Best N, Hornef MW, Savelkoul PH, Penders J. On the origin of species: factors shaping the establishment of infant's gut microbiota. Birth Defects Res C Embryo Today. 2015;105(4):240-251.
68 Bosch A, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of

upper respiratory tract microbiota in infancy is affected by mode of delivery. EBioMedicine. 2016;9:336-345.
69 Gerber GK, Onderdonk AB, Bry L. Inferring dynamic signatures of microbes in complex host ecosystems. PLOS Comput Biol. 2012;8(8):e1002624.
70 Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci USA. 2010;107(26):11971-11975.
71 Dalby MJ, Hall LJ. Recent advances in understanding the neonatal microbiome. F1000Res. 2020;9.
72 Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006;118(2):511-521.
73 Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence *in situ* hybridization (FISH). Anaerobe. 2011;17(6):478-482.
74 Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. CMAJ. 2013;185(5):385-394.
75 Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe. 2015;17(5):690-703.
76 Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. J Pediatr Gastroenterol Nutr. 2010;51(1):77-84.
77 Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Sears MR, et al. Infant gut microbiota and the hygiene hypothesis of allergic disease: impact of household pets and siblings on microbiota composition and diversity. Allergy Asthma Clin Immunol. 2013;9(1):15.
78 Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345(6200):1048-1052.
79 Penders J, Gerhold K, Stobberingh EE, Thijs C, Zimmermann K, Lau S, et al. Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood. J Allergy Clin Immunol. 2013;132(3):601-607.e608.
80 Holling CS. Resilience and stability of ecological systems. Annual Review of Ecology and Systematics. 1973;4(1):1-23.
81 Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev. 2013;77(3):342-356.
82 Sprockett D, Fukami T, Relman DA. Role of priority effects in the early-life assembly of the gut microbiota. Nat Rev Gastroenterol Hepatol. 2018;15(4):197-205.
83 Fukami T. Historical contingency in community assembly: integrating niches, species pools, and priority effects. Annual Review of Ecology, Evolution, and Systematics. 2015;46(1):1-23.
84 Verster AJ, Borenstein E. Competitive lottery-based assembly of selected clades in the human gut microbiome. Microbiome. 2018;6(1):186.
85 Litvak Y, Baumler AJ. The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. PLOS Pathog. 2019;15(2):e1007563.
86 Dethlefsen L, Eckburg PB, Bik EM, Relman DA. Assembly of the human intestinal microbiota. Trends Ecol Evol. 2006;21(9):517-523.
87 Darcy JL, Washburne AD, Robeson MS, Prest T, Schmidt SK, Lozupone CA. A phylogenetic model for the recruitment of species into microbial communities and application to studies of the human microbiome. ISME J. 2020;14(6):1359-1368.
88 Mallon CA, Elsas JDV, Salles JF. Microbial invasions: the process, patterns, and mechanisms. Trends Microbiol. 2015;23(11):719-729.
89 Booth DJ, Murray BR. Coexistence. Encyclopedia of Ecology, 2008. p. 664-668.
90 Trosvik P, Stenseth NC, Rudi K. Convergent temporal dynamics of the human infant gut microbiota. ISME J. 2010;4(2):151-158.
91 Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms. 2019;7(1).
92 Hou K, Wu ZX, Chen XY, Wang JQ, Zhang D, Xiao C, et al. Microbiota in health and diseases. Signal Transduct Target Ther. 2022;7(1):135.
93 Debray R, Herbert RA, Jaffe AL, Crits-Christoph A, Power ME, Koskella B. Priority effects in microbiome assembly. Nat Rev Microbiol. 2022;20(2):109-121.
94 Guittar J, Shade A, Litchman E. Trait-based community assembly and succession of the infant gut microbiome. Nat Commun. 2019;10(1):512.
95 Walter J, Maldonado-Gomez MX, Martinez I. To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. Curr Opin Biotechnol. 2018;49:129-139.
96 Fefferman NH, Price CA, Stringham OC. Considering humans as habitat reveals evidence of successional disease ecology among human pathogens. PLOS Biol. 2022;20(9):e3001770.
97 Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, et al. Anticipating critical transitions. Science. 2012;338(6105):344-348.
98 Mutshinda CM, O'Hara RB, Woiwod IP. What drives community dynamics? Proc Biol Sci. 2009;276(1669):2923-2929.
99 Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, et al. Signatures of ecological processes in microbial community time series. Microbiome. 2018;6(1):120.
100 Stein RR, Bucci V, Toussaint NC, Buffie CG, Ratsch G, Pamer EG, et al. Ecological modeling from

time-series inference: insight into dynamics and stability of intestinal microbiota. PLOS Comput Biol. 2013;9(12):e1003388.
101 Sloan WT, Nnaji CF, Lunn M, Curtis TP, Colloms SD, Couto JM, et al. Drift dynamics in microbial communities and the effective community size. Environ Microbiol. 2021;23(5):2473-2483.
102 Lahti L, Salojarvi J, Salonen A, Scheffer M, de Vos WM. Tipping elements in the human intestinal ecosystem. Nat Commun. 2014;5:4344.
103 Gonze D, Lahti L, Raes J, Faust K. Multi-stability and the origin of microbial community types. ISME J. 2017;11(10):2159-2166.
104 Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, et al. Regime shifts, resilience, and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics. 2004;35(1):557-581.
105 Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. Nature. 2001;413(6856):591-596.
106 Scheffer M, Carpenter S, Young B. Cascading effects of over-fishing marine systems. Trends Ecol Evol. 2005;20(11):579-581.
107 Potts LD, Douglas A, Perez Calderon LJ, Anderson JA, Witte U, Prosser JI, et al. Chronic environmental perturbation influences microbial community assembly patterns. Environ Sci Technol. 2022;56(4):2300-2311.
108 Daniels L, Budding AE, de Korte N, Eck A, Bogaards JA, Stockmann HB, et al. Fecal microbiome analysis as a diagnostic test for diverticulitis. Eur J Clin Microbiol Infect Dis. 2014;33(11):1927-1936.
109 Ye H, Sugihara G. Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. Science. 2016;353(6302):922-925.
110 Das P, Marcisauskas S, Ji B, Nielsen J. Metagenomic analysis of bile salt biotransformation in the human gut microbiome. BMC Genomics. 2019;20(1):517.
111 Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLOS One. 2014;9(7):e102451.
112 Rottjers L, Faust K. From hairballs to hypotheses-biological insights from microbial networks. FEMS Microbiol Rev. 2018;42(6):761-780.
113 Röttjers L, Faust K. Can we predict keystones? Nat Rev Microbiol. 2019;17(3):193.
114 Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol. 2014;5:219.
115 Power ME, Tilman D, Estes JA, Menge BA, Bond WJ, Mills LS, et al. Challenges in the quest for keystones. BioScience. 1996;46(8):609-620.
116 Pester M, Bittner N, Deevong P, Wagner M, Loy A. A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. ISME J. 2010;4(12):1591-1602.
117 Ze X, Duncan SH, Louis P, Flint HJ. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. ISME J. 2012;6(8):1535-1543.
118 Faust K. Open challenges for microbial network construction and analysis. ISME J. 2021;15(11):3111-3118.
119 Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun. 2011;2(2):589.
120 Hanemaaijer M, Roling WF, Olivier BG, Khandelwal RA, Teusink B, Bruggeman FJ. Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. Front Microbiol. 2015;6:213.
121 Wilmes P, Martin-Gallausiaux C, Ostaszewski M, Aho VTE, Novikova PV, Laczny CC, et al. The gut microbiome molecular complex in human health and disease. Cell Host Microbe. 2022;30(9):1201-1206.
122 San Roman M, Wagner A. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. PLOS Comput Biol. 2018;14(7):e1006340.
123 Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol. 2010;8(1):15-25.
124 Pianka ER. On r- and K-selection. The American Naturalist. 1970;104(940):592–597.
125 MacArthur RH, Wilson EO. The theory of island biogeography: Princeton University Press; 2001.
126 Hoek M, Merks RMH. Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. BMC Syst Biol. 2017;11(1):56.
127 Rabiu BA, Gibson GR. Carbohydrates: a limit on bacterial diversity within the colon. Biol Rev Camb Philos Soc. 2002;77(3):443-453.
128 Cummings JH, Macfarlane GT. The control and consequences of bacterial fermentation in the human colon. J Appl Bacteriol. 1991;70(6):443-459.
129 Arike L, Hansson GC. The densely O-glycosylated MUC2 mucin protects the intestine and provides food for the commensal bacteria. J Mol Biol. 2016;428(16):3221-3229.
130 La Rosa SL, Ostrowski MP, Vera-Ponce de Leon A, McKee LS, Larsbrink J, Eijsink VG, et al. Glycan processing in gut microbiomes. Curr Opin Microbiol. 2022;67:102143.
131 Sicard JF, Le Bihan G, Vogeleer P, Jacques M, Harel J. Interactions of intestinal bacteria with components of the intestinal mucus. Front Cell Infect Microbiol. 2017;7:387.
132 Chng KR, Ghosh TS, Tan YH, Nandi T, Lee IR, Ng AHQ, et al. Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. Nat Ecol Evol. 2020;4(9):1256-1267.
133 Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. Gut Microbes. 2012;3(4):289-306.
134 Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. Front Genet. 2015;6:81.
135 Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial

metabolites. Cell. 2016;165(6):1332-1345.
136 Machado D, Maistrenko OM, Andrejev S, Kim Y, Bork P, Patil KR, et al. Polarization of microbial communities between competitive and cooperative metabolism. Nat Ecol Evol. 2021;5(2):195-203.
137 Liu W, Jacquiod S, Brejnrod A, Russel J, Burmolle M, Sorensen SJ. Deciphering links between bacterial interactions and spatial organization in multispecies biofilms. ISME J. 2019;13(12): 3054-3066.
138 Leschine SB. Cellulose degradation in anaerobic environments. Annu Rev Microbiol. 1995;49: 399-426.
139 Tait K, Sutherland IW. Antagonistic interactions amongst bacteriocin-producing enteric bacteria in dual species biofilms. J Appl Microbiol. 2002;93(2):345-352.
140 Tan CH, Koh KS, Xie C, Zhang J, Tan XH, Lee GP, et al. Community quorum sensing signalling and quenching: microbial granular biofilm assembly. NPJ Biofilms Microbiomes. 2015;1:15006.
141 Vasse M, Fiegna F, Kriesel B, Velicer GJ. Killer prey: Ecology reverses bacterial predation. PLOS Biol. 2024;22(1):e3002454.
142 Bull CT, Shetty KG, Subbarao KV. Interactions between myxobacteria, plant pathogenic fungi, and biocontrol agents. Plant Dis. 2002;86(8): 889-896.
143 Raaijmakers JM, Vlami M, de Souza JT. Antibiotic production by bacterial biocontrol agents. Antonie Van Leeuwenhoek. 2002;81(1-4):537-547.
144 Haas D, Defago G. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat Rev Microbiol. 2005;3(4):307-319.
145 Askeland RA, Morrison SM. Cyanide production by *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*. Appl Environ Microbiol. 1983;45(6):1802-1807.
146 Gause GF. The struggle for existence: a classic of mathematical biology and ecology: Williams & Wilkins; 1934. 163 p.
147 Martinson JNV, Walk ST. *Escherichia coli* residency in the gut of healthy human adults. EcoSal Plus. 2020;9(1).
148 Lidicker WZ. A clarification of interactions in ecological systems. BioScience. 1979;29(8): 475-477.
149 Rodríguez-Martínez JM, Pascual A. Antimicrobial resistance in bacterial biofilms. Reviews in Medical Microbiology. 2006;17(3):65-75.
150 Russel J, Roder HL, Madsen JS, Burmolle M, Sorensen SJ. Antagonism correlates with metabolic similarity in diverse bacteria. Proc Natl Acad Sci USA. 2017;114(40):10684-10688.
151 Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. Nature. 2012;489(7415):231-241.
152 Araujo G, Montoya JM, Thomas T, Webster NS, Lurgi M. A mechanistic framework for complex microbe-host symbioses. Trends Microbiol. 2024.
153 Quaglio AEV, Grillo TG, De Oliveira ECS, Di Stasi LC, Sassaki LY. Gut microbiota, inflammatory bowel disease and colorectal cancer. World J Gastroenterol. 2022;28(30):4053-4060.
154 Wu K, Luo Q, Liu Y, Li A, Xia D, Sun X. Causal relationship between gut microbiota and gastrointestinal diseases: a mendelian randomization study. J Transl Med. 2024;22(1):92.
155 Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett. 2014;588(22):4223-4233.
156 Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med. 2015;21(8):895-905.
157 Murri M, Leiva I, Gomez-Zumaquero JM, Tinahones FJ, Cardona F, Soriguer F, et al. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. BMC Med. 2013;11:46.
158 Wacklin P, Kaukinen K, Tuovinen E, Collin P, Lindfors K, Partanen J, et al. The duodenal microbiota composition of adult celiac disease patients is associated with the clinical manifestation of the disease. Inflamm Bowel Dis. 2013;19(5):934-941.
159 Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology. 2014;146(6):1489-1499.
160 Wei S, Bahl MI, Baunwall SMD, Hvas CL, Licht TR. Determining gut microbial dysbiosis: a review of applied indexes for assessment of intestinal microbiota imbalances. Appl Environ Microbiol. 2021;87(11).
161 Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-1031.
162 Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65.
163 Levine JM, D'Antonio CM. Elton revisited: a review of evidence linking diversity and invasibility. Oikos. 1999;87(1).
164 Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-484.
165 Rajilic-Stojanovic M, de Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev. 2014;38(5): 996-1047.
166 Louis P, Young P, Holtrop G, Flint HJ. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA: acetate CoA-transferase gene. Environ Microbiol. 2010;12(2):304-314.
167 Crobach MJT, Ducarmon QR, Terveer EM, Harmanus C, Sanders I, Verduin KM, et al. The bacterial gut microbiota of adult patients infected, colonized or noncolonized by *Clostridioides difficile*. Microorganisms. 2020;8(5).
168 van Rossen TM, van Beurden YH, Bogaards JA, Budding AE, Mulder CJJ, Vandenbroucke-Grauls C. Fecal microbiota composition is a better predictor of recurrent *Clostridioides difficile* infection than clinical factors in a prospective, multicentre cohort study. BMC Infect Dis. 2024;24(1):687.

169 van den Heuvel TR, Jonkers DM, Jeuring SF, Romberg-Camps MJ, Oostenbrug LE, Zeegers MP, et al. Cohort profile: the Inflammatory Bowel Disease South Limburg Cohort (IBDSL). Int J Epidemiol. 2017;46(2):e7.
170 Zhang YZ, Li YY. Inflammatory bowel disease: pathogenesis. World J Gastroenterol. 2014;20(1):91-99.
171 Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15(3):382-392.
172 Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.
173 Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. Gut. 2016;66(5):813-822.
174 Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspith BN, Rayment N, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol. 2011;11:7.
175 Kolho KL, Korpela K, Jaakkola T, Pichai MV, Zoetendal EG, Salonen A, et al. Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. Am J Gastroenterol. 2015;110(6):921-930.
176 Papa E, Docktor M, Smillie C, Weber S, Preheim SP, Gevers D, et al. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PLOS One. 2012;7(6):e39242.
177 Andoh A, Kobayashi T, Kuzuoka H, Tsujikawa T, Suzuki Y, Hirai F, et al. Characterization of gut microbiota profiles by disease activity in patients with Crohn's disease using data mining analysis of terminal restriction fragment length polymorphisms. Biomed Rep. 2014;2(3):370-373.
178 Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, et al. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. Inflamm Bowel Dis. 2009;15(8):1183-1189.
179 Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, et al. Increased proportions of *Bifidobacterium* and the *Lactobacillus* group and loss of butyrate-producing bacteria in inflammatory bowel disease. J Clin Microbiol. 2013;52(2):398-406.
180 Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut. 2003;52(2):237-242.
181 Galazzo G, Tedjo DI, Wintjens DSJ, Savelkoul PHM, Masclee AAM, Bodelier AGL, et al. Faecal microbiota dynamics and their relation to disease course in Crohn's disease. J Crohns Colitis. 2019;13(10):1273-1282.
182 Kim M, Benayoun BA. The microbiome: an emerging key player in aging and longevity. Transl Med Aging. 2020;4:103-116.
183 Gibson GR, Roberfroid MB. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr. 1995;125(6):1401-1412.
184 Chen Y, Zhou J, Wang L. Role and mechanism of gut microbiota in human disease. Front Cell Infect Microbiol. 2021;11:625913.
185 Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, et al. Expert consensus document: the International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat Rev Gastroenterol Hepatol. 2017;14(8):491-502.
186 Jacobsen CN, Rosenfeldt Nielsen V, Hayford AE, Moller PL, Michaelsen KF, Paerregaard A, et al. Screening of probiotic activities of forty-seven strains of *Lactobacillus* spp. by *in vitro* techniques and evaluation of the colonization ability of five selected strains in humans. Appl Environ Microbiol. 1999;65(11):4949-4956.
187 Kristensen NB, Bryrup T, Allin KH, Nielsen T, Hansen TH, Pedersen O. Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials. Genome Med. 2016;8(1):52.
188 Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science. 2016;352(6285):586-589.
189 Ianiro G, Puncochar M, Karcher N, Porcari S, Armanini F, Asnicar F, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nat Med. 2022;28(9):1913-1923.
190 van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med. 2013;368(5):407-415.
191 Colman RJ, Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. J Crohns Colitis. 2014;8(12):1569-1581.
192 Glassner KL, Abraham BP, Quigley EMM. The microbiome and inflammatory bowel disease. J Allergy Clin Immunol. 2020;145(1):16-27.
193 El Hage Chehade N, Ghoneim S, Shah S, Chahine A, Mourad FH, Francis FF, et al. Efficacy of fecal microbiota transplantation in the treatment of active ulcerative colitis: a systematic review and meta-analysis of double-blind randomized controlled trials. Inflamm Bowel Dis. 2023;29(5):808-817.
194 Sood A, Singh A, Mahajan R, Midha V, Kaur K, Singh D, et al. Clinical predictors of response to faecal microbiota transplantation in patients with active ulcerative colitis. J Crohns Colitis. 2020.
195 Rees NP, Shaheen W, Quince C, Tselepis C, Horniblow RD, Sharma N, et al. Systematic review of donor and recipient predictive biomarkers of response to faecal microbiota transplantation in patients with ulcerative colitis. EBioMedicine. 2022;81:104088.
196 Schierova D, Brezina J, Mrazek J, Fliegerova KO, Kvasnova S, Bajer L, et al. Gut microbiome

changes in patients with active left-sided ulcerative colitis after fecal microbiome transplantation and topical 5-aminosalicylic acid therapy. Cells. 2020;9(10).
197 Costello SP, Hughes PA, Waters O, Bryant RV, Vincent AD, Blatchford P, et al. Effect of fecal microbiota transplantation on 8-week remission in patients with ulcerative colitis: a randomized clinical trial. JAMA. 2019;321(2):156-164.
198 Moayyedi P, Surette MG, Kim PT, Libertucci J, Wolfe M, Onischi C, et al. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. Gastroenterology. 2015;149(1):102-109.e106.
199 Schmidt TSB, Li SS, Maistrenko OM, Akanni W, Coelho LP, Dolai S, et al. Drivers and determinants of strain dynamics following fecal microbiota transplantation. Nat Med. 2022;28(9):1902-1912.
200 He R, Li P, Wang J, Cui B, Zhang F, Zhao F. The interplay of gut microbiota between donors and recipients determines the efficacy of fecal microbiota transplantation. Gut Microbes. 2022;14(1):2100197.
201 Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5(1):27.
202 Budding AE, Grasman ME, Eck A, Bogaards JA, Vandenbroucke-Grauls CM, van Bodegraven AA, et al. Rectal swabs for analysis of the intestinal microbiota. PLOS One. 2014;9(7):e101344.
203 Robinson CK, Brotman RM, Ravel J. Intricacies of assessing the human microbiome in epidemiologic studies. Ann Epidemiol. 2016;26(5):311-321.
204 Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data. 2019;6:190007.
205 Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. F1000Res. 2016;5:1492.
206 Mysara M, Vandamme P, Props R, Kerckhof FM, Leys N, Boon N, et al. Reconciliation between operational taxonomic units and species boundaries. FEMS Microbiol Ecol. 2017;93(4).
207 Ruscheweyh HJ, Milanese A, Paoli L, Sintsova A, Mende DR, Zeller G, et al. mOTUs: profiling taxonomic composition, transcriptional activity and strain populations of microbial communities. Curr Protoc. 2021;1(8):e218.
208 Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-2643.
209 Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42(Database issue):D633-642.
210 Vermeire S, Joossens M, Verbeke K, Wang J, Machiels K, Sabino J, et al. Donor species richness determines faecal microbiota transplantation success in inflammatory bowel disease. J Crohns Colitis. 2016;10(4):387-394.
211 McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Elife. 2019;8.
212 McLaren MR, Nearing JT, Willis AD, Lloyd KG, Callahan BJ. Implications of taxonomic bias for microbial differential-abundance analysis. bioRxiv. 2022.
213 Wright RJ, Comeau AM, Langille MGI. From defaults to databases: parameter and database choice dramatically impact the performance of metagenomic taxonomic classification tools. Microb Genom. 2023;9(3).
214 Zaura E, Pappalardo VY, Buijs MJ, Volgenant CMC, Brandt BW. Optimizing the quality of clinical studies on oral microbiome: A practical guide for planning, performing, and reporting. Periodontol 2000. 2021;85(1):210-236.
215 Weiss SJ, Xu Z, Amir A, Peddada S, Bittinger K, Gonzalez A, et al. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. PeerJ Preprints. 2015.
216 Carr A, Diener C, Baliga NS, Gibbons SM. Use and abuse of correlation analyses in microbial ecology. ISME J. 2019;13(11):2647-2655.
217 Kodikara S, Ellul S, Le Cao KA. Statistical challenges in longitudinal microbiome data analysis. Brief Bioinform. 2022;23(4):bbac273.
218 Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013;10(12):1200-1202.
219 Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
220 Leite MFA, Kuramae EE. You must choose, but choose wisely: model-based approaches for microbial community analysis. Soil Biology and Biochemistry. 2020;151.
221 Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol. 2016;26(5):322-329.
222 McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLOS Comput Biol. 2014;10(4):e1003531.
223 Jiang D, Armour CR, Hu C, Mei M, Tian C, Sharpton TJ, et al. Microbiome multi-omics network analysis: statistical considerations, limitations, and opportunities. Front Genet. 2019;10:995.
224 Xia Y, Sun J. Hypothesis testing and statistical analysis of microbiome. Genes Dis. 2017;4(3):138-148.
225 Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome. 2018;6(1):32.
226 Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. Science. 2012;336(6086):1255-1262.
227 Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: A Network Perspective. Trends Microbiol. 2017;25(3):217-228.

228 Zuo Y, Yu G, Tadesse MG, Ressom HW. Biological network inference using low order partial correlation. Methods. 2014;69(3):266-273.
229 Hirano H, Takemoto K. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinformatics. 2019;20(1):329.
230 Faust K, Raes J. CoNet app: inference of biological association networks using Cytoscape. F1000Res. 2016;5:1519.
231 Park SY, Ufondu A, Lee K, Jayaraman A. Emerging computational tools and models for studying gut microbiota composition and function. Curr Opin Biotechnol. 2020;66:301-311.
232 Gonze D, Coyte KZ, Lahti L, Faust K. Microbial communities as dynamical systems. Curr Opin Microbiol. 2018;44:41-49.
233 Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, McDonald D, et al. Rethinking "enterotypes". Cell Host Microbe. 2014;16(4): 433-437.
234 Gower JC. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika. 1966;53(3-4):325-338.
235 Coenen AR, Hu SK, Luo E, Muratore D, Weitz JS. A primer for microbiome time-series analysis. Front Genet. 2020;11:310.
236 Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature. 2018;562(7728): 583-588.
237 Kumar M, Ji B, Zengler K, Nielsen J. Modelling approaches for studying the microbiome. Nat Microbiol. 2019;4(8):1253-1267.
238 Mounier J, Monnet C, Vallaeys T, Arditi R, Sarthou AS, Helias A, et al. Microbial interactions within a cheese microbial community. Appl Environ Microbiol. 2008;74(1):172-181.
239 Kuntal BK, Gadgil C, Mande SS. Web-gLV: a web based platform for Lotka-Volterra based modeling and simulation of microbial populations. Front Microbiol. 2019;10:288.
240 Momeni B, Xie L, Shou W. Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. Elife. 2017;6:1-34.
241 Chan SHJ, Simons MN, Maranas CD. SteadyCom: Predicting microbial abundances while ensuring community stability. PLOS Comput Biol. 2017;13(5):e1005539.
242 Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, et al. Quantifying diet-induced metabolic changes of the human gut microbiome. Cell Metab. 2015;22(2):320-331.
243 Zengler K, Palsson BO. A road map for the development of community systems (CoSy) biology. Nat Rev Microbiol. 2012;10(5):366-372.
244 Diener C, Gibbons SM. More is different: metabolic modeling of diverse microbial communities. mSystems. 2023;8(2):e0127022.
245 Karkman A, Lehtimaki J, Ruokolainen L. The ecology of human microbiota: dynamics and diversity in health and disease. Ann N Y Acad Sci. 2017;1399(1):78-92.
246 Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, et al. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006;443(7114):950-955.
247 Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. 2010;20(7):947-959.
248 Riera J, Baldo L. Animal Microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. Microbiome. 2020;2:36.
249 Vemuri R, Martoni CJ, Kavanagh K, Eri R. *Lactobacillus acidophilus* DDS-1 modulates the gut microbial co-occurrence networks in aging mice. Nutrients. 2022;14.
250 Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLOS Comput Biol. 2012;8(9):e1002687.
251 Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLOS Comput Biol. 2015;11(5):e1004226.
252 Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLOS Comput Biol. 2012;8(7):e1002606.
253 Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. 2016;10(7): 1669-1681.
254 Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, et al. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome Biol. 2016;17(1):121.
255 Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST, Liu YY. Mapping the ecological networks of microbial communities. Nat Commun. 2017;8(1):2042.
256 Jones EW, Shankin-Clarke P, Carlson JM. Navigation and control of outcomes in a generalized Lotka-Volterra model of the microbiome. In: Kotas J, editor. Advances in nonlinear biological systems: modeling and optimal control. 11. USA: American Institute of Mathematical Sciences; 2020. p. 97-120.
257 Yodzis P. Introduction to theoretical ecology: Longman Higher Education; 1989.
258 Gause GF. Experimental studies on the struggle for existence. Journal of Experimental Biology. 1932;9(4):389-402.
259 Hardin G. The competitive exclusion principle. Science. 1960;131(3409):1292-1297.
260 Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13(4):260-270.
261 Burnham P, Gomez-Lopez N, Heyang M, Cheng AP, Lenz JS, Dadhania DM, et al. Separating the signal from the noise in metagenomic cell-free DNA sequencing. Microbiome. 2020;8(1):18.
262 Hindmarsh AC, Petzold LR. Algorithms and software for ordinary differential equations and differential algebraic equations, Part II: Higher-order methods and software packages. Computers in Physics 1995;9:148-155.

263 Soetaert K, Petzoldt T, Setzer RW. Solving differential equations in R: package deSolve. Journal of Statistical Software. 2010;33(9).
264 Goh K-II, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. PNAS. 2002;99(20):12583–12588.
265 Barabási A. Network science. 1st ed: Cambridge University Press; 2016.
266 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1995;57(1):289-300.
267 Sasaki Y. The truth of the F-measure. University of Manchester: School of Computer Science. 2007: 1-5.
268 Cougoul A, Bailly X, Vourc'h G, Gasqui P. Rarity of microbial species: In search of reliable associations. PLOS One. 2019;14(3):e0200458.
269 Gupta S, Hjelmso MH, Lehtimaki J, Li X, Mortensen MS, Russel J, et al. Environmental shaping of the bacterial and fungal community in infant bed dust and correlations with the airway microbiota. Microbiome. 2020;8(1):115.
270 Ma ZS. The P/N (positive-to-negative links) ratio in complex networks-a promising *in silico* biomarker for detecting changes occurring in the human microbiome. Microb Ecol. 2018;75(4):1063-1073.
271 Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer AC, et al. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. 2020;8(1):133.
272 Stone L, Roberts A. Conditions for a species to gain advantage from the presence of competitors. Ecology. 1991;72:947-957.
273 Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? Ecology. 2018;99(3):690-699.
274 Mayfield MM, Stouffer DB. Higher-order interactions capture unexplained complexity in diverse communities. Nat Ecol Evol. 2017;1(3):62.
275 Tilman D. The importance of the mechanisms of interspecific competition. The American Naturalist. 1987;129(5):769-774.
276 Pace ML, Cole JJ, Carpenter SR, Kitchell JF. Trophic cascades revealed in diverse ecosystems. Trends Ecol Evol. 1999;14(12):483-488.
277 Højsgaard S, Edwards D, Lauritzen S. Graphical models with R: Springer Science & Business Media; 2012.
278 Wootton J. Predicting direct and indirect effects: an integrated approach using experiments and path analysis. Ecology. 1994;75(1):151-165.
279 Succurro A, Ebenhoh O. Review and perspective on mathematical modeling of microbial ecosystems. Biochem Soc Trans. 2018;46(2): 403-412.
280 Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol. 2000;15(8):321-326.
281 Costantino RF, Desharnais RA, Cushing JM, Dennis B. Chaotic dynamics in an insect population. Science. 1997;275(5298):389-391.
282 Becks L, Hilker FM, Malchow H, Jurgens K, Arndt H. Experimental demonstration of chaos in a microbial food web. Nature. 2005;435(7046): 1226-1229.
283 Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. Nature. 2009;461(7260): 53-59.
284 Faust K, Raes J. Host-microbe interaction: rules of the game for microbiota. Nature. 2016;534(7606):182-183.
285 Rottjers L, Faust K. Manta: a clustering algorithm for weighted ecological networks. mSystems. 2020;5(1).
286 Claussen JC, Skieceviciene J, Wang J, Rausch P, Karlsen TH, Lieb W, et al. Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. PLOS Comput Biol. 2017;13(6):e1005361.
287 Blanchet FG, Cazelles K, Gravel D. Co-occurrence is not evidence of ecological interactions. Ecol Lett. 2020;23(7):1050-1063.
288 Steinway SN, Biggs MB, Loughran TP, Jr., Papin JA, Albert R. Inference of network dynamics and metabolic interactions in the gut microbiome. PLOS Comput Biol. 2015;11(5):e1004338.
289 Bjornstad ON, Grenfell BT. Noisy clockwork: time series analysis of population fluctuations in animals. Science. 2001;293(5530):638-643.
290 Rouyer T, Fromentin JM, Menard F, Cazelles B, Briand K, Pianet R, et al. Complex interplays among population dynamics, environmental forcing, and exploitation in fisheries. Proc Natl Acad Sci U S A. 2008;105(14):5420-5425.
291 Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc Natl Acad Sci U S A. 2013;110(31):12804-12809.
292 David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559-563.
293 Schwartz DJ, Langdon AE, Dantas G. Understanding the impact of antibiotic perturbation on the human microbiome. Genome Med. 2020;12(1):82.
294 Joseph TA, Pasarkar AP, Pe'er I. Efficient and accurate inference of mixed microbial population trajectories from longitudinal count data. Cell Syst. 2020;10(6):463-469 e466.
295 Shenhav L, Furman O, Briscoe L, Thompson M, Silverman JD, Mizrahi I, et al. Modeling the temporal dynamics of the gut microbial community in adults and infants. PLOS Comput Biol. 2019;15(6):e1006960.
296 Joseph TA, Shenhav L, Xavier JB, Halperin E, Pe'er I. Compositional Lotka-Volterra describes microbial dynamics in the simplex. PLOS Comput Biol. 2020;16(5):e1007917.
297 Rouyer T, Fromentin JM, Stenseth NC, Cazelles B. Analysing multiple time series and extending significance testing in wavelet analysis. Marine Ecology Progress Series. 2008;359:11-23.
298 Addison PS. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine, and finance: CRC Press; 2017.

299 Benincà E, van Boven M, Hagenaars T, van der Hoek W. Space-time analysis of pneumonia hospitalisations in the Netherlands. PLOS One. 2017;12(7):e0180797.
300 Ben-Ari T, Neerinckx S, Agier L, Cazelles B, Xu L, Zhang Z, et al. Identification of Chinese plague foci from long-term epidemiological data. Proc Natl Acad Sci U S A. 2012;109(21):8196-8201.
301 Teissier Y, Paul R, Aubry M, Rodo X, Dommar C, Salje H, et al. Long-term persistence of monotypic dengue transmission in small size isolated populations, French Polynesia, 1978-2014. PLOS Negl Trop Dis. 2020;14(3):e0008110.
302 Martin-Platero AM, Cleary B, Kauffman K, Preheim SP, McGillicuddy DJ, Alm EJ, et al. High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nat Commun. 2018;9(1):266.
303 Torrence C, Compo GP. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society. 1998;79(1):61-78.
304 Cazelles B, Cazelles K, Chavez M. Wavelet analysis in ecology and epidemiology: impact of statistical tests. J R Soc Interface. 2014;11(91):20130585.
305 Cazelles B, Chavez M, Berteaux D, Menard F, Vik JO, Jenouvrier S, et al. Wavelet analysis of ecological time series. Oecologia. 2008;156(2):287-304.
306 Cazelles B, Chavez M, Magny GC, Guegan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. J R Soc Interface. 2007;4(15):625-636.
307 Chavez M, Cazelles B. Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data. Sci Rep. 2019;9(1):7389.
308 Fowlkes EB, Mallows CL. A method for comparing two hierarchical clusterings. Journal of the American Statistical Association. 1983;78(383):553-569.
309 Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31(22):3718-3720.
310 Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson G, Solymos P, et al. vegan: community ecology package, R package version 2.6-4 2022.
311 Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems. 2017;2(2).
312 Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457-463.
313 Lahti L, Shetty S, Blake T, Salojarvi J. Tools for microbiome analysis in R. Version 1. 2017;507.
314 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS One. 2013;8(4):e61217.
315 van den Boogaart KG, Tolosana-Delgado R. "compositions": a unified R package to analyze compositional data. Computers & Geosciences. 2008;34(4):320-338.
316 Millard S. EnvStats: an R package for environmental statistics. New York: Springer; 2013.
317 Ward JH. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association. 1963;58(301):236-244.
318 Vandermeer J. Coupled oscillations in food webs: balancing competition and mutualism in simple ecological models. Am Nat. 2004;163(6):857-867.
319 Benincà E, Johnk KD, Heerkloss R, Huisman J. Coupled predator-prey oscillations in a chaotic food web. Ecol Lett. 2009;12(12):1367-1378.
320 Massoud EC, Huisman J, Benincà E, Dietze MC, Bouten W, Vrugt JA. Probing the limits of predictability: data assimilation of chaotic dynamics in complex food webs. Ecol Lett. 2018;21(1):93-103.
321 Wang Y, Niu Q, Zhang X, Liu L, Wang Y, Chen Y, et al. Exploring the effects of operational mode and microbial interactions on bacterial community assembly in a one-stage partial-nitritation anammox reactor using integrated multi-omics. Microbiome. 2019;7(1):122.
322 Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. ISME J. 2012;6(2):298-308.
323 Giovannoni SJ, Vergin KL. Seasonality in ocean microbial communities. Science. 2012;335(6069):671-676.
324 Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc Natl Acad Sci U S A. 2006;103(35):13104-13109.
325 Newton RJ, Kent AD, Triplett EW, McMahon KD. Microbial community dynamics in a humic lake: differential persistence of common freshwater phylotypes. Environ Microbiol. 2006;8(6):956-970.
326 Matulich KL, Weihe C, Allison SD, Amend AS, Berlemont R, Goulden ML, et al. Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. ISME J. 2015;9(11):2477-2489.
327 Carini P, Delgado-Baquerizo M, Hinckley ES, Holland-Moritz H, Brewer TE, Rue G, et al. Effects of spatial variability and relic DNA removal on the detection of temporal dynamics in soil microbial communities. mBio. 2020;11(1).
328 Flores GE, Caporaso JG, Henley JB, Rideout JR, Domogala D, Chase J, et al. Temporal variability is a personalized feature of the human microbiome. Genome Biol. 2014;15(12):531.
329 Vandeputte D, De Commer L, Tito RY, Kathagen G, Sabino J, Vermeire S, et al. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. Nat Commun. 2021;12(1):6740.
330 Kugathasan S, Denson LA, Walters TD, Kim MO, Marigorta UM, Schirmer M, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. Lancet. 2017;389(10080):1710-1718.
331 Petchey OL, Gaston KJ. Functional diversity: back to basics and looking forward. Ecol Lett. 2006;9(6):741-758.
332 Shetty SA, Kostopoulos I, Geerlings SY, Smidt H, de Vos WM, Belzer C. Dynamic metabolic interactions and trophic roles of human gut

microbes identified using a minimal microbiome exhibiting ecological properties. ISME J. 2022;16(9):2144-2159.
333 Louis P, Duncan SH, Sheridan PO, Walker AW, Flint HJ. Microbial lactate utilisation and the stability of the gut microbiome. Gut Microbiome. 2022;3.
334 Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 1969;37(3).
335 Friston KJ, Bastos AM, Oswal A, van Wijk B, Richter C, Litvak V. Granger causality revisited. Neuroimage. 2014;101:796-808.
336 Tsonis AA, Deyle ER, Ye H, Sugihara G. Convergent cross mapping: theory and an example. In: Tsonis AA, editor. Advances in Nonlinear Geosciences: Springer; 2018.
337 Ye H, Deyle ER, Gilarranz LJ, Sugihara G. Distinguishing time-delayed causal interactions using convergent cross mapping. Sci Rep. 2015;5:14750.
338 Dhamala M, Rangarajan G, Ding M. Estimating Granger causality from fourier and wavelet transforms of time series data. Phys Rev Lett. 2008;100(1):018701.
339 Olayeni OR. Causality in continuous wavelet transform without spectral matrix factorization: theory and application. Computational Economics. 2015;47(3):321-340.
340 Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Nino on the synchronous dengue epidemics in Thailand. PLOS Med. 2005;2(4):e106.
341 Pascual M, Cazelles B, Bouma MJ, Chaves LF, Koelle K. Shifting patterns: malaria dynamics and rainfall variability in an African highland. Proc Biol Sci. 2008;275(1631):123-132.
342 Constantin de Magny G, Guegan JF, Petit M, Cazelles B. Regional-scale climate-variability synchrony of cholera epidemics in West Africa. BMC Infect Dis. 2007;7:20.
343 Blauw AN, Benincà E, Laane RW, Greenwood N, Huisman J. Dancing with the tides: fluctuations of coastal phytoplankton orchestrated by different oscillatory modes of the tidal cycle. PLOS One. 2012;7(11):e49319.
344 Winder M, Schindler DE, Essington TE, Litt AH. Disrupted seasonal clockwork in the population dynamics of a freshwater copepod by climate warming. Limnology and Oceanography. 2009;54(6part2):2493-2505.
345 Ménard F, Marsac F, Bellier E, Cazelles B. Climatic oscillations and tuna catch rates in the Indian Ocean: a wavelet approach to time series analysis. Fisheries Oceanography. 2006;16(1):95-104.
346 Murdoch WW, Kendall BE, Nisbet RM, Briggs CJ, McCauley E, Bolser R. Single-species models for many-species food webs. Nature. 2002;417(6888):541-543.
347 Mondal D, Percival DB. Wavelet variance analysis for gappy time series. Annals of the Institute of Statistical Mathematics. 2008;62(5):943-966.
348 Foster G. Wavelets for period analysis of unevenly sampled time series. Astron J. 1996;112: 1709-1729.
349 Thiebaut C, Roques S. Time-scale and time-frequency analyses of irregularly sampled astronomical time series. EURASIP Journal on Advances in Signal Processing. 2005;2005(15).
350 Azuara J, Sabatier P, Lebreton V, Jalali B, Sicre M-A, Dezileau L, et al. Mid- to late-holocene mediterranean climate variability: contribution of multi-proxy and multi-sequence comparison using wavelet spectral analysis in the northwestern Mediterranean basin. Earth-Science Reviews. 2020;208.
351 Dong F-T, Gai N, Tang Y, Wang Y-F, Yi T-F. Evidence of quasi-periodic oscillation in the optical band of the blazar 1ES 1959+650. Research in Astronomy and Astrophysics. 2022;22(11).
352 Mouchet M, Guilhaumon F, Villéger S, Mason NWH, Tomasini JA, Mouillot D. Towards a consensus for calculating dendrogram-based functional diversity indices. Oikos. 2008;117(5):794-800.
353 Seyedian SS, Nokhostin F, Malamir MD. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. J Med Life. 2019;12(2):113-122.
354 Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474(7351):307-317.
355 Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2:17004.
356 Sha S, Xu B, Wang X, Zhang Y, Wang H, Kong X, et al. The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. Diagn Microbiol Infect Dis. 2013;75(3):245-251.
357 Andoh A, Kuzuoka H, Tsujikawa T, Nakamura S, Hirai F, Suzuki Y, et al. Multicenter analysis of fecal microbiota profiles in Japanese patients with Crohn's disease. J Gastroenterol. 2012;47(12):1298-1307.
358 Dey N, Soergel DA, Repo S, Brenner SE. Association of gut microbiota with post-operative clinical course in Crohn's disease. BMC Gastroenterol. 2013;13:131.
359 Luo A, Leach ST, Barres R, Hesson LB, Grimm MC, Simar D. The microbiota and epigenetic regulation of T helper 17/regulatory T cells: in search of a balanced immune system. Front Immunol. 2017;8:417.
360 Vester-Andersen MK, Prosberg MV, Jess T, Andersson M, Bengtsson BG, Blixt T, et al. Disease course and surgery rates in inflammatory bowel disease: a population-based, 7-year follow-up study in the era of immunomodulating therapy. Am J Gastroenterol. 2014;109(5):705-714.
361 Cade B, Noon B. A gentle introduction to quantile regression for ecologists. Front Ecol Environ. 2003.
362 Koenker R. Quantile regression. in: Econometric Society Monographs: Cambridge: Cambridge University Press; 2005.
363 Knight CA, Ackerly DD. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. Ecology Letters. 2002.
364 Mujagic Z, Ludidi S, Keszthelyi D, Hesselink M, Kruimel J, Lenaerts K, et al. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in

gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. Aliment Pharmacol Ther. 2014;40:288-297.
365 Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep. 2016;6:33721.
366 Geraci M, Bottai M. Linear quantile mixed models. Statistics and Computing 2014;24:461–479.
367 Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. J Physiol. 2009;587(Pt 17):4153-4158.
368 Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. Statistics and Computing. 2022.
369 Carpenter SR, Cole JJ, Pace ML, Batt R, Brock WA, Cline T, et al. Early warnings of regime shifts: a whole-ecosystem experiment. Science. 2011;332(6033):1079-1082.
370 Cade BS, Noon BR, Flather CH. Quantile regression reveals hidden bias and uncertainty in habitat models. Ecology. 2005;86(3):786–800.
371 Magro DO, Santos A, Guadagnini D, de Godoy FM, Silva SHM, Lemos WJF, et al. Remission in Crohn's disease is accompanied by alterations in the gut microbiota and mucins production. Sci Rep. 2019;9(1):13263.
372 Pisani A, Rausch P, Bang C, Ellul S, Tabone T, Marantidis Cordina C, et al. Dysbiosis in the gut microbiota in patients with inflammatory bowel disease during remission. Microbiol Spectr. 2022;10(3):e0061622.
373 Jackson MA, Goodrich JK, Maxan ME, Freedberg DE, Abrams JA, Poole AC, et al. Proton pump inhibitors alter the composition of the gut microbiota. Gut. 2016;65(5):749-756.
374 Molodecky NA, Kaplan GG. Environmental risk factors for inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2010;6(5):339-346.
375 Alam MT, Amos GCA, Murphy ARJ, Murch S, Wellington EMH, Arasaradnam RP. Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. Gut Pathog. 2020;12(1).
376 Zhang SL, Wang SN, Miao CY. Influence of microbiota on intestinal immune system in ulcerative colitis and its intervention. Front Immunol. 2017;8:1674.
377 Kumari R, Ahuja V, Paul J. Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. World J Gastroenterol. 2013;19(22):3404-3414.
378 Nishihara Y, Ogino H, Tanaka M, Ihara E, Fukaura K, Nishioka K, et al. Mucosa-associated gut microbiota reflects clinical course of ulcerative colitis. Sci Rep. 2021;11(1):13743.
379 Iljazovic A, Roy U, Galvez EJC, Lesker TR, Zhao B, Gronow A, et al. Perturbation of the gut microbiome by *Prevotella* spp. enhances host susceptibility to mucosal inflammation. Mucosal Immunol. 2021;14(1):113-124.
380 Maconi G, Camatta D, Cannatelli R, Ferretti F, Carvalhas Gabrielli A, Ardizzone S. Budesonide MMX in the treatment of ulcerative colitis: current perspectives on efficacy and safety. Ther Clin Risk Manag. 2021;17:285-292.
381 Debast SB, Bauer MP, Kuijper EJ, European Society of Clinical Microbiology and Infectious Diseases. European society of clinical microbiology and infectious diseases: update of the treatment guidance document for *Clostridium difficile* infection. Clin Microbiol Infect. 2014;20 Suppl 2:1-26.
382 Sokol H, Landman C, Seksik P, Berard L, Montil M, Nion-Larmurier I, et al. Fecal microbiota transplantation to maintain remission in Crohn's disease: a pilot randomized controlled study. Microbiome. 2020;8(1):12.
383 Rossen NG, Fuentes S, van der Spek MJ, Tijssen JG, Hartman JH, Duflou A, et al. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. Gastroenterology. 2015;149(1):110-118 e114.
384 van Lingen E, Nooij S, Terveer E, Crossette E, Prince A, Bhattarai S, et al. Fecal microbiota transplantation engraftment after budesonide or placebo in patients with active ulcerative colitis using pre-selected donors: a randomized pilot study. J Crohns Colitis. 2024;jjae043:1381-1393.
385 Terveer EM, Vendrik KE, Ooijevaar RE, Lingen EV, Boeije-Koppenol E, Nood EV, et al. Faecal microbiota transplantation for *Clostridioides difficile* infection: Four years' experience of the Netherlands Donor Feces Bank. United European Gastroenterol J. 2020;8(10):1236-1247.
386 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-359.
387 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i890.
388 Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.
389 Ruscheweyh HJ, Milanese A, Paoli L, Karcher N, Clayssen Q, Keller MI, et al. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. Microbiome. 2022;10(1):212.
390 Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Logratio analysis and compositional distance. Mathematical Geology. 2000;32(3):271-275.
391 Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. Journal of Open Source Software. 2021;6(63).
392 Morgan M. Dirichlet-multinomial mixture model machine learning for microbiome data. 2022.
393 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLOS One. 2012;7(2):e30126.
394 Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal. 2017;9(2): 378-400.

395 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 2015;67(1).
396 Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society. 1943;54(3): 426-482.
397 Lahti L, Shetty S. microbiome R package. 2012-2019.
398 Pinheiro J, Bates D, R Core Team. nlme: linear and nonlinear mixed effects models. 2022.
399 Wilson BC, Vatanen T, Cutfield WS, O'Sullivan JM. The super-donor phenomenon in fecal microbiota transplantation. Front Cell Infect Microbiol. 2019;9:2.
400 Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. Immunology. 2017;151(4):363-374.
401 Precup G, Vodnar DC. Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. Br J Nutr. 2019;122(2):131-140.
402 Lahti L, Salojärvi J, Salonen A, Scheffer M, de Vos WM. Tipping elements in the human intestinal ecosystem. Nature Communications 2014;4344.
403 Pinto S, Benincà E, Galazzo G, Jonkers D, Penders J, Bogaards JA. Heterogeneous associations of gut microbiota with Crohn's disease activity. Gut Microbes. 2024;16(1):2292239.
404 Caenepeel C, Deleu S, Vazquez Castellanos JF, Arnauts K, Braekeleire S, Machiels K, et al. Rigorous donor selection for fecal microbiota transplantation in active ulcerative colitis: key lessons from a randomized controlled trial halted for futility. Clin Gastroenterol Hepatol. 2024.
405 Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-180.
406 Vieira-Silva S, Sabino J, Valles-Colomer M, Falony G, Kathagen G, Caenepeel C, et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat Microbiol. 2019;4(11):1826-1831.
407 Tian Y, Zhou Y, Huang S, Li J, Zhao K, Li X, et al. Fecal microbiota transplantation for ulcerative colitis: a prospective clinical study. BMC Gastroenterol. 2019;19(1):116.
408 Crothers JW, Chu ND, Nguyen LTT, Phillips M, Collins C, Fortner K, et al. Daily, oral FMT for long-term maintenance therapy in ulcerative colitis: results of a single-center, prospective, randomized pilot study. BMC Gastroenterol. 2021;21(1):281.
409 Christensen L, Roager HM, Astrup A, Hjorth MF. Microbial enterotypes in personalized nutrition and obesity management. Am J Clin Nutr. 2018;108(4):645-651.
410 Lee SH, Yoon SH, Jung Y, Kim N, Min U, Chun J, et al. Emotional well-being and gut microbiome profiles by enterotype. Sci Rep. 2020;10(1):20736.
411 Roswall J, Olsson LM, Kovatcheva-Datchary P, Nilsson S, Tremaroli V, Simon MC, et al. Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. Cell Host Microbe. 2021;29(5):765-776 e763.
412 Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol. 2018;3(1):8-16.
413 Hanssen NMJ, de Vos WM, Nieuwdorp M. Fecal microbiota transplantation in human metabolic diseases: from a murky past to a bright future? Cell Metab. 2021;33(6):1098-1110.
414 Pinto S, Šajbenová D, Benincà E, Nooij S, Terveer EM, Keller JJ, et al. Dynamics of gut microbiota after fecal microbiota transplantation in ulcerative colitis: success linked to control of Prevotellaceae. J Crohns Colitis. 2024.
415 Olesen SW, Gerardin Y. Re-evaluating the evidence for faecal microbiota transplantation 'super-donors' in inflammatory bowel disease. J Crohns Colitis. 2021;15(3):453-461.
416 Danne C, Rolhion N, Sokol H. Recipient factors in faecal microbiota transplantation: one stool does not fit all. Nat Rev Gastroenterol Hepatol. 2021;18(7):503-513.
417 Peri R, Aguilar RC, Tuffers K, Erhardt A, Link A, Ehlermann P, et al. The impact of technical and clinical factors on fecal microbiota transfer outcomes for the treatment of recurrent *Clostridioides difficile* infections in Germany. United European Gastroenterol J. 2019;7(5): 716-722.
418 Xiao Y, Angulo MT, Lao S, Weiss ST, Liu YY. An ecological framework to understand the efficacy of fecal microbiota transplantation. Nat Commun. 2020;11(1):3329.
419 Schmidt TSB, Raes J, Bork P. The human gut microbiome: from association to modulation. Cell. 2018;172(6):1198-1215.
420 Grinspan AM, Kelly CR. Fecal microbiota transplantation for ulcerative colitis: not just yet. Gastroenterology. 2015;149(1):15-18.
421 Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of similar species. Proc Natl Acad Sci U S A. 2006;103(16):6230-6235.
422 Fuentes S, Rossen NG, van der Spek MJ, Hartman JH, Huuskonen L, Korpela K, et al. Microbial shifts and signatures of long-term remission in ulcerative colitis after faecal microbiota transplantation. ISME J. 2017;11(8):1877-1889.
423 Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a microbiome study. Cell. 2014;158(2):250-262.
424 Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018;24(4): 392-400.
425 Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol. 2014;32(8):822-828.
426 Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, et al. Universality of human microbial dynamics. Nature. 2016;534(7606): 259-262.

427 Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature. 2014;509(7500):357-360.
428 Tipton L, Muller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, et al. Fungi stabilize connectivity in the lung and skin microbial ecosystems. Microbiome. 2018;6(1):12.
429 Feng C, Jia H, Wang H, Wang J, Lin M, Hu X, et al. MicroNet-MIMRF: a microbial network inference approach based on mutual information and Markov random fields. Bioinform Adv. 2024;4(1):vbae167.
430 Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol Syst Biol. 2018;14(6):e8157.
431 de Vos MGJ, Zagorski M, McNally A, Bollenbach T. Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. Proc Natl Acad Sci USA. 2017;114(40):10666-10671.
432 Wilson WG, Lundberg P, Vázquez DP, Shurin JB, Smith MD, Langford W, et al. Biodiversity and species interactions: extending Lotka–Volterra community theory. Ecology Letters. 2003;6(10):944-952.
433 Haegeman B, Loreau M. A mathematical synthesis of niche and neutral theories in community ecology. J Theor Biol. 2011;269(1):150-165.
434 Dam P, Fonseca LL, Konstantinidis KT, Voit EO. Dynamic models of the complex microbial metapopulation of lake mendota. NPJ Syst Biol Appl. 2016;2:16007.
435 Jang SS, Oishi KT, Egbert RG, Klavins E. Specification and simulation of synthetic multicelled behaviors. ACS Synth Biol. 2012;1(8):365-374.
436 Scheffer M. *Personal Communication*. 2019.
437 Lyu R, Qu Y, Divaris K, Wu D. Methodological considerations in longitudinal analyses of microbiome data: a comprehensive review. Genes (Basel). 2023;15(1).
438 Picot A, Shibasaki S, Meacock OJ, Mitri S. Microbial interactions in theory and practice: when are measurements compatible with models? Curr Opin Microbiol. 2023;75:102354.
439 Pacheco AR, Pauvert C, Kishore D, Segre D. Toward FAIR representations of microbial interactions. mSystems. 2022;7(5):e0065922.
440 Dal Co A, van Vliet S, Kiviet DJ, Schlegel S, Ackermann M. Short-range interactions govern the dynamics and functions of microbial communities. Nature Ecology & Evolution. 2020;4(3):366-375.
441 Kim HJ, Boedicker JQ, Choi JW, Ismagilov RF. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proc Natl Acad Sci U S A. 2008;105(47):18188-18193.
442 Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun. 2017;8(1):1784.
443 Simren M, Barbara G, Flint HJ, Spiegel BM, Spiller RC, Vanner S, et al. Intestinal microbiota in functional bowel disorders: a Rome foundation report. Gut. 2013;62(1):159-176.
444 Zhang X, Pei YF, Zhang L, Guo B, Pendegraft AH, Zhuang W, et al. Negative binomial mixed models for analyzing longitudinal microbiome data. Front Microbiol. 2018;9:1683.
445 Ramaboli MC, Ocvirk S, Khan Mirzaei M, Eberhart BL, Valdivia-Garcia M, Metwaly A, et al. Diet changes due to urbanization in South Africa are linked to microbiome and metabolome signatures of Westernization and colorectal cancer. Nat Commun. 2024;15(1):3379.
446 Gellman RH, Olm MR, Terrapon N, Enam F, Higginbottom SK, Sonnenburg JL, et al. Hadza *Prevotella* require diet-derived microbiota-accessible carbohydrates to persist in mice. Cell Rep. 2023;42(11).
447 Levin J, Passmore J-A, Williams B, editors. Role of vaginal microbiota in genital inflammation and enhancing HIV transmission. International AIDS Conference; 2016; Durban, South Africa. www.natap.org/2016/IAC/IAC_20.htm.
448 Lebeer S, Ahannach S, Gehrmann T, Wittouck S, Eilers T, Oerlemans E, et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. Nat Microbiol. 2023;8(11):2183-2195.
449 Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio. 2014;5(2):e00889.
450 O'Keefe SJ, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, et al. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun. 2015;6:6342.
451 Marcelino VR, Welsh C, Diener C, Gulliver EL, Rutten EL, Young RB, et al. Disease-specific loss of microbial cross-feeding interactions in the human gut. Nat Commun. 2023;14(1):6546.
452 Kullberg RFJ, Wikki I, Haak BW, Kauko A, Galenkamp H, Peters-Sengers H, et al. Association between butyrate-producing gut bacteria and the risk of infectious disease hospitalisation: results from two observational, population-based microbiome studies. Lancet Microbe. 2024;5(9):100864.
453 Duizer C, de Zoete MR. The role of microbiota-derived metabolites in colorectal cancer. Int J Mol Sci. 2023;24(9).
454 Wortelboer K, Nieuwdorp M, Herrema H. Fecal microbiota transplantation beyond *Clostridioides difficile* infections. EBioMedicine. 2019;44:716-729.
455 Ducarmon QR, Kuijper EJ, Olle B. Opportunities and challenges in development of live biotherapeutic products to fight infections. J Infect Dis. 2021;223(12 Suppl 2):S283-S289.
456 Pamer EG. Fecal microbiota transplantation: effectiveness, complexities, and lingering concerns. Mucosal Immunol. 2014;7(2):210-214.
457 Kaiser JC, Verschoor CP, Surette MG, Bowdish DM. Host cytokine responses distinguish invasive from airway isolates of the *Streptococcus milleri/anginosis* group. BMC Infect Dis. 2014;14:498.
458 Besselink MG, van Santvoort HC, Buskens E, Boermeester MA, van Goor H, Timmerman HM, et al. Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised,

double-blind, placebo-controlled trial. Lancet. 2008;371(9613):651-659.
459 de Souza HSP, Fiocchi C, Iliopoulos D. The IBD interactome: an integrated view of aetiology, pathogenesis and therapy. Nat Rev Gastroenterol Hepatol. 2017;14(12):739-749.
460 Paramsothy S, Paramsothy R, Rubin DT, Kamm MA, Kaakoush NO, Mitchell HM, et al. Faecal microbiota transplantation for inflammatory bowel disease: a systematic review and meta-analysis. J Crohns Colitis. 2017;11(10):1180-1199.
461 Xu D, Chen VL, Steiner CA, Berinstein JA, Eswaran S, Waljee AK, et al. Efficacy of fecal microbiota transplantation in irritable bowel syndrome: a systematic review and meta-analysis. Am J Gastroenterol. 2019;114(7):1043-1050.
462 Taur Y, Coyte K, Schluter J, Robilotti E, Figueroa C, Gjonbalaj M, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med. 2018;10(460).
463 Schluter J, Peled JU, Taylor BP, Markey KA, Smith M, Taur Y, et al. The gut microbiota is associated with immune cell dynamics in humans. Nature. 2020;588(7837):303-307.
464 Yelin I, Flett KB, Merakou C, Mehrotra P, Stam J, Snesrud E, et al. Genomic and epidemiological evidence of bacterial transmission from probiotic capsule to blood in ICU patients. Nat Med. 2019;25(11):1728-1732.
465 Reijnders D, Goossens GH, Hermes GD, Neis EP, van der Beek CM, Most J, et al. Effects of gut microbiota manipulation by antibiotics on host metabolism in obese humans: a randomized double-blind placebo-controlled trial. Cell Metab. 2016;24(1):63-74.
466 Liu X, Salles JF. Lose-lose consequences of bacterial community-driven invasions in soil. Microbiome. 2024;12(1):57.
467 Liu X, Salles JF. Bridging ecological assembly process and community stability upon bacterial invasions. ISME J. 2024;18(1).
468 Mallon CA, Le Roux X, van Doorn GS, Dini-Andreote F, Poly F, Salles JF. The impact of failure: unsuccessful bacterial invasions steer the soil microbial community away from the invader's niche. ISME J. 2018;12(3):728-741.
469 Stecher B, Chaffron S, Kappeli R, Hapfelmeier S, Freedrich S, Weber TC, et al. Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. PLOS Pathog. 2010;6(1):e1000711.
470 Rodriguez-Verdugo A, Ackermann M. Rapid evolution destabilizes species interactions in a fluctuating environment. ISME J. 2021;15(2):450-460.
471 Shepherd ES, DeLoache WC, Pruss KM, Whitaker WR, Sonnenburg JL. An exclusive metabolic niche enables strain engraftment in the gut microbiota. Nature. 2018;557(7705):434-438.
472 Kearney SM, Gibbons SM, Erdman SE, Alm EJ. Orthogonal dietary niche enables reversible engraftment of a gut bacterial commensal. Cell Rep. 2018;24(7):1842-1851.
473 Makhalanyane TP, Bezuidt OKI, Pierneef RE, Mizrachi E, Zeze A, Fossou RK, et al. African microbiomes matter. Nat Rev Microbiol. 2023;21(8):479-481.
474 Vargas-Robles D, Morales N, Rodriguez I, Nieves T, Godoy-Vitorino F, Alcaraz LD, et al. Changes in the vaginal microbiota across a gradient of urbanization. Sci Rep. 2020;10(1):12487.
475 Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. mSystems. 2017;2(1).
476 Rossetto Marcelino V. Gut health: The value of connections. Elife. 2023;12.
477 Viertel TM, Ritter K, Horz HP. Viruses versus bacteria-novel approaches to phage therapy as a tool against multidrug-resistant pathogens. J Antimicrob Chemother. 2014;69(9):2326-2336.
478 Maslov S, Sneppen K. Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. Sci Rep. 2017;7:39642.
479 Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnology and Oceanography. 2000;45(6):1320-1328.
480 Tautz D, Domazet-Loso T. The evolutionary origin of orphan genes. Nat Rev Genet. 2011;12(10):692-702.
481 Malik AA, Martiny JBH, Brodie EL, Martiny AC, Treseder KK, Allison SD. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. ISME J. 2020;14(1):1-9.
482 Ballal SA, Gallini CA, Segata N, Huttenhower C, Garrett WS. Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts. Cell Microbiol. 2011;13(4):508-517.
483 Perez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, et al. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. Gut. 2013;62(11):1591-1601.
484 McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med. 2011;3(106):106ra106.
485 Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013;152(1-2):39-50.
486 Goers L, Freemont P, Polizzi KM. Co-culture systems and technologies: taking synthetic biology to the next level. J R Soc Interface. 2014;11(96).
487 Pearce SC, Coia HG, Karl JP, Pantoja-Feliciano IG, Zachos NC, Racicot K. Intestinal *in vitro* and *ex vivo* models to study host-microbiome interactions and acute stressors. Front Physiol. 2018;9:1584.
488 Shah P, Fritz JV, Glaab E, Desai MS, Greenhalgh K, Frachet A, et al. A microfluidics-based *in vitro* model of the gastrointestinal human-microbe interface. Nat Commun. 2016;7:11535.

489 Puschhof J, Pleguezuelos-Manzano C, Clevers H. Organoids and organs-on-chips: Insights into human gut-microbe interactions. Cell Host Microbe. 2021;29(6):867-878.
490 Min S, Kim S, Cho SW. Gastrointestinal tract modeling using organoids engineered with cellular and microbiota niches. Exp Mol Med. 2020;52(2):227-237.
491 Mirzayi C, Renson A, Genomic Standards Consortium, Massive Analysis and Quality Control Society, Zohra F, Elsafoury S, et al. Reporting guidelines for human microbiome research: the STORMS checklist. Nat Med. 2021;27(11): 1885-1892.
492 Chistoserdova L. Recent progress and new challenges in metagenomics for biotechnology. Biotechnol Lett. 2010;32(10):1351-1359.
493 Liu YX, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell. 2021;12(5):315-330.
494 Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
495 Tedersoo L, Kungas R, Oras E, Koster K, Eenmaa H, Leijen A, et al. Data sharing practices and data availability upon request differ across scientific disciplines. Sci Data. 2021;8(1):192.
496 Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol. 2015;16:276.
497 Wen H, Wang HY, He X, Wu CI. On the low reproducibility of cancer studies. Natl Sci Rev. 2018;5(5):619-624.
498 Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020;587(7834):448-454.
499 Langille MGI, Ravel J, Fricke WF. "Available upon request": not good enough for microbiome data! Microbiome. 2018;6(1):8.
500 Christian TM, Gooch A, Vision T, Hull E. Journal data policies: exploring how the understanding of editors and authors corresponds to the policies themselves. PLOS One. 2020;15(3):e0230281.

Appendices

Summaries (EN/NL)

English summary

The gut microbiota

The human body hosts countless microorganisms, with a significant portion residing in the digestive system. The bacteria and other microorganisms in our gut, such as fungi and viruses, are collectively referred to as the gut microbiota. A rich and diverse gut microbiota can contribute to good health in its host, for example, by suppressing harmful bacteria. Additionally, these bacteria assist with tasks such as nutrient digestion and training the immune system. The microorganisms constantly influence each other and their human host. Microbes adapt to the specific conditions of their host, with factors such as diet, lifestyle, hormonal regulation, and the immune system playing a role (Figure 1). As a result, the composition of the microbiota frequently changes, for instance, after foreign travel or during antibiotic treatment. Despite these adjustments, it is hoped that the microbiota does not lose its functions, ensuring, among other things, that the integrity of the intestinal wall remains intact.

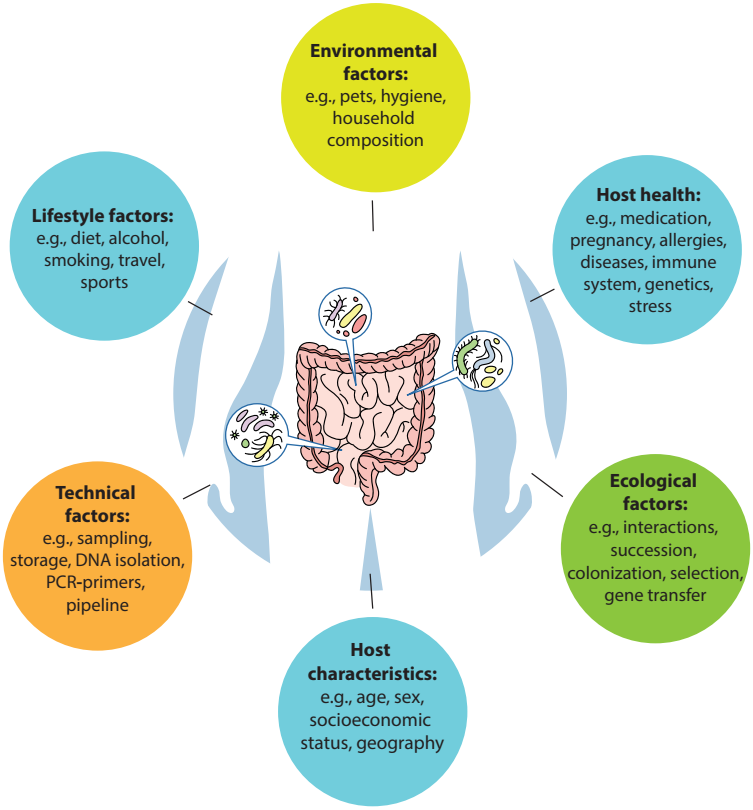


Figure 1. Factors influencing the composition of the gut microbiota.
While technical factors do not directly affect the composition of the gut microbiota itself, they do influence the types and the abundance of microorganisms that can be detected in fecal samples.

The significant role of the gut microbiota in our health has led to extensive research into this ecosystem. Fecal samples are often collected to study the composition and diversity of gut bacteria using DNA analysis techniques such as 16S rRNA sequencing. However, technical factors, such as sample quality, the DNA extraction method used, and the choice of primers, can influence which species are detected and how they are represented in the results (Figure 1). Additionally, incomplete databases, the vast microbial diversity, the dynamic nature of the microbiota, and the limited knowledge about many species make these analyses particularly challenging. The resulting data are complex and require advanced methods for accurate interpretation.

The dynamics and stability of microbial communities, in relation to health and disease, can be studied using techniques such as network analyses and time series models. Network analyses map which bacteria frequently coexist within an ecosystem, providing clues about how microbes might influence and interact with each other. Time series models help track changes in microbial communities over time and uncover patterns. Studying the microbiota is challenging because it cannot be directly observed in the body, and often only a limited number of samples are available. A small number of samples, both in terms of participants and measurement points, complicates the identification of robust associations and makes it difficult to distinguish between individual variation and general patterns.

In the project 'Ecology meets human health', we first examined the reliability of network analyses and alternative methods for mapping relationships between microbes. Next, we combined clinical, microbiological, and ecological concepts to better understand how microbial dynamics are linked to intestinal diseases, specifically Crohn's disease and ulcerative colitis, as well as the success of fecal microbiota transplantation (FMT) as a treatment. In this context, we investigated the ecological factors that influence the gut microbiota and the functioning of this complex ecosystem.

Ecological structure in the gut microbiota

The gut ecosystem consists of numerous species whose presence depends on variations in the environment and functional needs, such as breaking down food, producing vitamins, or combating pathogens. Additionally, interactions often occur between bacteria, which can have positive (beneficial) or negative (detrimental) effects on the species involved. Understanding these interactions is crucial for grasping ecological processes and changes within the microbiota. Correlation methods are often used to map these networks.

In Chapter 2, we investigated the reliability of correlation methods for inferring interaction networks. For our research, we used the generalized Lotka-Volterra (gLV) model to simulate bacterial communities. This model provides insights into microbial dynamics without requiring actual gut microbiota samples and allows for the adjustment of parameters such as bacterial growth rates. Moreover, it enabled us to use a known interaction network as a reference, which is not possible with real samples. We examined the effects of interindividual variation (differences in microbiota composition between individuals) and sample size (the amount of available data) on the accuracy of network reconstructions. While correlations in microbial abundances often indicate ecological interactions, we demonstrated that measurement noise, such as variations in sample processing, complicates the detection of true interactions.

Furthermore, correlations do not differentiate specific interaction types, making laboratory verification necessary to understand these relationships. The gLV model offers valuable insights but also underscores the indispensable role of studies using real data. However, many human microbiota studies are snapshots in time, meaning apparent correlations can be driven by external factors, such as fluctuations in nutrients, without indicating actual interactions between species. Longitudinal studies, which track microbial communities over time, provide a much better basis for understanding consistency and patterns. Such datasets are scarce, likely due to practical challenges, such as repeated sampling, which depend heavily on the host's consent and willingness to participate in research.

In Chapter 3, we demonstrated how the microbiota changes over time and what information these variations reveal about relationships between species. To do this, we analysed time series data from two individuals collected by researcher Caporaso et al. in 2011. These two individuals submitted stool samples almost daily for a year, providing a clear view of the variation within the microbiota. Using wavelet clustering, we uncovered patterns in these data. Wavelet clustering has already been established in ecological and epidemiological studies, and it has also proven particularly suitable for non-stationary microbiota time series, providing greater insight into the collective temporal behaviour of bacteria compared to conventional correlation methods. With wavelet spectra, we constructed 'trees' that depicted relationships between bacterial species. These trees showed significant differences from those based on correlation methods, such as a greater total branch length (indicating higher functional diversity) and distinct subgroups. This highlights that wavelet clustering is more sensitive to subtle differences in community structures than correlation-based methods. Our findings underscore the importance of the method chosen by researchers for analysing microbiota data.

Gut microbiota and inflammatory bowel disease

The interaction between humans and the microbiota is the result of over a billion years of co-evolution, leading to a symbiotic relationship. Our microbes are involved in numerous essential functions, and disturbances in their species composition, known as dysbiosis, have been linked to various diseases. Dysbiosis can weaken the mucus layer in the gut wall, a protective layer that covers the inside of the intestines and helps keep harmful substances and microorganisms out of the body. This can lead to colonization by harmful organisms, an increased risk of inflammation, and metabolic disruptions, putting the host's health at risk (Figure 2).

Inflammatory bowel diseases (IBD), including Crohn's disease and ulcerative colitis, are chronic inflammations of the intestinal mucosa. These conditions are associated with an altered composition and diversity of the gut microbiota. In Chapters 4 to 6, we explore the relationships between bacterial dysbiosis and the disease progression in patients with IBD. Although much research has been conducted in this area, findings across studies have often been inconsistent. The differences in findings regarding the involvement of microbes in IBD can likely be attributed to technical variations in research methods and diversity among patients, including variations in disease assessment, medication use, and lifestyle factors. Additionally, the variability of the microbiota in gastrointestinal disorders, such as natural fluctuations in composition, makes it challenging to identify consistent biological signals.

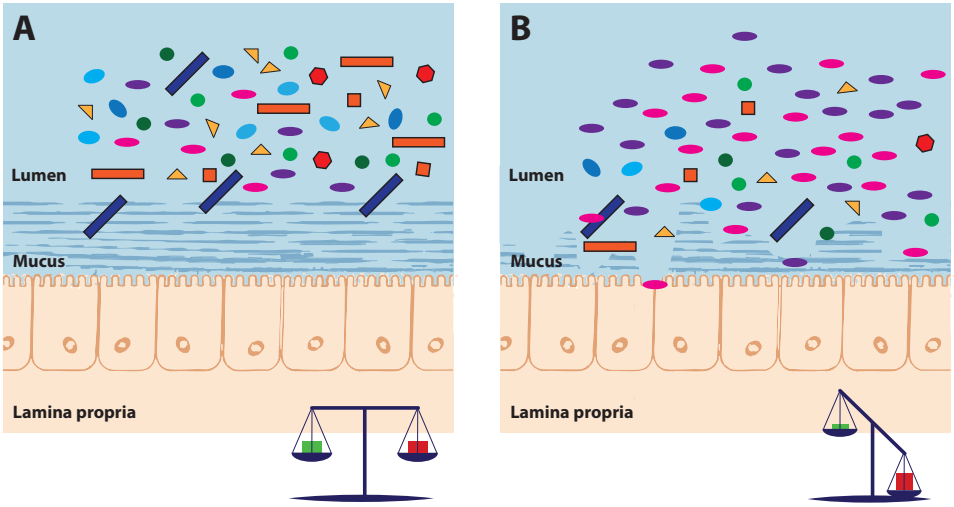


Figure 2. Illustration of the difference between a gut with a healthy microbiota and a disrupted microbiota. A) A healthy gut microbiota is diverse and balanced. The mucus layer provides nourishment for the bacteria while also protecting the gut cells. B) When the balance is disrupted, some bacteria can overgrow and displace other bacteria. Then, the mucus layer can be damaged, allowing bacteria to harm the gut wall.

In Chapter 4, we analysed stool samples from patients with Crohn's disease to describe associations between bacterial abundance and disease remission or exacerbation. We employed a quantile regression model to uncover relationships that go beyond the average response of all patients. Quantile regression allows for a more comprehensive view of the relationships between bacterial abundance and disease. Associations with specific bacterial families may only be observable in a minority of patients. While generic associations can also be identified using methods that focus on the average response, it is also essential to understand less common differences in the microbiota, as these may provide insight into personalized treatment approaches. We also correlated the relative abundance of bacterial families with known biomarkers of disease activity, such as fecal calprotectin and serum C-reactive protein. Our findings revealed significant negative associations between various bacterial families and disease, such as Pasteurellaceae and Ruminococcaceae. When comparing regressions with clinically defined exacerbation, we found that associations with fecal calprotectin were stronger than with other indicators. In summary, our research highlights the heterogeneity of Crohn's disease and its relationship with the gut microbiota.

Ecological determinants of FMT treatment success

Microbiota-related therapies aim to intentionally alter the microbiota of patients to shift it from a dysbiotic to a healthy state. Fecal microbiota transplantation (FMT), commonly known as a stool transplant, is an experimental treatment in which fecal material, including the microbiota from healthy donors, is transferred to the patient to restore the disrupted microbiota. FMT has established itself as a promising treatment for microbiota-related conditions, particularly for the treatment of recurrent *Clostridioides difficile* infection. In ulcerative colitis, the success of FMT is determined by achieving clinical remission.

One of the challenges of FMT is the inconsistency in outcomes regarding the final composition of the patient's microbiota. This means that each person responds differently to certain donor bacteria and that various immune responses can be activated in patients with different conditions. Hypothetical outcomes include a return to the original dysbiotic state, an alternative dysbiotic state, a long-term change with beneficial properties, or a persistent shift to a healthy microbiota with donor species. In our study, we investigated the microbial families associated with the success of FMT treatment in ulcerative colitis. This allowed us to explore the dynamics of the gut microbiota. This longitudinal approach provided insight into weekly changes (Figure 3), a perspective often missing in randomized controlled trials that mainly focus on clinical outcomes.

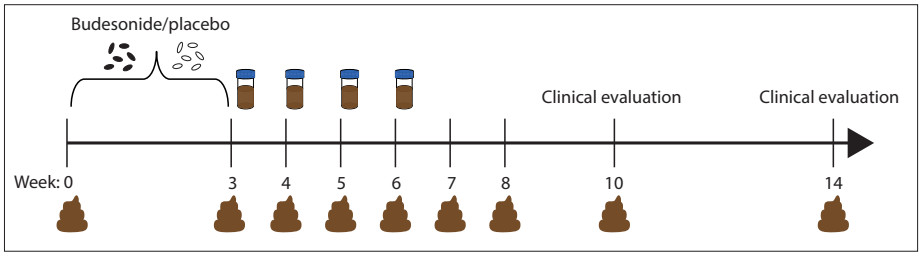


Figure 3. Design of the FECBUD-study (data used in Chapters 5 and 6). Patients were first pre-treated for three weeks with budesonide (n = 12), a medication commonly used to reduce inflammation, or a placebo (n = 12), an inactive substance with no therapeutic effect. After that, the patients received four consecutive fecal transplants from a healthy donor provided by the Netherlands Donor Feces Bank (NDFB). Treatment evaluation took place after 10 and 14 weeks from the start of the study. A subgroup of the patient group (9 out of 24 patients) achieved a successful combined clinical and endoscopic remission after the FMT treatment. Fecal samples from the patients were collected at the beginning of the study, after the pre-treatment, weekly after the fecal transplant, and at two, four, and eight weeks post-FMT.

In Chapter 5, we studied the associations related to the clinical success of FMT in patients and the development of the microbiota during and after treatment. We used a wide range of analytical techniques to investigate potential associations between bacterial families and clinical outcomes, including ordination analysis, Dirichlet multinomial mixture analysis, and longitudinal modeling. The use of these approaches allowed us to identify significant differences in microbial composition and diversity between patients who benefited from the treatment and those who did not. For example, we found that the success of FMT in patients with ulcerative colitis seems to be associated with limited growth of Prevotellaceae and the presence of the families Lachnospiraceae and Ruminococcaceae. Monitoring the dynamics of these microbial families could potentially provide early insight into the success of treatment during FMT.

It is widely believed that the colonization of donor species in the recipient's microbiota is a key mechanism behind the success of FMT. An interesting finding from our research in Chapter 5 is that we found no indication of a shift in the microbial composition of the recipient towards the donor microbiota among patients with clinical success of FMT. In Chapter 6, we therefore examined whether the donor-centered view of FMT holds true by analysing whether microbiota dynamics are related to achieving remission in patients after FMT treatment.

To do this, we categorized the species based on their origin and temporal presence: already present in the host before FMT, derived from the donor, or species newly introduced during the FMT treatment. We then modelled the number of species per category (host-associated, donor-derived, and novel) for patients who did or did not benefit from FMT. Our results show that patients who benefited from the treatment retained a higher number of host-associated species compared to patients who did not benefit from the treatment. Although donor species initially colonized more extensively in patients who did not benefit from the treatment, this colonization decreased over time, aligning with the level seen in patients who did benefit from the treatment. This suggests that a successful clinical response to FMT may be facilitated by a microbiota that is receptive to colonization without compromising the resident microbiota. We also discovered that host species with higher relative abundances before FMT are better able to persist after FMT.

In conclusion

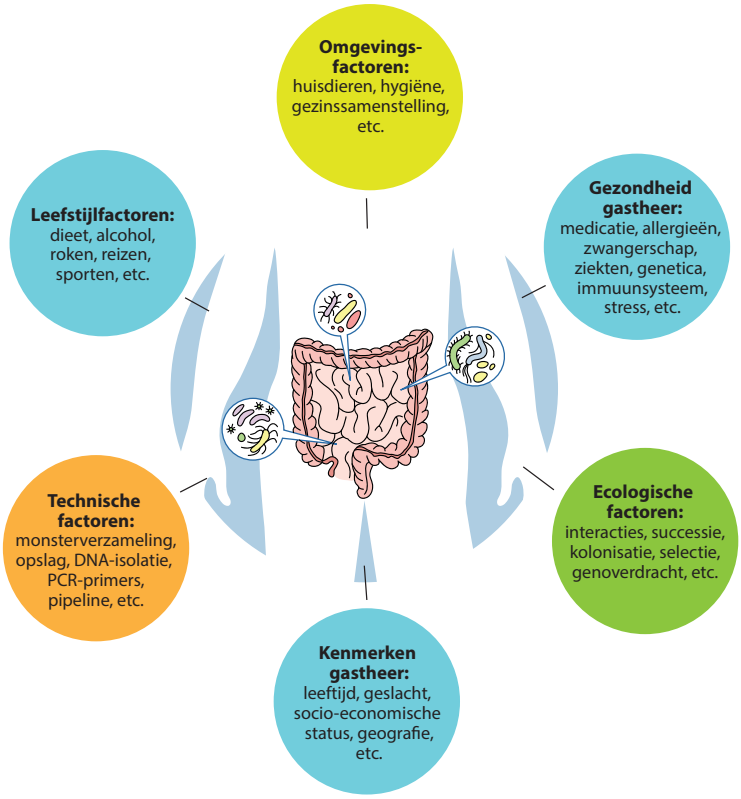
There are numerous environmental factors and habits (among others, diet and lifestyle) that influence the composition and function of microbes. This complexity can sometimes be overwhelming, but knowledge from various disciplines provides valuable insights. By understanding how systems respond to changes and regain balance, we deepen our knowledge of the complex dynamics within the human gut microbiota. These insights can improve microbiota-modulating strategies such as FMT and stimulate innovation in personalized therapeutic approaches. This leads to a new perspective, viewing the microbiota not just as a battleground against pathogenic microbes, but as a complex ecological community. Management strategies such as habitat restoration play an important role in this. To achieve this, extensive (longitudinal) datasets are needed that include a wide range of individuals with diverse backgrounds and lifestyles. Such data are essential for understanding the mechanisms that influence the dynamics of ecosystems in health and disease. In addition, suitable methods must be available to effectively investigate the complex microbiota data.

The limitations of a 'one-size-fits-all' approach are becoming increasingly evident, particularly due to the diversity, genotypes (genetic composition), and phenotypes (observable traits and characteristics) of the gut microbiota. This underscores the need for a more personalized approach in (clinical) scientific research. For instance, matching donor and recipient based on lifestyle and diet could increase the likelihood that transplanted bacteria successfully colonize and that the resident bacteria receive the right nutrients. This could contribute to a more effective improvement of the recipient's microbiota, tailored to the patient's unique needs. However, to truly understand microbial functioning, we need to keep the ecological perspective of bacteria in mind. This means recognizing that bacteria perceive their environment, resources, and interactions on a much smaller spatial scale, constantly adapting to their immediate surroundings. This implies that even within a single sample, different bacteria can exhibit different behaviours, functions, or interactions. Additionally, within the gut microbiota, various bacteria often have overlapping functions. This means that if a certain species is disrupted or removed, other species can take over that function. This mechanism is crucial for maintaining a healthy gut microbiota and remains an area that has not been sufficiently researched. An ecological approach, grounded in theory, is essential for interpreting the impact of the microbiota on health or disease. This perspective allows us to better understand the complex interactions within the microbiota, which is vital for developing effective therapeutic strategies.

Nederlandse samenvatting

De darmmicrobiota

Het menselijk lichaam herbergt als gastheer talloze micro-organismen, een groot deel daarvan bevindt zich in het spijsverteringsstelsel. De bacteriën en andere micro-organismen in onze darmen, zoals schimmels en virussen, worden samen de darmmicrobiota genoemd. Een rijke en diverse darmmicrobiota is in staat om bij te dragen aan een goede gezondheid van de gastheer, bijvoorbeeld door schadelijke bacteriën te onderdrukken. Daarnaast helpen de bacteriën ook bij de vertering van voedingsstoffen en het instrueren van het immuunsysteem. De micro-organismen beïnvloeden continu elkaar en hun menselijke gastheer. Microben passen zich aan de specifieke omstandigheden van hun gastheer aan, waarbij factoren zoals dieet, leefstijl, hormonale regulatie en het immuunsysteem een rol spelen (Figuur 1). Hierdoor verandert de samenstelling van de microbiota regelmatig, bijvoorbeeld na een buitenlandse reis of tijdens een antibioticabehandeling. Tijdens deze aanpassingen verliest de microbiota hopelijk geen functies, zodat de integriteit van de darmwand behouden blijft.



Figuur 1. Factoren die van invloed zijn op de samenstelling van de darmmicrobiota. Hoewel de technische factoren niet de samenstelling van de darmmicrobiota zelf beïnvloeden, hebben deze wel effect op zowel de soorten als de hoeveelheid micro-organismen die kunnen worden gevonden in de fecale monsters.

Doordat de darmmicrobiota zo belangrijk is voor onze gezondheid, wordt er uitgebreid onderzoek gedaan naar dit ecosysteem. Vaak worden ontlastingsmonsters verzameld om met DNA-analysetechnieken zoals 16S rRNA-sequencing de samenstelling en diversiteit van darmbacteriën te onderzoeken. Technische factoren, zoals de kwaliteit van het monster, de gebruikte DNA-extractiemethode en de keuze van primers, kunnen echter invloed hebben op welke soorten worden gedetecteerd en in welke mate ze worden gerepresenteerd in de resultaten (Figuur 1). Daarnaast maken incomplete databases, de enorme microbiële diversiteit, het dynamische karakter van de microbiota en de beperkte kennis over veel soorten de analyses extra uitdagend. De verkregen gegevens zijn hierdoor complex en vereisen geavanceerde methoden om nauwkeurig geïnterpreteerd te worden.

De dynamiek en stabiliteit van microbiële gemeenschappen, in relatie tot gezondheid en ziekte, is te onderzoeken met behulp van technieken zoals netwerkanalyses en tijdreeksmodellen. Netwerkanalyses brengen in kaart welke bacteriën vaak samen voorkomen binnen een ecosysteem. Dit kan een aanwijzing zijn voor hoe microben elkaar mogelijk beïnvloeden en samenwerken. Tijdreeksmodellen helpen om veranderingen in de microbiële gemeenschappen over de tijd te volgen en patronen te ontdekken. Het onderzoek van de microbiota is ook uitdagend, omdat het niet rechtstreeks in het lichaam kan worden bestudeerd en er vaak slechts een beperkt aantal monsters beschikbaar is. Een klein aantal monsters, zowel qua deelnemers als meetmomenten, bemoeilijkt het vaststellen van robuuste verbanden en maakt het lastig om onderscheid te maken tussen individuele variatie en algemene patronen.

Binnen het project 'Ecology meets human health' onderzochten we eerst de betrouwbaarheid van netwerkanalyses en alternatieve methoden om de relaties tussen microben in kaart te brengen. Vervolgens combineerden we klinische, microbiologische en ecologische concepten om beter te begrijpen hoe microbiële dynamiek samenhangt met darmaandoeningen, specifiek de ziekte van Crohn en colitis ulcerosa, en met het succes van fecale microbiota-transplantatie (FMT) als behandeling. Hierbij onderzochten we de ecologische factoren die de darmmicrobiota beïnvloeden en de werking van dit complexe ecosysteem bepalen.

Ecologische structuren in de darmmicrobiota

Het darmecosysteem omvat tal van soorten waarvan het voorkomen afhankelijk is van variaties in de omgeving en functionele vereisten, zoals de afbraak van voedsel, productie van vitamines of de bestrijding van ziekteverwekkers. Daarnaast zijn er ook vaak interacties tussen bacteriën, die positieve (winst) of negatieve (verlies) effecten hebben op de betrokken soorten. Inzicht in deze onderlinge interacties is belangrijk voor het begrijpen van ecologische processen en veranderingen in de microbiota. Vaak worden correlatiemethoden gebruikt om het netwerk in kaart te brengen.

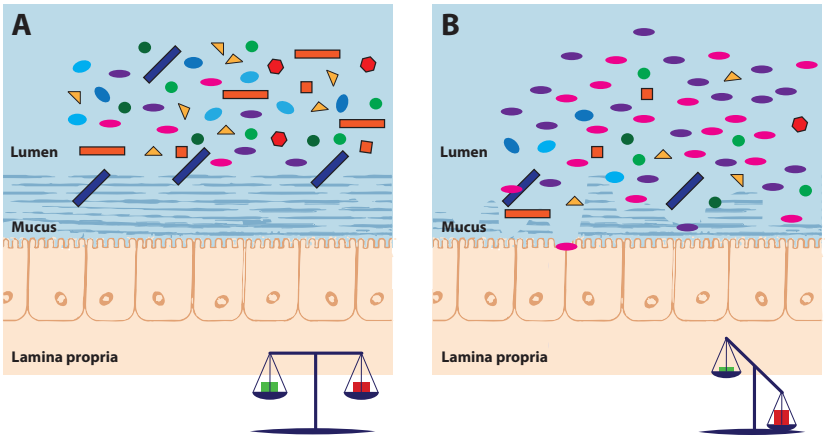
In Hoofdstuk 2 onderzochten we de betrouwbaarheid van correlatiemethoden bij het afleiden van de interactienetwerken. We hebben voor ons onderzoek het gegeneraliseerde Lotka-Volterra (gLV) model gebruikt om bacteriële gemeenschappen te simuleren. Dit model biedt inzicht in microbiële dynamieken zonder echte darmmicrobiota monsters nodig te hebben en maakt het mogelijk om parameters, zoals de groeisnelheid van de bacteriën, te variëren. Bovendien konden we een bekend interactienetwerk als referentie gebruiken, wat bij echte monsters niet mogelijk is.

We onderzochten onder andere de invloed van interindividuele variatie (verschillen in microbiota-samenstelling tussen individuen) en steekproefgrootte (de hoeveelheid beschikbare gegevens) op de nauwkeurigheid van netwerkreconstructies. Hoewel correlaties in microbiële aantallen vaak indicatief zijn voor ecologische interacties, toonden we aan dat meetruis, zoals variaties in monsterverwerking, het waarnemen van echte interacties bemoeilijkt. Daarnaast onderscheiden correlaties niet specifieke interactietypen, waardoor verificatie met laboratoriumstudies noodzakelijk blijft om de relatie te begrijpen. Het gLV-model biedt waardevolle inzichten, maar benadrukt ook dat studies met echte data onmisbaar zijn. Veel microbiota studies bij mensen zijn echter momentopnames waardoor schijnbare correlaties veroorzaakt kunnen worden door externe factoren, zoals schommelingen in voedingsstoffen (nutriënten), zonder dat er sprake is van echte interacties tussen soorten. Longitudinale studies, die microbiële gemeenschappen door de tijd volgen, stellen ons veel meer in staat om consistentie en patronen beter te begrijpen. Er bestaan echter maar weinig van dergelijke datasets, waarschijnlijk door praktische uitdagingen zoals herhaalde monsterafnames, aangezien dit sterk afhankelijk is van de toestemming en bereidheid om mee te werken aan onderzoek van de gastheer.

In Hoofdstuk 3 hebben we aangetoond hoe de microbiota in de loop van de tijd verandert en welke informatie deze variaties bevatten over verbanden tussen soorten. Hiervoor analyseerden we de tijdseries van twee individuen, verzameld door onderzoeker Caporaso en zijn collega's in 2011. Deze twee mensen hebben gedurende een jaar bijna dagelijks hun ontlasting ingeleverd, waardoor de variatie in de microbiota goed zichtbaar wordt. Met wavelet clustering hebben we vervolgens de patronen onthuld in deze gegevens. Wavelet clustering, al bekend in ecologische en epidemiologische studies, bleek bijzonder geschikt voor niet-stationaire tijdseries van microbiota en bood meer inzicht in collectief temporeel gedrag van de bacteriën dan gangbare correlatiemethoden. Met de waveletspectra bouwden we 'bomen' die verbanden tussen bacteriesoorten laten zien. Deze toonden aanzienlijke verschillen met bomen gebaseerd op correlatiemethoden, zoals een grotere totale taklengte (wijzend op meer functionele diversiteit) en duidelijke subgroepen. Dit laat zien dat wavelet clustering gevoeliger is voor subtiele verschillen in gemeenschapsstructuren dan correlatiemethoden. Onze resultaten benadrukken het belang van de methode die de onderzoeker kiest voor het analyseren van microbiota gegevens.

Darmmicrobiota en inflammatoire darmziekten

De interactie tussen mensen en de microbiota is het resultaat van meer dan een miljard jaar co-evolutie, wat heeft geleid tot een symbiotische relatie. Onze microben zijn betrokken bij tal van essentiële functies, waardoor verstoringen in hun soortensamenstelling, bekend als dysbiose, zijn gerelateerd aan verschillende ziekten. Dysbiose kan de mucuslaag in de darmwand verzwakken. Deze laag bedekt en beschermt de binnenkant van de darmen en helpt om schadelijke stoffen en micro-organismen buiten het lichaam te houden. Verzwakking van de mucuslaag kan leiden tot kolonisatie door schadelijke organismen, een verhoogd risico op ontstekingen en metabolische verstoringen. Dit brengt de gezondheid van de gastheer in gevaar (Figuur 2).



Figuur 2. Illustratie van het verschil tussen een darm met een gezonde microbiota en een verstoord microbiota. A) Een gezond darmmicrobiota is divers en in evenwicht. De mucus laag verzorgt de voeding voor de bacteriën en beschermt tegelijkertijd de darmcellen. B) Wanneer het evenwicht verstoord wordt kunnen sommige bacteriën uitgroeien en andere bacteriën verdrijven. Als het evenwicht verstoord is kan de mucuslaag worden aangetast, hierdoor kunnen bacteriën de darmwand beschadigen.

Inflammatoire darmziekten, waaronder de ziekte van Crohn en colitis ulcerosa, zijn chronische ontstekingen van de slijmvliezen van het darmkanaal. Deze zijn geassocieerd met een afwijkende samenstelling en diversiteit van de darmmicrobiota. In Hoofdstuk 4 tot en met 6 onderzoeken we de relaties tussen bacteriële dysbiose en het ziekteverloop van patiënten met inflammatoire darmziekten. In het verleden is hier al veel onderzoek naar gedaan, maar de resultaten tussen studies kwamen vaak niet met elkaar overeen. De verschillen in bevindingen over betrokkenheid van microben bij inflammatoire darmziekten kunnen waarschijnlijk worden toegeschreven aan technische variaties in onderzoeksmethoden en diversiteit tussen patiënten, inclusief variaties in ziektebeoordeling, medicatiegebruik en leefstijlfactoren. Daarnaast maakt ook de variabiliteit van de microbiota bij maag-darm aandoeningen, zoals natuurlijke fluctuaties in samenstelling, het uitdagend om consistente biologische signalen te identificeren.

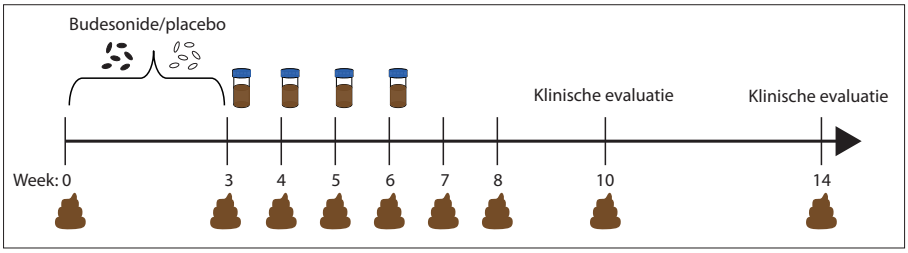
In Hoofdstuk 4 analyseerden we ontlastingsmonsters van patiënten met de ziekte van Crohn om associaties te beschrijven tussen bacteriële abundantie en remissie of exacerbatie van de ziekte. We gebruikten een kwantiel regressie model om relaties bloot te leggen die niet beperkt zijn tot de gemiddelde respons van alle patiënten. Kwantiel regressie maakt het mogelijk om een completer beeld te krijgen van de relaties tussen bacteriële abundantie en ziekte. Associaties met specifieke bacteriële families zijn mogelijk alleen waarneembaar bij een minderheid van de patiënten. Generieke associaties kunnen ook worden vastgesteld met methoden die zich richten op de gemiddelde respons, maar het is ook belangrijk om de minder algemene verschillen in de microbiota te begrijpen, omdat ze inzicht zouden kunnen geven in gepersonaliseerde benaderingen van behandeling. We koppelden de relatieve abundantie van bacteriële families ook aan bekende biomarkers van ziekteactiviteit, zoals fecale calprotectine en serum C-reactief proteïne. Onze bevindingen onthulden vooral significante negatieve associaties tussen verschillende bacteriële families en de ziekte, zoals Pasteurellaceae en Ruminococcaceae.

Bij het vergelijken van regressies met klinisch gedefinieerde verergering ontdekten we dat de associaties met fecale calprotectine sterker waren dan met de andere indicatoren. Samenvattend benadrukt ons onderzoek de heterogeniteit van de ziekte van Crohn en de relatie daarvan met het darmmicrobiota.

Ecologische bepalende factoren voor het succes van FMT-behandeling

Microbiota-gerelateerde therapieën zijn gericht op het doelgericht veranderen van de microbiota van patiënten, zodat deze van een dysbiotische naar een gezonde toestand gaan. Fecale microbiota-transplantatie (FMT), beter bekend als een poeptransplantatie, is een experimentele behandeling, waarbij fecaal materiaal, inclusief de microbiota van gezonde donoren aan de patiënt wordt overgebracht om het verstoorde microbiota te herstellen. FMT heeft zich gepositioneerd als een veelbelovende behandeling voor microbiota-gerelateerde aandoeningen, vooral in de behandeling van terugkerende *Clostridioides difficile*-infectie. Bij colitis ulcerosa wordt het succes van FMT bepaald door het bereiken van klinische remissie. Een van de uitdagingen bij FMT is de inconsistentie in de uitkomsten met betrekking tot de uiteindelijke samenstelling van de microbiota in de patiënt. Dit betekent dat elke persoon anders reageert op bepaalde donor bacteriën en dat er diverse immuunreacties kunnen worden geactiveerd bij patiënten met verschillende aandoeningen. Hypothetische uitkomsten zijn onder andere een terugkeer naar de oorspronkelijke dysbiotische toestand, een alternatieve dysbiotische toestand, een langdurige verandering met gunstige eigenschappen, of een blijvende verschuiving naar een gezonde microbiota met donorsoorten.

In ons onderzoek hebben we de microbiële families onderzocht die geassocieerd zijn met het succes van FMT-behandeling bij colitis ulcerosa. Hierdoor konden we de dynamiek van het darmmicrobiota onderzoeken. Deze longitudinale benadering bood inzicht in wekelijkse veranderingen (Figuur 3), een perspectief dat vaak ontbreekt in gerandomiseerde gecontroleerde proeven die voornamelijk gericht zijn op klinische uitkomsten.



Figuur 3. Opzet van de FECBUD-studie (gegevens gebruikt in Hoofdstukken 5 en 6). De patiënten werden eerst drie weken vooraf behandeld met budesonide (n = 12), een medicijn dat vaak wordt gebruikt om ontstekingen te verminderen, of een placebo (n = 12), een inactieve stof die geen therapeutische werking heeft. Daarna kregen de patiënten viermaal opeenvolgend van de Netherlands Donor Feces Bank (NDFB) een feces transplantatie van een gezonde donor. De evaluatie van de behandeling vond plaats na 10 en 14 weken na de start van de studie. Een subgroep van de patiëntengroep (9 van de 24 patiënten) bereikte een succesvolle gecombineerde klinische en endoscopische remissie na de FMT-behandeling. Fecesmonsters van de patiënten werden verzameld aan het begin van de studie, na de voorbehandeling, wekelijks na de feces transplantatie en twee, vier en acht weken na de FMT.

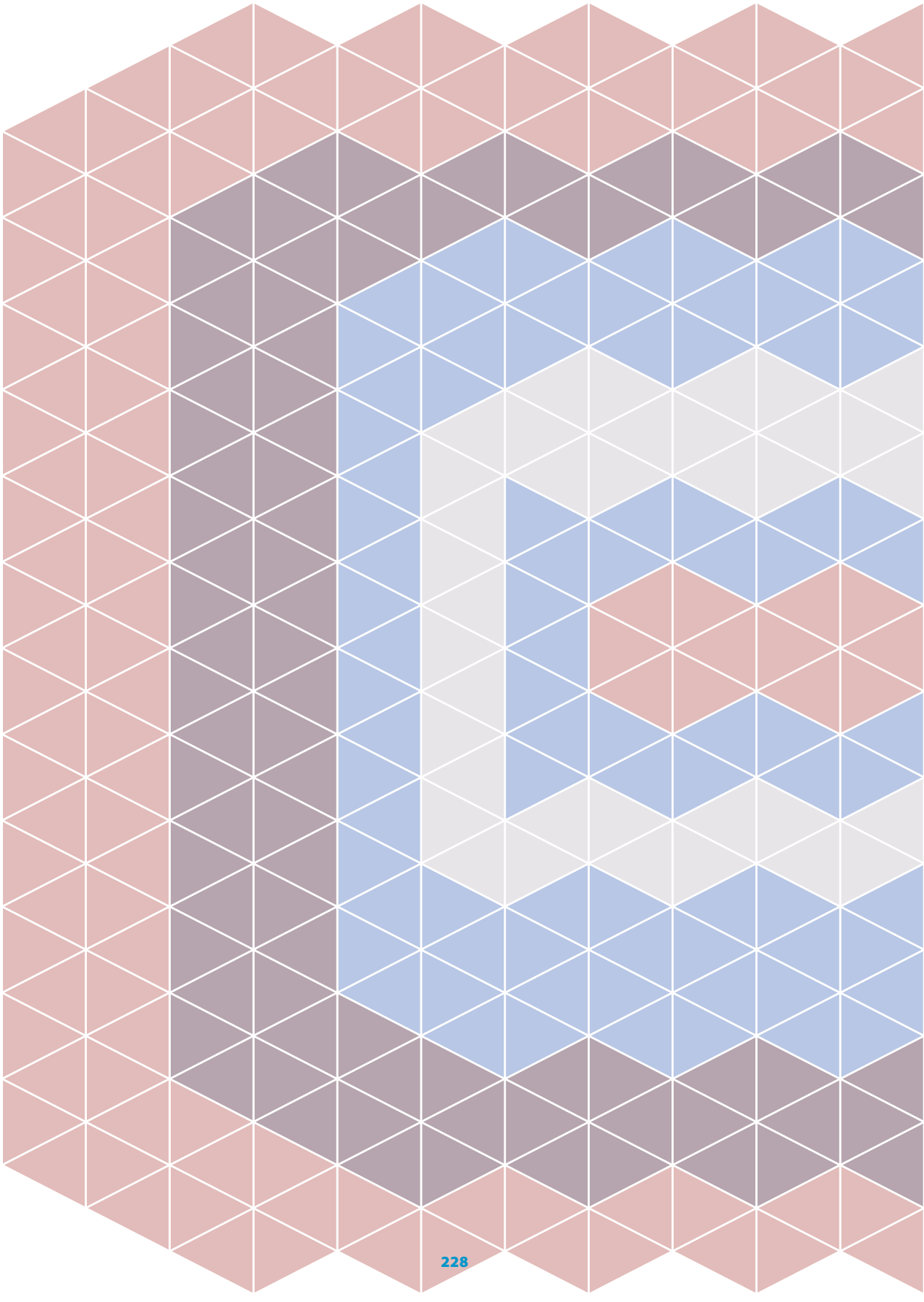
In Hoofdstuk 5 hebben we de associaties bestudeerd met betrekking tot het klinische succes van FMT bij patiënten en daarna de ontwikkeling van de microbiota tijdens en na de behandeling. We hebben een breed scala aan analytische technieken gebruikt om mogelijke associaties tussen bacteriële families en klinische uitkomsten te onderzoeken, waaronder een ordinatie analyse, Dirichlet multinomial mixture-analyse en longitudinale modellering. Het gebruik van deze benaderingen stelde ons in staat om significante verschillen in microbiële samenstelling en diversiteit te identificeren tussen patiënten die wel en geen baat hadden bij de behandeling. Zo vonden we dat het succes van FMT bij colitis ulcerosa patiënten lijkt samen te hangen met een beperkte groei van Prevotellaceae en de aanwezigheid van de families Lachnospiraceae en Ruminococcaceae. Het monitoren van de dynamiek van deze microbiële families zou mogelijk vroegtijdig inzicht kunnen geven in het succes van de behandeling tijdens FMT.

Er wordt algemeen aangenomen dat kolonisatie van donorsoorten in de microbiota van de ontvanger een sleutelmechanisme is achter het succes van FMT. Een interessante bevinding van ons onderzoek in Hoofdstuk 5 is dat we geen indicatie vonden voor een verschuiving in de microbiële samenstelling van de ontvanger naar de donormicrobiota onder de patiënten met een klinisch succes van FMT. In Hoofdstuk 6 onderzochten we daarom of de donor-gecentreerde visie van FMT klopt, door te analyseren of de microbiota-dynamiek gerelateerd is aan het behalen van remissie bij patiënten na FMT-behandeling. Hiervoor categoriseerden we de soorten op basis van hun oorsprong en temporele aanwezigheid: al aanwezig in de gastheer vóór FMT, afgeleid van de donor, of nieuwe soorten die tijdens de FMT-behandeling werden geïntroduceerd. Daarna modelleerden we het aantal soorten per categorie (gastheer-geassocieerd, donor-afgeleid en nieuw) voor patiënten die wel of geen baat hadden bij FMT. Onze resultaten tonen aan dat de patiënten die baat hadden bij de behandeling een hoger aantal gastheersoorten behielden in vergelijking met patiënten die niet profiteerden van de behandeling. Hoewel donorsoorten aanvankelijk meer koloniseerden bij de patiënten die geen baat hadden bij de behandeling, nam deze kolonisatie in de loop van de tijd af, waardoor het niveau gelijk werd aan dat van de patiënten die wel baat hadden bij de behandeling. Dit suggereert dat een succesvolle klinische reactie op FMT mogelijk wordt vergemakkelijkt door een microbiota die receptief is voor kolonisatie zonder de residentiële microbiota in gevaar te brengen. We ontdekten ook dat gastheersoorten met hogere relatieve abundanties vóór FMT beter in staat zijn om na FMT te blijven bestaan.

Tot slot

Er zijn talrijke omgevingsfactoren en gewoonten (voeding, leefstijl, enzovoort) die de samenstelling en functie van microben beïnvloeden. Deze complexiteit kan soms overweldigend zijn, maar kennis uit verschillende disciplines biedt waardevolle inzichten. Door te begrijpen hoe systemen reageren op veranderingen en weer in balans komen, vergroten we onze kennis van de complexe dynamiek binnen de menselijke darmmicrobiota. Deze inzichten kunnen microbiota-modulerende strategieën zoals FMT verbeteren en innovatie in persoonlijke therapeutische benaderingen stimuleren. Dit leidt tot een nieuw perspectief, waarbij de microbiota niet alleen wordt gezien als een strijdtoneel tegen pathogene microben, maar als een complexe ecologische gemeenschap. Beheerstrategieën zoals habitatherstel spelen hierbij een belangrijke rol. Om dit te bereiken zijn er uitgebreide (longitudinale) datasets nodig die een breed scala aan individuen met diverse achtergronden en levensstijlen omvatten. Dergelijke gegevens zijn essentieel om de mechanismen te begrijpen die de dynamiek van ecosystemen in gezondheid en ziekte beïnvloeden. Daarnaast moeten er geschikte methoden beschikbaar zijn om de complexe microbiota-data effectief te kunnen onderzoeken.

De beperkingen van een 'one-size-fits-all'-benadering worden steeds duidelijker, vooral door de diversiteit, genotypen (de genetische samenstelling) en fenotypen (de waarneembare eigenschappen en kenmerken) van de darmmicrobiota. Dit benadrukt de noodzaak van een meer op maat gemaakte benadering in (klinisch) wetenschappelijk onderzoek. Bijvoorbeeld het afstemmen van donor en ontvanger op basis van leefstijl en dieet kan de kans vergroten dat getransplanteerde bacteriën zich effectief vestigen en dat de aanwezige bacteriën de juiste voedingsstoffen krijgen. Dit kan bijdragen aan een effectievere verbetering van de microbiota van de ontvanger, aangepast aan de unieke behoeften van de patiënt. Maar om echt inzicht te krijgen in het functioneren van microben, moeten we het ecologisch perspectief van bacteriën in de gaten houden. Dit houdt in dat we begrijpen dat bacteriën hun omgeving, hulpbronnen en interacties waarnemen op een veel kleinere ruimtelijke schaal, met constante aanpassing aan hun directe omgeving. Dit betekent dat zelfs binnen een enkele steekproef verschillende bacteriën verschillende gedragingen, functies of interacties kunnen vertonen. Bovendien hebben binnen de darmmicrobiota verschillende bacteriën vaak overlappende functies. Dit houdt in dat als een bepaalde soort wordt verstoord of verwijderd, andere soorten die functie kunnen overnemen. Dit mechanisme is cruciaal voor het behoud van een gezonde darmmicrobiota en hiernaar is nog te weinig onderzoek gedaan. Een ecologische benadering, geworteld in theorie, is essentieel om de impact van de microbiota op gezondheid of ziekte te interpreteren. Dit perspectief stelt ons in staat om de complexe interacties binnen de microbiota beter te begrijpen, wat cruciaal is voor het ontwikkelen van effectieve therapeutische strategieën.



Appendices

Curriculum Vitae

Curriculum Vitae

Susanne Pinto was born on June 23, 1993, in Dordrecht, the Netherlands. She earned her Bachelor of Science degree in Biology at Leiden University in 2015. During her undergraduate studies, she conducted research on identifying larval Cholevinae (Coleoptera) species and developed an identification key for forensic entomology, aiding in post-mortem interval estimations. As part of her minor, she obtained a partial teaching qualification, which she later expanded to a full teaching degree during her master's studies. Education has consistently held a central place in her life. Since 2006, she has volunteered as a youth guide in Biesbosch National Park. From 2013 to 2016, she trained as a licensed swimming instructor and taught young children as part of local rescue teams, of which she has been a member since 2011.

From 2015 to 2018, Susanne pursued a Master of Science in Biology, also at Leiden University, specializing in Evolution, Biodiversity, and Conservation. Her research focused on ecological topics, including using mites as indicators of environmental changes in the Beekbergerwoud, both before and after its clearing in 1871 and during its restoration from 2005 onward. She also investigated the relationship between urban green spaces and the presence of macro-fauna in pharmaceutical cleanrooms at the Leiden Bio Science Park, including the Leiden University Medical Center, where she later conducted her PhD research. During her master's program, she broadened her studies with specializations in education and science communication and participated in the Leiden Leadership Programme of the Honours Academy. She especially appreciated the interdisciplinary interactions these tracks offered.

After completing her master's degree, Susanne pursued a PhD position that would combine her expertise in ecology with a focus on human health, while also providing opportunities to develop advanced data science and statistical skills. In September 2018, she started her PhD in the Department of Biomedical Data Science at Leiden University Medical Center. For the first two years, she was based at the Centre for Epidemiology and Surveillance of Infectious Diseases of the National Institute for Public Health and the Environment (RIVM), where dr. Hans Bogaards and dr. Elisa Benincà were her daily supervisors and closely guided her research. prof.dr. Ewout Steyerberg took on the role as her promotor during the COVID-19 pandemic and remained in this role until the completion of her PhD. Alongside her research, Susanne mentored two students, including a master's student, gaining valuable experience in teaching and supervision while deepening her appreciation for mentoring early researchers.

In 2023, Susanne continued advancing her academic career while completing her thesis. She joined dr. Maaike van Mourik's group at University Medical Center Utrecht (UMC Utrecht) to study automated outbreak detection systems.

Toward the end of her PhD, in 2024, Susanne was awarded a travel grant from the International Society for Microbial Ecology (ISME) and a grant from the Leiden University Fund (LUF)/ Slingelands Fund, which enabled her to attend her second ISME conference in Cape Town, South Africa. She had previously participated in the 2022 ISME18 conference in Lausanne, Switzerland. These conferences were major highlights of her academic journey, offering opportunities to present her research to an international audience, engage with leading experts in microbiome science, and gain new perspectives on global microbiome diversity.

In 2025, she returned to microbiome research as a postdoctoral researcher in the MetaHealth project under the supervision of prof.dr. Egija Zaura at the Department of Preventive Dentistry, Academic Centre for Dentistry Amsterdam (ACTA). Within this project, she also has the opportunity to supervise a PhD student while continuing to advance her own research and scientific learning.

Colofon

Thesis	**Ecology meets human health** Studies on human gut microbiota in health and disease
Author	Susanne Pinto
Cover & Layout	Jan Pinto (janpinto.nl)
Printing	Ridderprint, Alblasserdam

The work in this thesis was financially supported by the Dutch Research Council (NWO) through the program *Complexity in Health and Nutrition* (NWO grant 645.001.002; www.nwo.nl/onderzoeksprogrammas/complexiteit), with co-funding by the National Institute for Public Health and the Environment (RIVM), the Netherlands.