Intestinal microbiota assembly and dynamics in health and disease

General Discussion & Summary

General discussion & Summary

The human intestinal tract is a unique setting that supplies a nutrient-rich environment for its complex microbial community that counts up to 100 trillion of microbes. All these microbes, including bacteria, archaea, viruses and eukarya, have co-evolved with the host [1]. This mutual relationship provides the host with benefits such as metabolic balance[2-5], processing of nutrients (including fiber digestion), vitamin synthesis, colonization resistance against invading pathogens[6, 7] and maturation and homeostasis of the gastrointestinal lymphoid tissues[8]. The initial colonization of the human intestinal tract starts at birth with the rupture of the amniotic membranes and subsequent passage through the birth canal [9, 10]. Subsequently, microbial populations evolve as the host matures and the diet changes. The infant gut microbiota is very unstable, showing big fluctuations in composition during the first 2.5 years of life[11, 12]. Given the strong dynamics in the microbiota during early infancy and its strong effect on the maturation of the host's immune system, a comprehensive insight into the processes that shape the infant microbiota is of particular importance. At around school-age the microbiota stabilizes [13] and resembles the mature adult composition. Once matured, the gut microbiota has been shown to be relatively resilient [14-17]. Nevertheless, it can still undergo dramatic compositional shifts, a condition known as dysbiosis, due to stressors like profound changes in diet, antibiotics use or diseases.

The past years have been the golden age for microbiota research with technological improvements leading to a rapid expansion of knowledge on the ecological dynamics of gut microbiota and an exponential increase in large-scale cohort studies and publications [18]. Nevertheless, the investigation of the gut microbiota and its role in human health is still young and in need of more fine-tuned studies. The vast majority of microbiota research is still based upon cross-sectional studies that can only partially explain the role of the gut microbiota in health and disease.

Studying the microbiome using a longitudinal design is pivotal in order to detect fluctuations of the microbial community but more important to find the relationships between bacteria and external factors and move from cross-sectional to temporal associations.

Moreover, to improve microbiome studies it is also key to better account for (or eliminate) the intrinsic compositional structure of next-generation sequencing data.

To this end, my thesis aimed to implement proper mathematical approaches prior to the data analysis of longitudinal data, to ensure more reliable results that can provide a deeper understanding on the assembly and maturation of the complex gut microbiota in humans and on the association between microbial perturbations with the development or progression of diseases.

The importance of longitudinal studies on the human microbiome

Since the introduction of next-generation sequencing techniques, the microbiome research field has revolutionised, and the microbiome has been studied in association to a plethora of diseases and determinants. However, for a long time the majority of these studies were cross-sectional in nature, comparing the microbiota composition of individuals exposed to a certain determinant or suffering from a disease to that of individuals without such an exposure or disease.

Such cross-sectional studies, however, cannot capture the dynamics in microbial

ecosystems. In particular, shifts that might occur as a result of the introduction of a specific environmental exposure, dietary changes or development or progression of diseases might easily be overlooked. Shifts might include temporary blooms of specific species followed by stabilization of the microbiome into the original or an alternative state [19, 20]. More importantly, cross-sectional studies are prone to selection bias, confounding and reverse causation (i.e., microbiota perturbations are a consequence rather than a cause of a disease). Many examples exist of bias in cross-sectional microbiome studies with the confounding effect of antidiabetic medication as a classic example[21].

In order to advance microbiome research, we should therefore move towards longitudinal study designs in which the outcome of interest (e.g., disease or disease exacerbation) has not yet manifested in any of the participants at baseline. Prospective cohort studies have the potential to provide the strongest scientific evidence of all types of observational study designs. In addition, as prospective studies examine changes in microbial composition over time in association to a certain exposure or the manifestation of disease (exacerbation), each individual serve as its own control. This significantly reduces the number of potential confounding factors that could lead to either spurious or undetected associations.

For those reasons all the studies presented in this thesis have a longitudinal design. In **chapter 3**. we collected faecal samples from 98 infants repeatedly at 1-2, 4 and 8 weeks, as well as 4, 5, 6, 9, 11 and 14 months of age. In **chapter 4**, 1453 stool samples were collected from 440 children, at 5, 13, 21 and 31 weeks of age and once again at school-age (6–11 years). In **chapter 5**, we studied the dynamics and resilience of the microbiota among 106 travellers that experienced a bout of diarrhoea and subsequently did or did not develop post-infectious Irritable Bowel Syndrome. In **chapter 6**, we examined the stability of the microbiota composition in 15 healthy control subjects and 57 patients with Crohn's disease (CD). As 22 of the CD patients developed an exacerbation during the course of the follow-up, we could examine whether the microbiota stability was associated with the disease course.

In all studies multiple stool samples along with extensive metadata were collected enabling us to move from mere cross-sectional to temporal associations.

Removing the compositionality of microbiome sequencing data using quantitative methods

Next to failing to acknowledge limitations of cross-sectional designs, another common flaw observed in microbiome studies is to not take into adequate consideration the compositional nature of microbiome sequencing data [22]. One way to overcome the compositional nature of microbiome sequencing data is to make the data quantitative again. In **chapter 2**, we investigated the use of various approaches to quantitatively profile the microbiota, as opposed to the traditional relative microbiome profiling (RMP), to overcome the compositional structure of sequencing data. Next to the Quantitative Microbial Profiling using flow cytometry-based microbial load (QMP) as introduced by Vandeputte *et al.* 2017 [23], we additionally combined Propidium Monoazide pre-treatment with flow cytometry-based cell counting in order to profile only intact cells (QMP-PMA), and also performed Quantitative Microbial Profiling using qPCR to determine the microbial load (QMP-qPCR). Overall, our results suggested that QMP could be a promising and elegant approach to overcome the compositional structure of microbiome data

but is still far from perfect as the use of cell flow cytometry can introduce additional biases. Moreover, this technique is still too laborious to apply in large cohort studies. We therefore investigated the possibility to use different, more high-throughput, methods to quantify the microbial load of the samples such as qPCR.

Our findings confirmed the previous observation [23] that absolute abundance profiles differ significantly from those generated by relative approaches. When comparing RMP to QMP, sample rank order concordance within the 15 most abundant genera varied widely with the highest concordance observed for *Fusicatenibacter* and the lowest concordance for *Blautia*. The differences among relative and quantitative methods extended also to enterotypes. When using DMM clustering, we identified two enterotypes enriched in *Bacteroides* or *Prevotella* with a significant difference in microbial load between these enterotypes. However, this difference was absent when the microbial loads were determined using qPCR.

When moving the focus on the different quantitative profile methods, we found that technical sources of variability may introduce additional bias depending on the quantification method being used. Generating quantitative microbiome profiles revealed that profiles obtained after PMAxx-treatment remained highly similar to the standard QMP profiles, although the observed genus richness slightly decreased upon PMAxx-treatment. This indicates that free extracellular DNA does not significantly bias the traditional flow-cytometry-based QMP method, although it cannot be deduced whether the existing dissimilarities between QMP-PMA and QMP microbial profiles are due to the elimination of free extracellular DNA or merely due to the introduction of additional technical variation during sample handling.

Previous studies have advocated qPCR as a more suitable and accessible alternative for microbial quantification compared to Flow-cytometry based quantification, although direct comparisons between both methods were lacking [24]. Our analysis demonstrated that quantification of bacterial load by qPCR results in highly divergent microbiome profiles as well as a strong decrease in the observed genus richness when compared to standard QMP or QMP-PMA methods. We ruled out that these deviant QMP-qPCR based profiles were the result of a lack of precision or sensitivity of qPCR as we proved that quantification of microbial load based upon Droplet Digital PCR (ddPCR) correlated strongly with qPCR-based quantification. Together these results prove that qPCR-based quantification might not be an adequate approach for quantitative microbiome profiling compared to flow-cytometry based quantification. Although 16S rRNA gene copy-number correction to account for variable in copy-numbers between bacterial taxa was applied in our study, the added value of this approach has recently been questioned as gene-copy number normalization even failed to improve the classification of 16S rRNA sequenced simple mock communities [25]. Indeed, low predictive accuracy and substantial disagreement has been observed between gene-copy number prediction tools [26]. As gene copy-number normalization is also applied in the traditional flow-cytometry-based QMP, a more comprehensive catalogue of copy numbers or other methods to account for variance are urgently needed.

This implies that there is currently no high-throughput and accessible laboratory method available to eliminate the compositional problem of microbiome sequencing data at its root. Therefore, to overcome the problem while still being able to investigate the role of the intestinal microbiota in human health, we decided to apply a mathemat-

ical approach to remove the compositional component from the data and thus allowing us to apply standard analysis techniques.

Removing the compositionality and sparseness using mathematical methods to ensure more reliable results and allow the implementation of new methods

As mentioned before**,** compositional data belongs to the mathematical space called the Simplex. As a consequence, when an element in a vector of compositional data increases, all the other elements together must decrease as all elements sum up to a constrained value. A mathematical solution to this problem was presented by Aitchison[27] who demonstrated how to project the compositional data from the Simplex to the Euclidean space by taking the logarithm of the ratio between each value and the geometric mean of the values. Applying this approach, called centred log-ratio (*clr*) transformation, to microbiome data is challenging because of the intrinsic sparseness of the data. Therefore, it is important to remove the zeros in the data without introducing biases. For example, a common practice is to add a small pseudo counts to the zeros in order to make the computation of the *clr* possible. But as shown in many studies [28-31] , the choice of the exact value of the pseudo count can have a large impact on the results. We therefore decided to model the zeros using the Dirichlet distribution[32], prior the *clr* transformation.

In **chapter 3**, the implementation of the Aitchison transformation, allowed us to obtain more reliable results on the effect that various ecological principles, including dispersal (limitation), neutral processes and environmental filtering contribute to the assembly of microbial communities during early infancy. Frequently used distance-based ordination methods such as PCoA and dbRDA rely on metrics such as the Bray-Curtis distance or the Jensen-Shannon divergence and are therefore more affected by the sequencing depth than by the actual microbial composition of samples. This was extensively illustrated by McMurdie and Holmes [33] by benchmarking the performance of commonly used distances and dissimilarities. The authors simulated publicly available data to introduce variations in sequencing depth and subsequently tested the performance of clustering methods and differential abundance testing. The results showed that a decrease in sequencing depth led to a poorer performance of the clustering methods and an increase in the false discovery rate for the differential abundance testing [33]. Moreover, distance-based ordination methods based on relative abundance data will mostly discriminate samples based on the most dominant bacteria rather than the most variable ones. A clear demonstration of this mechanism was given by Gorvitovskaia *et al.* [34] who demonstrated that the relative abundance of *Bacteroides* and *Prevotella,* rather than the underlying microbial community structures, is driving the separation of samples in ordination plots. To circumvent these types of bias, we applied *clr* transformation on microbial count data which allowed us to use PCA rather than PCoA. Besides adequately addressing the compositionality of the data, PCA on Euclidean distances also allows to identify and visually represent the contribution of individual taxa in the variations in overall microbial community structure between samples.

When this approach could not be used, we applied analysis methods that were developed specifically for compositional data.

The approaches to overcome the compositional nature of microbiome data applied

in this thesis have been carefully selected to address the research questions. It is important to stress that different transformations may be preferred depending on the analytic modelling tool of choice. The application of the Aitchison transformation is recommended when performing a PCA, but different data preprocessing steps might be preferred when performing regression analysis, network analysis or machine learning techniques (e.g., Random Forest).

The establishment and maturation of the gut microbiota: factors shaping the process

During the past few years, a lively academic controversy has emerged on the existence of prenatal microbial communities and *in utero* colonization of the foetus[35]. Although several studies showed molecular microbial profiles and limited numbers of viable microbes in placental tissue, amniotic fluid and foetal meconium[36, 37], current evidence favours contamination during sample collection and handling as the most likely source of these microbial signals[35, 38, 39]. The existence of interacting microbial communities in the womb that facilitate in utero colonization is thus highly unlikely, which makes rupture of the amniotic membranes the starting point of microbial colonization. The richness and diversity of the gut right after birth is extremely low when compared to adults, but as demonstrated in **chapters 3 and 4** gradually increases throughout and over the first year of life. In **chapter 3** the median Shannon index of children aged 1-2 weeks was 1.77 compared to a median diversity of 4.04 in the maternal faecal samples collected at the same time. This difference *per se* might not seems particularly large but as those values are logarithmic an increase of one unit is associated with substantial differences in microbial diversity. Similar results were reported in **chapter 4** in which children were sampled starting from 5 weeks of age. The assembly of the gut microbiota then steadily proceeded with gradual increases in microbial diversity up to the age of 6-8 months postpartum where a drastic increase in richness and diversity occurred [**Chapter 3 and 4**]. Unsurprisingly, this time window coincides with the start of weaning, pointing towards the introduction of solid foods or the cessation of breast feeding as a driving force of significant increases in the infant microbial richness and diversity.

Our results in **chapter 3** showed that longer breastfeeding duration was most strongly linked to a delayed increase in microbial diversity. This indicates that, regardless of the introduction of other food substrates, the availability of human milk maintains a simple microbial community dominated by only a few genera.

When examining the overall microbial community structure in both **chapter 3 and 4**, we identified 6 microbial clusters. In **chapter 3**, the majority of samples collected at 1-2 weeks of age grouped into 3 clusters, dominated by *Bifidobacterium* (Actinobacteria) for cluster 1, *Escherichia (*Proteobacteri*a)* and *Streptococcus (*Firmicutes) for cluster 4 and *Bacteroides* (Bacteroidetes) for cluster 6. In **chapter 4**, for samples collected at 5 weeks postpartum, only 2 cluster were prevailing and were characterized by *Escherichia* for cluster 1, while cluster 3 was characterized by *Streptococcus* and *Veillonella.* The early dominance of Bacteroides in some of the infants in the LucKi cohort [**Chapter 3**] could be linked to a vaginal delivery as has also been shown in several previous studies [40, 41]. Interestingly Yassour *et al.* [40] reported that children born via C-section also harboured *Bacteroides* strains during the first week of life but lost these bacteria

upon which streptococci became dominant. Nayfach *et al.* [42] further substantiated these findings by tracking the vertical transmission of bacterial strains from mother to infant based upon the identification of rare Single Nucleotide Polymorphism (SNP) characteristic of bacterial strains in shotgun metagenomic data. In line with our results in **chapter 3**, the authors show that *Bacteroides* and *Parabacteroides* are among the most vertically transmitted bacteria in case of vaginal delivery while transmission of these bacteria among C-section delivered infants was lacking. Further evidence of vertical transmission of maternal gut microbiota comes from the study of Korpela *et al.* [43]. Using publicly available shotgun metagenomic sequencing the authors used rare SNPs that were not shared with samples from any non-family members as marker to track the transmission of bacterial strains. Their results not only showed that in the 87% of vaginal delivered infant the vertical transmission of gut microbiota involved mainly bacteria of the classes Actinobacteria and Bacteroidia, but also that the colonization of maternal strains was more persistent. Altogether, our findings that vaginal delivered infants shared a significantly higher proportion of *Bacteroides* ASVs with faecal microbiota of their mothers as compared to C-section delivered infants [**Chapter 3**] are in strong agreement with these previous studies. Together these results suggest that the neonatal microbial composition is more strongly influenced by exposure to the maternal gut microbiota than by the passage through the birth canal.

Indeed, causal evidence for the relationship between exposure to maternal faeces and persistence of *Bacteroides* has recently been demonstrated in a proof-of-concept study on maternal faecal microbiota transplantation in Caesarean-section delivered infants [44]. The relative Bacteroidales abundance observed in the first week of life decreased more than 100-fold by the age of 3 weeks postpartum in C-section delivered infants. In contrast, the abundance of Bacteroidales (mainly *Bacteroides*) increased over the first weeks of life in C-section born infants that received maternal FMT. This again confirms engraftment of maternal faecal *Bacteroides* strains. Moreover, together with our findings, these results further question the benefits of bacterial baptism or vaginal seeding approaches [45].It remains however unclear why we could observe a dominance of *Bacteroides* in some of the children in the Dutch Lucki cohort as early as 1-2 weeks postpartum, whereas a similar *Bacteroides* dominated group of neonates was lacking in the German PAPS study. This is even more surprising as the proportion of C-section delivered infants was even lower in the latter cohort (6.6 vs. 14.3%). Although technical variation due to differences in DNA isolation protocols and sample collection methods could be partly responsible for this difference, national differences in childbirth practices likely also play a role. In contrast to Germany, where inpatient hospital delivery is standard practice and mothers and their newborns generally remain hospitalized for 3-4 days after delivery, homebirth and outpatient hospital delivery are common practice in the Netherlands. Further research on the potential impact of delivery practices, bowel movements during labor and in- vs. outpatient childbirth on the persistent colonization of *Bacteroides* is currently needed.

More research on priority effects, a well-known ecological concept, might also shed additional light on the dynamics of the infant gut colonization. The dispersion of bacteria from the “regional” pool is known to be affected by exposure to maternal gut microbiota, environment of delivery and contact with other persons (e.g., hospital staff). Altogether those elements can affect the pioneer species colonizing the infant gut.

Moreover, those early colonizers might change the environmental condition of the infant gut, affecting the chances for other bacteria to colonize the infant gut. An example of this process is (nontoxigenic) *Bacteroides fragilis* that upon colonization the niche is resistant to colonization by the same, but not different, species [46].

Altogether these results suggest that the early colonization of the gut is a chaotic but not stochastic phase characterized by a low diversity and a high interindividual variability, likely driven by dispersal limitation and environmental selection. In particular, vaginal delivery sets up a unique initial microbial community by exposing the child to maternal faeces during delivery. This was further supported by the results of our neutral community modelling [**Chapter 3**], in which *Bacteroides fragilis* resulted under negative selection in case of caesarean section.

Once the gut is colonized the maturation of the gut microbiota proceeds gradually and steadily towards adulthood with a major influence of dietary factors [47]. For example, the origin and trajectory of the *Bifidobacterium*-dominated cluster 1 (**Chapter 3**) is likely the result of a combination of different dispersal mechanisms (e.g., seeding by the maternal milk microbiome), as compared to the other clusters, and subsequently selection and drift driven by the ability of *Bifidobacterium* species to degrade HMOs derived from breast milk. As early as of 4 weeks postpartum and onwards, dietary factors showed to have a greater impact on the microbial community structure than perinatal determinants. In particular, results from **chapter 3 and 4** show that cessation of breastmilk relate more with changes in microbial composition than introduction of solid food. However, in contrast to previous studies with limited number of sampling time-points and lack of detailed dietary data[48], our data in **chapter 3** do demonstrate that also the type and complexity of solid foods is important for maturation of the infant gut microbiota. Infants with a more adult-like omnivore dietary pattern, characterized by the consumption of rice, pasta, fish, and meat products, at the age of 9 months had the most mature gut microbiota composition with highest levels of *Faecalibacterium spp.* and lowest levels of *Enterococcus spp*. and *Staphylococcus* spp. This elegantly demonstrates the importance of longitudinal study designs and sufficient numbers of repeated samples when studying the effect of diet and other determinants on the highly dynamic infant microbiota.

Next to the impact of birth mode and diet, also dispersal from other individuals and companion animals appeared to affect the infant gut microbiota [**Chapter 3 and 4**]. This is in line with several previous studies [49-52], although the specific effects of sibship size and pet exposure differ between studies. On this regard, future studies should focus on more refined microbial profiling on a strain level (e.g., by whole metagenome shotgun sequencing) to track the microbial dispersal from siblings and pets to newborns. Nonetheless, the goal to fully explain the inter-individual microbiota variations could be unachievable as indicated by the increasing evidence on the central role of stochastic events on the assembly and maturation of the gut microbiota.

Longitudinal methods to link disturbances in the maturation of the gut microbiota to the onset of allergic manifestations

Many epidemiological studies [53-58] suggest that the infant gut microbiota plays an important role in manifestation of allergic diseases and asthma, although the results vary considerably between studies. The lack or early samples and different ages of sam-

ple collection, different microbial profile methods and insufficient control for potential confounders might contribute to the heterogeneity between study results[55-59]. Moreover, cross-sectional case-control studies cannot discriminate whether the change in microbiota composition is the cause or the result of the allergic manifestation. In this context and given the complex dynamics of gut microbiota assembly and maturation, longitudinal studies are important to allow analysis on the overall development of the infant gut microbiota. In the study presented in **chapter 4**, the collection of an adequate number of repeated samples during the first year of life and the follow-up up to school age together with recording of extensive metadata allowed us to implement, for the first time in a microbiome study, a Joint Modelling while correcting for potential confounders. The joint modelling combines how an exposure variable changes over time (longitudinal modelling of a risk factor) with the time at which an outcome event occurs (survival analysis). The results of the joint modelling showed that throughout the first year of life a higher microbial richness was associated with a lower risk and delayed onset of atopic dermatitis.

The collection of repeated samples during such an important time window also allowed us to define the stage of maturity of the infant gut microbiota using a machine learning approach. We subsequently examined to what extent microbiota maturation was linked to atopic dermatitis. Our results [**Chapter 4**} showed that children that developed atopic dermatitis and allergic manifestation have a more mature microbiota in the earliest time points when compared to infants that remain free from allergies. After the first half year of life this trend appears to be reverted with a less mature microbiota among infants who developed allergic symptoms.

Moreover, the longitudinal design of the study in **chapter 4** allowed us to identify microbial taxa that were differentially abundant throughout the infancy among infants who did or did not developed allergic manifestations. In line with previous studies [57, 60], the abundance of *Lachnobacterium* and *Faecalibacterium* was decreased among children who subsequently developed atopic dermatitis. The extensive duration of this decreased abundance throughout infancy suggests a protective role of those bacterial genera in preventing the development of atopic dermatitis. Altogether, we showed that applying various multivariable longitudinal models can provide additional insight into the temporal associations between the dynamic infant gut microbiota and the onset of non-communicable diseases, such as allergies.

By applying such methods, we can separate bacterial taxa likely involved in the pathophysiology of allergies from bacteria that merely shift as a consequence from the disease, its treatment or disease-related dietary restrictions. This greatly helps to unveil those bacteria, such as *Faecalibacterium*, that are suitable candidates for next-generation probiotics in the primary prevention of allergic diseases.

Gut microbiota in adult human health: the role dysbiosis in the onset and disease course of gastrointestinal disorders

The last decades have been characterized by numerous studies that underscored the role of the human gut microbiota in health and homeostasis. The microbiome is involved in metabolism regulation, immune maturation and response and protection against pathogens [61-64] just to cite a few examples. Given the involvement of the microbiome in all these pivotal functions, it is not surprising that microbial perturbations

(dysbiosis) have been linked to the onset and course of numerous diseases. Obesity[65], type II diabetes[66], activation of HIV[67], IBS[68], IBD[69], and atopy[53-55, 57, 59] are few examples of a long and growing list of diseases and disorders that have been linked to microbial perturbations. The studies presented in **Chapter 5 and 6** investigated the temporal associations between dysbiosis and Inflammatory Bowel Disease and Irritable Bowel Syndrome, respectively.

Studying the microbiota prior to disease onset is often impossible for diseases that manifest in adulthood as it would require extremely large sample sizes and extensive follow-up. Therefore, no study had previously been able to study the microbiota composition among individuals that subsequently developed IBS. In **Chapter 5** we describe the first study that has examined the baseline microbiota as well as its stability and resilience upon a bout of diarrhoea among intercontinental travellers that subsequently did or did not develop post-infectious IBS. We demonstrated that, as compared to control subjects, the microbial diversity and community structure were already significantly different among PI-IBS cases prior to disease onset. Longitudinal analyses moreover revealed differentially abundant genera, including increased *Bacteroides* levels, in cases as compared to controls from baseline onwards. These results clearly show for the first time that the microbiota diversity and composition can predispose to the development of PI-IBS after an episode of gastroenteritis. We can therefore eliminate the possibility that the microbiota alterations are merely the result of avoidance of specific food items that IBS patients link to worsening of symptoms or that microbiota alterations are epiphenomena linked to an unknown trigger of IBS. As these alternative explanations could not be completely ruled out in all previous cross-sectional studies, our study greatly enhances the evidence for a role of the microbiota in the pathogenesis of (PI-) IBS.

Despite the large literature on the role of the microbiota in IBD, the results on the role of the microbiota in disease flares are often inconsistent and sometimes even contradictory. Many cross-sectional studies compared patients the microbiota of patients with active Crohn 's Disease (CD) to that of patients in remission. The microbiota of CD patients with an active disease flare has been linked to increased levels of Enterobacteriaceae [70, 71] and *Bacteroides spp.* [70, 72], a reduction of *F. prausnitzii* [73-75] and *Clostridium coccoides* group [73, 76] in some but far from all studies. These inconsistencies may in large part be due the inter-individual variation in microbiota composition, confounding factors, and the heterogeneous nature of CD, that can only be partly accounted for in cross-sectional studies. Longitudinal studies are therefore important to shed further light on the causal relationship between dysbiosis and disease onset and course. The longitudinal design of the study in **Chapter 6** allowed us to compare the stability of the intestinal microbiota in healthy subjects as compared to CD patients. We showed that CD patients have a less stable faecal microbiota as compared to healthy subjects. This is in line with a study in which faecal samples were sequentially collected from patients with ulcerative colitis (UC) remaining in remission and with stable medication during a year of follow-up. Only one-third of the dominant taxa were persistently detected among UC patients in this study, while healthy individuals showed a remarkable microbiota stability [77] .

However, contrasting previous cross-sectional studies [70, 75], by profiling multiple samples from CD patients with changing or stable disease activity we did not observe

the stability of the gut microbiota to be associated with disease course. Two other longitudinal studies also did not observe a correlation between microbial composition and active inflammation [78, 79]. Halfvarson and colleagues profiled the microbiota in a cohort of 128 IBD patients, including 49 CD patients, and observed a distinct microbiota as compared to healthy controls but no association with calprotectin levels, as inflammatory marker, in the patient group [78, 79]. In a large cohort of over 2,000 a cohort of non-IBD and IBD faecal samples from four countries also observed a clear dysbiosis in CD patients but no link with disease activity [78, 79]. Together this sharp contrast between case-control studies and longitudinal cohort studies when examining the link between the microbiota and IBD disease activity highlights the risk of identifying spurious findings and bias in cross-sectional studies.

Future perspectives

The results presented in this thesis show how much the microbiome field can benefit from longitudinal study designs and appropriate statistical frameworks to answer biological questions. Despite the improvements the microbiome field witnessed during the past years, there is still a profound lack of analysis methods tailored specifically to handle microbiome data. Many tools have been borrowed from the ecology field and adjusted for microbiome studies, however those tools are not meant to handle the high dimensionality, sparseness and compositionality typical for the data generated in microbiome studies [33].

This lack of analysis methods results in the inability to answer some important questions on the microbe-microbe and microbe-host interactions and the mechanisms linking dysbiosis to diseases. Statistical challenges related to sparseness, high dimensionality and complexity are becoming even more evident when applying methods with higher taxonomic resolution such as whole metagenome sequencing or when integrating multiple -omics data. A possible focus of future efforts is the microbial profiling technology itself. High-throughput and 16S rRNA gene amplicon sequencing technologies have set the path to revolutionize the microbiome field to what we know it today. Nonetheless as the knowledge and experience in this field grow, the awareness of its intrinsic limitation grows as well. It is time to re-think sequencing technologies to create data that are not intrinsically compositional. Many efforts have been made already in this direction, one also presented in this thesis, but this task should ultimately be accomplished together with the gene sequencing industry.

On the subject of data analysis, a promising method that is becoming more and more present in microbiome studies is represented by network analysis. This method holds the potential to discover keystone species and key players in the food chain allowing a deeper understanding of what defines a healthy microbiota. Unfortunately, the majority of the studies do not account for the possible spurious correlations generated by correlation indexes, i.e., Spearman's correlation coefficient, commonly used to generate the interaction networks, which can lead to visually impressive interaction networks that are hampered by a limited validity.

Next to studying bacterial interactions, the integration of metabolomic data with metagenomic data such as microbial abundance and gene pathway abundance, as well as extensive (clinical) metadata, is also pivotal to reveal causal pathways and leads for disease prevention. Further development of statistical frameworks for multi-omics

data integration should therefore be an important research focus. It is also important to stress how much the microbiome field is in need of more fundamental research on the gut microbiota. To define what a healthy microbiota is and what the inter-bacterial dynamics are. Because only answering to these questions it will be possible to make successful interventions to improve patient's health.

With respect to whole metagenome shotgun sequencing, technologies are improving at an unprecedented pace, in particular with respect to the improved quality of long-read sequencing technologies such as nanopore sequencing[80]. Longer reads will significantly simplify metagenomic data processing and assembly, yet almost all bioinformatic pipelines are still focused on short-read sequencing data. Given the pace at which long-read sequencing is developing, time is pressing to also shift the focus from a bioinformatics point of view.

Last, but not least, establishing gold standards or guidelines for microbiome data analysis and reporting are crucially important to enable comparison of results between studies and facilitate more rapid sample analysis and throughput. This call was already made more than 10 years ago [81] and is to date still largely unaddressed.

Together with tailored analytical methods and appropriate standards, future microbiome studies will benefit of bigger cohorts, more frequent sampling, and longer follow up to unravel the long- and short-term fluctuations of the gut microbial communities and how those fluctuations can impact human health later in life.

In conclusion the studies presented in this thesis demonstrate the importance of longitudinal study designs and appropriate analytical methods to study the microbiota in health and disease. The application of such analytical methods and longitudinal study designs substantially decreases the possibility of false positive findings due to spurious correlations, confounding factors, and reverse causalities. Together these results aid in a stronger foundation for effective microbiota-based intervention strategies to prevent or reduce the burden of non-communicable diseases

References

1. Rawls, J.F., et al., Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell, 2006. 127(2): p. 423-433.
2. Backhed, F., et al., The gut microbiota as an environmental factor that regulates fat storage. Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(44): p. 15718-15723.
3. Duncan, S.H., P. Louis, and H.J. Flint, Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Applied and Environmental Microbiology, 2004. 70(10): p. 5810-5817.
4. Chen, F.D., W. Wu, and Y.Z. Cong, Short chain fatty acids regulation of neutrophil production of IL-10. Journal of Immunology, 2016. 196.
5. Sartor, R.B., Microbial influences in inflammatory bowel diseases. Gastroenterology, 2008. 134(2): p. 577-594.
6. Hooper, L.V., OPINION Do symbiotic bacteria subvert host immunity? Nature Reviews Microbiology, 2009. 7(5): p. 367-374.
7. Othman, M., R. Aguero, and H.C. Lin, Alterations in intestinal microbial flora and human disease. Current Opinion in Gastroenterology, 2008. 24(1): p. 11-16.
8. Takahashi, K., Interaction between the Intestinal Immune System and Commensal Bacteria and Its Effect on the Regulation of Allergic Reactions. Bioscience Biotechnology and Biochemistry, 2010. 74(4): p. 691-695.
9. Korpela, K., et al., Selective maternal seeding and environment shape the human gut microbiome. Genome Research, 2018. 28(4): p. 561-568.
10. Wampach, L., et al., Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. Nature Communications, 2018. 9.
11. Zhuang, L., et al., Intestinal Microbiota in Early Life and Its Implications on Childhood Health. Genomics Proteomics Bioinformatics, 2019. 17(1): p. 13-25.

12. Stewart, C.J., et al., Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature, 2018. 562(7728): p. 583-588.
13. Zhong, H.Z., et al., Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. Microbiome, 2019. 7.
14. Costello, E.K., et al., The application of ecological theory toward an understanding of the human microbiome. Science, 2012. 336(6086): p. 1255-62.
15. Gilbert, J.A. and S.V. Lynch, Community ecology as a framework for human microbiome research. Nat Med, 2019. 25(6): p. 884-889.
16. Donaldson, G.P., S.M. Lee, and S.K. Mazmanian, Gut biogeography of the bacterial microbiota. Nat Rev Microbiol, 2016. 14(1): p. 20-32.
17. Faust, K., et al., Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology, 2015. 25: p. 56-66.
18. Gilbert, J.A., et al., Current understanding of the human microbiome. Nature Medicine, 2018. 24(4): p. 392-400.
19. Walker, A.W., et al., Dominant and diet-responsive groups of bacteria within the human colonic microbiota. The ISME journal, 2011. 5(2): p. 220.
20. Dethlefsen, L. and D.A. Relman, Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A, 2011. 108 Suppl 1: p. 4554-61.
21. Forslund, K., et al., Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature, 2015. 528(7581): p. 262-266.
22. Lin, H. and S.D. Peddada, Analysis of microbial compositions: a review of normalization and differential abundance analysis. NPJ Biofilms Microbiomes, 2020. 6(1): p. 60.
23. Vandeputte, D., et al., Quantitative microbiome profiling links gut community variation to microbial load. Nature, 2017. 551(7681): p. 507-511.
24. Jian, C., et al., Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLoS One, 2020. 15(1): p. e0227285.
25. Starke, R., V.S. Pylro, and D.K. Morais, 16S rRNA Gene Copy Number Normalization Does Not Provide More Reliable Conclusions in Metataxonomic Surveys. Microb Ecol, 2021. 81(2): p. 535-539.
26. Louca, S., M. Doebeli, and L.W. Parfrey, Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome, 2018. 6(1): p. 41.
27. Aitchison, J., The Statistical Analysis of Compositional Data. Journal of the Royal Statistical Society: Series B (Methodological), 1982. 44(2): p. 139-160.
28. Egozcue, J.J., et al., Isometric logratio transformations for compositional data analysis. Mathematical Geology, 2003. 35(3): p. 279-300.
29. Greenacre, M., Measuring Subcompositional Incoherence. Mathematical Geosciences, 2011. 43(6): p. 681-693.
30. Costea, P.I., et al., A fair comparison. Nat Methods, 2014. 11(4): p. 359.
31. Paulson, J.N., H.C. Bravo, and M. Pop, Reply to: "a fair comparison". Nat Methods, 2014. 11(4): p. 359-60.
32. Fernandes, A.D., et al., ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. Plos One, 2013. 8(7).
33. McMurdie, P.J. and S. Holmes, Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol, 2014. 10(4): p. e1003531.
34. Gorvitovskaia, A., S.P. Holmes, and S.M. Huse, Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. Microbiome, 2016. 4: p. 15.
35. Blaser, M.J., et al., Lessons learned from the prenatal microbiome controversy. Microbiome, 2021. 9(1): p. 8.
36. Aagaard, K., et al., The placenta harbors a unique microbiome. Sci Transl Med, 2014. 6(237): p. 237ra65.
37. Rackaityte, E., et al., Viable bacterial colonization is highly limited in the human intestine in utero. Nat Med, 2020. 26(4): p. 599-607.
38. de Goffau, M.C., et al., Batch effects account for the main findings of an in utero human intestinal bacterial colonization study. Microbiome, 2021. 9(1): p. 6.
39. Hornef, M. and J. Penders, Does a prenatal bacterial microbiota exist? Mucosal Immunol, 2017. 10(3): p. 598-601.
40. Mitchell, C.M., et al., Delivery Mode Affects Stability of Early Infant Gut Microbiota. Cell Rep Med, 2020. 1(9): p. 100156.
41. Yassour, M., et al., Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci Transl Med, 2016. 8(343): p. 343ra81.
42. Nayfach, S., et al., An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res, 2016. 26(11): p. 1612-1625.
43. Korpela, K., et al., Selective maternal seeding and environment shape the human gut microbiome. Genome Res, 2018. 28(4): p. 561-568.
44. Korpela, K., et al., Maternal Fecal Microbiota Transplantation in Cesarean-Born Infants Rapidly Restores Normal Gut Microbial Development: A Proof-of-Concept Study. Cell, 2020. 183(2): p. 324-334 e5.
45. Mueller, N.T., et al., Bacterial Baptism: Scientific, Medical, and Regulatory Issues Raised by Vaginal Seeding of C-Section-Born Babies. J Law Med Ethics, 2019. 47(4): p. 568-578.
46. Lee, S.M., et al., Bacterial colonization factors control specificity and stability of the gut microbiota. Nature, 2013. 501(7467): p. 426-9.
47. Yatsunenko, T., et al., Human gut microbiome viewed across age and geography. Nature, 2012. 486(7402):

p. 222-7.
48. Backhed, F., et al., Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe, 2015. 17(5): p. 690-703.
49. Azad, M.B., 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): p. 15.
50. Hasegawa, K., et al., Household siblings and nasal and fecal microbiota in infants. Pediatr Int, 2017. 59(4): p. 473-481.
51. Lane, A.A., et al., Household composition and the infant fecal microbiome: The INSPIRE study. Am J Phys Anthropol, 2019. 169(3): p. 526-539.
52. Martin, R., et al., Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. PLoS One, 2016. 11(6): p. e0158498.
53. Azad, M.B., et al., Infant gut microbiota and food sensitization: associations in the first year of life. Clin Exp Allergy, 2015. 45(3): p. 632-43.
54. Dzidic, M., et al., Aberrant IgA responses to the gut microbiota during infancy precede asthma and allergy development. J Allergy Clin Immunol, 2017. 139(3): p. 1017-1025 e14.
55. Penders, J., et al., The role of the intestinal microbiota in the development of atopic disorders. Allergy, 2007. 62(11): p. 1223-36.
56. Simonyte Sjodin, K., et al., Temporal and long-term gut microbiota variation in allergic disease: A prospective study from infancy to school age. Allergy, 2019. 74(1): p. 176-185.
57. Stokholm, J., et al., Maturation of the gut microbiome and risk of asthma in childhood. Nat Commun, 2018. 9(1): p. 141.
58. Fujimura, K.E., et al., Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med, 2016. 22(10): p. 1187-1191.
59. Zimmermann, P., et al., Association between the intestinal microbiota and allergic sensitization, eczema, and asthma: A systematic review. J Allergy Clin Immunol, 2019. 143(2): p. 467-485.
60. Arrieta, M.C., et al., Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. J Allergy Clin Immunol, 2018. 142(2): p. 424-434 e10.
61. O'Hara, A.M. and F. Shanahan, The gut flora as a forgotten organ. EMBO Rep, 2006. 7(7): p. 688-93.
62. Sekirov, I., et al., Gut microbiota in health and disease. Physiol Rev, 2010. 90(3): p. 859-904.
63. Vyas, U. and N. Ranganathan, Probiotics, prebiotics, and synbiotics: gut and beyond. Gastroenterol Res Pract, 2012. 2012: p. 872716.
64. Guarner, F. and J.R. Malagelada, Gut flora in health and disease. Lancet, 2003. 361(9356): p. 512-9.
65. Clarke, S.F., et al., The gut microbiota and its relationship to diet and obesity: new insights. Gut Microbes, 2012. 3(3): p. 186-202.
66. Qin, J., et al., A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature, 2012. 490(7418): p. 55-60.
67. Hofer, U. and R.F. Speck, Disturbance of the gut-associated lymphoid tissue is associated with disease progression in chronic HIV infection. Semin Immunopathol, 2009. 31(2): p. 257-66.
68. Ghoshal, U.C., et al., The gut microbiota and irritable bowel syndrome: friend or foe? Int J Inflam, 2012. 2012: p. 151085.
69. Li, Q., et al., Molecular-phylogenetic characterization of the microbiota in ulcerated and non-ulcerated regions in the patients with Crohn's disease. PLoS One, 2012. 7(4): p. e34939.
70. Kolho, K.L., et al., Fecal Microbiota in Pediatric Inflammatory Bowel Disease and Its Relation to Inflammation. Am J Gastroenterol, 2015. 110(6): p. 921-30.
71. Papa, E., et al., Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PLoS One, 2012. 7(6): p. e39242.
72. Andoh, A., 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): p. 370-373.
73. Sokol, H., et al., Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis, 2009. 15(8): p. 1183-9.
74. Swidsinski, A., et al., Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm Bowel Dis, 2008. 14(2): p. 147-61.
75. Wang, W., et al., Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. J Clin Microbiol, 2014. 52(2): p. 398-406.
76. Seksik, P., et al., Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut, 2003. 52(2): p. 237-42.
77. Martinez, C., et al., Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. Am J Gastroenterol, 2008. 103(3): p. 643-8.
78. Halfvarson, J., et al., Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol, 2017. 2: p. 17004.
79. Pascal, V., et al., A microbial signature for Crohn's disease. Gut, 2017. 66(5): p. 813-822.
80. Maghini, D.G., et al., Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. Nat Protoc, 2021. 16(1): p. 458-471.
81. Chistoserdova, L., Recent progress and new challenges in metagenomics for biotechnology. Biotechnol Lett, 2010. 32(10): p. 1351-9.