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

Pinto, S.

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.

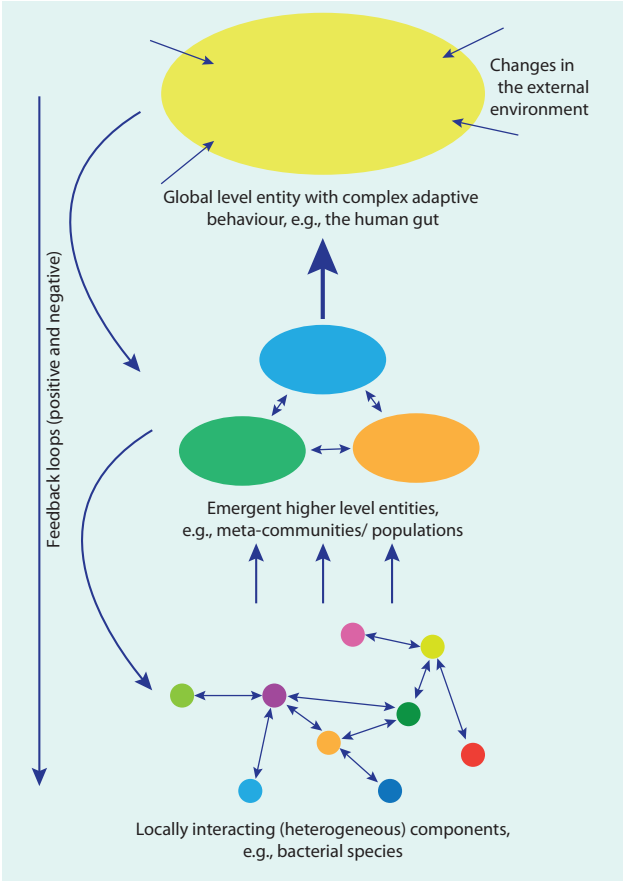

Figure 1.1 - Diagram illustrating the interactions and relationships within a complex system across different scales. In the human gut microbiota, emergence refers to the phenomenon where the overall functionality of the microbial community arises from the interactions among individual microbial species. 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]

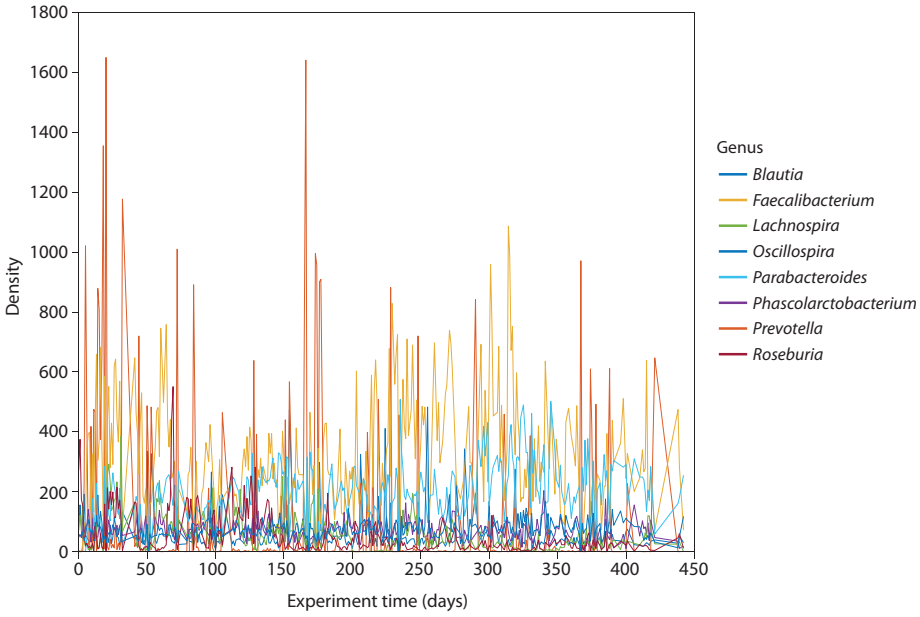

Figure 1.2 - Time series of the gut microbiota of one healthy male individual.[46] The time series shows temporal fluctuations on shorter timescales and overall stability over extended periods.[23]

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]

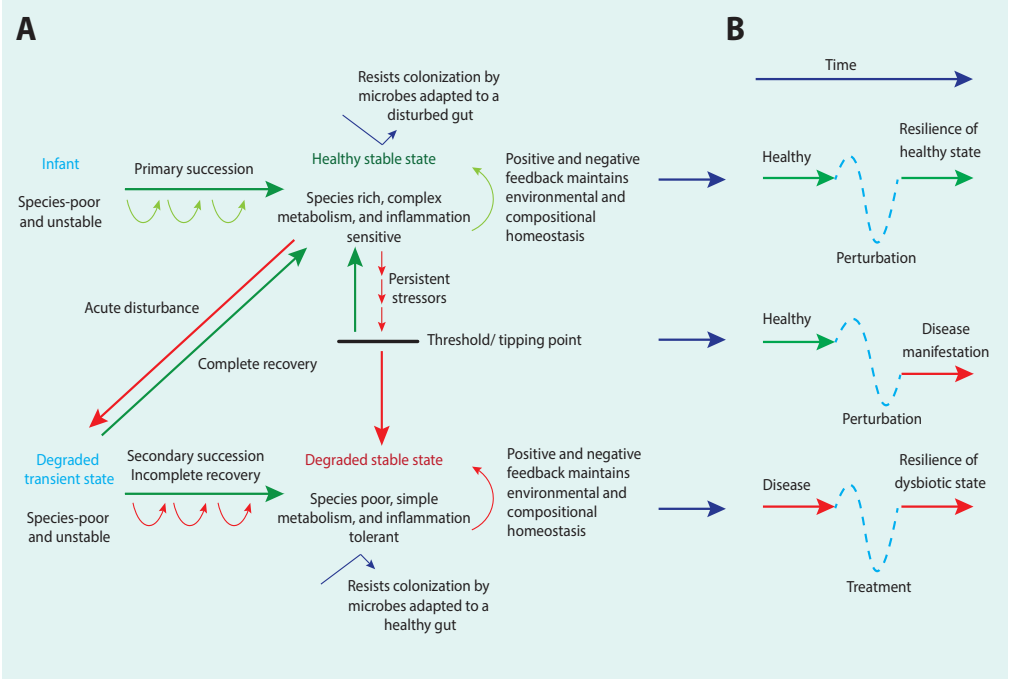

Figure 1.3 - Transitions in the composition of the human gut microbiota.
A) Positive and negative feedback loops probably have a role in driving succession and maintaining equilibrium states (resilience). B) Perturbations are often followed by an unstable state, which may return to a previous stable 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]

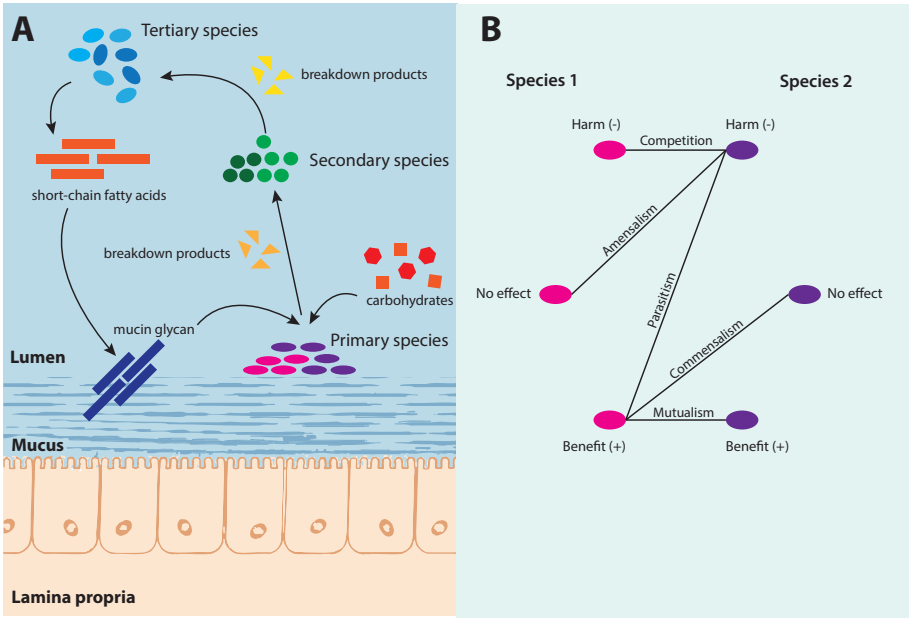

Figure 1.4 - A schematic representation of the gut, showing primary, secondary, and tertiary consumers and their potential interactions. A) At the start of the bacterial food chain are the bacteria that consume the primary nutrient sources, such as polysaccharides, oligosaccharides, proteins, sugars, 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.