JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Study of eu- and dysbiosis in the intestinal microbiome of very preterm infants in the PRIMAL study cohort

Dissertation
zur Erlangung des Grades
"Doktor der Naturwissenschaften"

am Fachbereich Biologie
der Johannes Gutenberg-Universität
in Mainz

Jonas Klopp
geboren in Trier

Mainz, den 13.01.2022

Jonas Klopp
Study of eu- and dysbiosis in the intestinal microbiome of very preterm infants in the PRIMAL study cohort
Doctoral Thesis
Reviewers: Prof. Dr. rer. nat. Thomas Hankeln and Prof. Dr. med. Stephan Gehring
Johannes Gutenberg-Universität Mainz
Universitätsmedizin
AG Pädiatrische Immunologie und Infektiologie
Zentrum für Kinder- und Jugendmedizin
Obere Zahlbacher Str. 63
55131 Mainz

Contents

1 Introduction

1.1 The microbiome

To understand the biological meaning of the term microbiome, it is useful to first understand its etymological origins. Microbiome is a combination of two words: microbiota and genome (Backhed, 2005). While the specific expression "microbiome" was originally proposed in the early 2000 by Lederberg et al. to be used "to signify the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space" (Lederberg and McCray, 2001) it has since experienced a shift in meaning. In more recent times - especially since the inception of the Human Microbiome Project (HMP) - the concept has generally been used to encapsulate the entire genetic material that is derived from a microbial community in a given environment (Turnbaugh et al., 2007).

The microbiome of humans consists of many genes from microorganisms such as bacteria, fungi and protozoans as well as viruses. While the taxonomic composition of these different species may vary drastically between individuals at times (Rinninella et al., 2019), relevant genetic pathways provided by microbes are present in most healthy adults (Huttenhower et al., 2012). The metabolic and physiological capacities provided by these genes have major implications for the host's overall health and homeostatic mechanisms in the body (Amon and Sanderson, 2017). Especially in the early stages of microbial colonization, interactions between the microbiota and the host's immune system are essential for facilitating the establishment of effective immune responses in the future life of the host (Gensollen et al., 2016). Later in life, bacteria and other commensals in the gut provide individual metabolic capacities that are either absent or inadequate in the host genome. A non-comprehensive list of examples for genetic contributions of the microbiome is shown in table 1.1.

Tab. 1.1: Example of genetic capabilities provided by the microbiome

Genetic contribution	Reference
aid in defending the host against pathogen infections	(Deng et al., 2018)
breakdown of dietary fiber	(Tasse et al., 2010)
metabolization of xenobiotics	(Maurice et al., 2013)
biosynthesis of vitamins (e.g. Vitamin B_6, Vitamin B_{12}, Vitamin K_1)	(Morowitz et al., 2011)
production of health promoting short-chain fatty acids	(Nicholson et al., 2012)
crosstalk with the immune system	(Zoumpopoulou et al., 2009)
aid in defending the host against pathogen infections	(Deng et al., 2018)

1.1.1 A short history of microbiome research

Even though the term microbiome and its associated extensive research is a rather modern invention, interest in the commensal relationship that humans have with microorganisms reach as far back as the ancient times way before any theory of microbial live forms were postulated. The earliest records of human use of the genetic and metabolic capabilities of microorganisms date back to 10,000 - 7,000 *bc* in the Middle East, Egypt, Greece and Italy when bacteria were used to ferment milk into yogurt and cheese (Gasbarrini et al., 2016). Gaius Plinius Secundus, called Pliny the Elder, even suggested in his encyclopedia *Naturalis Historia* (a treatise concerning "the natural world, or life") the use of fermented milk as a treatment against gastrointestinal infections (Secundus, 0077).

Starting with the work of pioneers and founders of modern microbiology, namely scientist like Robert Koch, Louis Pasteur, Theodor Escherich (Shulman et al., 2007) and others in the late 19^{th} and early 20^{th} century the contextual framework to understand the interaction of microbiota with the human host was boosted to a new level. A scientist who built on this knowledge and contributed notably to the field of early microbiology was the Russian nobel laureate, Élie Metchnikoff.
Metchnikoff proposed an early form of the idea that a healthy gut microbiome might be responsible for the overall health of the human host - an idea he called "orthobiosis". He based his hypotheses on Charles Darwin's evolutionary theory and assumed a form of intra-organismic Darwinism. In his view the administration of beneficial bacteria such as *Bacillus bulgaricus* could positively influence the native colon flora (Podolsky, 2012). While some of Metchnikoffs ideas like the notion that the colon was a useless - if not toxin spreading - remnant of "unfinished" intra-organismic natural selection have fallen out of favor within the scientific community, others

have not. His early attempts of working on a theory concerning the microbial-host interaction have witnessed a major revival with the advent of high throughput sequencing techniques in the early 2000s.

Interest in this kind of holistic study between the microbiome and diseases had temporarily lost momentum due to the efficacy of antibiotics, which rendered many problems in that field seemingly solved. After a hiatus of a couple decades, two main reasons shifted this paradigm and paved the way for a return of large scale microbiome studies:

1. Antibiotic resistance became a major concern for global health (Dyar et al., 2017).
2. Following the revolutionary advances in the field of sequencing nucleic acids dubbed Next-Generation-Sequencing, the cost to analyze huge amounts of DNA has plummeted (see fig. 1.1) (Mardis, 2011). This development allowed scientists to widen their research questions which could be experimentally answered and opened up new and interesting opportunities to study the microbiome in its seemingly endless complexity.

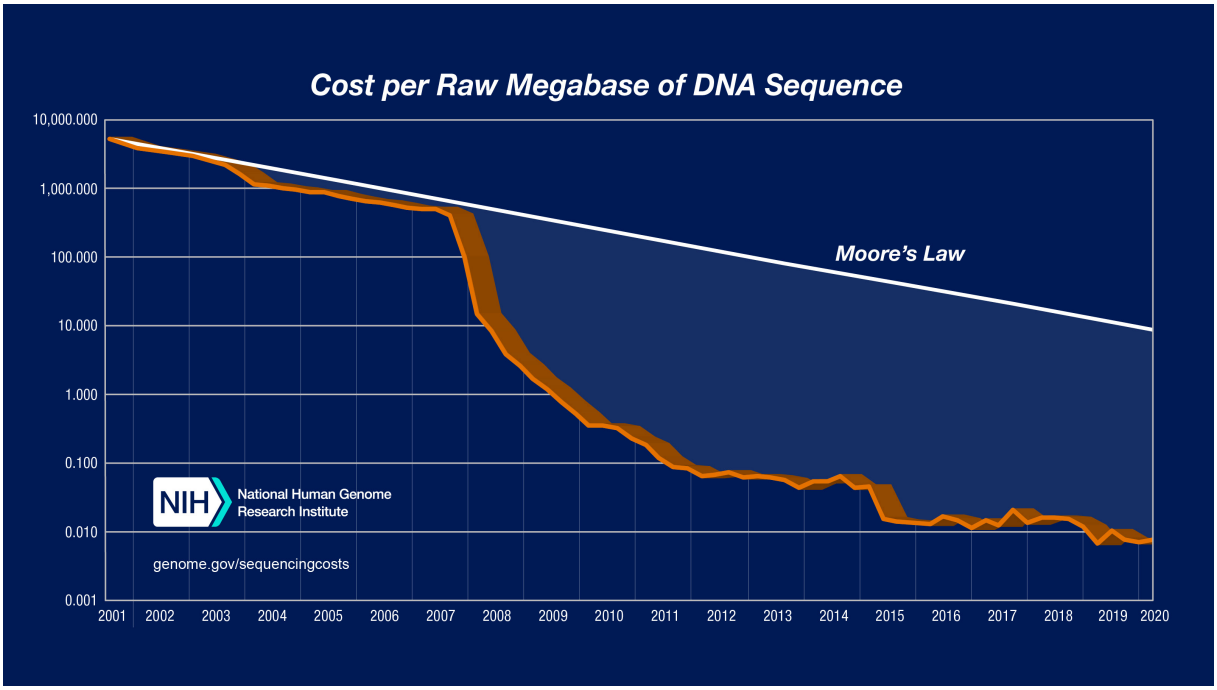


Fig. 1.1: Cost in US $ of DNA sequencing has been reduced significantly in the last 20 years since the beginning of the 21th century. The straight line in the graph refers to "Moore's law" (G. Moore, 1998). Moore's original observation was that the number of transistors in a dense integrated circuit (and thus the computing power) doubles about every two years. This reference goes to show that the cost of sequencing dropped at an even higher exponential rate than the already impressive rise in computing power over the last 20 years since 2000. Graphic from (Wetterstrand, 2020)

In the tradition of groundbreaking international research cooperations like the Human Genome Project (International Human Genome Sequencing Consortium, 2001) or the ENCODE project (The ENCODE Project Consortium, 2004) a new project was founded in 2007 - The Human Microbiome Project (HMP). This international, interdisciplinary effort set out to study the role

that microbes and their respective genetic make up play in human health and disease. One major aim was to describe on a large scale the entirety of given microbiomes on the skin, in the gut and other niches thus combining the fields of ecological and medical microbiology (Turnbaugh et al., 2007). The main idea that guided the HMP was the notion that we, as humans, have to be considered as a 'supraorganism'. That is, a commensal community consisting of human cells, microbial cells like bacteria, protozoa, fungi and viruses. As a conceptual follow up to the Human Genome project, the HMP wanted to explore genes that are essential for our physiology but have been evolved by our symbiotic partners rather than ourselves (Gill et al., 2006).

This first phase of the HMP (also called HMP1) was followed by a second phase called Integrative Human Microbiome Project (iHMP) (The NIH HMP Working Group et al., 2009). The second phase aimed at providing integrated longitudinal datasets that described the human microbiome at the major body sites through an array of "omics" approaches. In 2019 the NIH Human Microbiome Portfolio Analysis Team published a 10 year review of the results that have been produced by the NIH's overall 728,000,000 $ investment in microbiome research (NIH Human Microbiome Portfolio Analysis Team, 2019). The report lists tremendous advances in the field such as the development of a wide array of bioinformatical algorithms for the analyses of all kinds of multi-omics microbiome datasets, new strategies to culture previously "unclutivable" bacteria, newly sequenced whole genomes of thousands of bacteria and - among many others - a more detailed view on the different interactions of the microbiome with the human host in health and disease.

1.1.2 The infant gut microbiome

As presented in the prior section, great progress has been made in the field of microbiomes in adults. However, the microbiome of children and especially infants poses distinct and different problems. Particularly the initial seeding and timing of the development of the early infant gut microbiome remain understudied (R. E. Moore and Townsend, 2019). The current dogma in the field is that a neonate's development *in utero* takes place in a sterile environment. This hypothesis has already been proposed more than 100 years ago (Kendall, 1918) and still largely holds true today (de Goffau et al., 2019; Lauder et al., 2016) although there is considerable controversy (Blaser et al., 2021). A major source of this controversy in the discussion around the "sterile-womb" hypothesis stems from the fact that some scientist seem to have found evidence of bacteria in meconium samples and even in utero (Stinson, Boyce, et al., 2019). Meconium is the first pass stool of newborns and primarily comprised of intestinal epithelium and material that has been ingested *in utero* (Jiménez et al., 2008). While prebirth colonization might explain bacterial DNA in meconium, another possible explanation for the phenomena is in line with the "sterile-womb" hypothesis and postulates a colonization of infant guts with microbes right after and even during birth. Thus the meconium could already be "contaminated" by external bacterial contact (Perez-Muñoz et al., 2017).

Either way, infants start their life without a well diversified microbiome. The ecological niche, which is the infant gut, is soon readily occupied and colonized by microbial life forms. The process is commonly understood as an ecological succession event with the main driver being niche-based competition but - especially in the very early stages - stochastic effects also play a major role (Costello et al., 2012). Main drivers that have been identified of the non-random assemblage of the microbiome are breastfeeding, antibiotic use and delivery mode (Bäckhed et al., 2015). Strong evidence for the ecology-based explanation model is the existence of human milk oligosaccharides (HMOs). HMOs are a group of approximately 200 unique carbohydrate polymers consisting of 3-22 monosaccharides only found in breast milk of *Homo sapiens* (German et al., 2008). It has long been speculated that typical colonizers of the infant gut like *Bifidobacterium* and HMOs have co-evolved over the course of millions of years of human evolution, since these oligosaccharides are not digestible by human infants and can instead solely be broken down by specific members of *Bifidobacterium* and other bacteria generally regarded as beneficial to host-health (Duranti et al., 2019; F. Turroni et al., 2018; Katayama, 2016; Garrido et al., 2013).

A landmark paper investigating the temporal development of the gut microbiome in infants after the third month of life has been published in 2018 in the context of "The Environmental Determinants of Diabetes in the Young" (TEDDY) study (Stewart et al., 2018). In this study the authors collected multi-center-sourced microbiome data from 903 children aged 3 to 46 months. They sequenced multiple time points per child ($n_{16s\ rRNA} = 12005$, $n_{metagenomic} = 10867$) to identify the temporal progression of the emerging microbiome. Beginning right after the first three months of life starts a developmental phase in which a comparatively low alpha-diversity is found within the gut and *Actinobacteria* especially a host of *Bifidobcateria* dominate. The time between months 15 and 30 is a transitional phase in which the diversity in the gut steeply increases with the introduction of many bacterial species. Finally, the child gut enters a stable phase at around 3 years of age where a durable microbial composition has been established. Key members of the gut microbiome at this stage are primarily from the phyla *Actinobacter*, *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Verrucomicrobia*. The primary driver of microbiome structure identified was the consumption of human breast milk which led - in line with the HMO co-evolution hypothesis - to a drastic increase in the relative abundance of *Bifidobacterium* species.

Stark contrasts in the earliest microbes that infants encounter have also been described for different delivery modes. Vaginal delivery will expose the children to a wide array of microbes found in women's vaginal microbiomes such as *Lactobacillus*, *Prevotella* or *Sneathia spp.*. Caesarean Section, on the other hand, will primarily introduce bacterial genera typically found on the skin microbiome such as *Staphylococcus*, *Corynebacterium* and *Propionibacterium spp*. (Dominguez-Bello et al., 2010). Fig. 1.2 shows an aggregated summary of the results of various microbiome studies in infants displaying the differences in microbes between vaginally and via c-section delivered infants. To account for this mismatch in initial bacterial species the practice of "vaginal seeding" or "bacterial baptism" has been adopted by many pediatricians (Simpson, 2018). In this procedure a thin gauze is incubated in a woman's vagina one hour before a scheduled C-Section. Immediately after birth the mouth, face and body of the newborn is rubbed with the

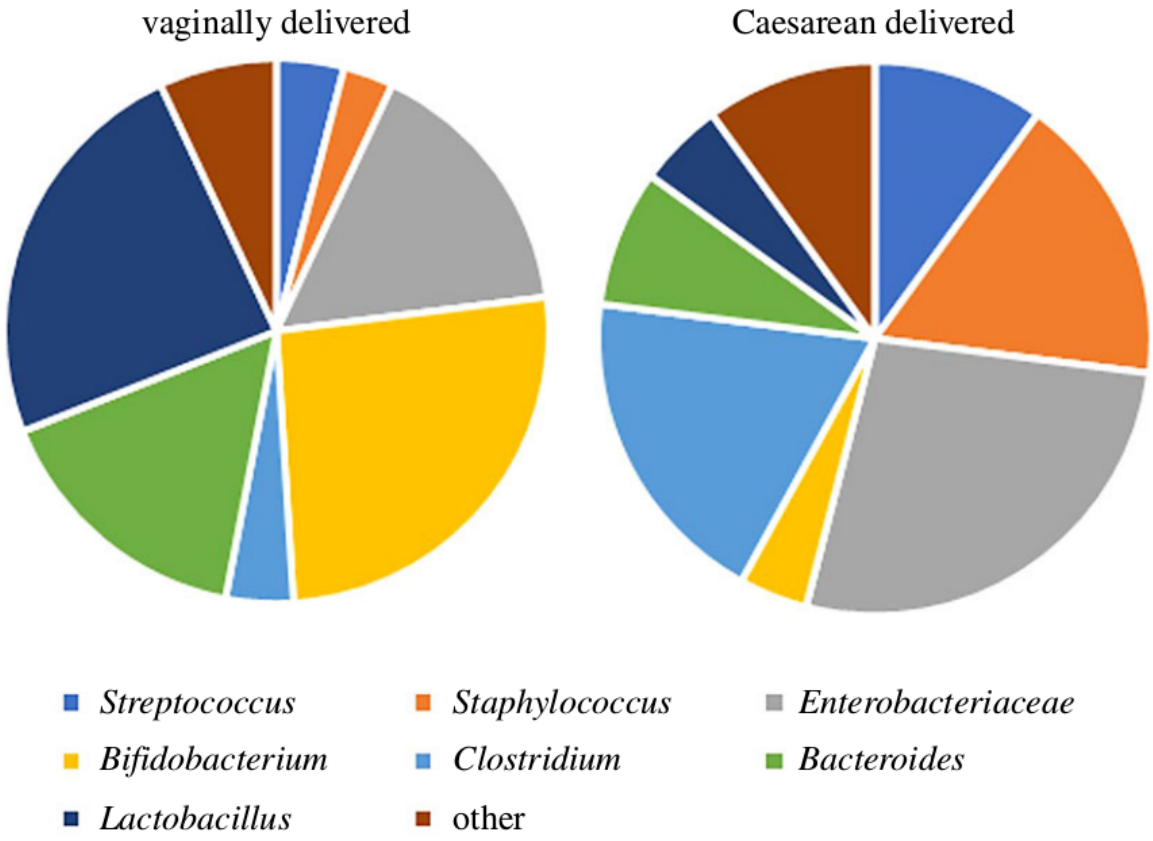


Fig. 1.2: Reproduced from (R. E. Moore and Townsend, 2019). Study result aggregation: Comparison of the gut microbiome in vaginally born infants and infants born via Caesarean section

vaginal secretions to imitate the microbial exposure during a vaginal birth (Mueller et al., 2019). Opponents of this practice have noted that without the availability of rigorously conducted clinical trials that show the efficacy, the potential risk should invoke a lot more reluctance in adopting it (Eschenbach, 2018). Potential negative health outcomes for the infant include infections with group B *Streptococcus*, *herpes simplex* or *Chlamydia trachomatis* (Cunnington et al., 2016). A causal role of changes in the vaginal microbiota in the context of C-Section is also being discussed (Kervinen et al., 2019), which leaves the question whether the vaginal microbiome during an artificial birth process is even appropriate for seeding of the newborn (Stinson, Payne, et al., 2018).

Another crucial co-variate in the explanation of microbial differences in the infant gut is the geographical location of the infant and mother. In a study from 2019 Kimberly A. Lackey and colleagues investigated the community structure in gut microbiomes of 377 infants from all around the world (Lackey et al., 2019). Children were included from rural and urban Ethiopia, Kenya, Gambia, Ghana through Peru, Spain, Sweden and the United States of America in an effort to identify what a "normal" infant microbiome looks like. Their data suggests that there are core bacterial genera present in all infant gut microbiomes irrespective of country of origin including the already mentioned *Bifidobacterium*, *Lactobacillus* and *Bacteroides*. Substantial differences in the relative abundance of these taxa could be noted with *Bacteroides* and *Clostridium senus stricto*

1 being more promiment in Europe and the USA compared to rural Africa and the oppsite was found for *Lactobacillus* (see fig. 1.3). Interestingly, the overall diversity within the microbiomes shows no significant changes between the continents.

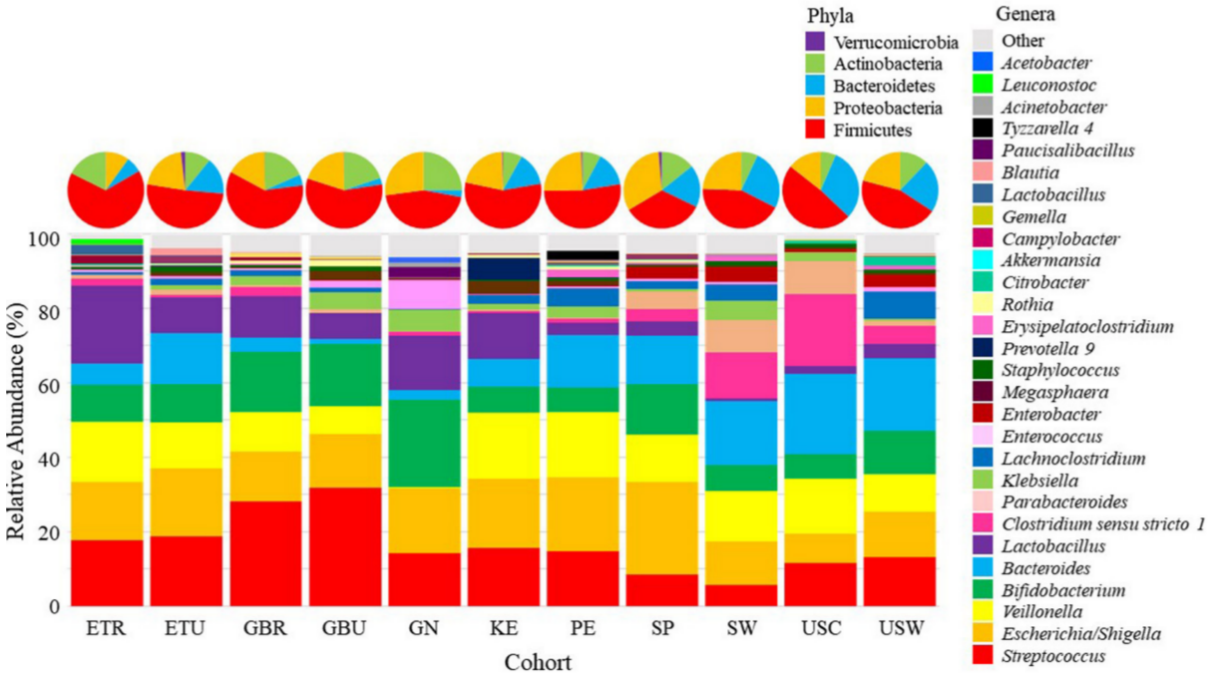


Fig. 1.3: Reproduced from (Lackey et al., 2019). Distribution of the 10 most prominent bacterial genera in infant gut microbiomes around the world: ETR, rural Ethiopia; ETU, urban Ethiopia; GBR, rural Gambia; GBU, urban Gambia; GN, Ghana; KE, Kenya; SP, Spain; SW, Sweden; PE, Peru; USC, California (United States); USW, Washington (United States)

1.1.3 The (very) preterm infant gut microbiome

An infant is classified as being prematurely born if the birth occurs between the 20th and 37th week of pregnancy (Keelan and Newnham, 2017). Premature births are responsible for at least 35% of global neonatal deaths every year. While the burden of disease is higher in the developing world (ca. 15%), the rates in developed nations are still around 1 in 10 births (Keelan and Newnham, 2017). In Germany around 8.6% of births occur prematurely (as of 2017) (Berger et al., 2019). In recent years the scientific community has directed its focus to the role of the microbiome in preterm birth (Chu et al., 2018). Evolutionarily speaking, surviving preterm infants are a novel development in the human history with unique challenges such as immature immune systems, frequent exposure to antibiotics and prolonged hospital stays right at the time when the microbiome is first forming (Underwood et al., 2020).

In the Norwegian Microbiota Study (NoMIC) the gut colonization and health outcomes of 519 children (160 preterm) were studied. The authors concluded that special factors unique to very preterm infants lead to their microbiome being different from that of term neonates - even after controlling for co-variates such as breastfeeding, mode of delivery and duration of

antibiotic exposure (Dahl et al., 2018). They found that in preterm infants the bacterial diversity was reduced compared to their term counterparts and the relative abundance of *Proteobacteria* and *Enterococcus* was increased (Dahl et al., 2018). An earlier study by La Rosa et al. showed a non-random pattern in which the gut of premature infants is colonized. In the first days of life, starting off with somewhat random pioneer species, the gut starts to transition through three distinct phases dominated by *Bacilli*, *Gammaproteobacteria* and finally a combination of *Gammaproteobacteria* and *Clostridia*, respectively, during the course of the first 40 days (La Rosa et al., 2014). The speed in which these phases are passed are significantly influenced by the birth mode (Ho et al., 2018).

On the genus level one of the main differences in the microbiome of term and preterm infants is the relative abundance of *Bifidobacterium*. These higher abundances of *Bifidobacteriaceae* are also associated with higher concentrations of acetate and a reduced ph-level in the stool of term infants which is in turn associated with better host health (Tauchi et al., 2019).

1.1.4 Neonatal dysbiosis - the 'unhealthy' microbiome

Conventionally the microbiome has been described in terms of homeostasis (i.e. a healthy microbiome) and dysbiosis (i.e. a diseased microbiome) (Hooks and O'Malley, 2017). These states are usually defined by the identification of relative changes in the structure between healthy and sick individuals (Petersen and Round, 2014). Dysbiosis is then broadly defined as: "[...] Any change to the composition of resident commensal communities relative to the community found in healthy individuals (Petersen and Round, 2014)." Many critics have noted that such a definition entails certain biases (Brüssow, 2020). For example: By comparing the microbiome of sick and healthy individuals the implicit assumption is, that any significant change in the microbiomes will be linked to the disease state, either causally or consequentially (Hooks and O'Malley, 2017). There are, however, a multitude of factors not necessarily linked to any disease state including temporal changes, environmental factors, diet choices etc. These have a modulating effect on the human microbiome and might therefore confound results in comparative microbiome studies if not adequately controlled for (H. Yang, 2019). Many studies have reported significant differences between the microbiomes of diseased and healthy control individuals in a wide array of diseases (DeGruttola et al., 2016; Wilkins et al., 2019).

In their paper, "Dysbiosis is not an answer", Scott W. Olesen and Eric J. Alm argue that the very concept of a vague dysbiosis definition might actually hinder the progress of microbiome research into a translational discipline. Their reasoning is that the concept relies on an undefined and broad notion of microbial imbalance lacking scientific verifiability. The authors claim that observing the fact that ill and healthy people have different microbiomes 'is no longer a novel or useful observation' and that 'we need to show that differences in the microbiota can be used to predict or ameliorate disease, and not just show that differences exist' (Olesen and Alm, 2016).

Despite the criticism, the concept of dysbiosis can be valuable as has been demonstrated through several studies. For complications like necrotizing enterocolitis (NEC) the advent of certain specific microbiota changes - such as a markable increase in gram-negative bacteria like *Enterobacteriaceae*, *Vibrionaceae*, and *Pseudomonadacea* - predates the manifestation of the disease by days to weeks (Pammi et al., 2017). Thus these changes are debated as potential biomarkers for risk-stratification of the illness in preterm infants (Fundora et al., 2020). Another example illustrating the gain of actionable insights is given by a randomized, double-blind, placebo-controlled trial including 4,556 infants in rural India that showed a significant reduction of sepsis and deaths in an infant population when administered a synbiotic containing *Lactobacillus plantarum* ATCC-202195 alleviating the dysbiotic state in the patients (Panigrahi, 2017).

A different approach of defining dysbiosis attempts to first extensively describe a normal state of a given microbiome niche. Many investigators including the Human Microbiome Project Consortium (The NIH HMP Working Group et al., 2009) have put immense effort and resources into mapping the normal landscape of microbiomes in healthy adults. This approach aims at identifying and characterizing the major patterns in the space of global human microbiomes. Attempts have been made to cluster the vast heterogeneity of microbiomes into enterotypes to facilitate the better understanding of normality in the space of human-microbiota relationship (MetaHIT Consortium (additional members) et al., 2011; Costea et al., 2018). This normal state is often referred to as homeostasis, eubiosis or normobiosis (Hooks and O'Malley, 2017). Based on these states, major deviations can be assigned the label dysbiotic even if they are not directly linked to a specific disease. Thus it might be possible that the current health status of a person may not be equivalent to a healthy microbiome. Sonnenburg et al. for example argue that a western lifestyle including a diet of low microbiota-accessible carbohydrates may select for a microbiome that is distinctly different from that of our ancestors. Given the possibility that co-evolutionary forces shaped our genome and our microbiome in tandem, disturbances in composition of the latter may lead to compatibility issues in the form of disease (E. D. Sonnenburg and J. L. Sonnenburg, 2014).

Over the years, common unifying traits in almost all abnormal microbiota definitions have been postulated (see tab. 1.2): First, an overall loss of biodiversity in the microbiome. Second, a loss of keystone beneficial taxa. Third, over proportional abundance of pathobionts. Some studies have proposed additional considerations such as a shift in the metabolic capacity of the metagenome (Vangay et al., 2015).

Tab. 1.2: Recurrent features of dysbiosis in the literature

(Petersen and Round, 2014)	(Vangay et al., 2015)	(Levy et al., 2017)
loss of overall diversity	overall loss of biodiversity	loss of alphadiversity
loss of beneficial microbial organisms	loss of keystone taxa	loss of commensal
expansion of pathobionts	blooms of pathogens and pathobionts	bloom of pathobionts
	shifts in metabolic capacity	

1.1.5 Antibiotic resistance genes in the infant gut

Increasing abundance of antibiotic resistances in human pathogens is a major global health concern. In the earlier days of antibiotic treatments it was seen as beneficial to use broad-spectrum antibiotics to have the highest chance of targeting the pathogenic agent (Fischbach and C. T. Walsh, 2009). Emerging evidence on the proliferation of antibiotic resistances genes (ARGs) lead to the development of the concept of "Antibiotic stewardship", which is a loosely defined set of guidelines for using antibiotics responsibly (Dyar et al., 2017). Some of these guidelines include:

- Making accurate diagnoses (instead of using broad-spectrum antibiotics)
- Regularly reviewing the need for therapy
- Monitoring antimicrobial use and resistance

Instead of broad-spectrum antibiotics some groups have developed compounds to narrowly target the growth of specific taxa of bacteria such as *Staphylococcus* (Liu et al., 2008). While this might be a promising route for the future, the use of broad-spectrum antibiotics, especially in neonatology and pediatrics, still remains crucial.

A common complication in the neonatal intensive care units is severe bacterial sepsis caused by bacteria such as *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Escherichia coli* and among many others several *Enterobacter* and *Pseudomonas* species (Sharma, 2013). In light of the imminent threat to a child due to a sepsis combined with the lack of reliable and fast assessments of the causative agents, physicians often opt to use empirical broad-spectrum antibiotics at the first sign of a suspected sepsis (N. Tripathi et al., 2012). This in turn leads to an increase in the evolutionary pressure on all microorganisms in the host to adapt to broad-spectrum antimicrobial agents. A study by Forslund et al. from the year 2014 used metagenomic sequencing data to assess the antibiotic resistome (the entirety of ARGs within an environment) of human gut microbes.

Human fecal samples from 832 individuals from 10 different countries were analyzed and an increase of ARGs caused by human activity (antibiotic usage in medical and food production context) could be demonstrated (Forslund et al., 2014). There is some evidence that this increase is especially prominent in the gastrointestinal tracts of extremely preterm infants compared to healthy infants, in large part due to the extensive use of broad-spectrum antibiotics in this population (Gasparrini et al., 2019). This reservoir of ARGs potentially offers a wide variety of mobilizable genes to be transfered from harmless commensal bacterial species to pathogens (M. O. A. Sommer et al., 2009) and thereby worsen the chance of treating vulnerable patients like infants.

Stecher et al. used a mouse model to investigate if and to what extend horizontal gene transfer (HGT) happens in the mammalian gut *in vivo* (Stecher et al., 2012). The authors could show that under normal physiological conditions the commensal microbiota of the mammalian host was able to inhibit contact-dependent conjunction between *Enterobacteriaceae*, which is one common way of HGT between species. However, if the intestinal tract was inflamed due to pathogentic activity, the conjugative transfer of genes reached extraordinary high rates opening up the opportunity for the pathogens to obtain genes to increase their evolutionary fitness such as ARGs from the surrounding microbiome. Another study by Karami et al. followed an infant from birth up to one year of age, sampling his gut microbiota and β-Lactamases genes (Karami et al., 2007). The authors found that in the first month of life the infant's gut harbored two distinct strains of *E. Coli*, dubbed 29A and 29B, being resistant to ampicillin and being susceptible to it, respectively. After the infant suffered a urinary tract infection, an ampicillin/amoxicillin treatment was used to kill off the pathogens. During this process *E. Coli* strain 29B acquired a resistance gene carrying plasmid, pNK29, from strain 29A in the gut and thus a resistance against β-Lactam based antibiotics. Further, a mutation in an associated promoter region of the resistance gene lead to an even higher level of resistance. Karami and colleagues then verified that it was in fact the same gene and reproduced the horizontal gene transfer between the strains *in vitro*, showing impressively how HGT and the use of broad-spectrum antibiotics in infants can lead to a spread of ARGs in the microbiome.

Few studies have so far investigated the extent to which ARGs are present in the preterm infant gut. Notably one study by Rose et al. investigated the resistome of 11 healthy preterm infants and found tremendous amounts and diversity of ARGs (Rose et al., 2017). Earlier studies also showed immense antibiotic resistance capacity in the gut microbiome of infants (Molly K. Gibson et al., 2016). Using metagenomic recombinant libraries in *E.Coli* followed by sequencing Moore et al. demonstrated that resistances against 14 of 18 tested antibiotics were present, functionally active and transferable in the gut microbiome of 22 tested infants (A. M. Moore et al., 2013). In light of the enormous potential for harm caused by a readily accessible reservoir of ARGs in the developing infant gut, a comprehensive study of the genetic elements that convey resistances is paramount to tackle these problems in the future. Metagenomic sequencing techniques combined with amplicon based marker gene analyses are promising ways to fill in

the gaps in the current understanding of infant gut resistome dynamics (Molly K Gibson et al., 2015).

1.2 Microbiota altering techniques - Probiotics and Fecal microbiota transplantation

The discovery of discrepancies between the microbiome structures of different medical states of interest has renewed the appreciation of methods for a modulation of these communities. 'Probiotics' is often used as an umbrella term for the tool with which these changes are meant to be conferred. The currently accepted definition of a Probiotic is: "Live microorganisms that, when administered in adequate amounts, confer a health benefit on the host" (C. Hill et al., 2014). As described in 1.1.1 the history of the use of bacterial species believed to beneficially influence is tightly linked to the study of the bacterial-host symbiosis. Early attempts in understanding beneficial effects of bacterial colonizers have already given rise to the use of fermented products to elicit desired positive changes in the well-being of patients (Gasbarrini et al., 2016). However, these rudimentary interventions lack rigorous scientific frameworks and are explicitly excluded from the current definition of Probiotics, which only includes the administration of defined entities (C. Hill et al., 2014).

The immense complexity of interactions between the host and the microbiome, external parameters as well as interactions within the microbiome itself make mechanistic studies of probiotics extremely resource intensive. A quantitative research approach has only been possible since the advent of the Next-Generation sequencing techniques and thus the mechanistic insights into the modus operandi of probiotics are still limited (Suez, 2019). Although large scale scientific studies for their efficacy are still pending, the global probiotics market is already valued at around USD 46.20 Billion and is predicted to reach USD 75.90 Billion by the year 2026 (Zion Market, 2020). Among physicians and especially gastroenterologists prescribing non-strain-specific probiotics is a widely used practice for the promotion of general bowl health - often on patient request (Draper et al., 2017).

While physicians' practice patterns in the context of probiotics do not consistently match the evidence-based recommendations on the subject, most physicians believe that probiotics do not confer harm and can therefore be administered on request even without a clear indication (Williams et al., 2010). A review by P. Cani and colleagues from 2020 concludes that, while probiotics hold great promises for the treatment of a diverse range of illnesses from bowel diseases to psychopathologies, current studies in adults fail to report any impact on the microbiota (Wieërs et al., 2020). The authors speculate that any beneficial effects seen through the use of probiotics are potentially due to a direct delivery of genes or metabolites to challenged microbiota or an immediate interaction with the epithelial and immune cells of the host rather than a colonization of the gut. This assumption is further strengthened by findings that mice and humans posses

individualized colonization resistance to many traditional probiotic strains (Zmora et al., 2018).

Despite these sobering results in regards to the use of probiotics in adults, an alternative option to alter the gut microbiome has been approved by the U.S. Food and Drug Administration (FDA) in 2013: Fecal microbiota transplantation (FMT) for the treatment of recurrent *Clostridium difficile* infection (FDA, 2013). During this procedure a stool sample of approximately 50 g from a healthy donor is transferred into the gastrointestinal tract of a diseased patient to replace the native microbiota with the donor microbiome (J.-W. Wang et al., 2019). A prominent randomized controlled trial published in the New England Journal of Medicine comparing the effects of FMT and a conventional treatment with vancomycin in patients with recurrent *C. difficile* infection found a 81% resolution in the FMT group compared to only 31% in the conventional treatment group ($p<0.001$) (van Nood et al., 2013). These results led to the early termination of the trial after the interim analyses, since it was regarded as unethical to continue and withhold FMT from the control group. However, the wide adoption of FMT to treat *C. difficile* infections has been met with harsh resistance. Critics argue that the longterm consequences of microbiome alterations on this scale and in this 'black-box' manner bears great risks (Tan and S. Johnson, 2019). More and more potential indications for FMT are being discussed, such as Inflammatory Bowel Disease (IBD) like Crohn's disease (CD) and ulcerative colitis (UC), multiple sclerosis (MS), Parkinson disease and even obesity (Vindigni and Surawicz, 2017). The risk/reward ratio of such endeavors remains a pressing topic, since the studies supporting such indications are far less conclusive than in the case of *C. difficile* infections and the potential adverse and long-term effects of FMT numerous (see tab. 1.3) (Vindigni and Surawicz, 2017; Tan and S. Johnson, 2019)

Tab. 1.3: Potential adverse effects of FMT (modified from (Vindigni and Surawicz, 2017))

Common	Severe	Long-term
Nausea/vomiting	Infection	Persistent undesirable changes in the microbiome
Abdominal pain	Sepsis	Risk of chronic disease
Bloating	Inflammatory bowel disease flare	
Diarrhea	Postinfectious irritable bowel syndrome	
Fever		

1.2.1 Probiotic use in infants

Contrary to the comparatively stable adult microbiome configuration, the early infant microbiome is highly susceptible to changes through probiotics. In a small placebo controlled cohort of 66 breastfed, full term infants Frese et al. studied the persistence of the probiotic species *Bifidobac-*

terium longum subsp. *infantis* (Frese et al., 2017). 34 of the 66 infants received 1.8–2.8 $\times$ 10^{10} colony forming units (CFU) of *B. infantis* mixed with breast milk for 21 days following the first week of life, after the safety and tolerability of the probiotic had been established (Smilowitz et al., 2017). The levels of *B. infantis* in the treatment group remained elevated in fecal samples compared to the controls for over 30 days and up to one year after the supplementation had ended. Further indication for colonization of the gut was the significantly decreased concentration of human milk oligosaccharides (HMO) in the treatment group indicating metabolic activity of the probiotic strain.

While the number of high quality studies investigating changes in the microbiome of infants upon receiving probiotics remain relatively sparse, studies looking at clinical outcomes are numerous. The overall consensus in the field is that probiotics are a generally safe, cheap and effective way to address common potential clinical complications that can arise in infants and especially preterm infants such as feeding intolerance, necrotizing enterocolitis (NEC) or late-onset sepsis (LOS) (Athalye-Jape and Patole, 2019). One noteworthy exception to this overarching assessment is seen in the PiPS study (Costeloe et al., 2016). In this UK based, multicentre, randomized placebo-controlled phase 3 clinical trial 1315 preterm infants were included. Half of the participants received a daily probiotic supplementation of *Bifidobacterium breve* BBG-001 for the time of their stay in hospital. The scientists were not able to detect significant changes in the occurrence of NEC or LOS. Many possible explanations and criticism (e.g. statistical problems, low dosage of probiotics, random biological variation) have been raised in regards to the results of the study (McKinlay et al., 2016). The authors themselves state that one major limitation was the cross-colonization of the placebo group with the probiotic strain which may have masked any beneficial effect that might have occurred in the treatment group. While *B. breve* could be detected in 84% of the probiotic group, the same strain was also identified in 49% of infants in the placebo group. However, studies with negative findings in regards to the use of probiotics in infants remain outliers. In fact, a very large number of meta-analyses (Athalye-Jape, Deshpande, et al., 2014; Sawh et al., 2016; Rao et al., 2016; Olsen et al., 2016; Dermyshi et al., 2017; Deshpande et al., 2017) of randomized controlled trials and observational studies with a collective tens of thousand of infants have revealed the effects of probiotic interventions to be so great, that some researchers even see them as "the miracle cure of this [21st] century [...] in preterm infants" and proclaim the current day and age to be a "golden age of probiotics" in neonatology (Dermyshi et al., 2017).

As the clinical studies seem to point in the direction of a widespread adaptation of probiotic interventions in infants, many questions still remain unanswered:

- There have been a few sporadic reports of bacteremia caused by probiotic supplementation (Bertelli et al., 2015; Esaiassen, Cavanagh, et al., 2016). Although this issue does not appear to happen very frequently, questions of how to implement an effective surveillance strategy should be answered.
- The optimal dosage, timing and selection of probiotic agents (Olsen et al., 2016)

- (long-term) effects on the microbiome structure of infants (Aceti et al., 2018)

Overall, there is a need to deepen the theoretical knowledge base of probiotic use and its interactions with the developing microbiome to lead the field out of a black-box modus operandi into a phase of evidence-based probiotic intervention design. (Aceti et al., 2018).

1.3 Priming immunity at the beginning of life - PRIMAL

To address the open questions in the field of neonatal microbiome development and host interactions the "**Pr**iming **im**munity **a**t the beginning of **l**ife" (PRIMAL) study was initiated as a Germany-wide, large scale research endeavor (Marißen et al., 2019). PRIMAL is a multicentre, prospective, randomized, double blind, placebo controlled trial to study the effects of probiotic supplementation on very preterm infants (VPI, gestational age 28-32 weeks). Over 650 VPIs will be included and microbiome samples (in the form of stool samples) will be collect from all participants at day 0, 28 and 365 of life. For the first 28 days of life the infants will receive a daily probiotic food supplement (Probactiol ®infantis) consisting of *Bifidobacterium animalis* subsp. lactis (BB-12), *Bifidobacterium infantis* and *Lactobacillus acidophilus* (La-5) (each as $1.5*10^9$ CFU) or a placebo. For a subset of the study cohort associated stool samples from the mother will be collected to enable the comparison of mother-infant microbiome structure.

PRIMAL is designed to gain a deeper understanding of the interactions between the developing microbiome and an infant's immune system. To investigate this, one needs to first identify the early microbiome structure of VPI in great detail. This will be accomplished using conventional, culture based methods as well as Next-Generation sequencing techniques (16s rRNA and metagenomic whole genome sequencing, see 1.4.2 and 1.4.3). With the high resolution data obtained from culturing and sequencing a multitude of questions can be addressed:

- Influence of probiotic supplementation on the VPI microbiome
- Deeper understanding of the concept of microbial dysbiosis in very early life
- Interactions between the VPI immune system and the developing microbiome

Besides the core objective of high resolution microbiome analysis of VPI upon probiotic intervention, PRIMAL also includes several basic science projects such as a VPI microbiome mouse model, immunological studies of the dynamic interactions of bacteria and leukocytes and advanced immunological phenotyping.

1.4 Bacterial Identification

The exact identification of commensal and pathogenic bacteria constitute the backbone of modern microbiology. Since the inception of microbiology, a number of methods have been developed to identify bacteria. In the early days, the microscope was the main tool to detect differences and similarities in the phenotype of different bacteria and thus classifying them into groups (Woo et al., 2008). From metabolic activity to surface structure, nearly every feature of bacterial species has been used to differentiate specimens. Since the Next-Generation sequencing revolution, new nucleic acid based methods have been utilized to allow the large scale characterization of microbes based on their genetic material. In the following paragraphs the most widely used methods for bacterial identification will be introduced.

1.4.1 KRINKO and conventional, culture-based microbiology

The German committee for hospital hygiene and the prevention of infectious disease (Kommission für Krankenhaushygiene und Infektionsprävention, KRINKO) of the Robert Koch Institute regularly issues recommendations specifying what kind of bacteria are deemed especially harmful pathogens and therefore indicate a special surveillance. Often times not entire genera of bacteria are pathogeneic but instead only certain members of a specific genus, as is the case with e.g. *Clostridioides* (Gibbs, 2009). KRINKO recommendations pay extra attention to bacteria that show a resistance to multiple of the four main groups of regularly used antibiotics. These are *cephalosporins*, *acylureidopenicillins*, *fluorochinolone* and *carbapenems* (Koch-Institut, 2013). For gram-negative bacteria, the number of resistances to different antibiotic groups is classified by the MRGN system (multi-resistant gram-negative bacteria) by the labels 2MRGN to 4MRGN. Usually, in order to identify the previously mentioned bacteria, conventional microbiological methods are being used. From an anal swipe or a stool sample, living bacteria are cultured on agar plates, identified via methods such as MALDI-TOF (see 1.4.2) or "Bunte Reihe" and tested for antibiotic resistances. This method yields data about single species of bacteria in regards to their presence, viability and concrete resistances. The method is thus a feasible way to identify predefined, selected pathogens in a low throughput manner.

1.4.2 Non-sequencing based bacterial identification

A commonly used method that identifies bacterial isolates without the help of sequencing is the *Matrix Assisted Laser Desorption/Ionization Mass Spectrometry with Time-of-Flight detector* (MALDI-TOF MS) (Dingle and Butler-Wu, 2013). Using this technique bigger molecules or entire cells are subjected to different steps (Karas and Krüger, 2003):

1. The research object is fixated to a carrier with the help of special matrix-materials
2. A pulsing laser shot on the object detaches molecules in a process known as "ablation"

3. The detached molecules are then ionized by the discharge or absorption of protons
4. The ionized objects are accelerated in a mass spectrometer and their "time of flight" is recorded
5. The measurement taken from the object of interest is compared to a reference database of matrix spectral measurements to assess the most likely identity of the object in question - in this case a bacterial species

1.4.3 Sequencing based bacterial identification

Apart from the conventional methods of bacterial identification, new methods based on next-generation sequencing techniques have emerged in the last two decades and are rapidly becoming the standard method for high throughput characterization of entire microbiomes. There are two main innovations driving this development:

1. Amplicon based sequencing techniques
2. Whole metagenomic shotgun sequencing

Amplicon based sequencing techniques rely on a two step process: First, a genomic region of interest is amplified using polymerase chain reaction (PCR). Second, the enriched DNA fragments are sequenced. To be useful as a tool for bacterial identification this method is depended on the availability of so called "universal primers". The genomic region, that the universal primers target, has to meet certain criteria:

- it needs to have conserved regions in most bacterial species for the primers to bind (universality)
- in between the conserved regions should reside a highly variable region that provides information for discriminatory bacterial classification
- the entire region should be ideally short enough to be readily sequenced by common short read sequencing machines

Such a genomic region has been identified for the bacterial kingdom in the late 1970s by Carl R. Woese and George E. Fox: The 16s rRNA gene (Woese and Fox, 1977). Providing the genetic information for the small subunit of the ribosome, the 16s rRNA gene is present in all bacteria because it is essential in the translation apparatus of every cell. Interestingly, this gene does not code for a protein but a RNA that gains its function through its unique structural arrangement. Since some parts of the 16s rRNA sequence are more important for its structural integrity than others, evolution has put immense pressure on these functionally paramount regions to be highly conserved. Less relevant regions of the 16s rRNA have up to 1000 times higher rates of evolutionary change, leading to highly conserved regions interspersed by hypervariable regions (Van de Peer, 1996).

Overall, there are 9 hypervariable regions in the 16s rRNA gene flanked by conserved regions suitable as targets for universal primers (see fig. 1.4). There is an ongoing debate which of these

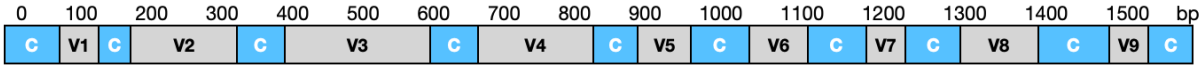


Fig. 1.4: Stylized primary structure of the 16s rRNA gene in bacteria. C indicates highly conserved regions. V1-9 indicate hypervariable regions

hypervariable regions is best for microbial studies (B. Yang et al., 2016). In the practice, all of the regions have their advantages and disadvantages and basically every hypervariable region and possible combination that can be sequenced on a high-throughput machine is being used by scientists to study infant microbiomes (Milani et al., 2017). Recent findings suggest that the regions V3, V4, V5 and V6 might yield superior results in regards to bacterial discriminatory power (Mancabelli et al., 2020). Some exemplary amplicon based studies:

- V1-V2 region (Alcon-Giner, Dalby, et al., 2020)
- V4 region (Pannaraj et al., 2017; Subramanian et al., 2014)
- V4-V5 region (C. J. Hill et al., 2017)
- V3-V4 region (Kim et al., 2020)
- V3-V5 region (Cortes et al., 2019)

When analyzing the resulting 16S rRNA gene based amplicon datasets there are essentially two different methods to process reads:

1. Operational taxonomic units (OTUs) based
2. Amplicon sequence variants (ASVs) based

In an OTU based approach, reads are clustered by a fixed similarity threshold (typically 97% identity). The resulting clusters are called operational taxonomic units and provide the basis for further analysis and taxonomic assignment (Westcott and Schloss, 2015). This approach allows for a faithful representation of the distances between sequences but has some major drawbacks: First, the clustering is computationally extremely demanding and scales exponentially with the number of sequences in a study (Wei et al., 2021). Second, the clustering is inherently dependent on all other sequences in the study, and thus OTUs might change if more or less data points are involved in the analysis, thus losing information about the biological ground truth (Benjamin J. Callahan et al., 2017). Third, a ring trial of 13 laboratories found considerable differences in the analysis of mock communities based on their choice of tools. Especially the use of custom databases and operational taxonomic units (OTUs) with stringent cut-offs contributed to problems in the reproducibility and comparability of analyses (D. M. O'Sullivan et al., 2021). Taken together, these drawbacks render OTUs unsuitable for cross-study meta-analysis.
To overcome some of these challenges, amplicon sequence variants (ASVs) can be used. In this approach, exact sequences that reflect the underlying biological sequences are statistically

distilled from the raw data. ASVs do not rely on clustering, making it easier to perform taxonomic and statistical comparisons during meta-analysis (Benjamin J. Callahan et al., 2017).
Both of the described methods follow up by comparing the resulting sequences to bacterial 16S rRNA gene reference databases to asses the taxonomic affiliation.

The second NGS based method is the whole metagenomic shotgun sequencing. With this approach, one does not amplify any marker gene but instead sequences every DNA strand found in an environmental sample (e.g. water, soil, stool). Contrary to 16s rRNA gene sequencing it is not a targeted approach (thus the term 'shotgun') and is in theory able to provide information about entire genomes of microbes irrespective of phylogenetic kingdom - bacteria, archaea, eukaryotes and viruses (Quince et al., 2017). The technique was pioneered by J. Craig Venter and first published in a 2004 study in Science (Venter, 2004). Venter and his collaborators used it to sequence a total of more than 1 billion bps from marine samples near Bermuda to gain insights into the microbial diversity and gene content of sea water and could show that the microbial life in the oceans is far greater than earlier anticipated. The functional component that comes with information about individual genes provides a major benefit. It allows the identification of especially relevant genes in the microbiome such as antibiotic resistance genes (Donia et al., 2014) and does not rely on taxonomic proxies to infer these functions.

Metagenomic sequencing is not without its drawbacks though. In a host-related microbiome setting (i.e. stool) most of the reads will belong to the host, since no filter is applied to the sequencing. This potentially leads to rare microbes not being detected, if the sequencing depth is not deep enough. Furthermore, assembly of metagenomes from short reads introduces an array of potential problems (Quince et al., 2017). Further, quality requirements for the DNA are way higher for metagenomic than amplicon based sequencing (Lindemann et al., 2017). In theory, many of the problems could be solved through increasing the sequencing depth, which unfortunately translates into prohibitively high costs for large scale microbiome projects. In practice, a combination of both methods is the most comprehensive and feasible way to investigate a microbiome of interest.

1.5 Aim of this thesis

Many studies have described changes in the microbiome compositions that are associated with disease states in patients and especially in preterm infants. The status of these microbiomes is often referred to as dysbiosis in contrast to a proposed healthy status called eubiosis.
The aims of this study are to comprehensively characterize the taxonomic composition of the gut microbiome of the PRIMAL study cohort and to critically evaluate the utility of the concepts eu- and dysbiosis in the developing preterm infant gut microbiome through the application of various bioinformatical methods.

1. To generate an accurate baseline characterization of the microbiome landscape in the PRIMAL study cohort, the taxonomic composition of preterm infants at three time points (day 0, day 28 and day 365) and a subset of their mothers will be compiled and compared on the basis of 16S rRNA gene sequencing
2. Studies have revealed beneficial effects on the health status of preterm infants through the supplementation of probiotic bacteria. Potential changes in the microbiome of the studied PRIMAL preterm infant cohort based on treatment group affiliation will be analyzed.
3. Results from conventional, culture based microbiological analyses and next-generation sequencing approaches (16S rRNA gene amplicon based sequencing and metagenomic whole genome shotgun sequencing) will be compared with a focus on dysbiosis, taxonomic classification and antibiotic resistance capacity
4. The microbiome compositions of preterm infants from the PRIMAL study cohort will be compared to publicly available reference microbiome datasets from preterm and full term infants
5. Potential interactions of taxa within the gut microbiome of preterm infants will be analyzed to generate hypothesis about the probiotic mechanism responsible for empirically observed positive health outcomes
6. Technical and conceptual challenges in the sequencing based analysis of microbiomes will be discussed in detail
7. A critical assessment of the utility of the dysbiosis concept in the context of disease development and alleviation will be performed

2 Methods

2.1 Stool sample collection

The infants included in this study are a part of the PRIMAL study cohort (Marißen et al., 2019). In short, for the PRIMAL study 654 preterm infants were recruited from 17 German children's hospitals (see tab. 2.1 for a list of all sites). Stool samples from the infants were collected at day 0 (visit 1), day 28 (visit 2) and day 365 (visit 3) of life. From a subset, corresponding stool samples of the mothers were also collected at day 0 of the infant's life. The recruitment of visit 1 samples took place in the time frame from April 2018 up to September 2020. Stool samples were collected on site and transferred to a -80 °Celsius storage place immediately. Specimens were then transported on dry ice to the Fecal Core Center Mainz (FCCM), where they were also stored at -80 °Celsius until further use. Visit 1 samples were examined in regards to color, texture and stickiness to asses their meconium status.

Tab. 2.1: Study Sites involved in the collection of PRIMAL samples and their abbreviations

Abbreviation	Study Site
AB	Aschaffenburg
BN	Bonn
BO	Bochum
E	Essen
FR	Freiburg
H	Hannover
HB	Bremen
HD	Heidelberg
HH	Hamburg
HL	Lübeck
HOM	Homburg
HRO	Rostock
J	Jena
K	Köln
PB	Paderborn
SN	Schwerin
TÜ	Tübingen
WI	Wiesbaden

2.2 DNA extraction, amplification and sequencing

After thawing, the DNA in meconium and stool sample was extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Germany) following the manufactures guidelines. An additional step consisting of bead-beating at 30 hertz for 2*7 minutes was performed on a TissueLyser (QIAGEN, Germany) to ensure the disruption of all necessary cell parts in gram-negative and especially gram-positive bacteria. Following the extraction, the DNA concentration was assessed using a Qubit Fluorometer (Thermo Fisher Scientific). A 251 bps long segment (V4 region) of the 16S rRNA gene present in all bacteria was amplified using the primers 515f and 806r (forward: 5'-GTGCCAGCMGCCGCGGTAA-3', reverse: 5'-GACTACHVGGGTWTCTAATCC-3') and indices were added using the Nextera XT DNA Library Preparation Kit (Illumina, Inc., USA). Up to 385 samples were sequenced in parallel on a MiSeq (Illumina, Inc., USA) sequencing machine using the V2 chemistry yielding a theoretical average of 13 million paired-end reads per run with an average length of 2x250 bps. Every step in the process was accompanied by negative controls (steril, nuclease free, DEPC-treated water) to reduce bias. Some samples were mixed with different amounts of Spike-in DNA (ZymoBIOMICS^{TM} Spike-in Control I (High Microbial Load)). Every sequencing run further included positive controls (ZymoBIOMICS^{TM} Microbial Community DNA Standard) and 25% PhiX-DNA to overcome the challenges of cluster generation, phasing and prephasing during sequencing. Technical metadata such as amount of extracted stool, DNA concentration after extraction and amplification, kits used, the date of extraction and the sample material (stool, meconium, water etc.) were recorded using Microsoft Excel. The Excel file was transformed into a comma-separated values (CSV) file and unique sample IDs were created for every sample prior to sequencing. This was achieved through eliminating every special character in the sample label and adding the number of the sequencing run in front of the label and a capitalized letter (in alphabetical order) behind the label, if the label was not unique. The necessary sample sheet for the sequencing machine was compiled with *create_SampleSheet_for_Miseq.py*.

Fig. 2.1: Primer binding sites in the V4 region of the 16s rRNA Gene of *E.coli* (as a surrogate for all members of the kingdom bacteria). The color red indicates the forward primer position and blue the reverse primer position

2.2.1 Taxonomic composition of positive controls

To ensure technical and quality consistency across different sequencing runs, positive controls were processed alongside the study samples. ZymoBIOMICS^{TM} Microbial Community DNA Standard is comprised of eight bacterial species in equal amounts and two eukaryotic organisms

(yeasts). To some samples, predefined bacterial DNA was added to allow an approximation of the absolute cell count. ZymoBIOMICS^{TM} Spike-in Control I (High Microbial Load) consists of two extremophile bacterial species in equal amounts of cells that do not naturally occur in the human gut microbiome (see tab. 2.3). The exact composition of these controls is listed in tab. 2.2 and tab .

Tab. 2.2: Taxonomic Composition of ZymoBIOMICS^{TM} Microbial Community DNA Standard

Species	Avg. GC (%)	Gram Stain	gDNA Abun. (%)
Pseudomonas aeruginosa	66.2	-	12
Escherichia coli	56.8	-	12
Salmonella enterica	52.2	-	12
Lactobacillus fermentum	52.8	+	12
Enterococcus faecalis	37.5	+	12
Staphylococcus aureus	32.7	+	12
Listeria monocytogenes	38.0	+	12
Bacillus subtilis	43.8	+	12
Saccharomyces cerevisiae	38.4	Yeast	2
Cryptococcus neoformans	48.2	Yeast	2

Tab. 2.3: Taxonomic Composition of 20 μl ZymoBIOMICS^{TM} Spike-in Control I (High Microbial Load)

Species	Cells (%)	16S rRNA Copies	Total DNA (ng)
Imtechella halotolerans	2 x 10^7	6.0 x 10^7	67.2
Allobacillus halotolerans	2 x 10^7	1.4 x 10^8	58.2

2.2.2 Whole genome shotgun metagenomics of study sample subset

DNA from a subset of study samples (maternal samples N=31, visit 2 samples N=30 and visit 1 samples N=4) were sent to the sequencing core facility at the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany for whole genome shotgun metagenomic sequencing. From the raw data, resistome datasets with phylogenetic origins of the respective antibiotic resistance genes were generated and provided by Dr. Thea Van Rossum (Structural and Computational Biology Unit, Bork Lab, EMBL, Heidelberg, Germany) on the basis of the Comprehensive Antibiotic Resistance Database (CARD, https://card.mcmaster.ca/).

2.3 16S rRNA gene read processing

All data analyses were performed on an Esprimo P957/E85+ (Fujitsu-Siemens ,64 GB RAM, 8 Intel Core i7 CPU) under Linux (Ubuntu LTS 16.04, Xenial Xerus) using the python programming

language (version 3.7.3) (van Rossum, 1995), BASH (version 4.3.48) (Gnu, 2007) and R (version 3.4.4) (Team, 2018) if not stated otherwise. A list of the relevant packages and libraries can be found in tab. 2.4. After sequencing, demultiplexing of the reads per sample was performed

Tab. 2.4: Python packages used

package	version	publication (if applicable)
anaconda-client	1.7.2	(Anaconda Inc., 2021)
anaconda-navigator	1.10.0	(Anaconda Inc., 2021)
biopython	1.73	(Cock et al., 2009)
bokeh	2.3.1	(Bokeh Development Team, 2018)
colorhash	1.0.2	
conda-build	3.21.4	(Anaconda Inc., 2021)
conda-package-handling	1.7.2	(Anaconda Inc., 2021)
conda-verify	3.4.2	(Anaconda Inc., 2021)
conda	4.10.1	(Anaconda Inc., 2021)
jupyter-client	5.2.4	(Kluyver et al., 2016)
jupyter-console	6.0.0	(Kluyver et al., 2016)
jupyter-contrib-core	0.3.3	(Kluyver et al., 2016)
jupyter-core	4.4.0	(Kluyver et al., 2016)
jupyter-nbextensions-configurator	0.4.1	(Kluyver et al., 2016)
jupyterlab-pygments	0.1.2	(Kluyver et al., 2016)
jupyterlab-server	0.2.0	(Kluyver et al., 2016)
jupyterlab-widgets	1.0.0	(Kluyver et al., 2016)
jupyterlab	0.35.6	(Kluyver et al., 2016)
matplotlib	3.0.3	(Hunter, 2007)
networkx	2.5	(Hagberg et al., 2008)
numpy	1.16.3	(Harris et al., 2020)
nxviz	0.6.2	
pandas	0.24.2	(The pandas development team, 2020)
pip	19.2.1	
plotly	3.9.0	(Plotly Technologies Inc., 2015)
scikit-learn	0.21.2	(Pedregosa et al., n.d.)
scipy	1.2.1	(SciPy 1.0 Contributors et al., 2020)
seaborn	0.9.0	(Waskom, 2021)
sinfo	0.3.1	
spyder	3.3.6	
statannot	0.2.3	
tqdm	4.59.0	(da Costa-Luis and Nechaev, 2021)
urllib3	1.26.4	
wget	3.2	
zipp	3.4.1	

directly on the MiSeq using the build-in MiSeq Reporter software with default settings. The number of produced reads was evaluated and a sanity check was performed for the number of PhiX reads with *total_reads_gz_in_Miseq_run.sh*. Quality checking of the reads was performed using the fastQC software with default parameters (Andrews, 2010). The raw reads were imported into R and then processed using a suit of tools from the Divisive Amplicon Denoising Algorithm 2 package in R (DADA2 (Benjamin J Callahan et al., 2016)).

2.3.1 Divisive Amplicon Denoising Algorithm 2 (DADA2)

DADA2 is a statistical model-based, open-source software package for correcting amplicon errors in the sequencing process without constructing traditional operational taxonomic units (OTUs) and instead producing base-exact amplicon sequence variants (ASVs) leading to results with intrinsic biological meaning (Benjamin J Callahan et al., 2016). DADA2 constitutes of six main steps:

1. Quality filtering of reads
2. Learning of an error model
3. Dereplication of multiple identical sequences
4. Sample inference
5. Merging of paired end reads
6. Chimera identification
7. Taxonomic classification

First, the raw reads (forward and reverse) are filtered and trimmed. The first 20 basepairs from each side of the read are trimmed off due to typically high rates of error in this region. Sequences with more than 0 unambiguous assigned basecalls (N) are discarded. All reads that map to the genome of PhiX are filtered out. Since the expected sequencing length given the primers is around 251 bps, reads are truncated to 240 bps - with a margin of safety of ca. 10 bps because the quality of sequencing tends to decline near the end of reads. After truncation, reads that statistically still contain more than 2 errors (DADA2 default) in their sequence are also discarded (see an explanation of the parameters in tab. X) with the `filterAndTrim()` command (see listing 2.1). For a second analysis, the investigation of the relative abundance of mitochondrial reads in the samples, the truncLen parameter was omitted at this step to allow reads with less than the expected amplicon size (mitochondrial region: ca. 200 bp) to be kept for analysis.

```
# reads are quality filtered and trimmed
filterAndTrim(fnFs, filtFs, fnRs, filtRs,
              maxN=0, maxEE=c(2,2),
              truncQ=2, truncLen = c(240,240), trimLeft = c(20,20),
              rm.phix=TRUE,
              compress=TRUE, multithread=TRUE)
```

Listing 2.1: Quality filtering

Second, DADA2 deploys a computational step to learn the specific error rates of the given sequences. It does this by alternately inferring sample composition and estimating error rates until both converge. The starting point is the assumption that only the most abundant sequence is correct and all deviations are errors. The algorithm then optimizes to a solution consistent with Phred-scores (Phred-Score is defined as the negative decadic logarithm of the probability for a technical error in the base call procedure during sequencing $-10 \log 10(x)$). Meaning, changes

from every base to every other possible base should be statistically consistent with the predicted error rate given the Phred-score (e.g. Phred-Score 30 means a 1 in 1000 chance that a base has been called incorrectly). For that purpose, all reads with more than 10% difference in their sequence are aligned using the Needleman-Wunsch algorithm. The most abundant sequence is chosen as reference. The model then works out the conditional probability that a given base originates from true biological differences rather than technical errors given the consensus Phred-Score. The algorithm continues by subdividing the alignment clusters and repeating the steps until a consensus between expected error rate and sample composition is reached. Fig. 2.2 offers a visualization of the error learning step. The computation is implemented using the

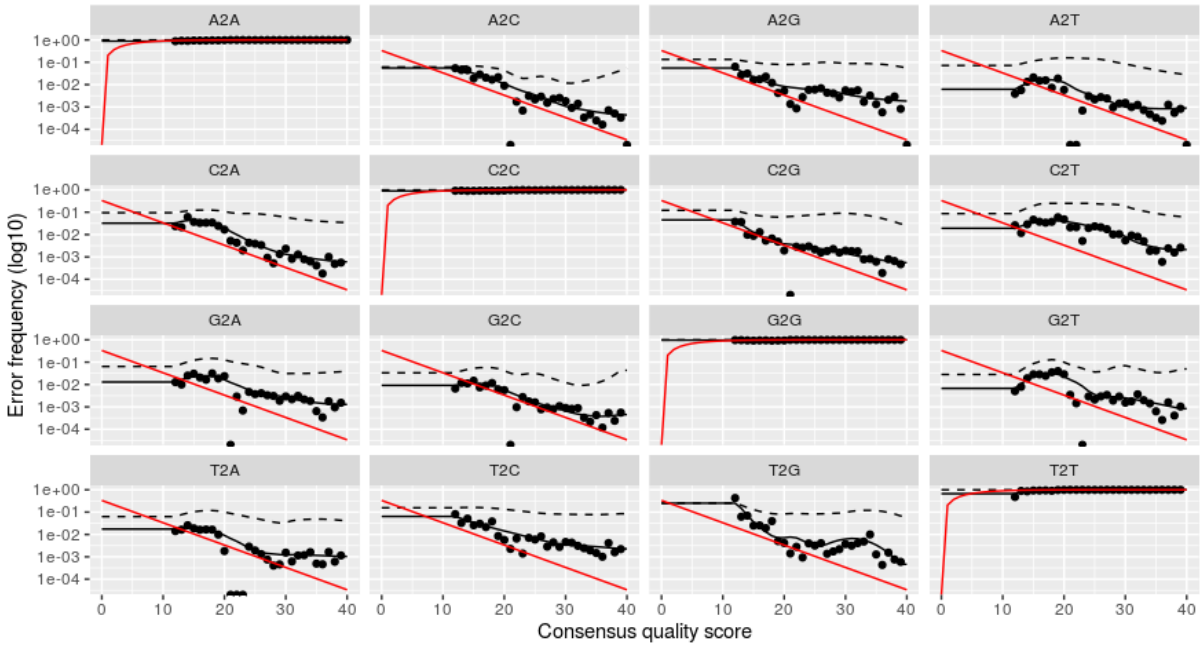


Fig. 2.2: Example of a calculated DADA2 error model. Red lines indicate the expected error rate given the consensus Phred-Score of a sequence. Dashed lines indicate the initial guess for error rates. Black dots indicate observed error rates and the black lines indicate non-linear approximations of the observed (learned) errors. In this example red and black lines show the same (expected) trend: decreasing chance of error with increasing consensus quality score in a given sequence.

`learnErrors()` method (see listing 2.2).

```
# Error Models for forward and reverse reads (errF, errR)
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)

plotErrors(errF, nominalQ=TRUE)
```

Listing 2.2: Error model learning

Third, since amplicon sequencing produces many identical sequences that would lead to major computational overhead in analyzing the data, reads are dereplicated. This means that only one representative sequence is kept for analyses and the number of identical replicates is stored for later use. The command `derepFastq()` is used (see listing 2.3).

```
# dereplication of forward and reverse reads (derepFs, derepRs)
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)
```

Listing 2.3: Dereplication of reads

In the fourth step, the core algorithm of DADA2 (`dada()`, see listing 2.4) is used to infer sample composition from the sequencing reads. ASVs are identified for every sample in the sequencing run. This means, the trained statistical model is used to differentiate real biological sequences from sequencing errors to reveal members of the studied microbiome.

```
# true biological sequences are revealed from the sequencing data (dadaFs, dadaRs)
dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)
```

Listing 2.4: sample composition inference

The forward and reverse reads are then merged to create consensus sequences with the `mergePairs()` function. In the case of the V4 region of the 16s rRNA gene, the expected gene region has a length of 251 bps. Given the sequencing length of 251 bps of the illumia MiSeq V2 chemistry, this allows for an excellent overlap and thus robust merging results. In a next step, only sequences that match the expected length of the majority of 16s rRNA genes are kept for further analyses (usually 249-254 bps) (see listing 2.5).

```
# merge forward and reverse reads
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)

# remove reads that do not conform to the expected length
seqtab <- seqtab[,nchar(colnames(seqtab)) %in% seq(249,254)]
```

Listing 2.5: read merging

In the next step, chimerical sequences are removed from the data set. Chimeras are artifacts created during the PCR steps in which an incomplete amplification product is being used as a primer for another PCR step leading to artificial sequences that do not represent real underlying biological sequences but combinations of such. This problem especially arises during amplicon based approaches that use highly similar gene regions such as the 16s rRNA gene. It is of great importance to account for these sequences, since they would otherwise artificially inflate the observed diversity and might even facilitate the "discovery" of new (fabricated) species (Haas et al., 2011). The function `removeBimeraDenovo()` removes chimeric sequences using a consenus approach in which the chimerical status of each sequence is assessed through a statistical majority voting (see listing 2.6).

```
# remove chimeric sequences
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=
    TRUE)
```

Listing 2.6: Chimera removing

2.3.2 Taxonomic assignment

The identified ASVs represent biological sequence variability in the 16s rRNAs of the studied samples. To get a deeper understanding of the microbiome composition, every unique sequence is processed using the `assignTaxonomy()` function which implements a naive bayesian classifier to rapidly assign bacterial taxonomy to 16s rRNA sequences (Q. Wang et al., 2007) (see listing 2.7).

```
taxa <- assignTaxonomy(seqtab.nochim, "silva_nr_v138_train_set.fa.gz", multithread = TRUE)
taxa <- addSpecies(taxa, "silva_species_assignment_v138.fa.gz")
```

Listing 2.7: Taxonomic assignment

As reference database for the taxonomic classification the SILVA database version 138 (Quast et al., 2012) was used which has over 9,4 million high quality ribosomal full length RNA entries including their taxonomic assignment.After the taxonomy has been assigned, the R data frames (ASV table, taxonomic assignment and meta data) are converted into a phyloseq-object for further processing including read decontamination.

2.3.3 Phyloseq and decontam

Phyloseq is an open-source software package to analyze high-dimensional microbiome data actively developed since 2013 (McMurdie and Holmes, 2013). It includes many convenience functions to assess and manipulate microbiome data (see fig. 2.3). The core of phyloseq is the

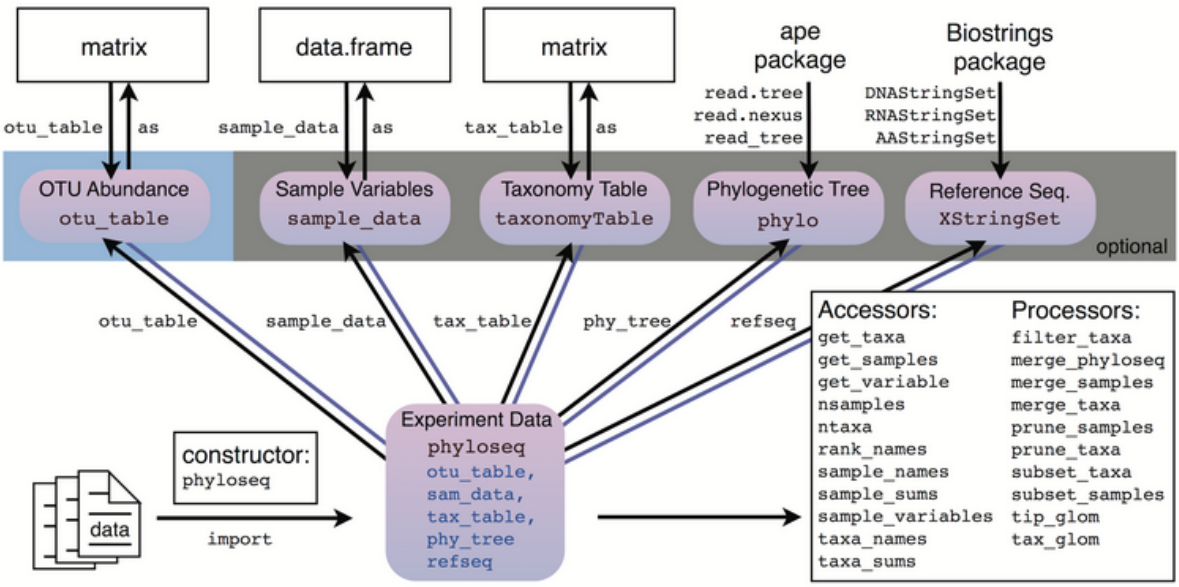


Fig. 2.3: Reproduced from https://joey711.github.io/phyloseq/: Overview of the functionality of the phyloseq package in R

multi-faceted phyloseq object class which represents microbiome data on an experiment level and is capable of efficiently storing and working with ASVs or OTUs, meta data, phylogenetic data and taxonomic data. It offers a wide variety of adjunct functions such as over 40 different distance metrics calculations and common algorithms for the reduction of high-dimensional data such as principal component and coordinate analyses (PCA, PCoA) or non-metric multidimensional scaling (NMDS). Phyloseq version 1.22.3 was used for all analyses. After import in phyloseq, another package was used to address potential contamination in the microbiome data: decontam version 1.1.2 (Davis et al., 2018). Decontam is a novel computational approach to statistically account for ASV contamination that can occur in every step of the microbiome pipeline. While appropriate laboratory practices can mitigate contamination, it can never be completely prevented. Two assumptions lie at the core of the decontam algorithm:

- contaminating sequences likely also appear in negative controls
- contaminating sequences have a higher frequency in low-concentration samples than samples with high amounts of DNA

In a first decontamination step, all reads are discarded that have been taxonomically assigned to anything but the expected kingdom, bacteria, meaning reads mapping to archaea and eukaryota. Since the primers are specifically designed to amplify bacteria, other taxonomic classifications are most likely the result of erroneous amplification during PCR. Next, information from the meta data on Control/Sample status and DNA concentration in ng/μl is collected to be used in the decontam algorithm. The tool assesses for every identified ASV whether it is more likely to be a contaminant or not. The chosen p-value is 0.05 as a threshold for when the null-hypothesis (ASV is not a containment) should be rejected in favor of the alternative hypothesis (ASV is a contaminant) (see listing 2.8).

```
# remove Archae and Eukaryota sequences and all bacteria that have no classification beyond
    kingdom
ps <- subset_taxa(ps, !Kingdom %in% c("Archaea","Eukaryota") & !is.na(Phylum) & !Phylum %in%
     c("","uncharacterized"))

# create column indication which samples are negative controls
sample_data(ps)$is.neg <- sample_data(ps)$Sample_or_Control == "Control"

# decontaminate reads
contamdf.com <- isContaminant(ps,conc = "dna_quant_ng_ul", neg = "is.neg", method = "
    combined", threshold = c(0.05))

ps.noncontam <- prune_taxa(!contamdf.com$contaminant, ps)
```

Listing 2.8: Read decontamination

2.3.4 Phylogenetic tree calculation

To infer phylogeny-aware distance metrics like weighted unifrac distances for beta diversity calculations between the microbiomes, phylogenetic trees were calculated in R. First, a multiple

sequence alignment was performed using the DECIPHER package (version 2.6.0) with default parameters. Gap-opening penalties were -18 for highly similar sequences and -16 for sequences that were very distant. Gap extension penalties were -2 and -1 for similar and distant sequences, respectively.

Second, the phangorn package (version 2.5.5) was used to calculate a phylogenetic tree. A neighbor-joining tree ("The Neighbor-Joining Method" 1987) was initially calculated as a basis for a generalized time-reversible with gamma rate variation (GTR+G+I) maximum likelihood tree (Ben J. Callahan et al., 2016).

2.4 Conventional microbiology

Conventional microbiological analysis were performed for visit 2 samples by the institute for medical microbiology and hygiene of the University Medical Center of the Johannes Gutenberg University Mainz, Germany under management of Dr. med. Ekkehard Siegel. Stool samples were cultured on different media (CNA-medium for gram-positive bacteria and CPS-medium for gram-negative bacteria) and studied for the presence of potential pathogenic species. The species classification was performed with MALDI-TOF and the antibiotic resistance capacity (against the antibiotics penicillin, cephalosporin, carbapenem, fluoroquinolone, gentamycin) was screened for via a VITEK machine (bioMérieux, Inc.). A full list of all bacteria that the samples were screen for and under what condition they were reported can be found in tab. 2.5

Tab. 2.5: List of bacterial taxa that were screened for

Genus	Species	reporting criteria
Staphylococcus	*aureus*	always
Serratia	*marcescens*	always
Morganella	*morganii*	if >= 2MRGN
Klebsiella	*pneumoniae*	always
Klebsiella	*oxytoca*	always
Escherichia	*coli*	if >= 2MRGN
Enterobacter	*cloacae*	if >= 2MRGN
Enterobacter	*aerogenes*	if >= 2MRGN
Citrobacter	*freundii*	if >= 2MRGN
Acinetobacter	*baumannii*	if >= 2MRGN

2.4.1 Sanger sequencing and read processing of cultured bacteria

The V4 region of the 16S rRNA gene of cultured isolates of *Enterobacter cloacae* and *Klebsiella pneumoniae* was amplified using the method described in 2.2. The DNA was then shipped to and sequenced by Eurofins Scientific. Resulting files were quality checked in SnapGene Viewer,

version 4.1.7 (SnapGene software (from Insightful Science; available at snapgene.com)) and aligned using the multiple sequence comparison by log-expectation algorithm (MUSCLE) (Edgar, 2004) with default parameters. Prior to alignment, forward and reverse sequences of the same read were merged to a consensus sequence using bbmerge.sh with default parameters (Bushnell et al., 2017).

The taxonomic identity of single sequences was further analyzed using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) with default parameters.

2.5 Statistical analysis

To test non-parametrically for statistical differences in the variance of independent samples, Wilcoxon-Mann-Whitney-U-Tests (Wilcoxon, 1945) (for two groups) and Kruskal-Wallis-Tests (Kruskal and Wallis, 1952) (for two or more groups) were performed as implemented in the statannot and scipy libraries in python. Multiple testing p-value corrections were either performed with the Bonferroni correction or the Benjamini-Hochberg Procedure.

For multivariate statistics the vegan package (version 2.5.6) (Dixon, 2003) in R was used. Permutational multivariate analysis of variance (PERMANOVA) was performed using the adonis and adonis2 functions. Permutation tests for homogeneity of multivariate dispersions were performed with the *betadisper* followed by the *permutest* function.

Diversity calculations and multidimensional scaling techniques were performed in R using the phyloseq package version 1.22.3 (McMurdie and Holmes, 2013). Diversity calculations included alpha and beta diversity calculations (Shannon diversity index (C. E. Shannon, 1948) and weighted Unifrac distances (Lozupone et al., 2011), respectively). Multidimensional scaling techniques consisted of Principal Coordinate Analysis (PCoA) and non-metric multidimensional scaling (NMDS). The R random seed was set to 711 for reproducibility.

Co-occurrence network analyses were performed in the julia programming language (version 1.6.1) making use of the FlashWeave package (Tackmann, 2019). Discretized mutual information tests (alpha=0.01) were performed within this package to identify statistically significant correlations between different members of the microbiome. The networks were further processed with the Cytoscape software package (P. Shannon, 2003) (version 3.7.2).

2.5.1 Differential antibiotic resistance gene analysis using machine learning

To determine antibiotic resistance genes that were more prevalent in mother or infant samples, a machine learning algorithm was deployed. First, NMDS was performed to visually inspect clusters of antibiotic resistance genes within sample types. Second, a random forest classifier

(N estimators = 100) was trained on 33 % of the antibiotic resistance gene table stratified by sample type with a random state of 42. The trained model was used to predict the affiliation of individual samples to sample type. Third, the feature importance attributes were extracted and ordered by weight to assess the antibiotic resistance genes that were most relevant in determining group affiliation.

To determine the importance of KRINKO relevant genera (*Staphylococcus*, *Serratia*, *Klebsiella*, *Escherichia-Shigella*, *Enterobacter*, *Citrobacter*) as well as *Bifidobacterium* and *Lactobacillus* on the antibiotic resistance gene clustering, the Axis 1 and 2 coordinates of the PCoA were statistically correlated with the relative abundance of the genera in question using the spearman correlation coefficient. The Benjamini/Hochberg procedure was used to correct for multiple testing.

2.6 Obtaining microbiome reference datasets

Relevant microbial metagenomic reference sequencing datasets were either downloaded from the European Nucleotide Archive (ENA) or requested from the authors. Appropriate studies were chosen based on the cohort in question being age matched, healthy full term or preterm infants that had their microbiome analyzed by 16S rRNA sequencing. Associated metadata and clinical data was either downloaded from the ENA repository or was requested from the authors directly. The studies that were chosen as reference are described below and a summary can be found in tab 2.6. One study from Los Angeles, California, USA by Pannaraj et al. investigated the relationship between the breast milk and infant gut microbiome using the V4 region of the 16S rRNA gene. In their cohort the authors looked at 107 healthy infants and reported on their microbiome (Pannaraj et al., 2017). A second study from Hill et al. used the V4-V5 region of the 16S rRNA gene to track changes in the infant gut microbiome over a period of 24 weeks. Their INFANTMET cohort consists - among others - of full-term, spontaneous vaginally delivered infants. An observational study called BAMBI from Alcon-Giner et al. from the UK followed a cohort of 234 infants of whom 101 were supplemented with a probiotic formulation of *Bifidobacterium* and *Lactobacillus*. 16S rRNA sequencing and tight collection of clinical metadata was performed.

Tab. 2.6: Reference studies that provided data sets. The number associated to the 16S rRNA gene region corresponds to the range of sequencing read length.

publication	N	age	16s region	accession
(Pannaraj et al., 2017)	107	1 month +- 10 days	V4 (249,254)	request from authors
(C. J. Hill et al., 2017)	192	1 month	V4-V5 (320,327)	PRJNA339264
(Alcon-Giner, Dalby, et al., 2020)	133	1 month +- 10 days	V1-V2 (267,278)	PRJEB31653

The raw sequencing data were processed using the same analysis pipeline described in 2.3.1 with parameter changes to the *in-silico* cut out at the seqtab step according to region of the 16S rRNA gene used (see tab. 2.6).

3 Results

Partial analyses of data from the following result section, especially of the meconium and mother samples, have been published in the journal *mSphere* on January 12, 2022 (Klopp et al., 2022). The full research article can be found in the appendix of this thesis.

3.1 PRIMAL samples

As part of the PRIMAL study 1353 fecal samples from infants and 290 samples from mothers were collected across 17 hospitals, transported to Mainz and processed (see 2.2). These samples were collected across three different visits (visit 1 : day 0 N=565, visit 2: day 28 N=529 and visit 3: year 1 N=259) per infant and for a subset of infants one sample of the mother (N=290). Different set sizes could be obtained for different families (see fig. 3.1). 131 complete sets (visit 1-3 and a mother sample) were analyzed. For 518 infants, visit 1 and 2 could be processed and 284 samples had at least visit 1 and a corresponding mother sample. Metadata regarding the

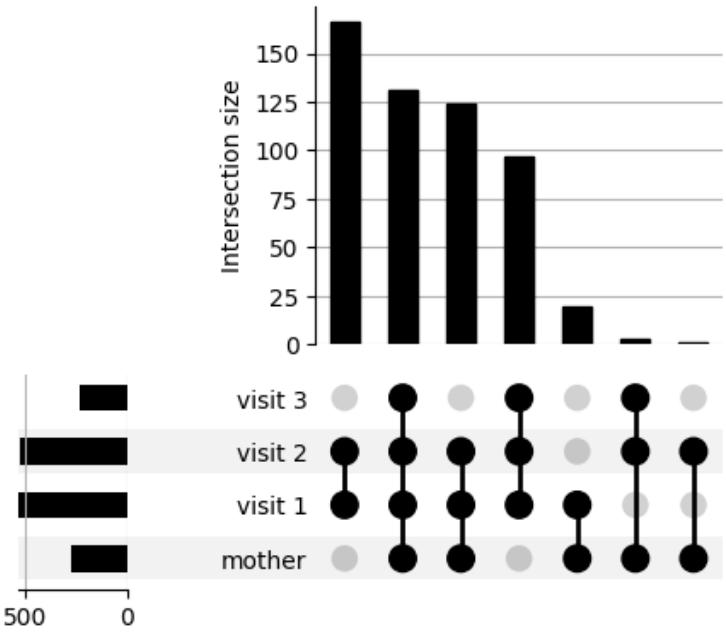


Fig. 3.1: Number of sets (different samples from one participating family) that have been analyzed in the study.

infants' sex, gestational age and hospital collection site were recorded (see fig. 3.2). The different infant groups were approximately the same size in regards to gestational age (28 to 30 weeks = 47.6%, >30 to 32 weeks = 52.4%) and sex (female = 46.4%, male = 53.6%) but differed in the frequencies of their collection site.

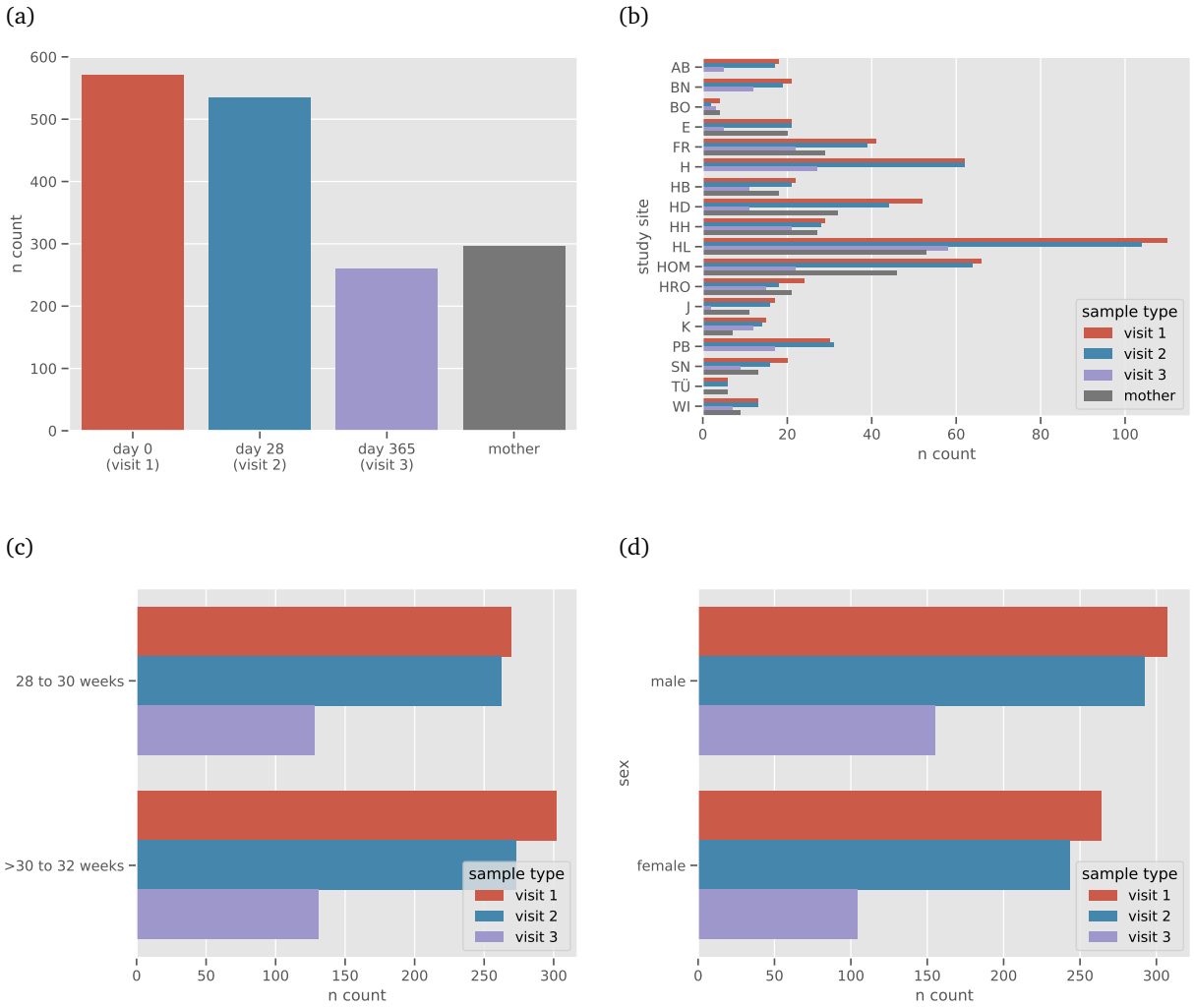


Fig. 3.2: Distribution and count of (a) different samples types that have been collected and processed in the study (b) by study site (c) by gestational age (d) by sex.

3.2 16S rRNA Sequencing and read processing

All samples were sequenced across 12 Illumina Miseq sequencing runs with an average read yield of 17 million reads per run (min: 12 million, max: 22 million). The average amount of PhiX reads per run was 32.81 % with a maximum of 46.33 % and a minimum of 22.82 % (see fig. 3.3). Every sequencing run included negative controls and every run (except one) positive controls to ensure consistent quality across sequencing runs. A steady microbiome composition of the

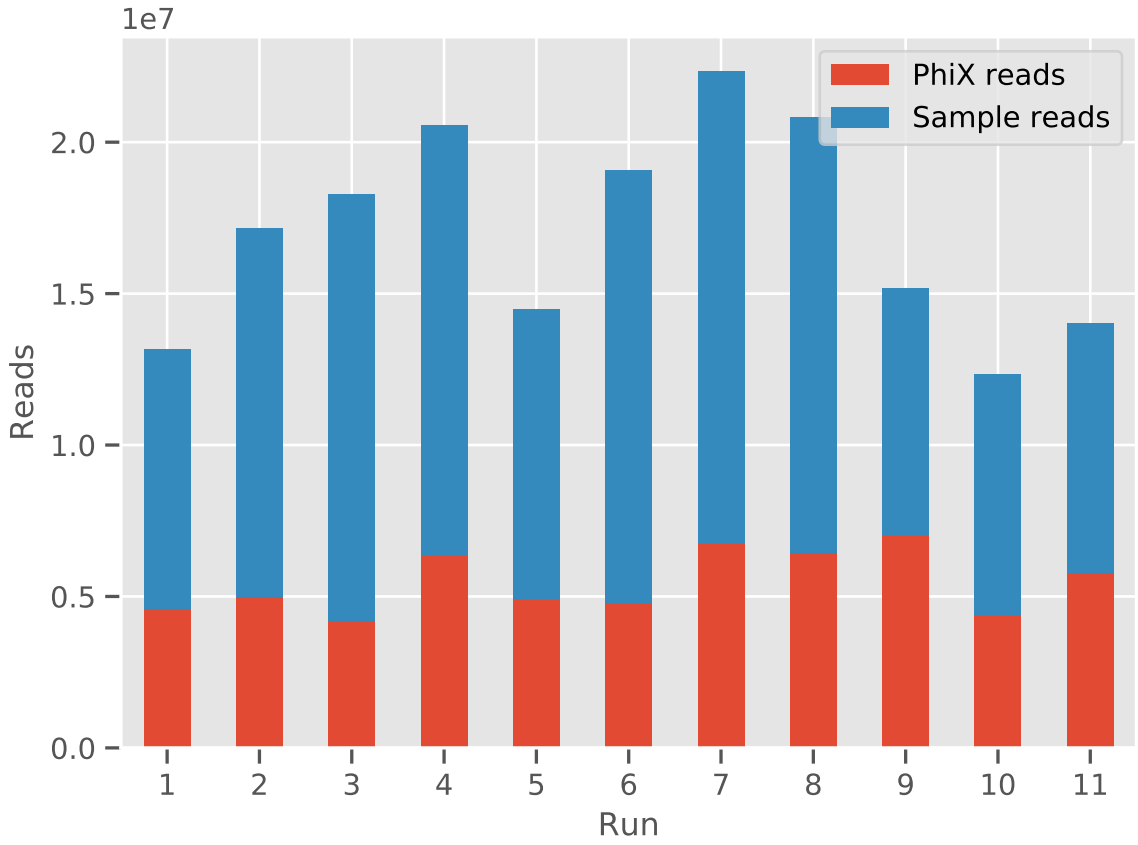


Fig. 3.3: Distribution of the read yield across sequencing runs

positive controls across runs in accordance with the manufacture's information was observed (see fig. 3.4 and tab. 2.2).

Different processing steps in the analysis pipeline discarded reads due to various criteria. In a first step, reads were discarded that did not meet the required quality score or expected length of approximately 250 bp. This usually affected less than 10 % of reads, but stark differences were observed in the negative controls and the meconium samples where the majority of reads did not pass the thresholds. Sequences were discarded if the corresponding reads in the forward and reverse files could not be merged, which affected only a minority of sequences. A last quality step consisted of the removal of chimeric sequences, PCR artifacts and sequences statistically classified as contaminations (see tab. 3.1). Different sample types varied significantly in the amounts of final reads (see fig. 3.5). Visit 1 samples in particular, which primarily consisted of meconium samples, showed very low yields of high quality bacterial reads in contrast to their average raw read counts.

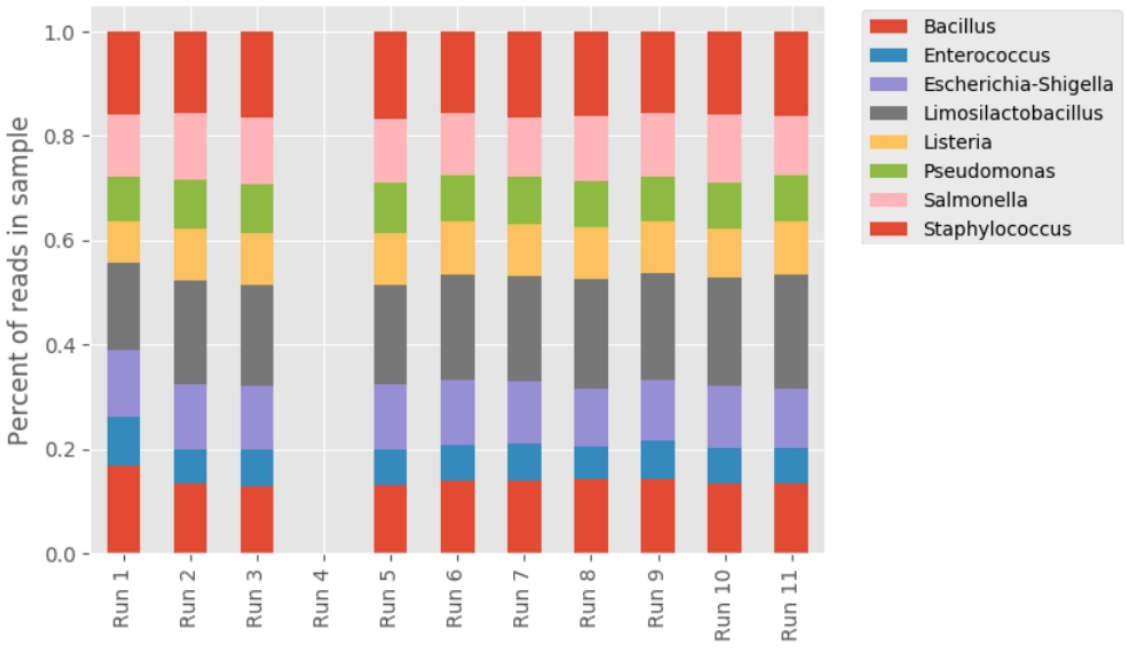


Fig. 3.4: Taxonomic distribution (genus level) of the bacterial content in the positive controls ZymoBIOMICS^{TM} Microbial Community DNA Standard across Miseq sequencing runs

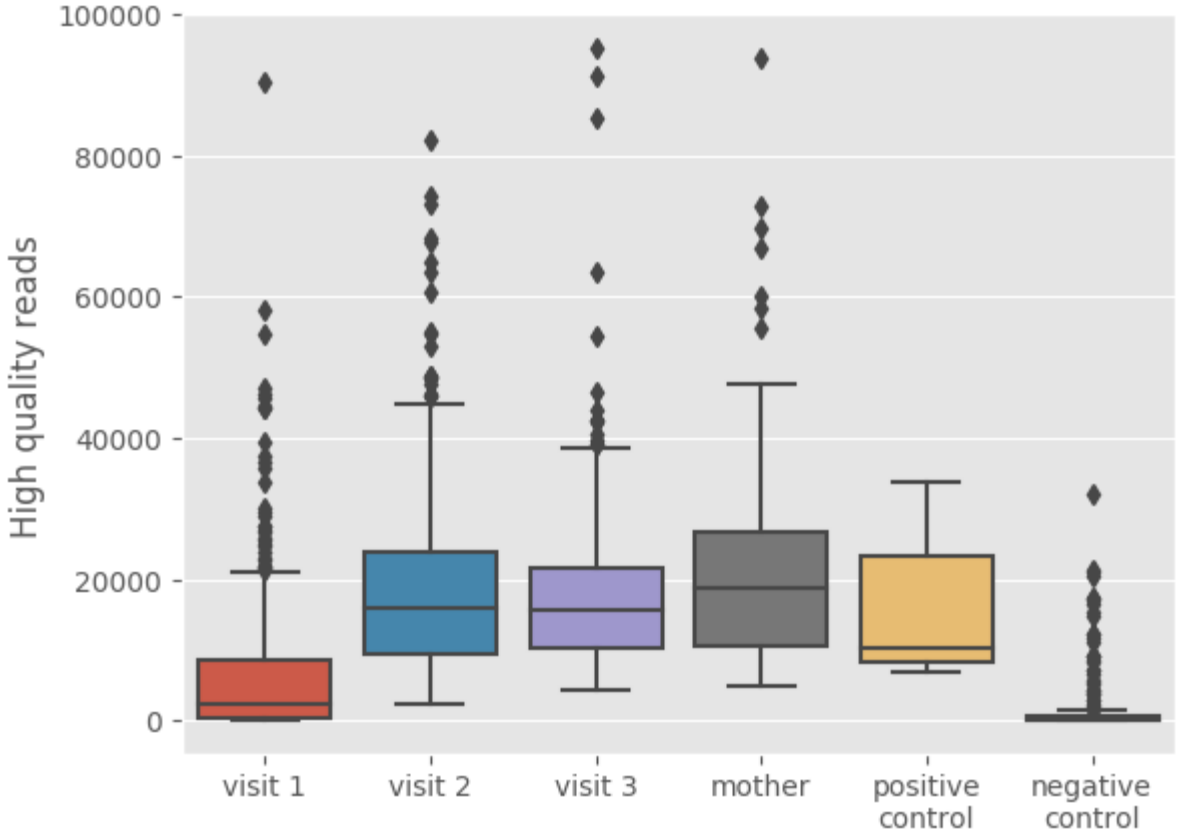


Fig. 3.5: Number of high quality reads remaining after the DADA2 pipeline across different sample types. 7 samples with more than 100000 reads (visit 1: n=1, visit 2: n=3, visit 3: n=2 and mother: n=1) are not displayed for better legibility.

Tab. 3.1: Statistics of how many reads passed various quality filtering steps in the analysis pipeline

(a) Positive controls

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	22659.5	21144.8	19571.4	15864.3	70 %
min	9557.0	9078.0	8071.0	6947.0	72.6 %
median	14978.0	14000.5	12622.5	10408.5	69.5 %
max	48264.0	44090.0	41131.0	33742.0	69.9 %

(b) Negative controls

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	9587.22	2159.65	2103.59	1899.2	19.8 %
min	6.00	4.00	0.00	0.0	0 %
median	4843.00	175.00	149.00	103.0	2.1 %
max	67394.00	36147.00	34726.00	32041.0	47.5 %

(c) Visit 1

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	24923.8	8163.73	7932.97	7575.10	30.4 %
min	23.0	9.00	0.00	0.00	0 %
median	21670.0	2727.00	2451.00	2451.00	11.31 %
max	168997.0	122079.00	121825.00	121745.0	72 %

(d) Visit 2

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	24117.46	21719.94	21207.22	18978.07	78.7 %
min	2647.00	2409.00	2377.00	2356.00	89 %
median	20197.00	18110.00	17654.00	16110.00	79.8 %
max	284478.00	236140.00	229047.00	177336.00	62 %

(e) Visit 3

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	23167.03	21163.09	20303.51	19101.61	82.4 %
min	5573.00	4891.00	4602.00	4295.00	77 %
median	18886.00	17331.00	16528.00	15727.00	83 %
max	260863.00	234489.00	226474.00	205236.00	78.7 %

(f) Mother samples

	Raw reads	Quality filtered	Merged	Decontamination	Ratio (retained)
mean	25967.19	23526.68	22499.94	21116.64	81.3 %
min	5732.00	5199.00	4998.00	4755.00	83 %
median	23324.00	21247.50	20326.00	19294.50	82.7 %
max	166367.00	141748.00	136834.00	125624.00	75.5 %

3.3 Mitochondrial sequences are co-amplified in the study samples

Amplification based sequencing strategies like the 16S rRNA gene sequencing can lead to the amplification and sequencing of off-target regions. During quality filtering, a majority of reads in the visit 1 samples were filtered out due to an insufficient length not corresponding to the target length of the V4 region of the 16S rRNA gene (ca. 250 bp). Analysis of these DNA sequences with a length of ca. 200 bp revealed that they originate from the 12S region of the human mitochondrial genome. The relative abundance of these sequences was significantly higher in visit 1 samples than at later time points (Kruskal-Wallis, p-value < 0.001). Visit 1 sample microbiomes consisted of 49.2 % mitochondrial reads on average (median: 50.5 %). This proportion decreased significantly at one month of age with an average of only 1.2 % (median: 0.1 %). During the course of the microbiome maturation in the first year of life, it further decreased to an amount of 0.04 % on average (median: 0 %) and was thus comparable to the relative abundance of mitochondrial reads in the adult samples that was also near zero with 0.3 % on average (median: 0 %) (see fig. 3.6).

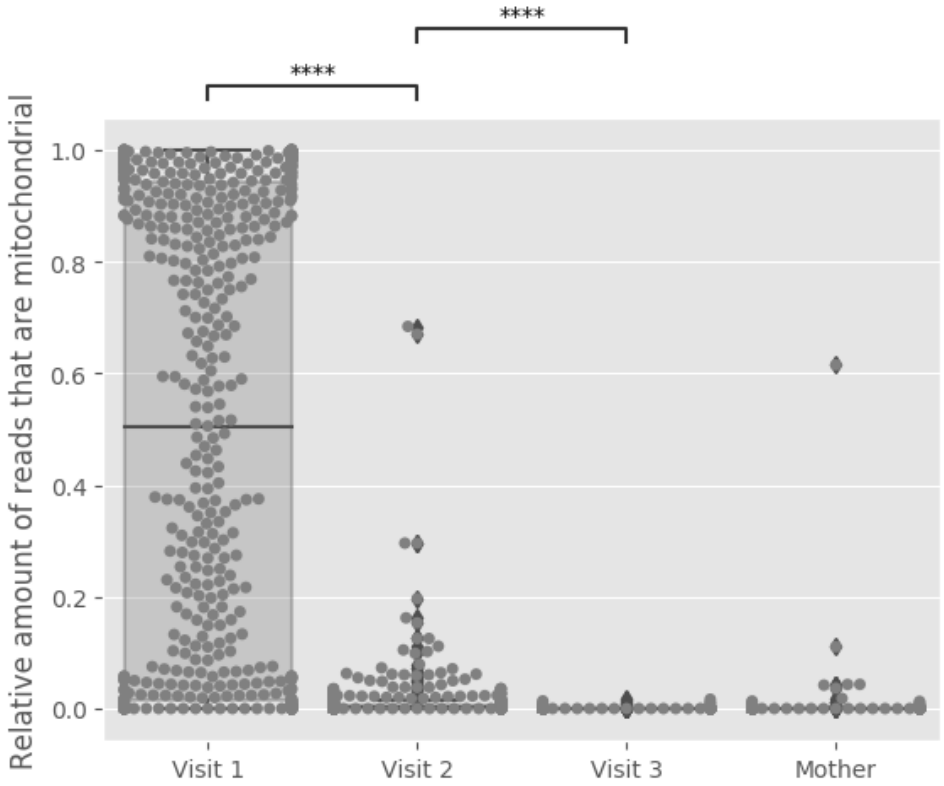


Fig. 3.6: Relative amount of mitochondrial reads across sample types. Every point represents one sample and the boxplots show the distribution across different samples types. Kruskal-Wallis tests with Bonferroni correction were performed between visit 1, 2 and 3. P-value annotation legend: ****, highly significant: p <= 1.00e-04.

One exemplary visit 1 sample was sequenced in triplicate to analyze the effects of varying levels of bacterial load in the source sample. The original sample consisted of >95 % mitochondrial reads. That amount dropped drastically when a bacterial Spike-in of 2 μl (12.54 ng) bacterial DNA (ZymoBIOMICS^{TM} Spike-in Control I consisting of *Imtechella halotolerans* and *Allobacillus halotolerans*) was introduced and decreased further when the amount of Spike-in was increased to 10 μl (62.7 ng). In the latter case nearly the entire microbiome consisted only of bacterial species from the Spike-in (see fig. 3.7).

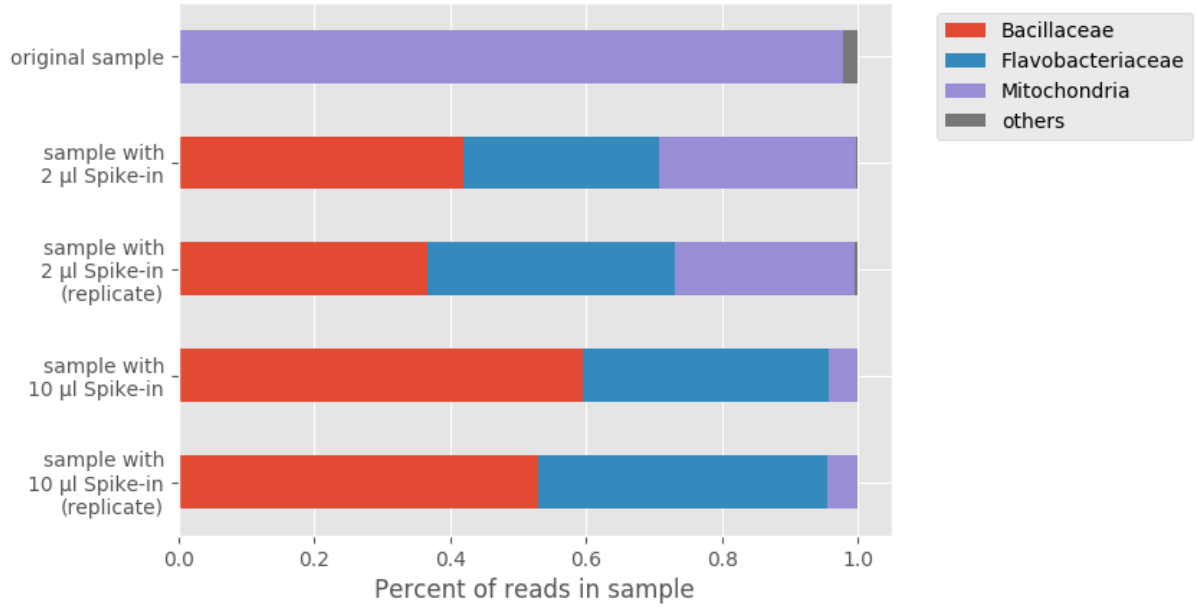


Fig. 3.7: From (Klopp et al., 2022): Microbiome composition on the family level of one meconium sample that has been sequenced in triplicate: 1. The original sample without Spike-in DNA 2. with 2 μl ZymoBIOMICS^{TM} Spike-in Control I (consisting of *Imtechella halotolerans*, Family Flavobacteriaceae and *Allobacillus halotolerans*, Family Bacillaceae) added and 3. with 10 μl added.

3.4 Taxonomic composition of study samples

16S rRNA taxonomic microbiome profiles were compiled for all samples. 5887 distinct ASVs (bacterial 16S rRNA gene sequences approximately with genus resolution) were observed across all rarefied samples (mean per sample: 46, min: 1, max: 242). The taxonomic composition and quantity of ASVs differed drastically across the different sample types (see fig. 3.8 and sections 3.4.1, 3.4.2, 3.4.3 and 3.4.4). For a full summary statistic of the most abundant ASVs see supplementary table Most_abudant_ASVs_statistics.csv.

Genera known to possibly originate from technical artifacts such as *Delftia, Flavobacterium, Pseudomonas, Burkholderia, Sphingomonas, Corynebacterium* and *Propionibacterium* (Douglas et al., 2020) were assessed in the sequenced samples and were only marginally present in the

microbiomes (see tab. 3.2). Alpha diversity - the diversity within samples - increases significantly

Tab. 3.2: Abundance of known technical contaminants in the samples. All data is shown as percentage of total microbiome per sample (relative abundance).

Genus	mean	min	25%	50%	75%	max
Corynebacterium	3.28	0.0	0.0	0.0	0.0	98.80
Pseudomonas	0.53	0.0	0.0	0.0	0.0	18.76
Sphingomonas	0.00	0.0	0.0	0.0	0.0	0.63
Delftia	0.00	0.0	0.0	0.0	0.0	0.04
Flavobacterium	0.00	0.0	0.0	0.0	0.0	0.08
Burkholderia	0.00	0.0	0.0	0.0	0.0	0.00

in accordance with the biological age across samples. Visit 1 samples show the lowest overall alpha diversity (measured by the Shannon diversity index). The diversity increases in visit 2 and visit 3 samples and mother samples statistically have the highest alpha diversity (see fig. 3.9).

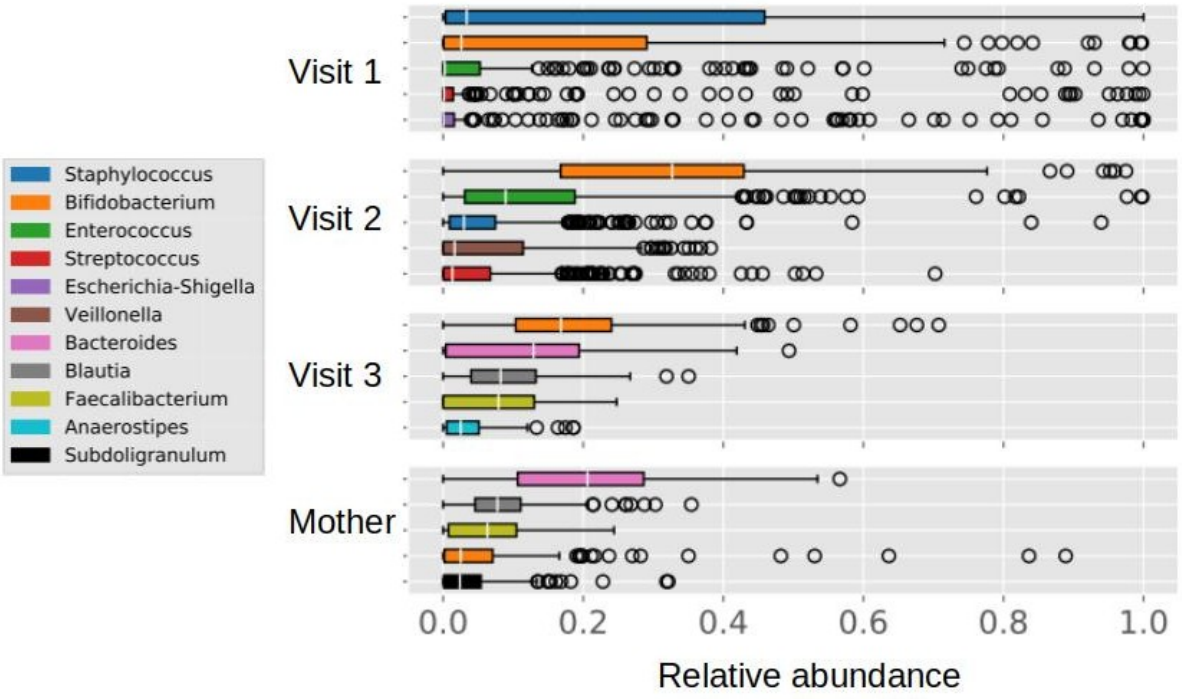


Fig. 3.8: Top 5 most abundant genera per sample type by median and their relative abundance across all sample types in the rarefied samples

These distinct microbiome differences result in a clustering based on sample type (see fig. 3.10). Every point in the graph represents the microbiome of one sample and their position relative to each other represents the overall similarity in their respective microbiome composition. The closer two points are, the more alike the microbiome analyzed. Maternal stool samples form a distinct cluster on the right side of the graph indicating stable, comparatively homogeneous adult microbiomes. Visit 2 samples from a bimodal cluster on the left side of the graph (see for more details 3.6.2). Visit 3 samples fall in between visit 2 samples and mother samples and

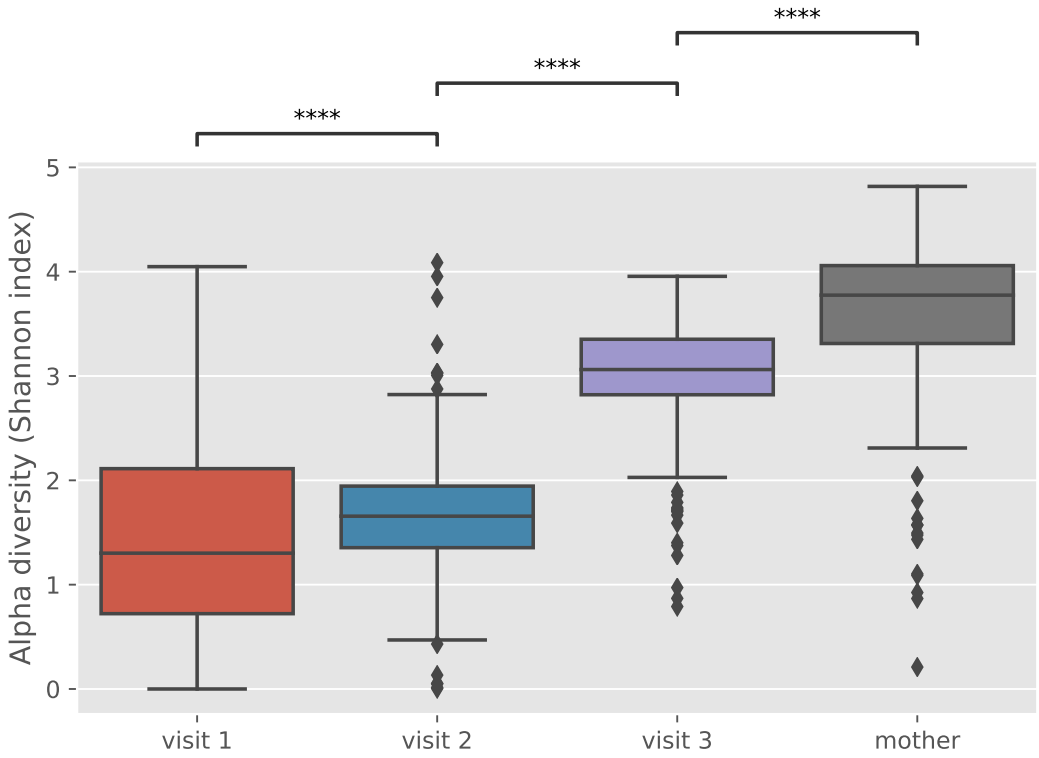


Fig. 3.9: The alpha diversity of the gut microbiome increases in accordance to the age of different sample types. Kruskal-Wallis tests with Bonferroni correction were performed to test for statistical differences. P-value annotation legend: ns: 5.00e-02 < p <= 1.00e+00, *: 1.00e-02 < p <= 5.00e-02, **: 1.00e-03 < p <= 1.00e-02, ***: 1.00e-04 < p <= 1.00e-03, ****: p <= 1.00e-04

their microbiomes represent an intermediate transition state between infancy and adulthood. Visit 1 samples have microbiomes with very high inter-sample variability that do not cluster well at any specific point in the graph. Thus, in contrast to the alpha diversity, the beta diversity - the diversity between different samples (based on unifrac distances) - is the highest for visit 1 samples and lowest in mother samples. After statistically controlling for differences in sample type and hospital collection site (PERMANOVA, adonis2 formula = distance weighted unifrac sample type + hospital site), 29.87 % of the variability in the microbiome between samples was statistically significantly explained by the different sample type (PERMANOVA: p-value < 0.001) and 1.62 % were explained by the hospital site at which the samples were collected (PERMANOVA: p-value < 0.001). Figure 3.11 shows the top 3 most abundant genera by sample type across the different hospital sites.

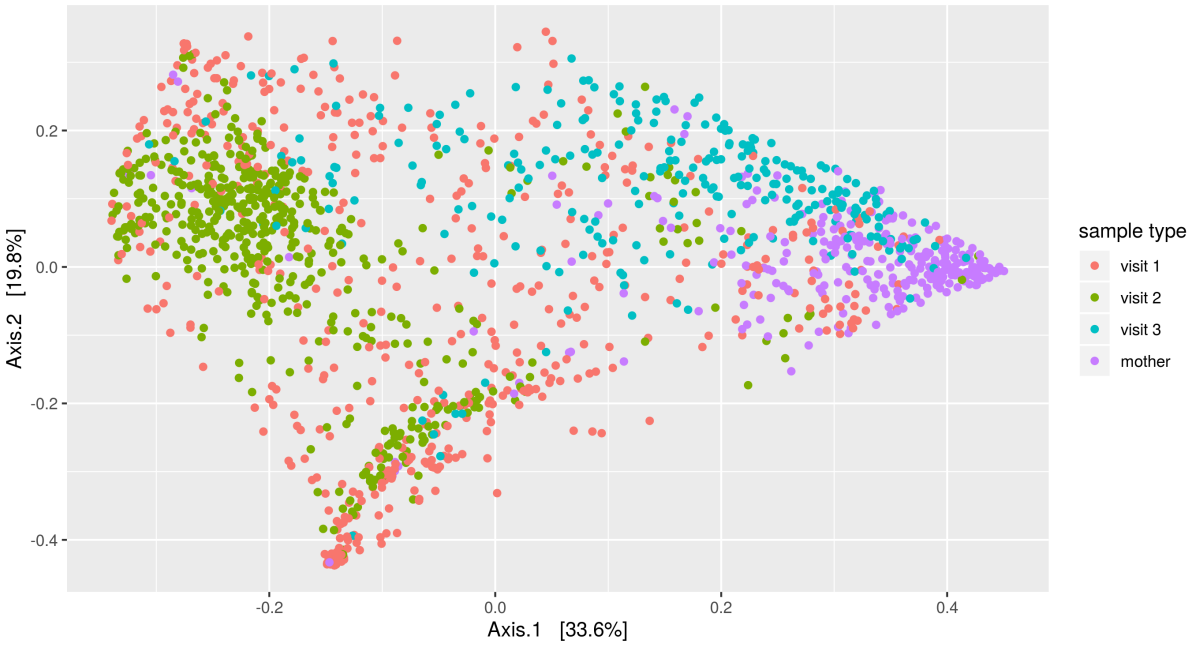
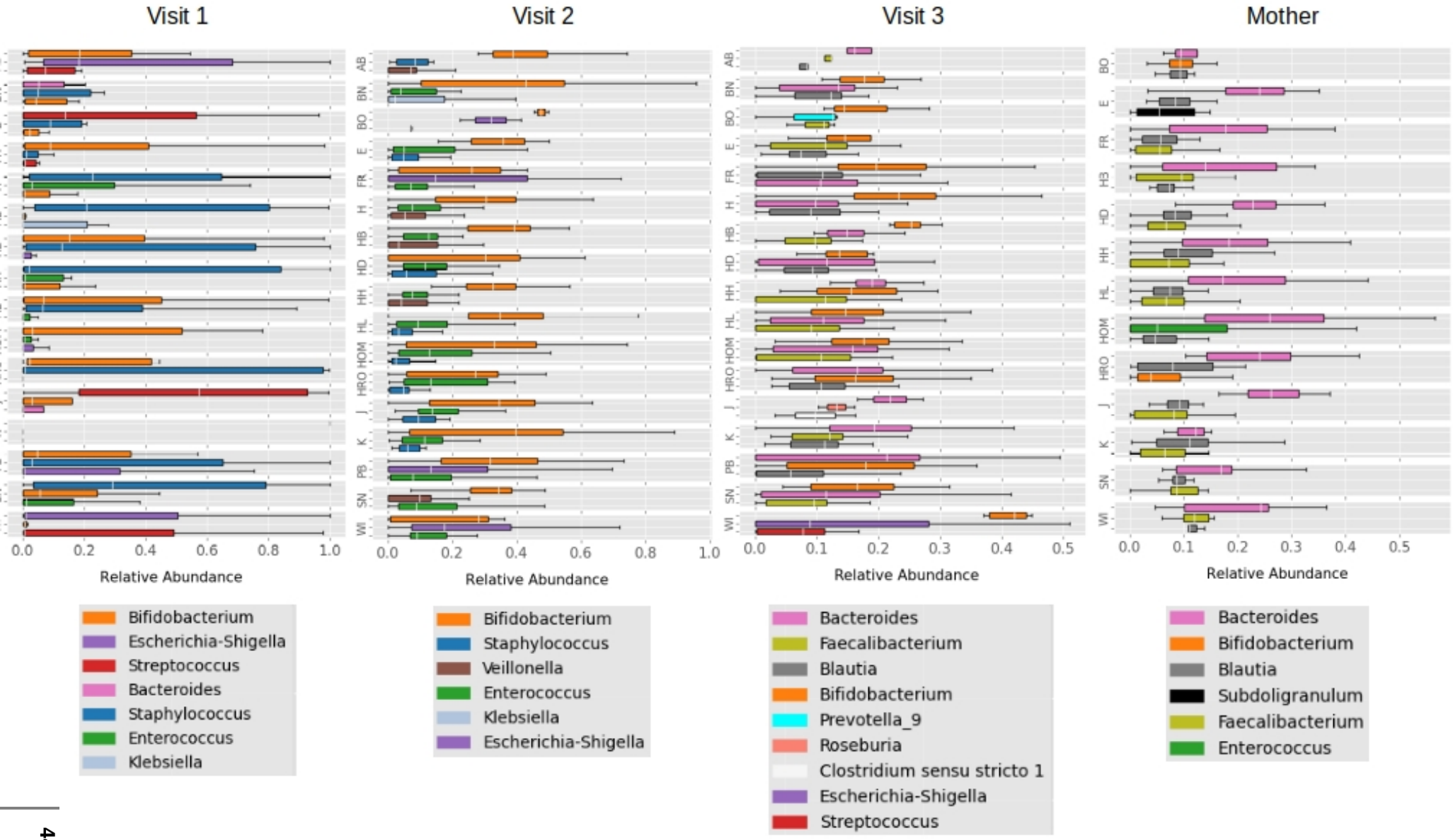


Fig. 3.10: Beta diversity across all sample types. Every point represents the microbiome of one sample. The closer two points, the higher the similarity in microbiome composition of their original sample. The percent values on the X- and Y-axis indicate the amount of variance explained by this principal coordinate.


Visit 1
Visit 2
Visit 3
Mother
AB
BN
BO
E
FR
H
HB
HD
HH
HL
HOM
HRO
J
K
PB
SN
WI
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.3
0.5
Relative Abundance
Bifidobacterium
Escherichia-Shigella
Streptococcus
Bacteroides
Staphylococcus
Enterococcus
Klebsiella
Veillonella
Prevotella_9
Roseburia
Clostridium sensu stricto 1
Faecalibacterium
Blautia
Subdoligranulum

3.4.1 Visit 1

Visit 1 fecal samples (N=565) were collected in the first days of the infants' lives. The main phyla that constituted the microbiomes were Firmicutes, Actinobacteriota, Proteobacteria, and Bacteroidota (see fig. 3.12). Some of the samples were dominated by only members of one phylum or even genus. The 5 most abundant genera were *Staphylococcus* (median 3.35 %, mean 26.5 %), *Bifidobacterium* (median 2.6 %, mean 8.27 %), *Enterococcus* (median 0.20 %, mean 8.26 %), *Streptococcus* (median 0.12 %, mean 7.11 %) and *Escherichia-Shigella* (median 0.04 %, mean 9.74 %) (see fig. 3.8). Microbiome composition of the infants was not significantly associated with the gestational age or the infants' biological sex. The hospital collection site had a significant impact on taxonomic composition and explained 9.17 % of the variance in the visit 1 samples (PERMANOVA p-value < 0.001). However, a permutation test for homogeneity of multivariate dispersions was also significant (p-value $= 0.046$) meaning that the different groups possibly have such a high internal dispersion that it interferes with the PERMANOVA analyses. Infants' biological sex did not have a significant influence on the microbiome.

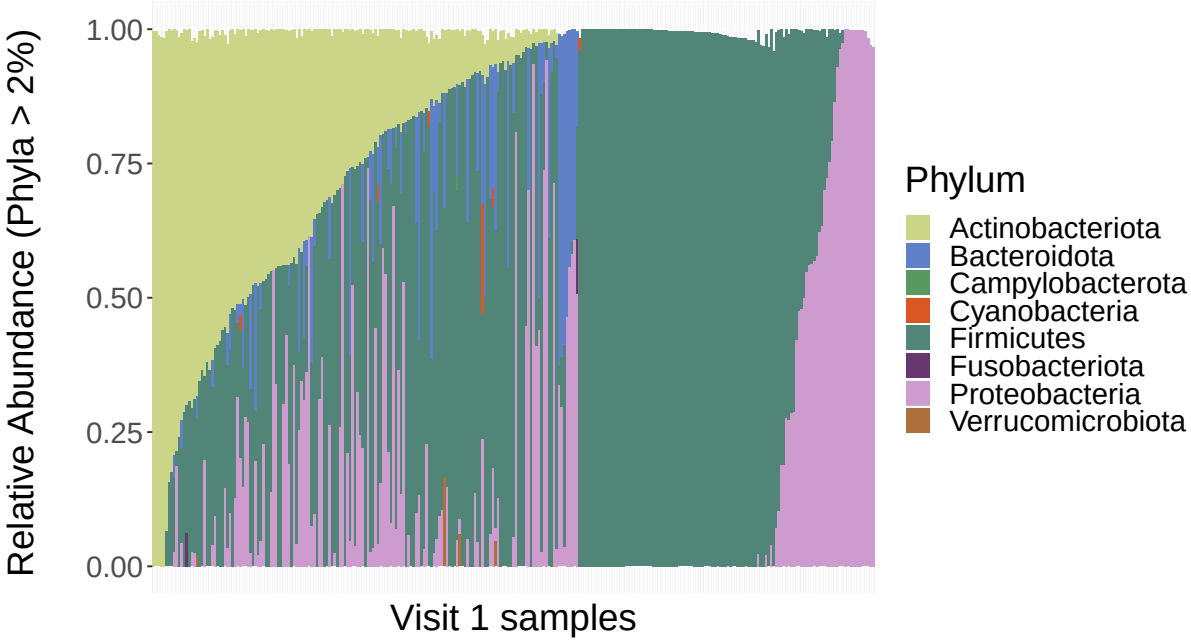


Fig. 3.12: Taxonomic composition of the visit 1 samples on the phylum level. Only Phyla with an overall abundance of more than 2 % across all samples are shown (N=565).

3.4.2 Visit 2

Visit 2 fecal samples (N=529) were collected after the study participants had either received probiotics or a placebo for 28 days. The microbiomes were principally composed of bacteria from the Phyla Firmicutes, Actinobacteriota, Proteobacteria and Bacteroidota (see fig. 3.13). Most abundant genera included *Bifidobacterium* (median 32.68 %, mean 30.56 %), *Enterococcus* (median 8.92 %, mean 13.45 %), *Staphylococcus* (median 2.96 %, mean 5.99 %), *Veillonella*

(median 1.64 %, mean 6.57 %) and *Streptococcus* (median 1.3 %, mean 5.33 %) across all visit 2 samples (see fig. 3.8). The degree of prematurity showed a significant effect (effect size = 1.5 %) on the microbiome at visit 2 (PERMANOVA p-value < 0.001). Hospital collection site was still significantly associated with changes in the microbiomes and explained 7.56 % of variance (PERMANOVA p-value < 0.001). The associated beta dispersion p-value was 0.008. Biological sex had no effect.

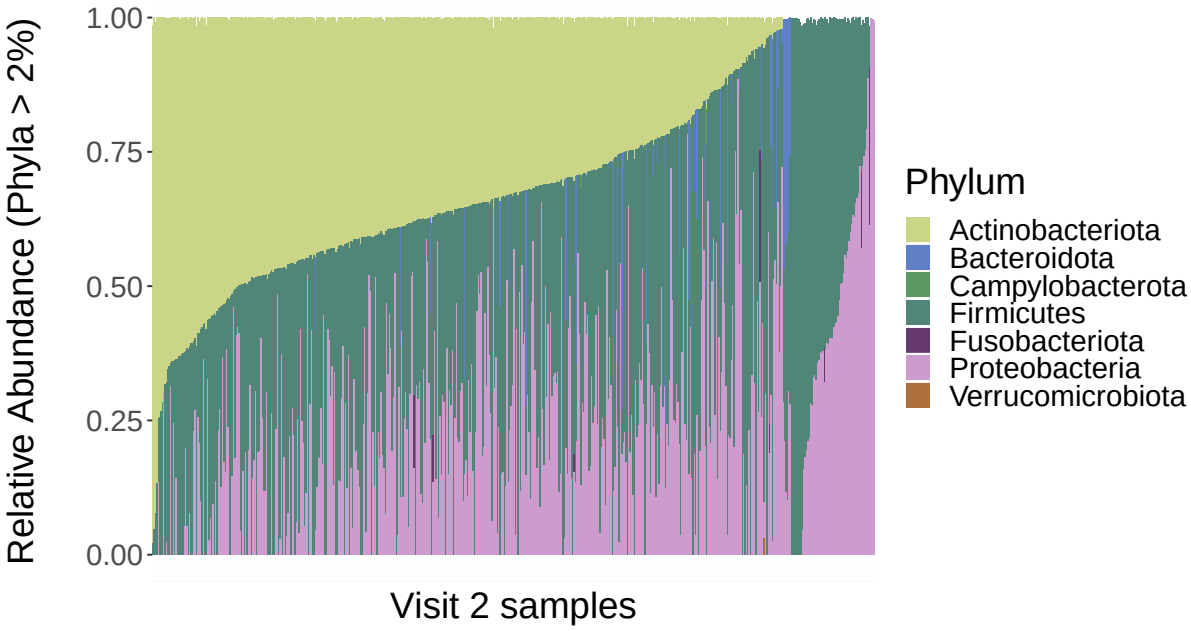


Fig. 3.13: Taxonomic composition of the visit 2 samples on the phylum level. Only Phyla with an overall abundance of more than 2 % across all samples are shown (N=529).

3.4.3 Visit 3

Visit 3 fecal samples (N=259) were collected 1 year after the infants' admission to the study. Major phyla in the microbiome were Firmicutes, Actinobacteriota, Bacteroidota and Proteobacteria (see fig. 3.14). Main genera were*Bifidobacterium* (median 16.82 %, mean 18.34 %), *Bacteroides* (median 12.89 %, mean 12.58 %), *Blautia* (median 8.21 %, mean 8.96 %), *Faecalibacterium* (median 7.89 %, mean 7.69 %) and *Anaerostipes* (median 2.5 %, mean 3.3 %) (see fig. 3.8). Hospital collection site explained 14 % of the differences in the microbiomes of one year old children (PERMANOVA p-value < 0.001). The associated beta dispersion p-value was 0.004. Gestational age had no longer a significant effect. Likewise, biological sex had no statistical significant impact.

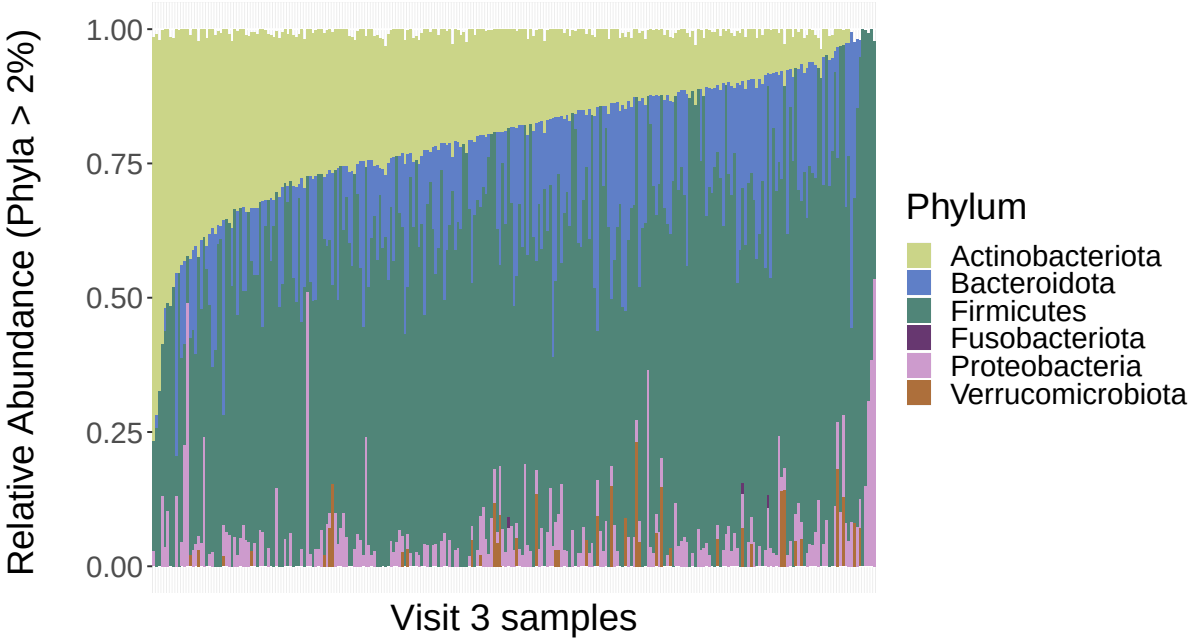


Fig. 3.14: Taxonomic composition of the visit 3 samples on the phylum level. Only Phyla with an overall abundance of more than 2 % across all samples are shown (N=259).

3.4.4 Mother Samples

Fecal samples of the infants' mothers were collected for a subset of infants (N=290) at the same time as visit 1 samples. Mothers' microbiomes resembled that of normal adults (Rinninella et al., 2019) and showed less inter-sample variation in the taxonomic configuration than the infant samples. The main phyla were Firmicutes, Bacteroidota, Actinobacteriota and Proteobacteria (see fig. 3.15). Top genera were *Bacteroides* (median 20.61 %, mean 20.11 %), *Blautia* (median 7.74 %, mean 8.23 %), *Faecalibacterium* (median 6.31 %, mean 6.59 %), *Bifidobacterium* (median 2.49 %, mean 5.6 %) and *Subdoligranulum* (median 2.39 %, mean 3.73 %) (see fig. 3.8). The microbiomes showed significant differences by hospital collection site. 11.43 % of variance was explained by the collection site (PERMANOVA p-value < 0.001). The associated beta dispersion test had a p-value of 0.024.

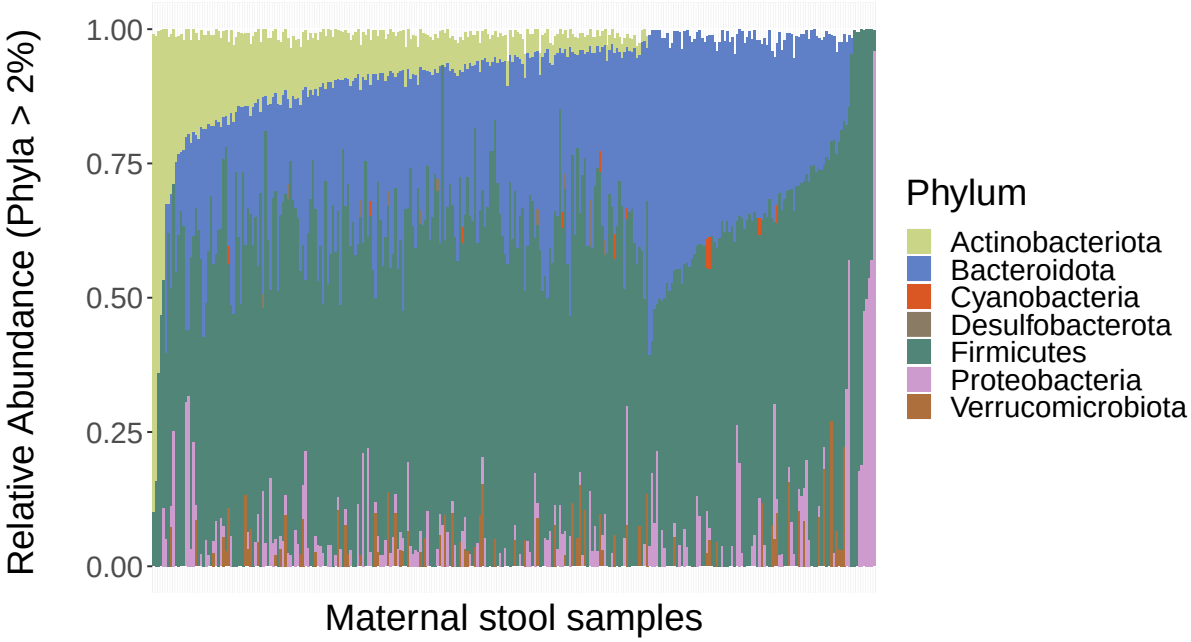


Fig. 3.15: Taxonomic composition of maternal stool samples on the phylum level. Only Phyla with an overall abundance of more than 2 % across all samples are shown.

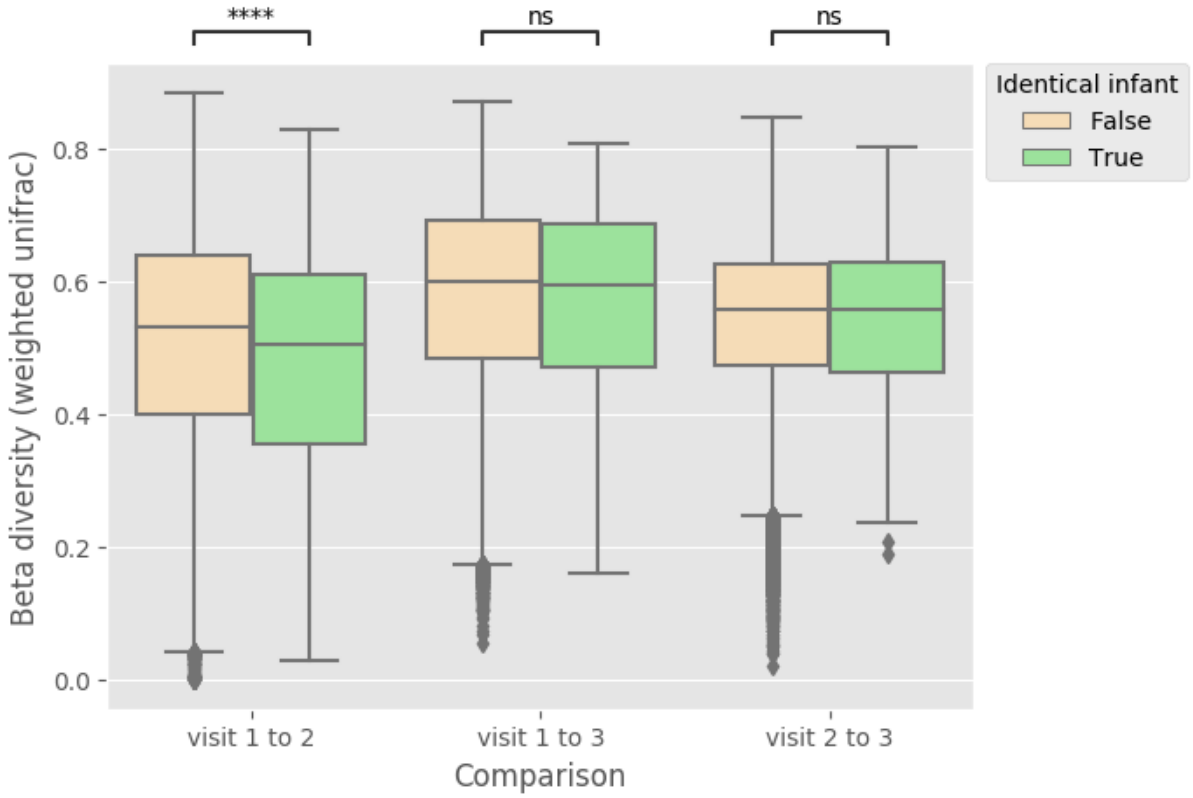
3.5 Similarities and differences in the microbiomes across sample types

The beta diversity between different visits and sample types was calculated to assess the likeness of their microbiomes. A beta diversity measure of 0 is defined as a complete identity of the microbiomes in regards to taxonomic composition. Microbiomes from visit 1 and visit 2 were significantly more similar in samples from the same infant than in samples from different infants (Kruskal-Wallis, p-value < 0.001). This effect was no longer visible in comparisons between microbiomes from visit 1 and 3 or visit 2 and 3 (see fig. 3.16 a).

To see if the maternal gut microbiome had an impact on the infant's microbiomes across visits, beta diversity was calculated for every pair of mother and infant (related and unrelated). Relatedness had no significant effect on the beta diversity between infant and mother samples (see fig. 3.16 b).

Comparisons between visit 1 samples and maternal samples and comparisons between visit 1 samples from different hospital collection sites were performed on the ASV level to assess potential transfer of microbes. Infants from the same hospital shared a significantly higher proportion of their ASVs than infants from different locations (Kruskal-Wallis, p-value < 0.001). The same effect could be observed for unrelated infants and mothers from different hospital sites (Kruskal-Wallis, p-value < 0.0001). The proportion of shared ASVs between mothers did not show a significant association with the hospital collection site. A strong and significant association was found between the relatedness of mothers and infants and the amount of ASVs they shared within a given hospital site (Kruskal-Wallis, p-value < 0.001).

(a)



(b)

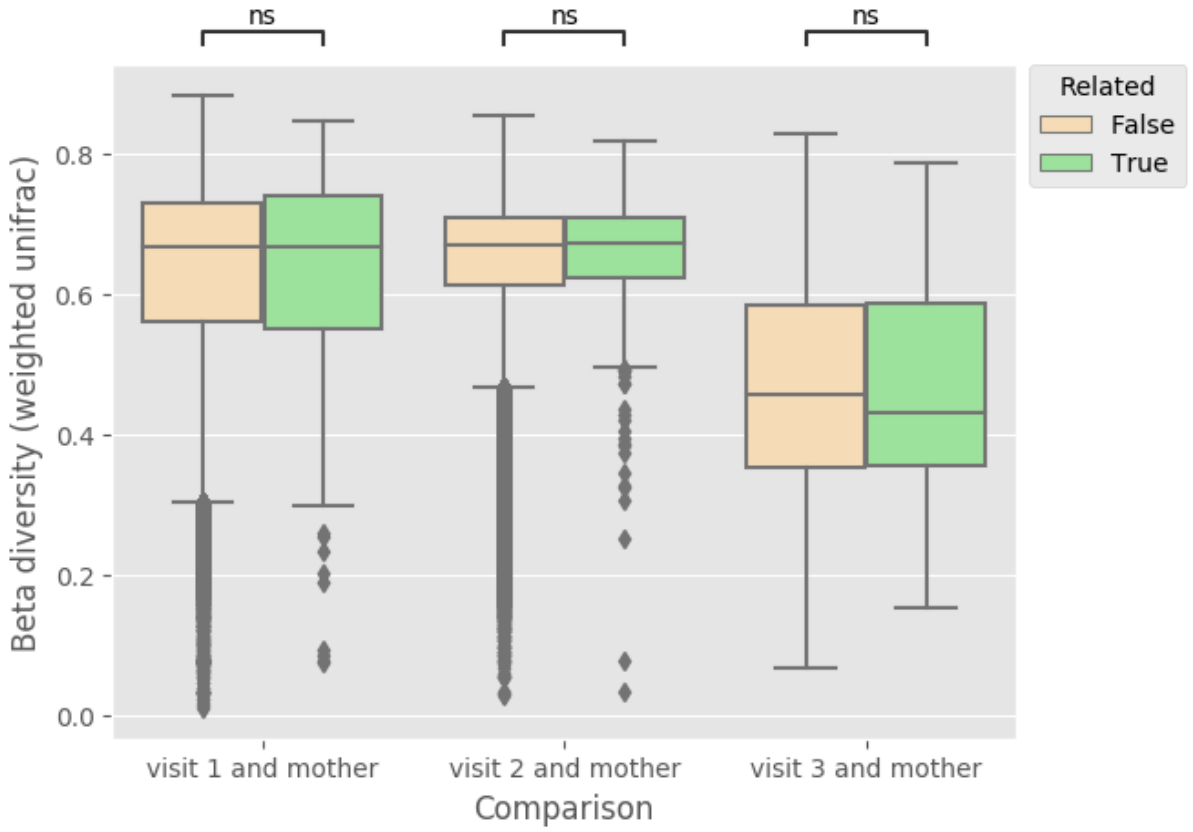


Fig. 3.16: Beta diversity between microbiomes (a) from different visits of identical and different infants and (b) from different visits and the associated mother sample of related and unrelated infants. Kruskal-Wallis tests were performed between all groups. P-value annotation legend: ns, not significant: $p > 0.05$, ****, highly significant: $p <= 1.00e-04$.

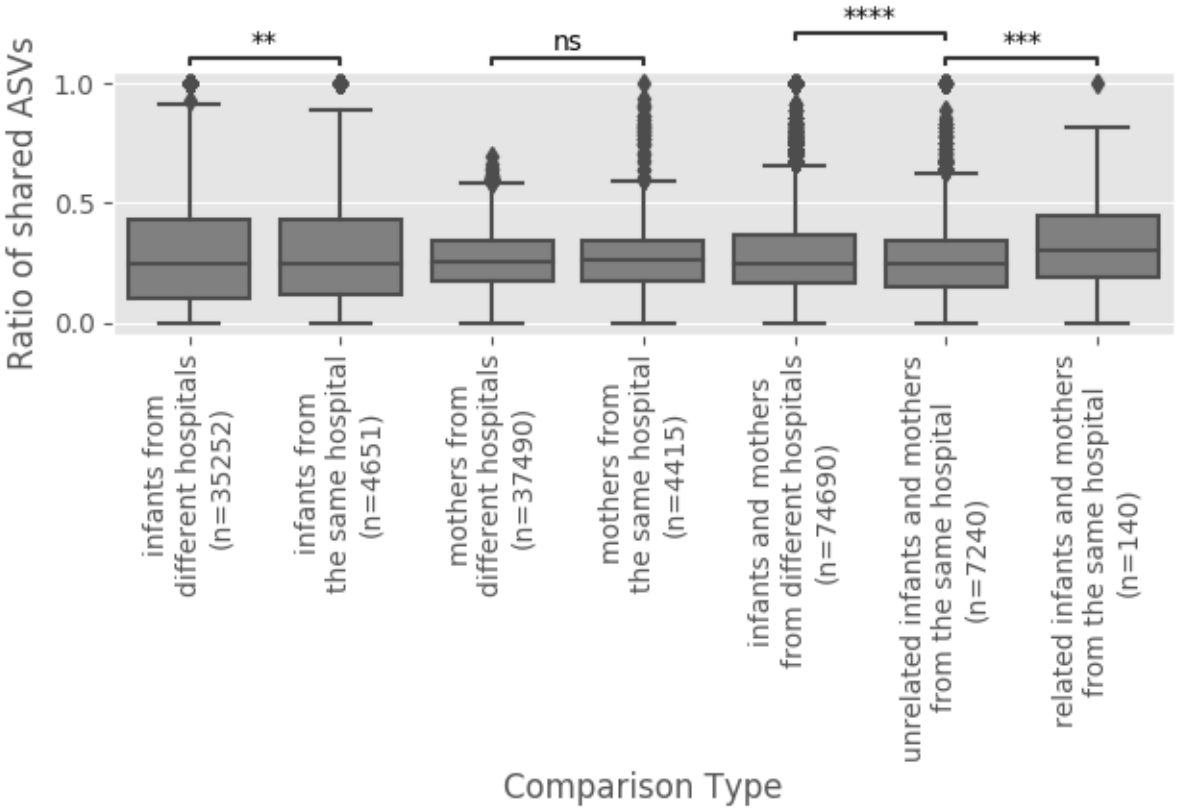


Fig. 3.17: Proportion of shared ASVs for different sample sets composed of mother and visit 1 samples. Each boxplot represents one type of comparison. On the x-axis the number (n) of pairwise comparisons for each comparison type is listed. Kruskal-Wallis tests with Bonferroni correction were performed between groups. P-value annotation legend: ns, not significant: $p > 0.05$, ** significant: $p < 1.00e\text{-}03$, *** significant: $p < 1.00e\text{-}04$, ****, highly significant: $p <= 1.00e\text{-}04$.

3.6 Analysis of the probiotic impact during the study

Three species (*Bifidobacterium animalis* subsp. lactis (BB-12), *Bifidobacterium infantis* and *Lactobacillus acidophilus* (La-5)) belonging to two probiotic genera, *Lactobacillus* and *Bifidobacterium*, were administered daily to half of the infants during the first 28 days of life. The other half of the population received a placebo preparation. To assess the effects of the probiotic supplementation, first the microbial content of representative verum and placebo capsules was analyzed. Second, the infants' microbiomes were investigated in regards to changes through the probiotic genera.

3.6.1 Comparison of placebo and probiotic capsules

9 verum capsules and 12 placebo capsules were analyzed in regards to their DNA content and their microbial composition. Placebos and verums differed drastically in DNA content and DNA concentration with average total amounts of 0 ng and 18755.55 ng, respectively (see fig. 3.18). Verum samples almost exclusively consisted of the two expected genera, *Lactobacillus* and *Bifidobacterium*. Placebo samples had vanishingly low counts of sequencing reads and showed a diverse set of bacterial genera. However, the composition closely mirrored that of negative controls (see fig. 3.19).

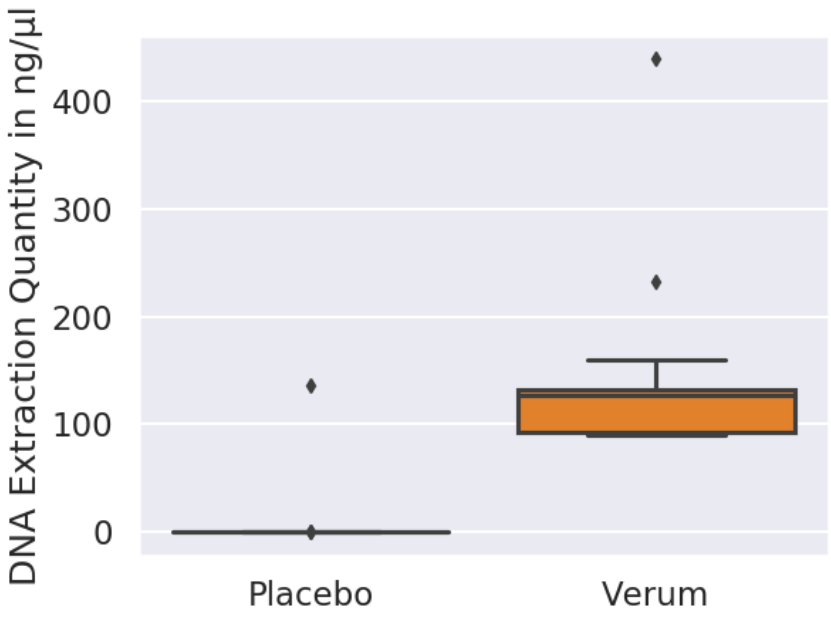


Fig. 3.18: Differences in DNA concentration in verum and placebo capsules.

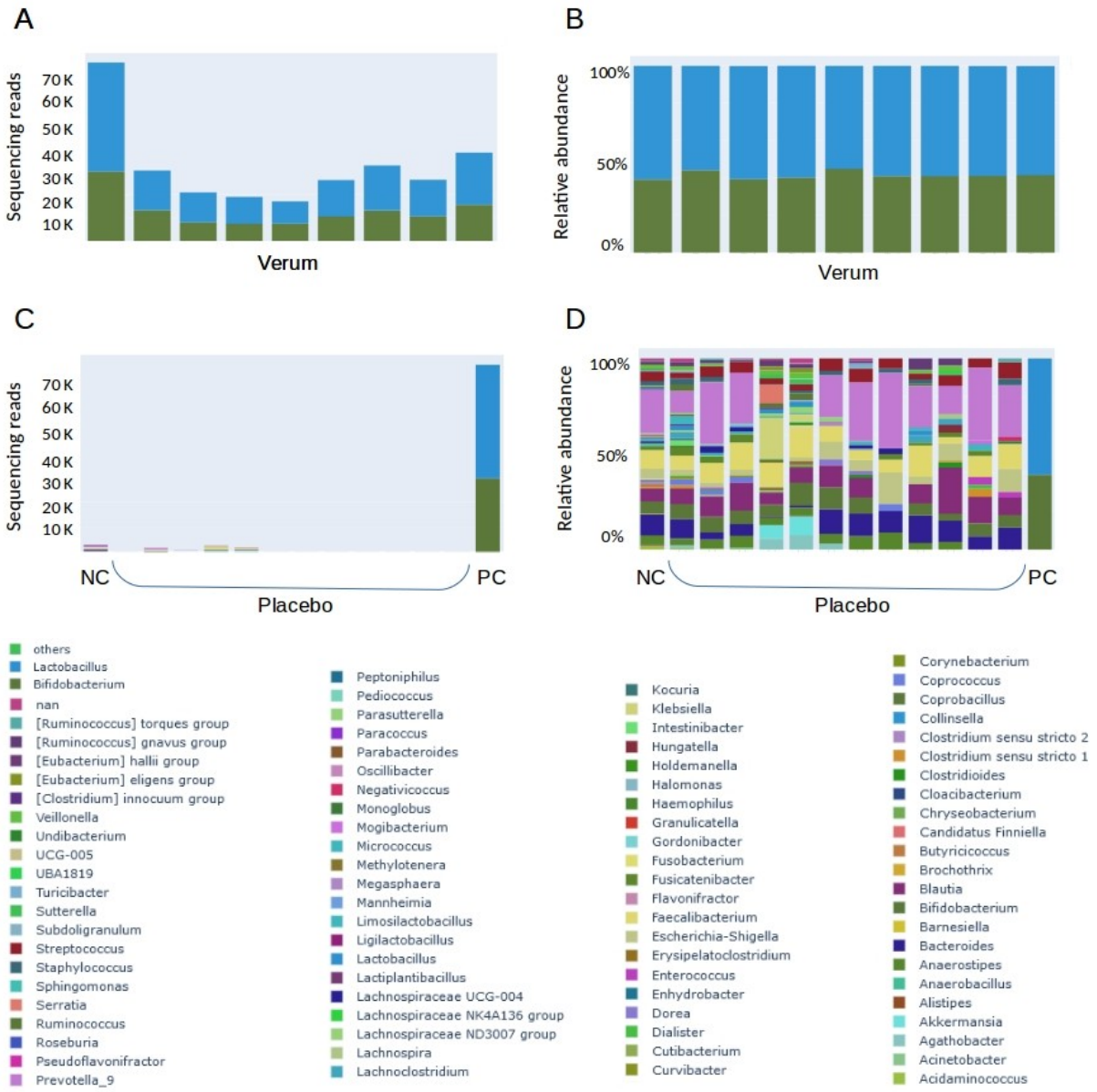


Fig. 3.19: Microbiome composition of verum and placebo capsules and controls (A) absolute read count of the 9 verum samples (B) relative composition of the microbiomes in the verum samples (C) absolute read count in the 12 placebo samples, a negative control (NC, water) and a positive control (PC, verum) (D) relative composition of the microbiomes in the placebo and control samples.

3.6.2 Prevalence of the probiotic taxa

In the population of visit 1 samples most do not show any prevalence of *Lactobacillus* and a highly uneven distribution of *Bifidobacterium*. The relative abundance of *Bifidobacterium* ranges from 0 % to 100 % (median 10.55 %, 25th percentile 0.31 %, 75th percentile 37.39 %). At visit 2 the relative abundance and prevalence of *Lactobacillus* only increases marginally with 18.33 % of samples possessing relative abudances of greater than 1 % of the taxon (mean 1.39 %). In contrast, *Bifidobacterium* showed a bi-modal abundance distribution with 15 % of samples having very few members of the taxon (< 1 %) and 85 % displaying a normal distribution. The abundance of *Bifidobacterium* ranges from 0 % to 97.69 % (median 32.8 %, 25th percentile 17.03 %, 75th percentile 43.02 %). *Lactobacillus* is virtually not present at visit 3 and in mother samples (mean 0.11 and 0.22, respectively). At visit 3 *Bifidobacterium* is an established member of the infants microbiome (median 16.65 %), while its prevalence is diminished in the adult samples (median 2.35 %) (see fig. 3.20).

For a subset of samples, information about the group affiliation was available (A or B, one of them being the probiotic group the other the placebo group). Highly significant differences in the microbiome of visit 2 samples were found on the basis of their group affiliation when controlling for hospital collection site (see fig. 3.21 ,PERMANOVA p-value < 0.001). The major driver for these differences was the relative abundance of *Bifidobacterium*. This taxa is strongly correlated with the principal variance explaining dimension of the PCoA plot (see fig. 3.22, spearman correlation coefficient 0.8, p-value < 0.0001). Interestingly, very low abundances of *Bifidobacterium* are almost exclusively found in the treatment group B, while treatment group A shows a wide range of *Bifidobacterium* abundance.

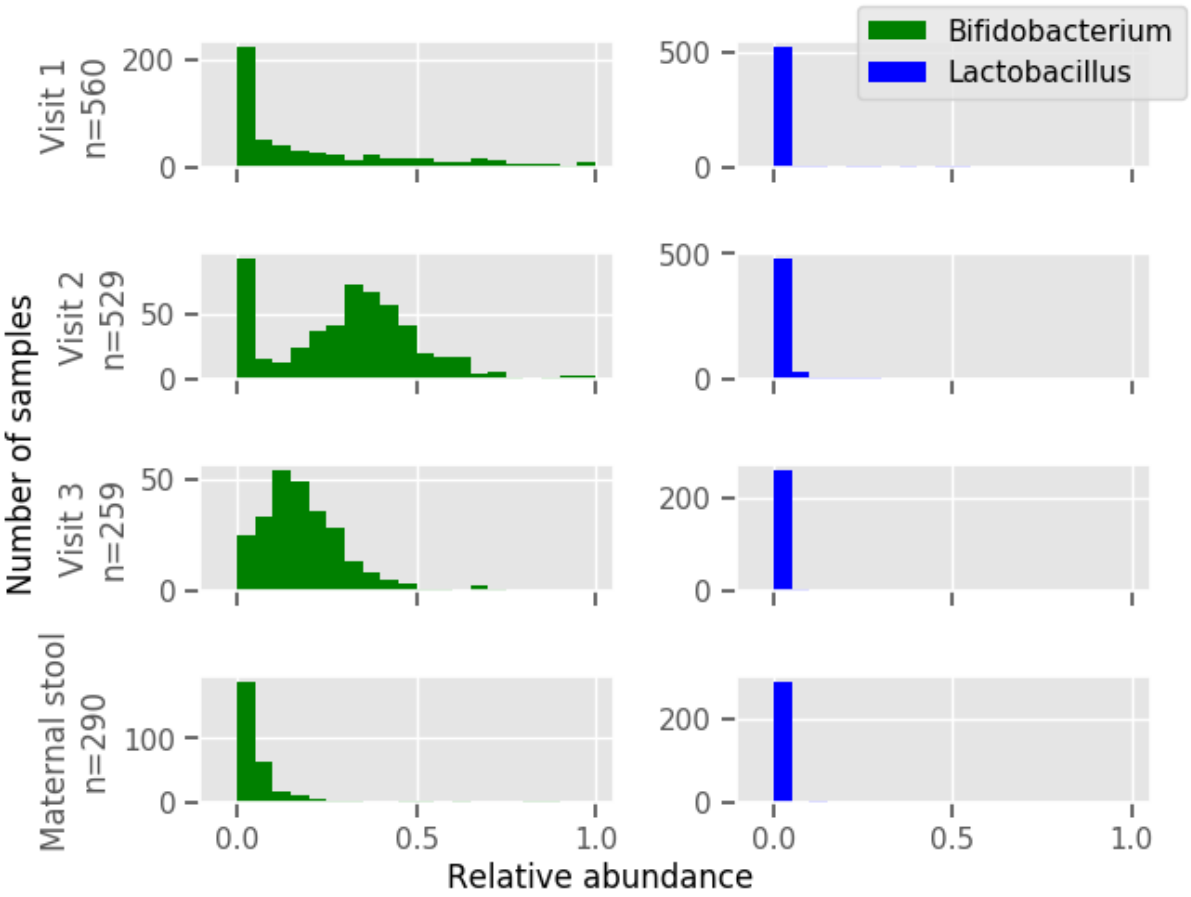


Fig. 3.20: Distribution of the relative abundance of the probiotic taxa, *Lactobacillus* and *Bifidobacterium*, across different sample types

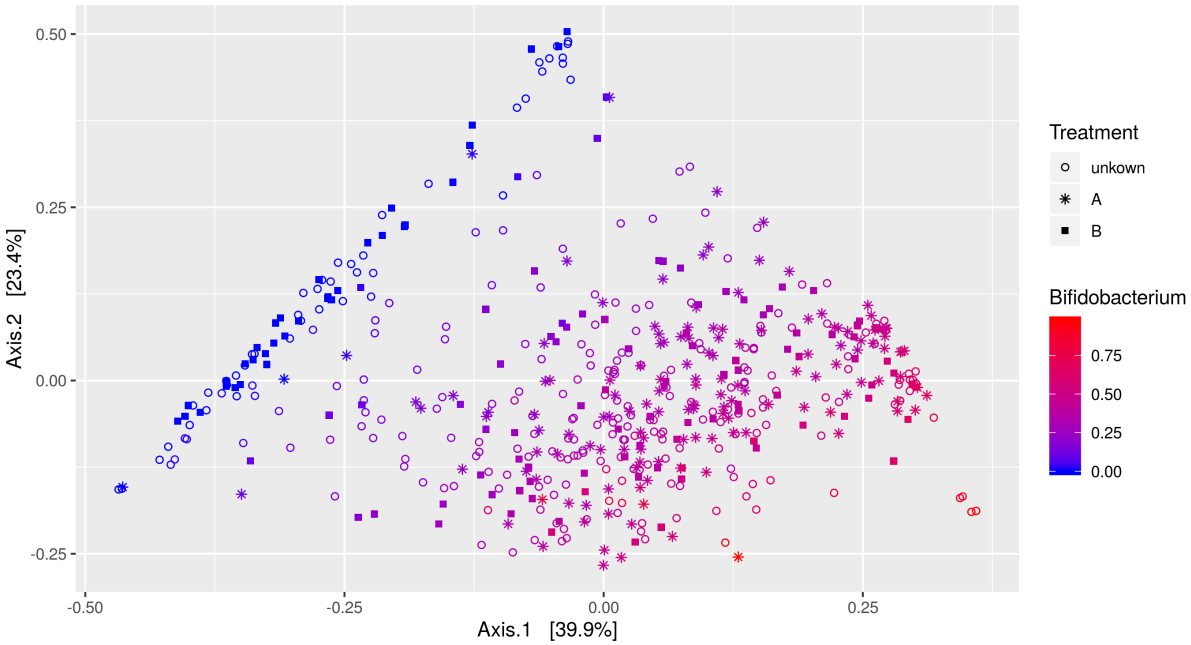


Fig. 3.21: Principal coordinate analysis of the beta diversity (weighted unifrac distances) of visit 2 samples. Every point represents the microbiome of one sample. Samples are colored by their relative abundance of *Bifidobacterium* in the microbiome. Different shapes indicate information about the administered treatment.

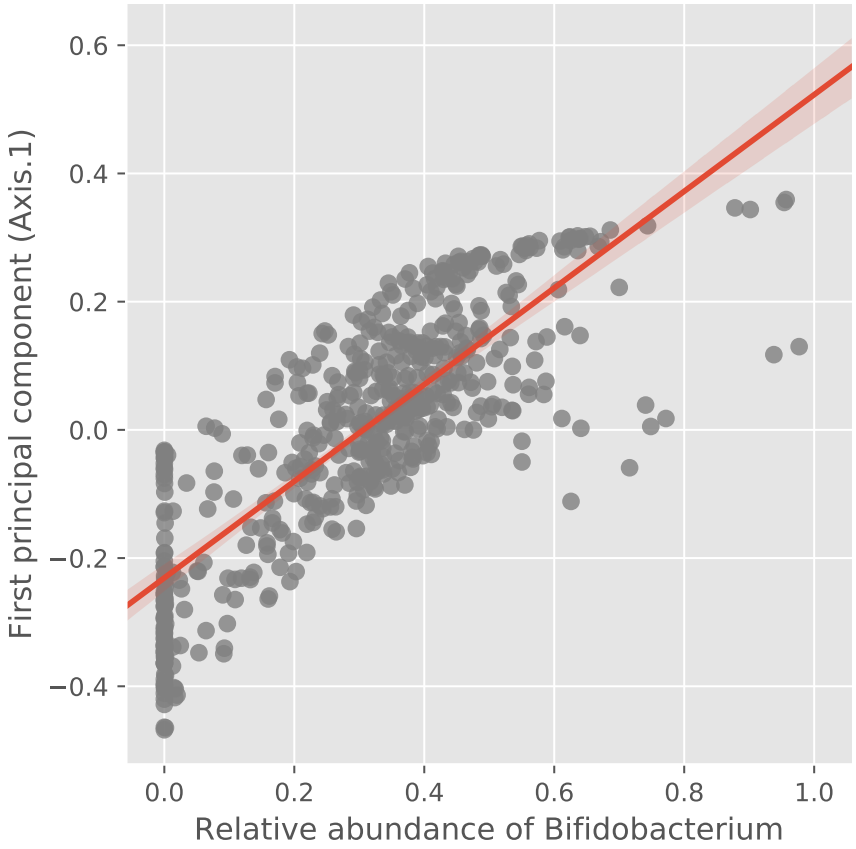


Fig. 3.22: Significant correlation of the relative abundance of *Bifidobacterium* and the first principal component explaining 39.9 % of difference in the visit 2 samples. Spearman correlation coefficient 0.8, p-value < 0.0001

3.6.3 Effect of the treatment on diversity and taxonomic composition

For a subset of PRIMAL samples, blinded data about the treatment group affiliation was available. 141 visit 2 samples were classified as treatment A and 128 as B. For visit 3 samples, 68 were classified as treatment A and 58 as B. It was assessed if any of the three pillars most often cited as signs of dysbiosis (loss of overall diversity, loss of beneficial microbial bacteria and bloom of pathobionts) were significantly different between the treatment groups.

There was no significant difference in the alpha diversity of either visit 2 or visit 3 samples (see fig. 3.23). In visit 2 samples, the relative abundance of the commensal bacteria *Bifidobac-*

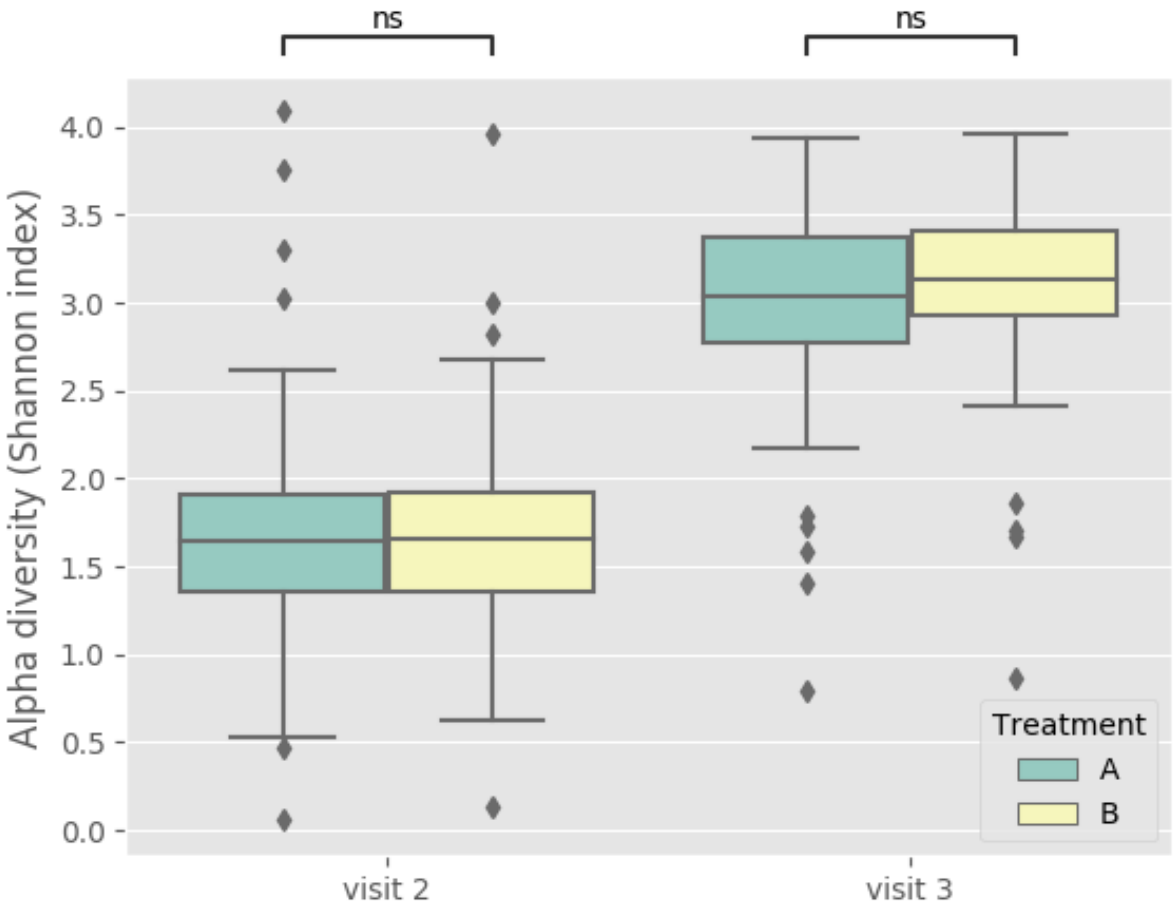
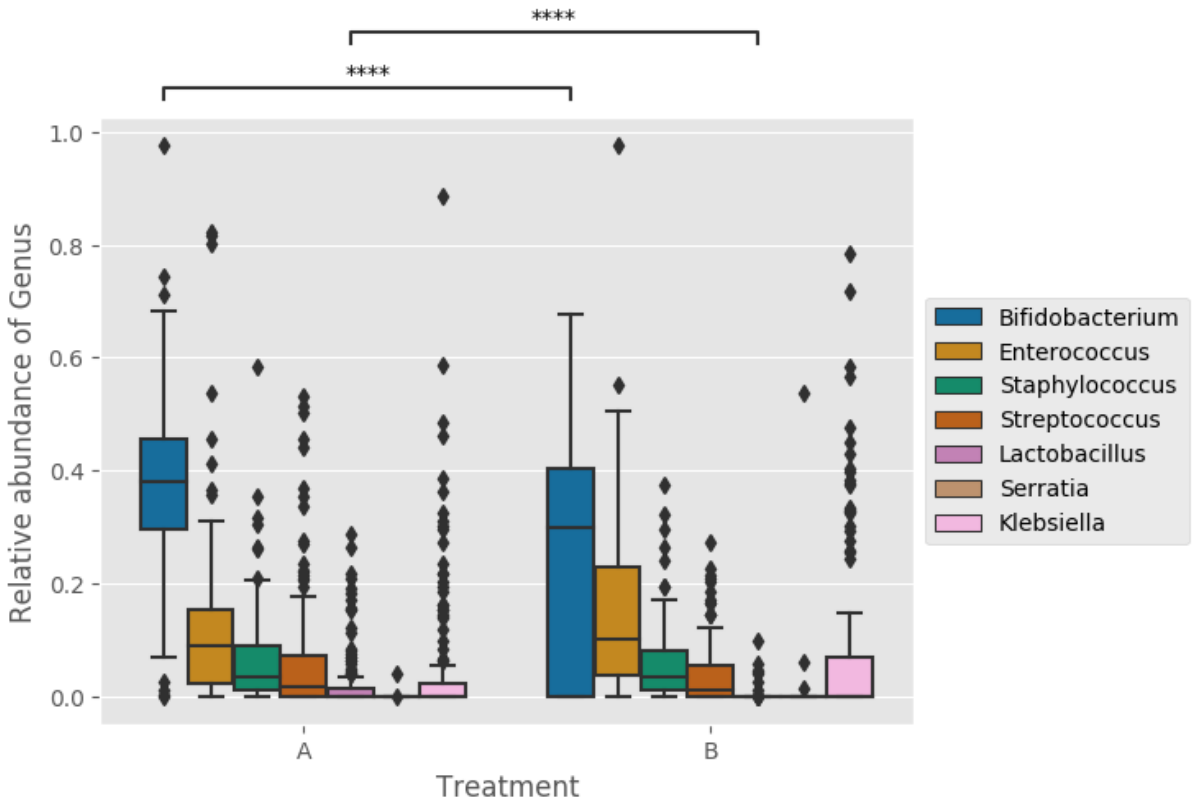


Fig. 3.23: Comparison of the alpha diversity between the two different PRIMAL treatments. Kruskal-Wallis tests with Bonferroni correction were performed between all groups. P-value annotation legend: ns, not significant.

terium and *Lactobacillus* was significantly increased in the treatment A group. None of the genera with potential pathogenic members (*Enterococcus*, *Staphylococcus*, *Streptococcus*, *Serratia* and *Klebsiella*) were differentially abundant. In visit 3 samples, none of the tested taxa displayed a significant difference between treatment groups.

(a)



(b)

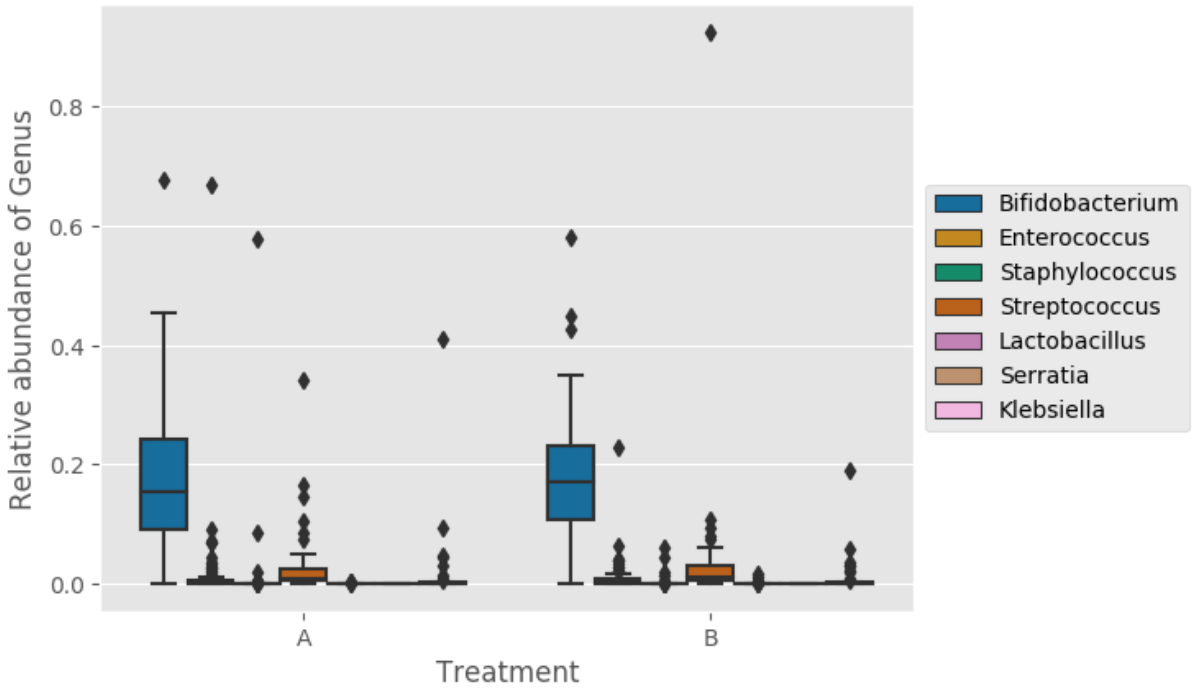


Fig. 3.24: Differences in relative abundances of the most important bacterial taxa (the two probiotic taxa and taxa with potentially pathogenic members) between the two different PRIMAL treatments in (a) visit 2 and (b) visit 3 samples. Kruskal-Wallis tests with Bonferroni correction were performed between all groups. Only significant differences are displayed for legibility. P-value annotation legend. ****, highly significant: p <= 1.00e-04.

3.7 Dysbiosis

To investigate possible dysbiotic states in the infants' microbiomes, different approaches were utilized. First, data from conventional microbiological analysis based on the KRINKO guidelines were used to look for specific pathogens and associated antibiotic resistance capacities in the visit 2 samples. Second, amplicon target sequencing and whole genome shotgun sequencing data was used to assess the taxonomic composition, interactions of microbes and the presence of antibiotic resistance genes. Third, publicly available sequencing data from healthy, term-born infants was used as a reference to compare the composition of microbiomes.

3.7.1 Data from conventional microbiology in comparison to NGS

Conventional microbiological analyses and DNA-based taxonomic profiling were performed for visit 2 samples. Microbiological investigations looked for specific species from six different bacterial genera (*Staphylococcus*, *Serratia*, *Klebsiella*, *Escherichia*, *Enterobacter*, *Citrobacter*). Their results were compared with findings from 16S rRNA gene sequencing (see fig. 3.25). The majority of analyzed samples produced a negative signal with both methods. A positive finding from the conventional microbiology approach was most associated with a detection of the identified genus through 16S rRNA gene sequencing in most instances. Notable exceptions occurred in the genera *Enterobacter* and *Klebsiella* were in 18.11 % and 1.89 % of cases the 16S rRNA gene sequencing was not congruent in its results, respectively. Sanger sequencing of pure cultures of the representative species (*Klebsiella pneumoniae* and *Enterobacter cloacae*) investigated by the conventional microbiology of the aforementioned genera revealed a 100 % identity of their respective 16S rRNA V4 regions that were used to discriminate taxa (see fig. 3.26). Detection of a genus in 16S rRNA gene sequencing was often not accompanied by a positive finding through conventional methods. This was especially relevant in the genera *Staphylococcus*, *Escherichia* and *Klebsiella*, which were detected in the microbiomes of 85.47 %, 41.5 % and 27.35 % of samples without being reported in the microbiological analysis, respectively.

There was no clear association between the relative abundance of any genus and the respective result from the microbiology (see fig. 3.27). The amount of detection of genera through 16S rRNA gene sequencing likewise was not a good predictor for microbiology results. This can be seen in the cases of *Escherichia* and *Klebsiella*, which have approximately the same rate of discovery through sequencing (45 % and 47 %) but differ strongly in the rate of discovery through conventional methods (3 % and 21 %).

The conventional methodology also had no predictive power in regards to the alpha diversity or relative abundance of *Bifidobacterium* or *Lactobacillus* in a given sample (see fig. 3.28). The groups stratified by positive and negative results were not statistically different (p-value > 0.05).

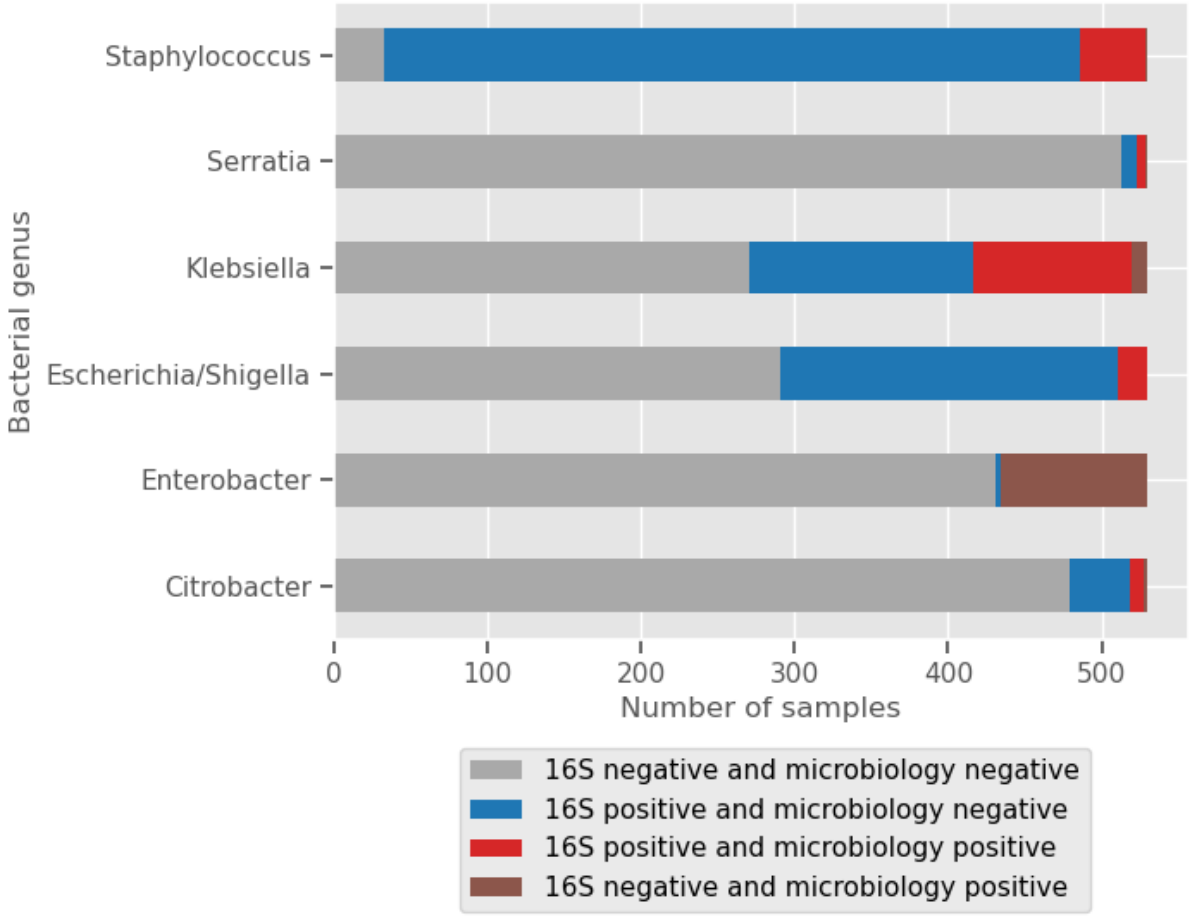


Fig. 3.25: Results from the analysis of visit 2 samples (N=530) using two different methods, conventional microbiology and 16S rRNA gene sequencing. 16S rRNA analysis was considered positive, when at least 0.1% of reads in a given microbiome were assigned to the respective genus.

```
K.pneumoniae      GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGTTACTGGGGCGTAAAGC
E.cloacae         GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGTTACTGGGGCGTAAAGC
                  ************************************************************

K.pneumoniae      GCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCAT
E.cloacae         GCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCAT
                  ************************************************************

K.pneumoniae      TCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAA
E.cloacae         TCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAA
                  ************************************************************

K.pneumoniae      TGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
E.cloacae         TGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGC
                  ************************************************************

K.pneumoniae      TCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCCGGTAGTC
E.cloacae         TCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCCGGTAGTC
                  **************************************************
```

Fig. 3.26: 100 % identity of the multiple sequence alignment of the V4 region of the 16S rRNA genes of *Klebsiella pneumoniae* and *Enterobacter cloacae*. The amplification/sequencing primers are colored in red (forward) and blue (reverse).

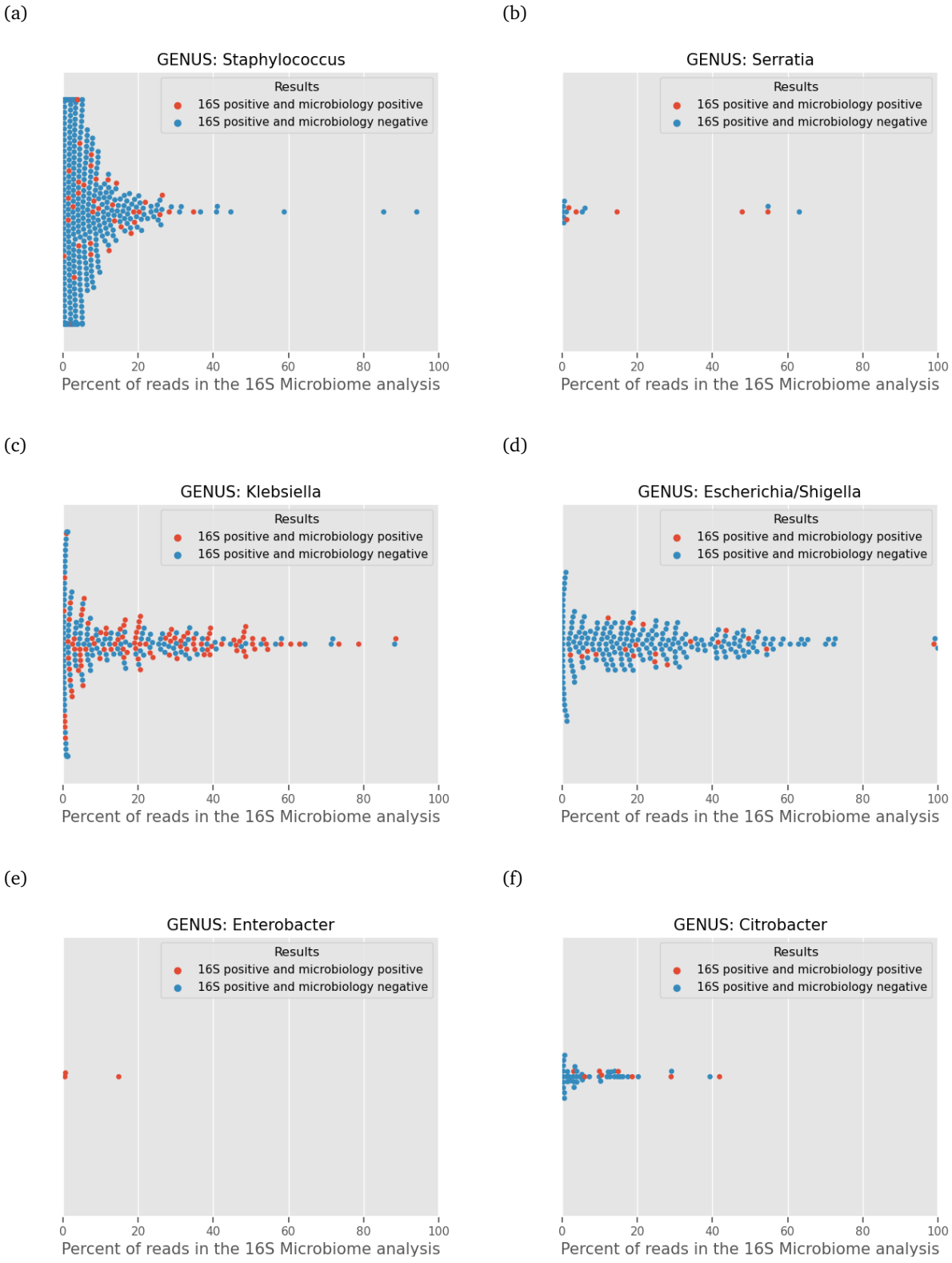
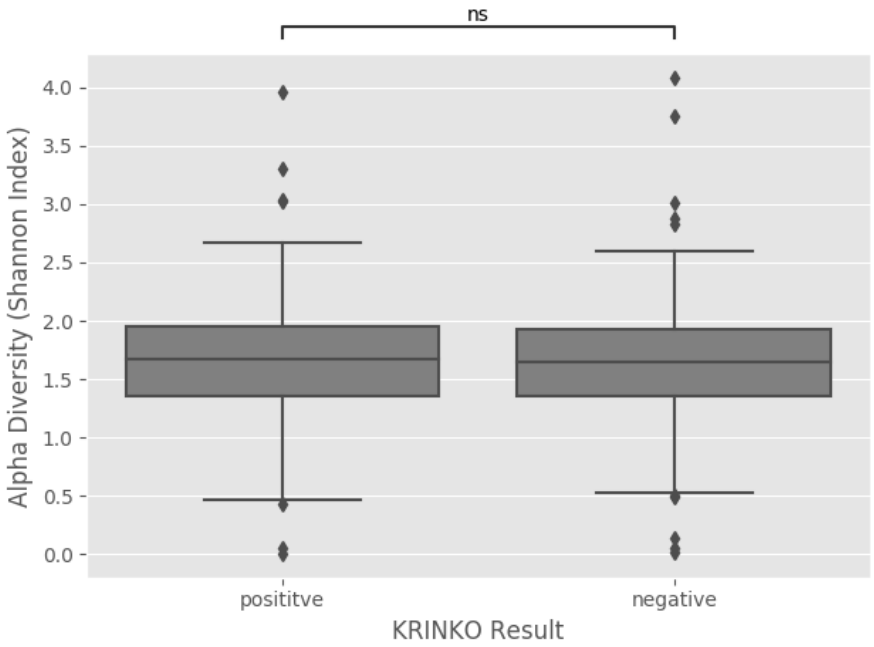
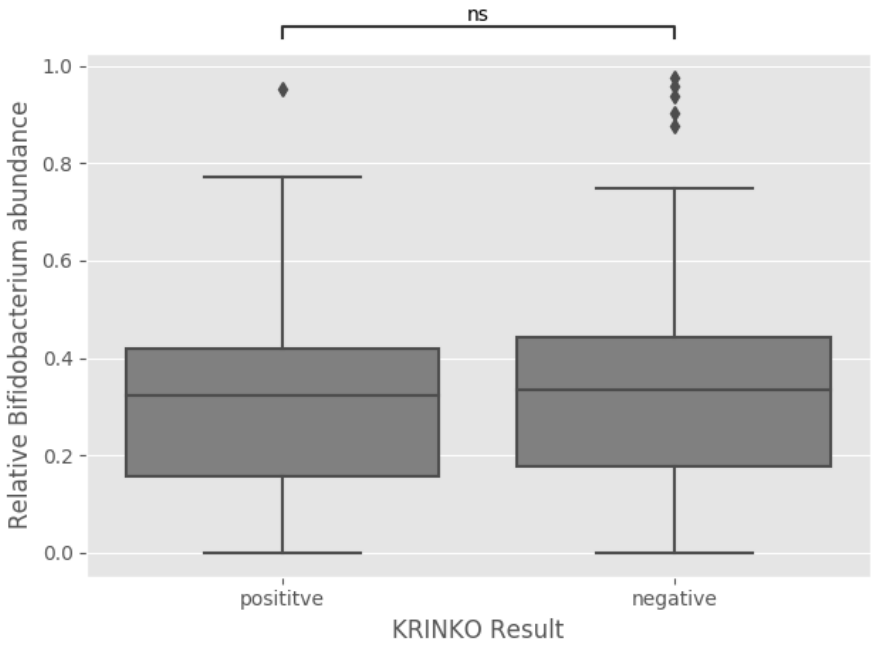


Fig. 3.27: (Dis)agreement of 16S rRNA gene sequencing and conventional microbiological taxa identification. Every dot represents one sample in which the respective genus was found through 16S sequencing (relative abundance > 0.1 %). Red dots indicate agreement between the methods; Blue dots disagreement.

(a)


ns
Alpha Diversity (Shannon Index)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
posititve
negative
KRINKO Result

(b)


ns
Relative Bifidobacterium abundance
1.0
0.8
0.6
0.4
0.2
0.0
posititve
negative
KRINKO Result

(c)

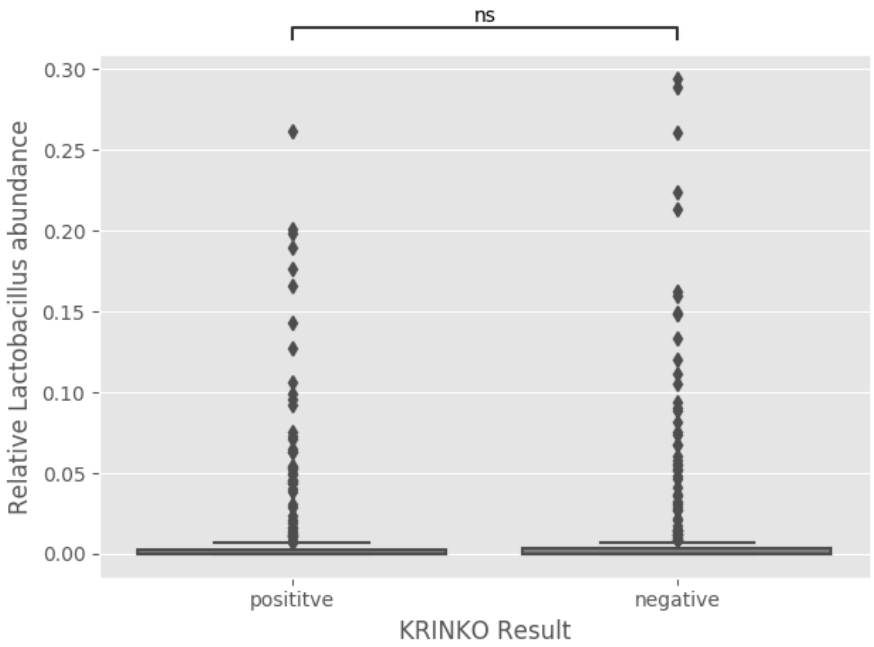


Fig. 3.28: (a) Alpha diversity (Shannon diversity index) and (b) relative abundance of *Bifidobacterium* and (c) relative abundance of *Lactobacillus* of samples stratified by positive and negative results using conventional microbiological methods. Kruskal-Wallis tests were performed between all groups. P-value annotation legend: ns, not significant: $p > 0.05$

3.7.2 Antibiotic resistance capacities

To assess the antibiotic capacity within PRIMAL study samples, visit 2 samples were subjected to conventional microbiological analysis to identify specific resistances (see 2.4) which yielded antibiograms. A subset of study samples (maternal samples N=31, visit 2 samples N=30 and visit 1 samples N=4) were sequenced using a whole genome shotgun sequencing approach to gain deeper insights into the genetic features providing protection against antimicrobial agents in the microbiomes. A comparison of the antibiotic capacities identified by the two methods is shown in fig. 3.29. All resistances that were screened for using conventional methods were also found with the sequencing approach. On top of that, many genes that convey resistances against the studied antibiotics were detected in the microbiome of most samples. The metagenomic capacity for resistances against penicillin, cephalosporin and fluoroquinolone was found in every sequenced sample and resistances against carbapenem and gentamycin in all but 4 (13.3 %) and 5 (16.6 %) samples, respectively.

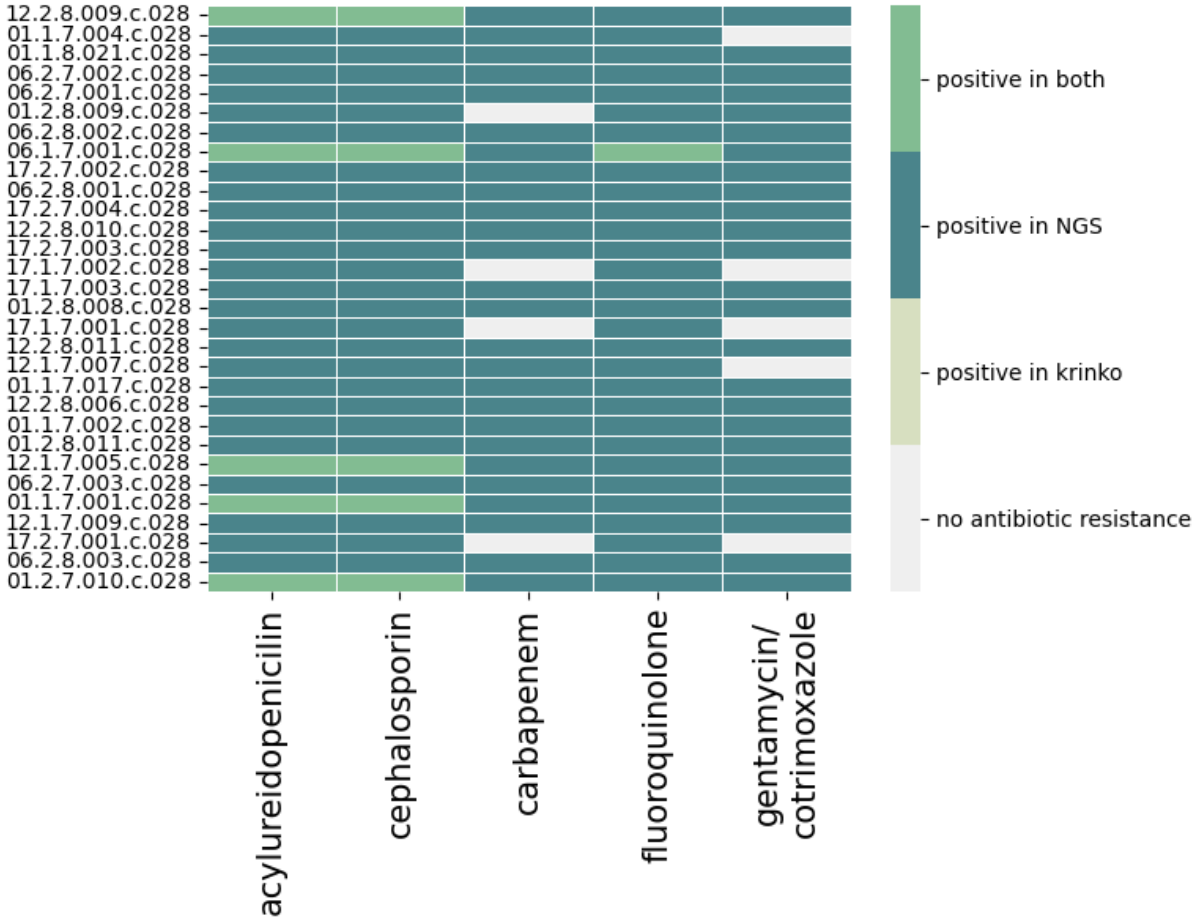


Fig. 3.29: Comparison of the resistance capacity against five antibiotics (relevant to KRINKO) in visit 2 samples determined by whole genome shotgun sequencing and conventional microbiology

The main parameters that drove clustering in the resistomes were the sample type (axis 1) and the number of antibiotic resistance genes (axis 2) in a sample (see fig. 3.30). The relative abundance of *Staphylococcus* was significantly associated with the clustering of mother and infant

samples (spearman correlation coefficient 0.78, p-value < 0.0001 (multiple testing corrected)). There were significant differences in the composition of antibiotic resistance genes across sample types. Association to a specific sample type explained 22.71 % of variance (PERMANOVA: p-value < 0.001). The antibiotic resistance genes that contributed the most to the significant differences are listed in tab. 3.3.

Tab. 3.3: Top 5 antibiotic resistance genes that contribute to the difference in the resistomes of mother and infant samples

Antibiotic resistance gene	more prevalent in	conveys resistance to	variance explained
lnuC	mother	lincosamide	0.0896
ErmG	mother	lincosamide macrolide streptogramin	0.0867
mgrA	visit 2	acridine dye cephalosporin fluoroquinolone penam peptide tetracycline	0.0667
dfrF	mother	diaminopyrimidine	0.0421
ErmF	mother	lincosamide macrolide streptogramin	0.0393

The number of antibiotic resistance genes explained 18.31 % of variance in the samples (PERMANOVA: p-value < 0.001). There was no significant difference in the number of antibiotic resistance genes across sample types or probiotic treatment group (PERMANOVA: p-value > 0.3).

Based on assembled contigs, the phylogenetic origin of some of the antibiotic resistance genes could be determined. 10095 (79.7 %) antibiotic resistance contigs could be assigned to the kingdom bacteria, while the origin of 2570 (20.3 %) could not be classified. At the genus level, the vast majority of antibiotic resistance genes (N = 12015, 94.9 %) could not be unambiguously assigned. Of the antibiotic resistance genes for whom the phylogenetic origin could be determined, most came from *Bifidobacterium* (N = 136, 1 %), *Bacteroides* (N = 135, 1 %) and *Clostridium* (N = 133, 1 %) (see fig. 3.31). Antibiotic resistance genes whose origin could be traced to *Bifidobacterium* were primarily resistant to mupirocin. Other genes products coded against rifampicin, tetracycline and erythromycin. None of the resistances was for antibiotics that are classified as essential for pediatric use by the KRINKO classification at the microbiology department at Mainz.

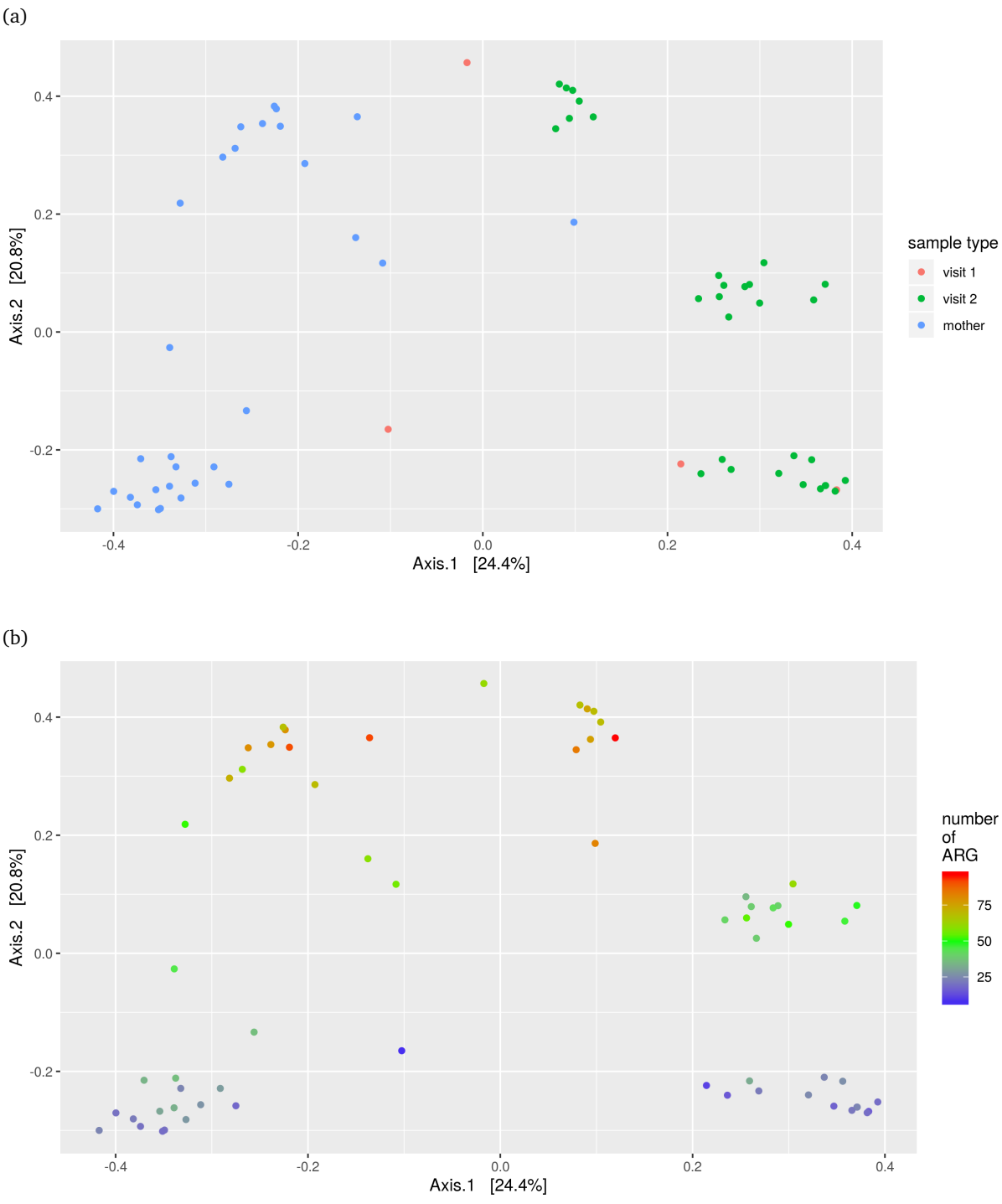
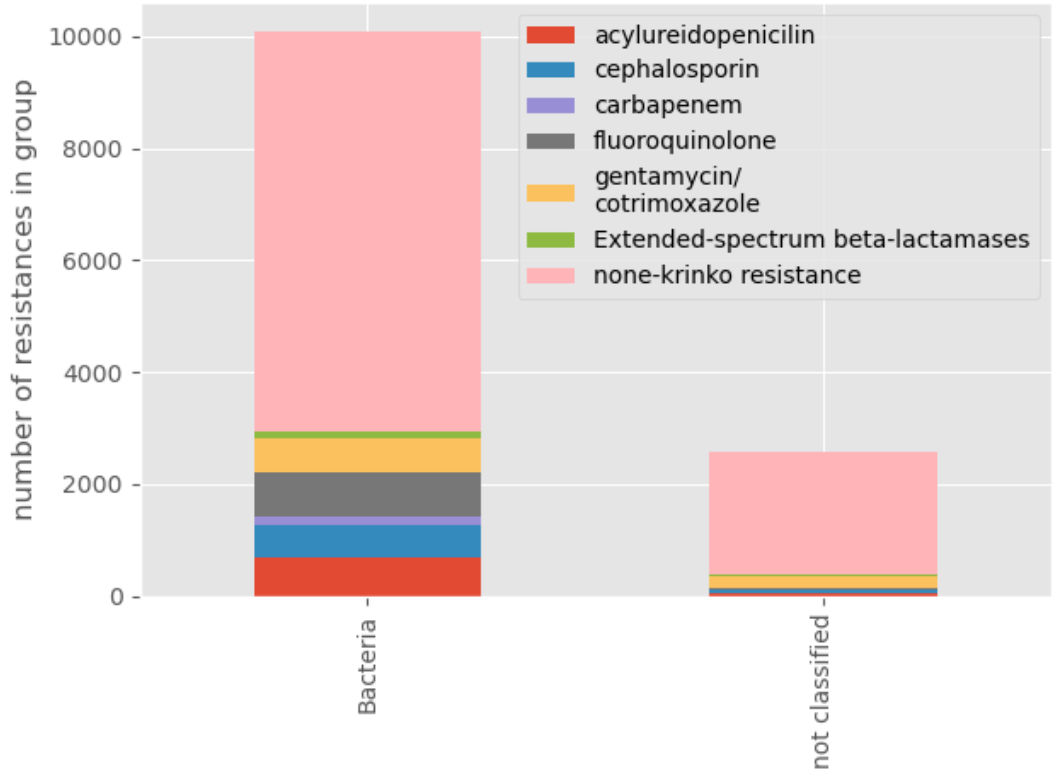


Fig. 3.30: Differences in the resistome (entirety of antibiotic resistance genes) across all samples that were whole genome sequenced measured by a PCoA of the jaccard distance matrix. Every point represents the resistome of one sample. The points are colored by (a) sample type and (b) number of antibiotic resistance genes in a sample.

(a)



(b)

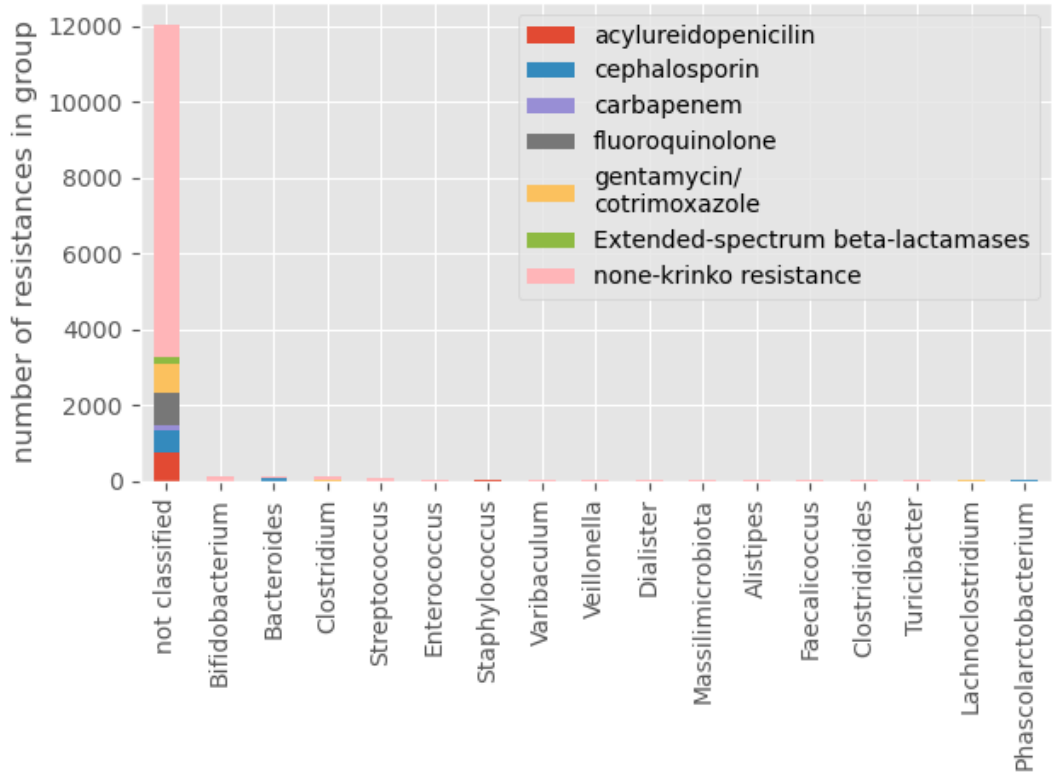


Fig. 3.31: Phylogenetic origin of the antibiotic resistance genes in the PRIMAL samples on the (a) kingdom level and (b) genus level. None-Krinko resistance contains all resistances that are not specifically screened for by the microbiology department in Mainz.

3.7.3 PRIMAL samples in the context of other studies

Dysbiosis is often described in terms of the abundance of commensal species, the bloom of pathogens or the diversity. Some studies have already been performed on the early life microbiome of preterm and full term infants and have produced data on infants of approximately 4 weeks of age. To assess similarities and differences between the PRIMAL samples and the ones from the literature, datasets of full term (two studies) and preterm infants (one study) have been downloaded and comparisons in regards to diversity and bacteria abundance have been performed.

There were significant differences in the alpha diversity of preterm reference samples and PRIMAL samples (Shannon index mean: 1.32 and 1.65, respectively, Kruskal-Wallis, p-value < 0.001). No significant differences were found between full term reference samples and PRIMAL samples (Shannon index mean: 1.65 and 1.65, respectively) (see fig. 3.32).

Compared to preterm reference samples, PRIMAL samples showed no significant difference in

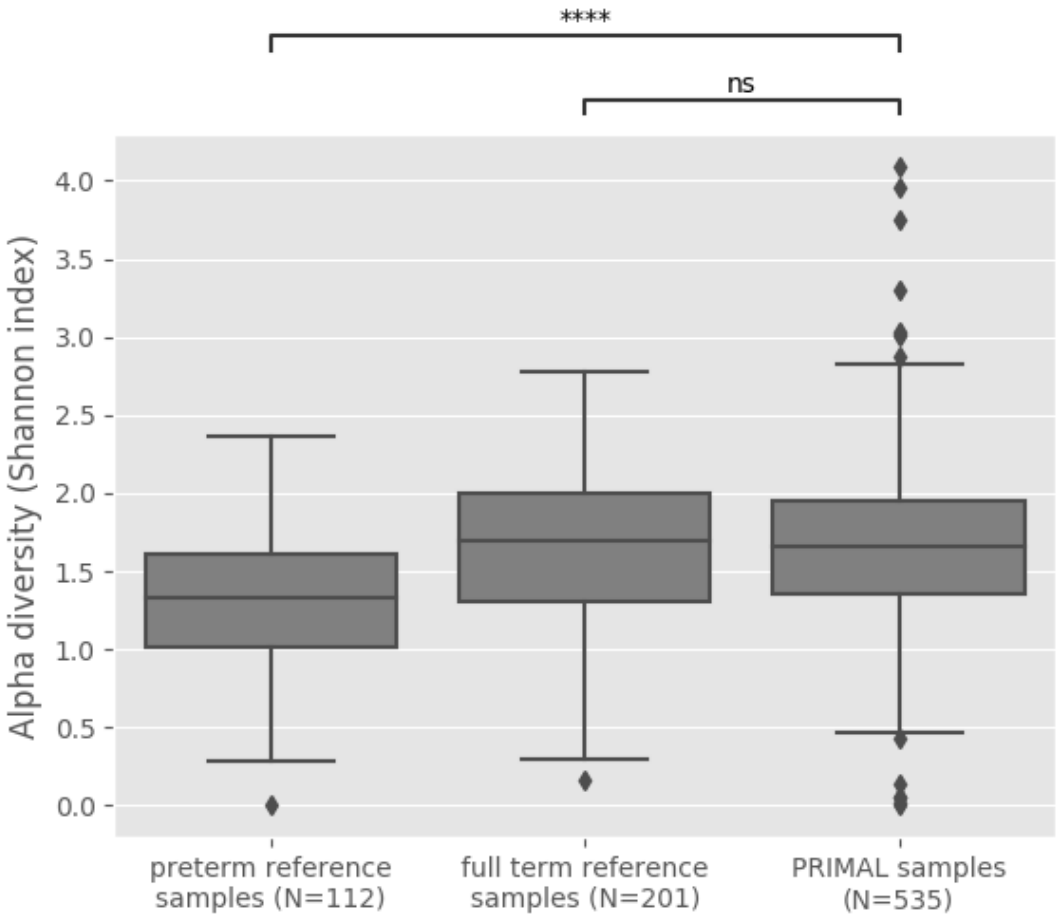


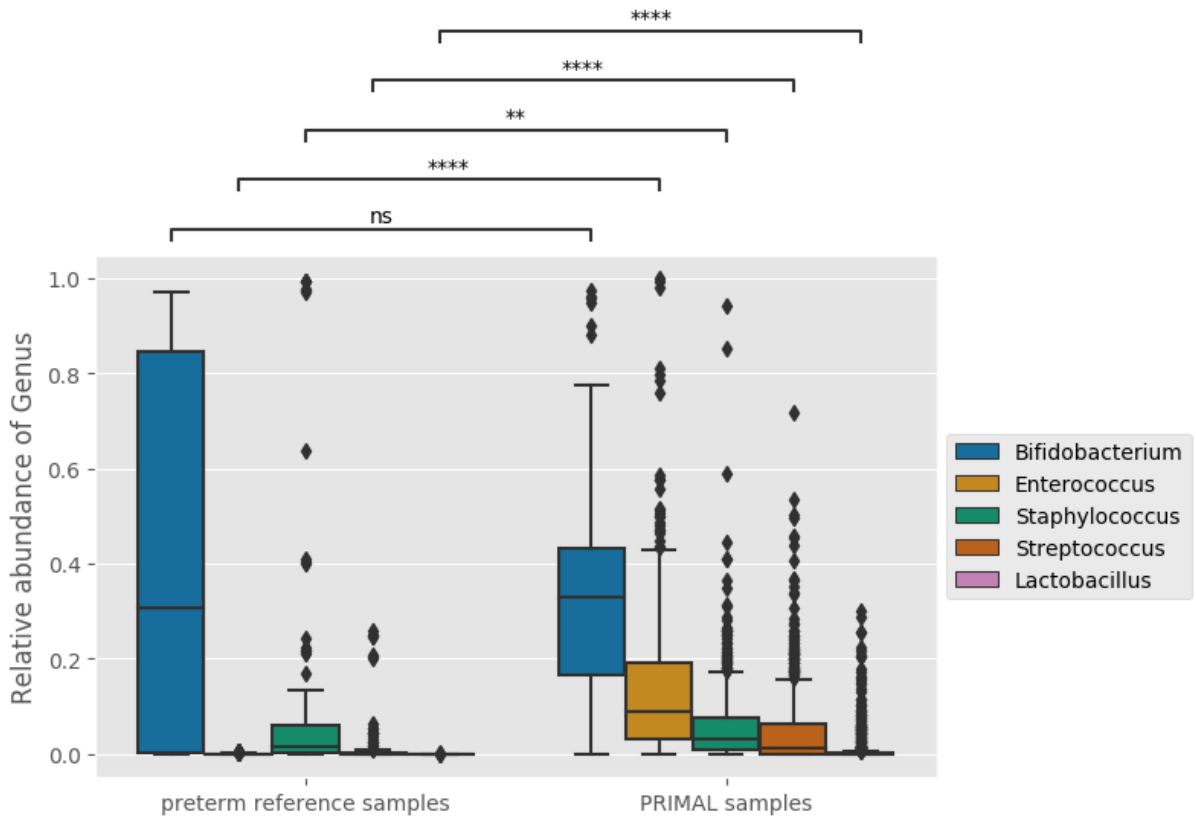
Fig. 3.32: Comparison of the alpha diversity between reference samples (preterm and full term) and PRIMAl samples. Kruskal-Wallis tests with Bonferroni correction were performed between all corresponding groups. P-value annotation legend: ns, not significant: p > 0.05, ****, highly significant: p <= 1.00e-04.

the relative abundance of the genus *Bifidobacterium* (median: 31 % and 33 %, respectively). However, the usually beneficial genus *Lactobacillus* and the potential pathogenic genera *Enterococcus*, *Staphylococcus* and *Streptococcus* were significantly enriched in PRIMAL samples (Kruskal-Wallis, p-value < 0.01).

The relative abundance of *Lactobacillus* and *Streptococcus* did not differ in PRIMAL samples when compared to full term infants from the literature. Highly significant changes could be observed for *Enterococcus* and *Staphylococcus* (Kruskal-Wallis, p-value < 0.001) as well as for *Bifidobacterium* (median: 42 % in the reference and 33 % in PRIMAL, Kruskal-Wallis, p-value < 0.001) (see fig. 3.33).

The two prior analyses revealed a normal distribution of alpha diversity in the samples and a bimodal distribution of the relative *Bifidobacterium* abundance. For the preterm reference samples, additional information was available on the use of probiotics: Infloran (*Lactobacillus acidophilus* und *Bifidobacterium infantis*) was administered twice daily until the age of 34 weeks (Alcon-Giner, Dalby, et al., 2020) in the probiotic group. Based on literature that suggests a low abundance of commensal bacteria (such as *Bifidobacterium*) and a low alpha diversity within samples to be a sign of dysbiosis, the group affiliation of PRIMAL samples and reference samples within the 10^{th} percentile of each parameter was determined (see fig. 3.34). Strikingly, no clustering by study was observed based on alpha diversity and relative *Bifidobacterium* abundance. Samples from all studies (PRIMAL samples with either treatment, full term reference samples and preterm reference samples that did not receive the probiotic) fell within the "dysbiotic" scope.

(a)



(b)

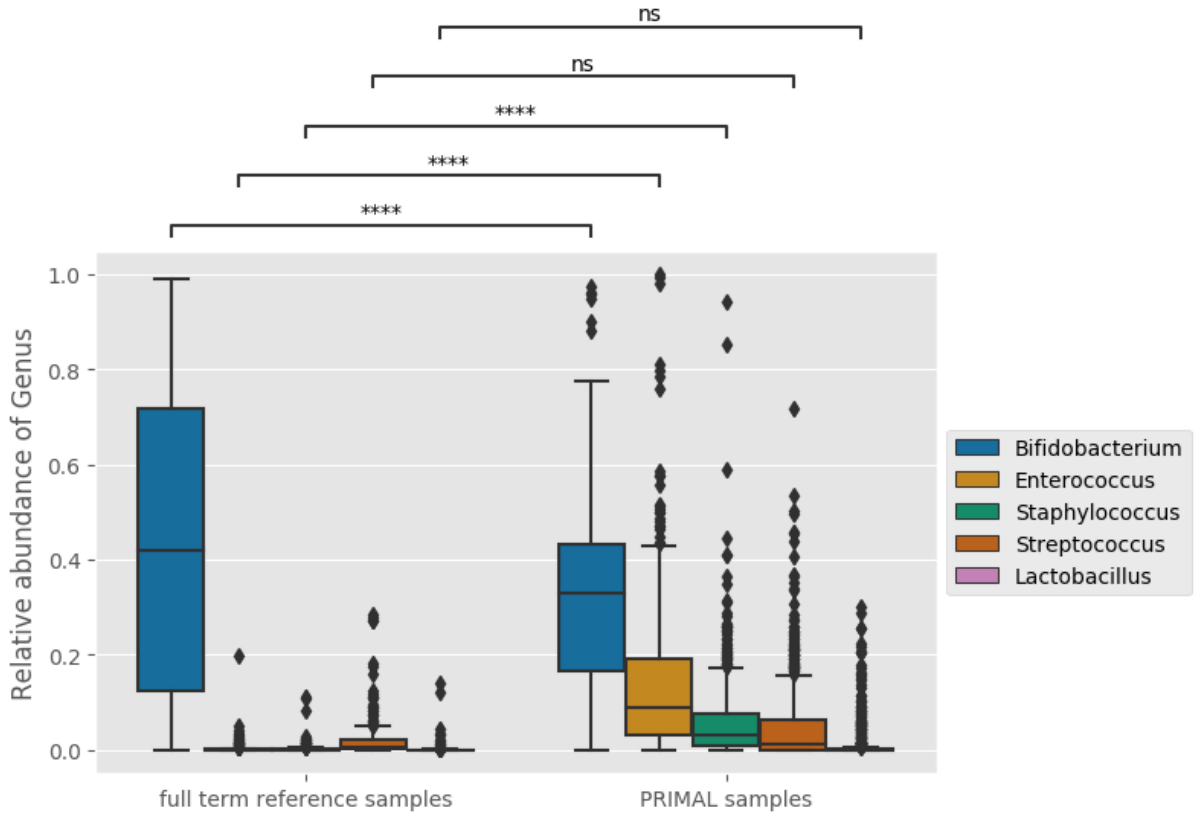


Fig. 3.33: Differences in the relative abundance of 5 bacterial genera (three potential pathogens and two beneficial) in reference samples (full term and preterm infants) and primal samples. Kruskal-Wallis tests with Bonferroni correction were performed between all corresponding groups. P-value annotation legend: ns, not significant: $p > 0.05$, **, significant: $p <= 1.00e-02$, ****, highly significant: $p <= 1.00e-04$.

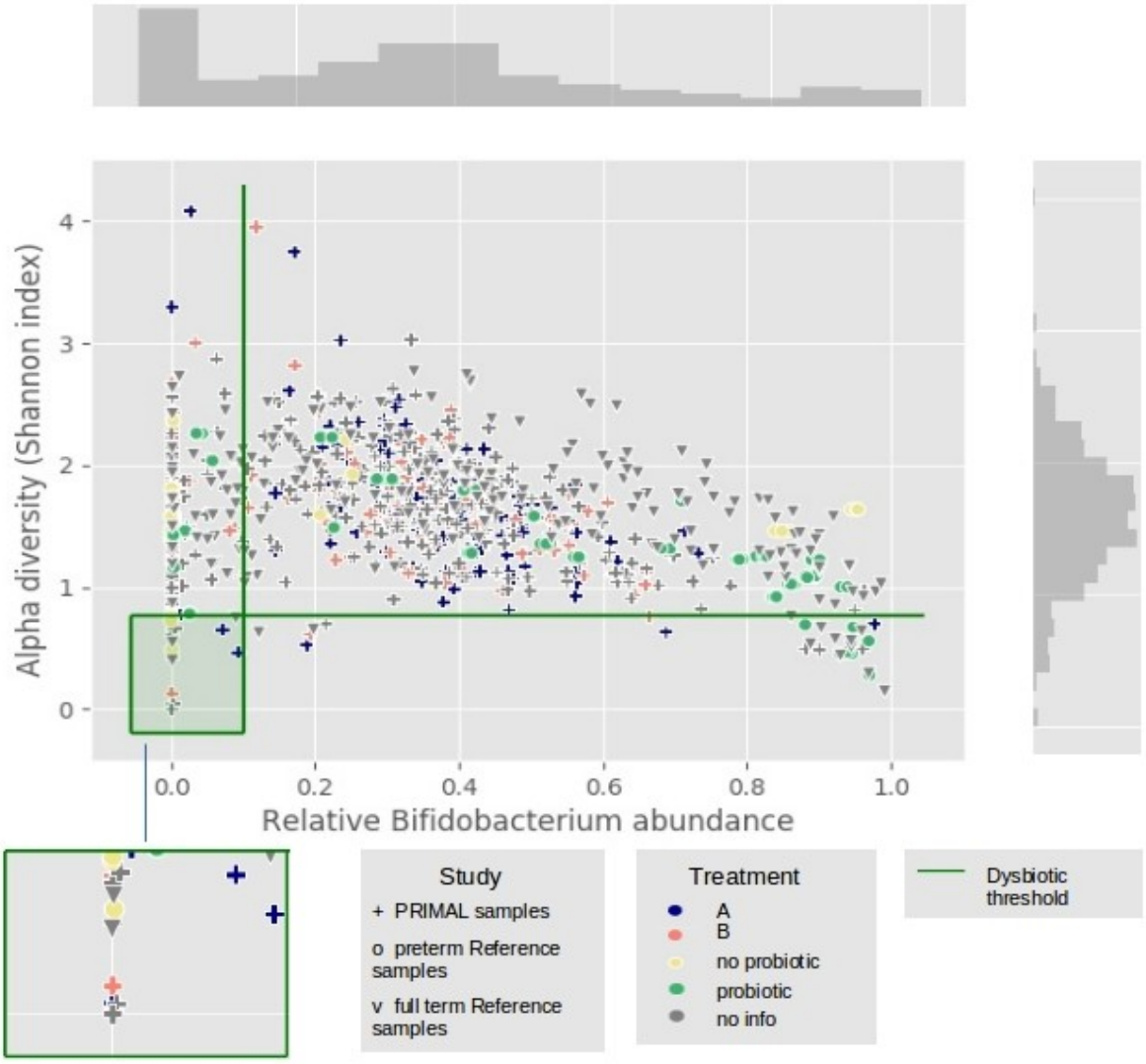


Fig. 3.34: Alpha diversity and relative abundance of *Bifidobacterium* in PRIMAL and reference samples (preterm and full term infants). Every point represents one sample and is colored by the meta information available (Treatment A or B for PRIMAL samples and probiotic or no probiotic for preterm samples). The green lines indicate the 10th percentile of each parameter. The two plots on the margins show the overall distribution of the two parameters. The box with the green frame shows a zoomed-in version of the "dysbiotic threshold" - the area of the plot that is below the 10th percentile of each parameter.

3.7.4 Co-occurrence networks

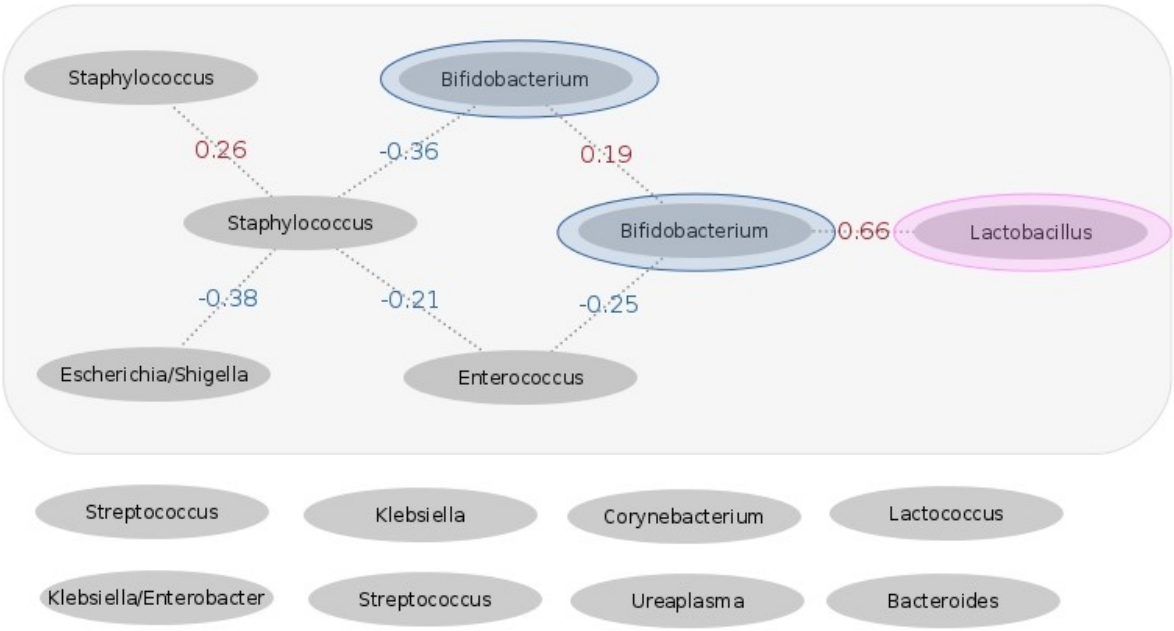
To study potential interactions in the microbiomes of PRIMAL samples, co-occurrence network analyses of the 15 most abundant ASVs measured by median and mean (taxa approximately on the species level) were performed. A range of statistically significant correlations were identified across sample types. The complexity of the co-occurrence networks increased with the age of the studied samples.
The co-occurrence network of the visit 1 samples showed only low levels of complexity, with the primary correlations being positive ones between two species of *Bifidobacterium* and between *Lactobacillus* and *Bifidobacterium*. The genus *Bifidobacterium* was negatively correlated to *Enterococcus* and *Staphylococcus*, which was in turn also negatively correlated to *Enterococcus* and *Escherichia/Shigella*.

The only positive correlation in visit 2 samples was detected between two ASVs that were both classified as *Veillonella*. *Bifidobacterium* had the most connections within the network and showed negative correlations to 5 taxa (*Clostridium sensu stricto 1*, *Enterococcus*, *Klebsiella*, *Enterobacter* and *Finegoldia*). Interestingly, although a *Lactobacillus* ASV was part of the 15 most abundant species, it did not show any significant correlation with any other taxon.

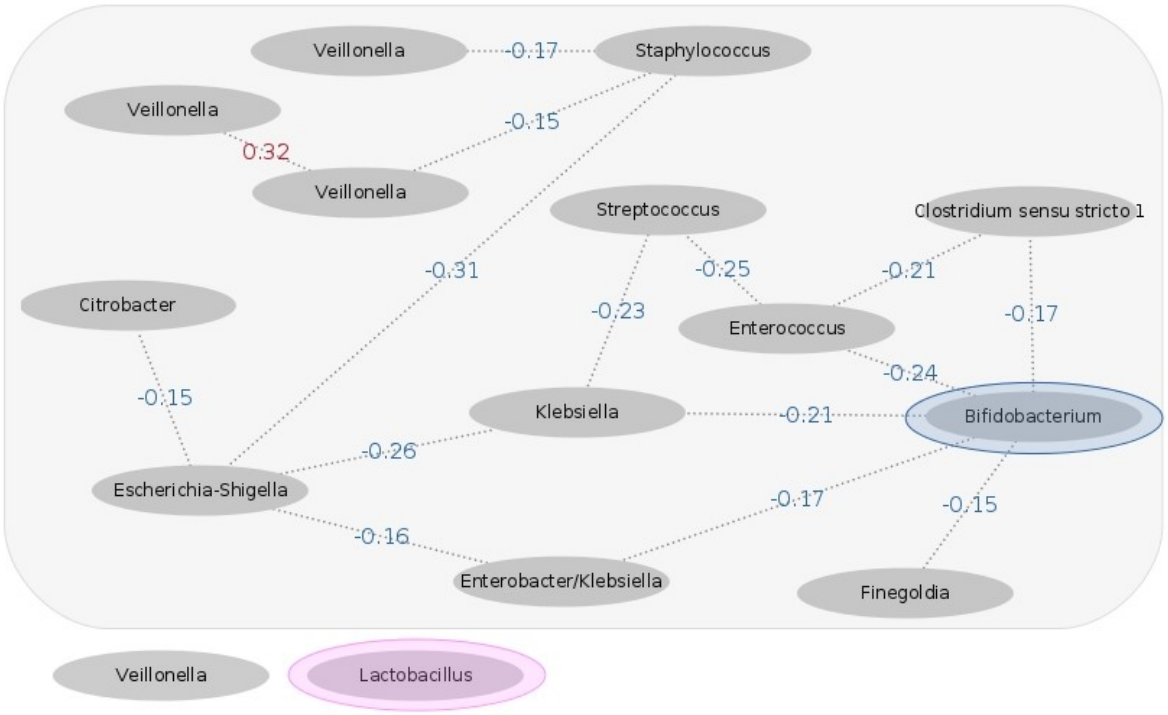
The level of complexity in the network further increased in visit 3 samples. *Bifidobacterium* only showed two significant co-occurrences, one with *Clostridium sensu stricto 1* and one with *Blautia*. The latter taxa gained in centrality in the network and was negatively correlated with three different taxa (*Escherichia/Shigella*, *Ruminococcus* and *Bifidobacterium*) and positively correlated to *Anaerostipes*. One ASV that was classified as *Bifidobacterium* showed no connection to any other taxa in the visit 3 network.

Similar to visit 3 samples, significant negative correlations between the genera *Bifidobacterium* and *Blautia* can be seen in mother samples. *Streptococcus* showed the most correlations in the network and was negatively associated to 6 ASVs from 5 different genera (*Blautia*, *Bacteroides*, *Faecalibacterium*, *Subdoligranulum*). With the exception of *Dorea* and *Anaerostipes*, most of the correlations in the mother samples were negative.

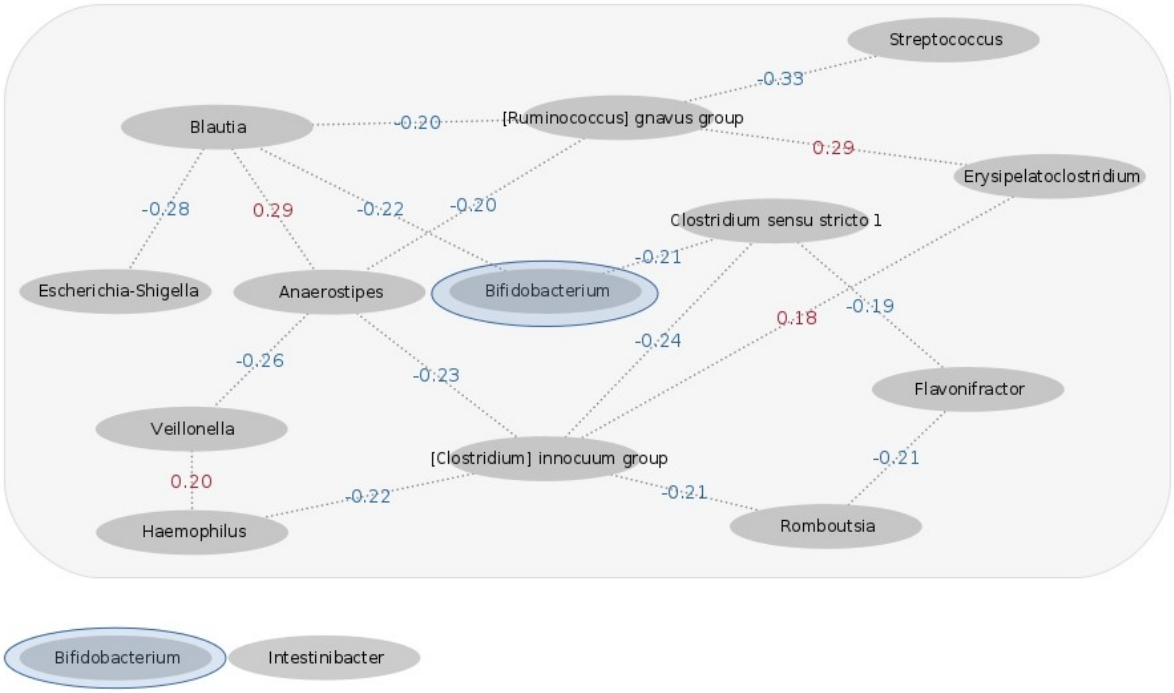
(a)


Staphylococcus
Bifidobacterium
0.26
-0.36
0.19
Staphylococcus
Bifidobacterium
0.66
Lactobacillus
-0.38
-0.21
-0.25
Escherichia/Shigella
Enterococcus
Streptococcus
Klebsiella
Corynebacterium
Lactococcus
Klebsiella/Enterobacter
Streptococcus
Ureaplasma
Bacteroides

(b)


Veillonella
-0.17
Staphylococcus
Veillonella
-0.15
0.32
Veillonella
Streptococcus
Clostridium sensu stricto 1
-0.31
-0.25
-0.21
Citrobacter
-0.23
Enterococcus
-0.17
-0.24
-0.15
Klebsiella
-0.21
Bifidobacterium
-0.26
Escherichia-Shigella
-0.17
-0.15
-0.16
Enterobacter/Klebsiella
Finegoldia
Veillonella
Lactobacillus

(c)



(d)

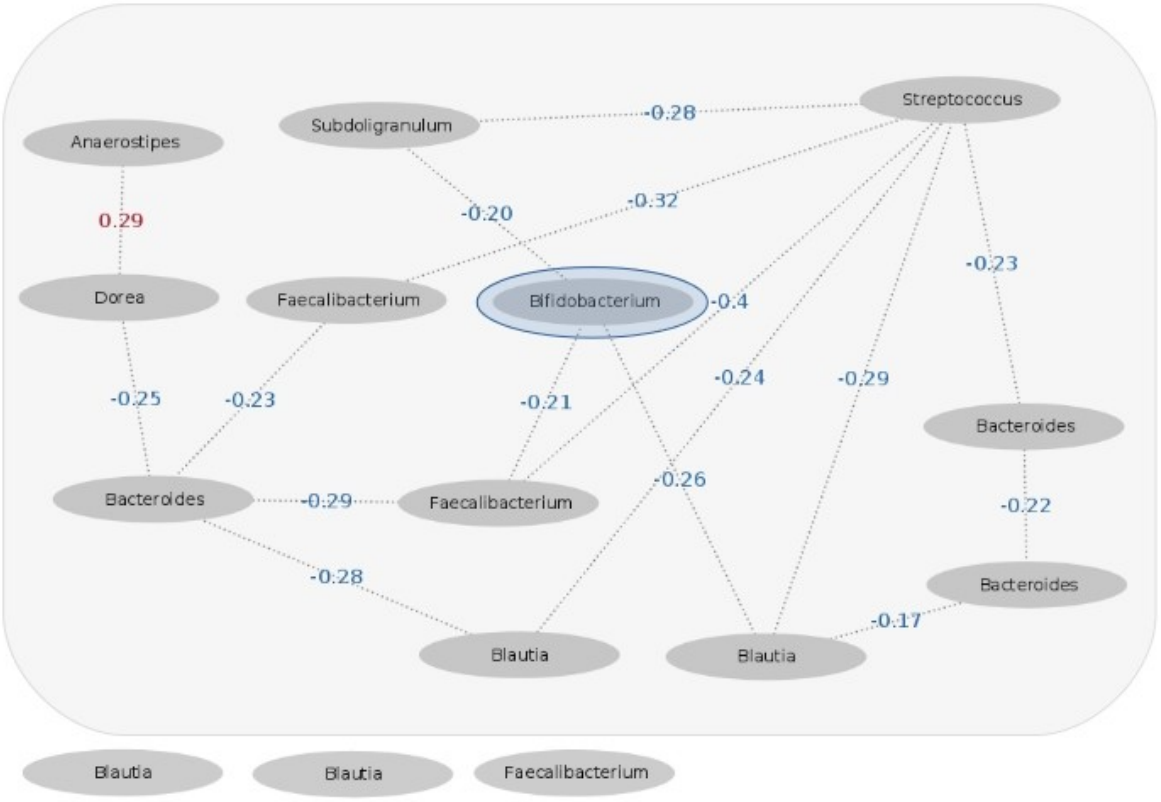


Fig. 3.35: Co-occurrence networks of the 15 most abundant ASVs in the (a) visit 1 samples (b) visit 2 samples (c) visit 3 samples and (d) mother samples. Nodes are ASVs (taxa approximately on the species level) labeled by their genus affiliation. Edges are statistically significant correlations between ASVs and their color coded labels display the correlation coefficients (red scale: positive, blue scale: negative). Discretized mutual information tests as implemented by FlashWeave were performed. The statistical significance level is alpha=0.01. The probiotic taxa *Bifidobacterium* and *Lactobacillus* are indicted in blue and magenta, respectively.

4 Discussion

4.1 Technical challenges in the sequencing based study of microbiomes

The use of next-generation-sequencing techniques has revolutionized our understanding of the richness and diversity of microbial ecosystems, especially the ones that live within us. To fully appreciate the extend and scope of this revolutionary approach but also its limitations, it is instructive to understand in detail the current methodology that is being used.
The primary method used in this study was the targeted 16S rRNA gene amplicon-based sequencing. In this approach, conserved regions of the 16S rRNA gene from bacteria are used to amplify hyper-variable interspersed regions that can then be mapped to specific bacterial genera to allow a large-scale analysis of the taxonomic composition of entire microbial ecosystems. Many choices and biological realities affect the final result and every taxonomic, functional and comparative analysis in this study has thus to be seen in the light of the following caveats.

4.1.1 Primer selection

One major challenge that arises is the selection of the "correct" region of the 16S rRNA gene to amplify. Many different regions have been used in the past, but a focus has lately been placed on the use of the V4 region of the 16S rRNA gene as it offers some advantages. First, the relatively short length of this region (around 250 bp) compared to other regions allows for a amplification with maximal overlap of the forward and reverse reads using Illumina's V2 chemistry on a MiSeq. This leads to a lower error rate than the sequencing of other regions (Q. Wang et al., 2007; Kozich et al., 2013). Given a typical error rate during Illumina sequencing of a 1 in 1000 probability that a base is wrongly assigned, a full coverage of the sequence by a forward and a reverse read reduces the probability of a spurious base calling to one in a million ($\frac{1}{1000} * \frac{1}{1000} = \frac{1}{1000000}$). Correct base calling is especially critical in amplicon sequencing, since - in some instances - differences of only a single basepair are used to differentiate between bacterial taxa. The probability of detecting spurious taxa thus increases with the amount of reads that are sequenced (Schloss et al., 2011). The V4 region has become an established target to fight this bias. Owing to its popularity, the V4 region has a second advantage over other regions as it can be used to propel robust meta analysis of microbiome studies. Therefore, the V4 region of the 16S rRNA gene was selected to amplify the microbiome during the current study.

In the present study, it was discovered that the forward and reverse primers of the V4 region show a 89.47 % and a 85.71 % sequence identity, respectively, to regions in the human mitochondrial 12S rRNA gene that span a 199 bp region, which has not been published prior. While this does not pose a challenge to the study of mature microbiomes (see fig. 3.6), the use of these primers led to the amplification of numerous mitochondrial sequences in developing microbiomes with low bacterial content. One possible explanation for this observation might be based on a mechanism of primer competition. In settings with very low bacterial content such as meconium, the amount of mitochondrial sequences probably far outnumbers the number of bacterial sequences. Purely based on statistic movement, primers preferentially targeting the 16S rRNA gene of bacteria might amplify the homologue region in the 12S rRNA gene of human mitochondria due to an inadequate number of bacterial targets. In later stages of the infant development, the microbiome is home to a myriad of bacteria that - by then - far outnumber the amount of mitochondrial sequences, so that the statistic effect allowing for a amplification of mitochondria instead of bacteria is no longer detectable (see fig. 4.1).

Further evidence for this hypothesis can be derived from the observation that the spiking

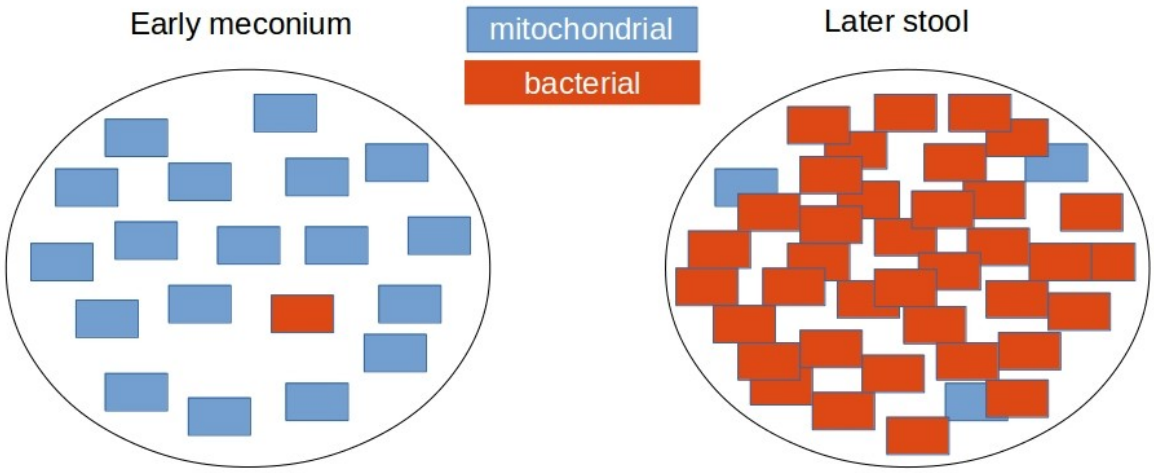


Fig. 4.1: Proposed competition driven mechanism by which mitochondrial sequences are erroneously amplified in microbiomes that have a very low bacterial content.

of meconium samples with minimal bacterial DNA (ca. 60 ng) leads to an almost complete suppression of mitochondrial signals (see fig. 3.7). The amounts of spike-in DNA that were used were only a tenth and half, respectively, of the usual amount that would be used in an adult microbiome to have approximately 5 % of the resulting microbiome be assigned to the spiked-in taxa according to the manufacturer (Zymo, 2021). This implies a primer competition that strongly favors the 16S rRNA gene and only in the most extreme cases leads to a noteworthy amplification of off-targets. These dynamics would also explain the overall absence of other sequences that should be in competition for the primers (such as fungal mitochondrial sequences and eukaryotic 18S rRNA genes).

4.1.2 Conceptual challenges

Microbial communities are inherently complex and can thus rarely be studied in their entirety. As a consequence, only subsets of the theoretically examinable material and data can be studied. The resulting sequencing data further demand the use of dimensionality reduction methods to extract biologically relevant insights. One of the major implications is, that metagenomic data is compositionally and reflects only relative abundances and not absolute abundances (Fang and Subedi, 2020). While the calculation of diversity metrics, such as alpha or beta diversity, allows for large scale comparability between different microbiomes, some of the biological nuances are naturally lost through this data aggregation.

Biological limitations

Besides the purely technical limitations, inherent limitations of biological systems were revealed during the current study. In the context of taxonomic studies through the V4 region of 16S rRNA genes, this was for example reflected in the fact that some bacterial taxa such as *Enterobacter cloacae* and *Klebsiella pneumoniae* were found to share a 100% sequence identity in the target gene although they are phylogenetically diverse (see fig. 3.26). The same holds true for members of the genera *Escherichia* and *Shigella* which are not differentiated in the SILVA database.

The principal conceptual challenges described above (high complexity and compositionality) are also present in the whole genome shotgun approach.

4.2 Differences and similarities in the taxonomic composition across sample types

Since the inception of the Human Microbiome Project (The NIH HMP Working Group et al., 2009) many insights have been gained in regards to the nature, development and functions of human-host associated microbiomes. However, further baseline characterization of the ecology, taxonomic composition and interactions still remains a pressing field of inquiry, especially in (very preterm) infant populations. To facilitate translational research that aims to understand and rationally intervene in the complex development of microbiome structures, vast amounts of descriptive research remains to be done (A. Tripathi et al., 2018). These fundamental discoveries of facts around the microbiome are still the basis for meaningful hypothesis-driven microbiome research (A. Tripathi et al., 2018).

Many factors can confound the study of microbiomes, such as contamination and collection site biases. During the current project, the microbiome structures of 565 children over the course of three visits and a subset of their mothers were analyzed. To rule out contamination as much as possible, positive controls (a defined set of microbes) and negative controls were processed along side the study samples. There are reports in the literature about systematic contaminations in microbiome studies (especially in low biomass settings) caused by specific bacteria in the kits used to process the samples - the so called "kitome" (Douglas et al., 2020). After applying a suite of quality control and decontamination algorithms, the prevalence of these taxa was negligible in the PRIMAL study samples (see 3.4), strengthening the confidence in the taxonomic profiles.

Significant changes have been found in the studied PRIMAL samples based on the hospital site at which they had been collected. This is in line with the recent understanding that hospitals house microbial ecosystems and posses specific microbiomes (Arnold, 2014). Even short stays at hospitals have the potential to profoundly alter the microbiome structure of patients (Ravi et al., 2019).

The complexity or alpha diversity of the studied microbiomes increased significantly with the age of the studied individual. Samples that were collected within the first few days of an infant's life did not possess a highly complex microbiome and were sometimes dominated by only one phylum or even genus. Over the course of the first 28 days, the diversity markedly increased and was still gaining complexity after one year. The highest diversity could be found in the microbiome of the adult mother samples. This observation confirms earlier studies about the development of the gut microbiome, that found an ongoing remodeling and increasing complexity in the gut microbiome until a stable phase is reached at about 31 to 46 month of age (Stewart et al., 2018). Contrary to the alpha diversity within samples, the beta diversity between samples decreased with age. This means that the inter-sample variation of e.g. visit 1 samples right after birth was way higher than at later time points which is in accord with earlier studies (Yatsunenko et al., 2012). Given a framework of a microbiome development that starts off "messy" and changes over time to form a stable taxonomic composition, it makes sense that random initial colonization by bacteria from the maternal skin, vagina and other environmental sources (Arrieta et al., 2014) would result in high inter-sample variability. When clustered by similarities between their microbiomes, infant samples in this study tended to group closer to the mother samples the older the infants were - again highlighting a microbiome development towards a somewhat stable adult state (see fig. 3.10).

Evidence for huge individual differences that decrease over time further comes from the observation that the beta diversity between visits 1 right after birth and 2 after one month of life was significantly lower when observing the identical infant rather than distinct ones, yet this effect does not hold at visit 3 after one year. In other words: Differences in the initial random seeding of bacteria make way for a niche based selection in the further development of the microbiome over the course of the first year as has been described in the literature (Ferretti et al., 2018).

In the PRIMAL cohort, there was a significant difference in the number of shared taxa between related and unrelated mother and infant pairs at visit 1 (right after birth). Other studies on the transmission of bacterial strains from mother to infant found that maternal microbes are the primary source of bacteria in the newborn (Ferretti et al., 2018), possibly explaining these significant differences by relatedness in the early stages of the microbiome. The high abundance of typical skin colonialist taxa in the earliest samples of the PRIMAL neonates is further evidence for a transmission not only of stool microbes from the mother but also other locations like the skin or the vagina (Ferretti et al., 2018).

4.3 A re-imagined concept of dysbiosis in infants

To approach a revised interpretation of the concept of dysbiosis in infant populations, multiple interconnected analyses were performed. The conventional, microbiological definition of dysbiosis as stated in the KRINKO (presence of known multiresistant pathogens) was set in the context of results from NGS methods. The traditional hallmarks of microbiome dysbiosis (loss of alpha diversity, loss of commensal bacteria and bloom of pathobionts (Levy et al., 2017)) were compared between PRIMAl study subjects and age-matched preterm and full term reference samples. Further, statistical significant co-abundances of taxa in the microbiomes were analyzed to draw conclusions on possible eu- or dysbiotic interactions in the ecosystem "infant gut".
The effect of probiotic supplementation on the taxonomic composition of the developing infant microbiome was assessed to identify drivers of reported health outcomes associated with their use in the literature (Härtel et al., 2017).

4.3.1 Traditional microbiological methods compared to next-generation sequencing

Almost all methods that strive to identify dysbiotic states do so in order to anticipate future diseases and provide a conceptual framework of how best to avoid them.
The KRINKO defines a suite of specific bacteria that partially show resistances to multiple antibiotics to be considered a thread to an infant's health ("Hygienemaßnahmen bei Infektionen oder Besiedlung mit multiresistenten gramnegativen Stäbchen" 2012). Some of the bacteria such as *Klebsiella* and *Serratia* are mainly considered dangerous because of their ability to rapidly spread in intensive care units. Conventionally, these bacteria are identified by culturing them from patient samples on specific culture media and analyzing their taxonomic identity and antimicrobial capacity with an array of laboratory techniques, such as MALDI-TOF. In contrast, NGS methods identify entire microbiomes culture-independently in a high-throughput fashion. Both methods have their own advantages and disadvantages and the following comparison of the two methods based on data from this study will illuminate these.

Taxonomic classification

One of the main differences between the 16S rRNA gene based analysis and the culture based techniques is the taxonomic resolution that can theoretically be achieved. Owing to the relative short read length that a typical high-throughput sequencing run produces (around 250 bp) the variability in the target region that is being amplified allows for a taxonomic resolution up to the genus level and does not allow further classifications on species or strain level (J. S. Johnson et al., 2019). In rare cases only the family or even more basic (order, phylum) affiliation can confidently be assigned to reads. On the other hand, culture based methods aim to identify bacteria on the species level (Dingle and Butler-Wu, 2013). However, there is a consensus in the literature that - in practice - methods for the identification of bacteria (such as MALDI-TOF) have comparable limitations to the identification of bacteria at the species level (Strejcek et al., 2018). This is caused by inherent similarities of the target features between taxa that are used to differentiate bacteria (Rychert, 2019).
All methods that try to taxonomically identify bacteria have to deal with the biological reality of evolution that immensely hampers the neat classification of microbial lifeforms into the framework of distinct species (Van Rossum et al., 2020) (later discussed in detail).

For the most part, every genus that was identified in the PRIMAL cohort using conventional methods was also found using 16S rRNA gene sequencing (see fig. 3.25). An exception to this was only present in the genera *Enterobacter* and *Klebsiella*. A follow up analysis revealed that the sequence of the V4 region of the 16S rRNA gene was identical between *Enterobacter cloacae* and *Klebsiella pneumoniae* - two species that the conventional microbiology screened for. The genera *Enterobacter* and *Klebsiella* are closely related and both belong to the family *Enterobacteriaceae*. The differentiation of some members of the genera is an ongoing debate and their nomenclature has a complicated history with some bacteria having homotypic synonyms in both genera as is the case with *Enterobacter aerogenes* and *Klebsiella mobilis* (Tindall et al., 2017). It is thus not surprising that a sequence based differentiation could not be performed for these genus affiliations. However, this methodological difference in the reporting of the presence of genera could only be detected if no other member of the microbiome belonged to the same genus as these instances would have concealed the incongruence.

On the other hand, there were a lot of samples for which the 16S rRNA produced a positive result for a given genus (e.g. *Staphylococcus* or *Escherichia*) and the conventional microbiology did not. This was probably due to the untargeted nature of the sequencing based approach which identifies all genera and is not limited to a predefined subset that is specifically screened for as in the conventional approach. Another contributing factor could be that sequencing is also able to detect DNA from dead organisms (Boers et al., 2019), while culturing can only be done with living specimens.

There was no strict correlation between a positive result in the conventional analysis and the relative abundance of the genus in the microbiome (see fig. 3.27). This underscores the

ability of the conventional methods to culture bacteria from samples even if their relative abundance was very low. Further, there was no significant difference in the relative abundance of beneficial genera (such as *Bifidobacterium* and *Lactobacillus*) or the alpha diversity of a sample and the result from the conventional analysis (see fig. 3.28). This illustrates the limited scope of information a KRINKO analysis can provide in the context of traditional measures of dysbiosis such as microbiome diversity.

Antibiotic resistance capacity

Increasing resistance of bacteria to a wide array of antimicrobial agents is an emerging global health problem (Wood et al., 2020). The (over) use of antibiotics and antibiotic resistances is often discussed in the context of dysbiosis in children. A meta-analysis that screened 4668 relevant publications found antibiotic exposure to be associated with a decreased diversity in the microbiomes of children aged 0-18 years and a striking shift in the abundances of bacterial taxa (McDonnell et al., 2021). To combat the overuse, an exact understanding of the underlying biology of antibiotic resistance in a microbial ecosystem is paramount to find rational ways to apply antibiotics. Conventional tools and sequencing based technologies can be used to identify antimicrobial resistances (Vasala et al., 2020). During the current study, the antibiograms produced by conventional microbiology and whole genome shotgun sequencing were compared for the PRIMAL cohort. In all instances in which a positive result was found with the conventional methodology, NGS confirmed the result by identifying genes in the microbiome of the sample that produced products effective against the antibiotic in question. Further, for most PRIMAL samples, NGS identified an array of antibiotic resistance genes (often times all that were principally tested by the conventional method) to be present in the genetic horizon of the studied microbiome.

One point to note is, that the conventional methods and NGS look for categorically different things: The culture based method identifies single bacterial species and tests for a predefined set of phenotypic resistances against antimicrobial agents. Sequencing on the other hand identifies all antibiotic resistance genes in a given microbiome irrespective of their bacteria of origin. Further, genomic sequencing can not differentiate between antibiotic resistance genes that are really expressed and exert a phenotypic resistance or are just present as genetic information.

This is especially relevant for possible conclusions about microbial fitness based on the presence of antibiotic resistance genes. The common notion holds that bacteria that express antibiotic resistance genes in a milieu in which anthropogenic antibiotics are not present have a significant fitness disadvantage compared to wild type bacteria. Their main function is seen in conferring antimicrobial drug resistances against the three major antibiotic targets (bacterial cell wall, protein synthesis and DNA replication and repair) (C. Walsh, 2000). Thus, the reduction of antibiotic use should lead to a decrease in the abundance of antibiotic resistance genes. However, a study by Cardoso and colleagues could show that the complex interactions in certain individual dysbiotic states in the gut microbiome of mice can lead to a vast range of possible fitness outcomes - and

are by no means limited to negative outcomes - for the resistant species even if host genetics and diet is identical (Leónidas Cardoso et al., 2020). This implies more complex roles of antibiotic resistance genes.

Naturally occurring antibiotics play a crucial role in the formation of microbial ecosystems and are emitted in this context as cell-to-cell signaling molecules (Baquero et al., 2013). Consequently, antibiotic resistance genes are the counter part during this two way communication. The non-acquired intrinsic resistome, genes whose normal function is not the protection against anthropogenic antibiotics, can account for up to 3% of a bacterial genome (Fajardo et al., 2008). The wide range of functions that antibiotic resistance genes exert is further underscored by their presence in environmental samples from locations that did not come into contact with anthropogenic antibiotics such as Alaskan soil (Allen et al., 2009), 30,000 year old permafrost (D'Costa et al., 2011) or a cave system isolated for four million years (Bhullar et al., 2012).

In order to conceptualize dysbiotic states based on antibiotic resistance genes it is important to note that the human microbiome can serve as a diverse reservoir of antibiotic resistance genes that can be readily transferred from one class of bacteria to the next given the right selective pressure e.g. during antibiotic treatment (M. O. Sommer et al., 2010). Inflammatory processes in the gut, which are a common health outcome ascribed to gut microbiome dysbiosis, can further boost the lateral gene transfer between commensal and pathogenic bacteria such as *Enterobacteriaceae* (Stecher et al., 2012) aggravating the patients situation by opening up routes to the spread of antibiotic resistance genes in the microbiome.

To assess the likely origin of antibiotic resistance genes in the PRIMAL samples, two methods were used: First, KRINKO relevant genera were correlated with the clusters in the resistome analysis and second, read contigs were mapped to different phylogenetic levels. The clustering based on sample type was strongly correlated with the relative abundance of *Staphylococcus*. This might indicate that the *Staphylococcus* species in the samples could be carriers of many antibiotic resistance genes that differentiate children from mothers. On the other hand, *Staphylococcus* might just be a proxy of correlation but not the causative parameter for the clustering. The taxonomic assignment of antibiotic resistance genes revealed that the majority of antibiotic resistance genes could be traced to be part of known bacterial genomes. However, around one fifth of the genes were not attributable to any specific bacteria. This might be due to inherent challenges when using short read sequencing to link resistance genes to specific genomes (Rose et al., 2017) but might also represent underlying biological features. It is well known that mobile genetic elements, such as plasmids, can be carriers of antibiotic resistance genes and can transfer genetic material across different bacteria and facilitate horizontal gene transfer (Partridge et al., 2018). Antibiotic resistance genes on plasmids are not the only route from which said genes can be accessed by members of the microbiome. Other mobile genetic elements such as transposons allow the transferal from genomes to plasmid and vice versa (Babakhani and Oloomi, 2018), opening up other ways to spread antibiotic resistance genes. Almost none of the antibiotic resistance genes could be traced to more precise taxonomic groups than kingdom (e.g. family,

order, genera etc.). The ones that could be traced were for the most part not relevant in the context of KRINKO such as mupirocin, tetracycline or erythromycin. Interestingly, these were also found during the current study on *Bifidobacterium* genomes, indicating that even conventional beneficial members of the infant microbiome might be sources of antibiotic resistance genes.

Other studies have already described rich resistomes in the gut microbiome of preterm infants (Rose et al., 2017). Taken together with the results of the PRIMAL samples studied it is also abundantly clear, that the resistomes of infants and mothers vary greatly and a definition of dysbiosis that incorporates antibioitc resistance genes in some form needs to be infant specific.

Summary of the strengths and weaknesses of conventional culture based methods and NGS approaches

Conventional microbiology and NGS based methods both have advantages and disadvantages when it comes to assessing "dysbiosis". The culture based bacterial identification is a great way to identify easy-to-culture, living microbes and it can provide information on the presence of antibiotic capabilities against specified antibiotics of the studied species. On the downside, it is limited in its throughput rate and by the predefinition of bacteria and antibiotic resistances that can be found. Especially during KRINKO, it is unclear how relevant the identified bacteria are from a clinical perspective, since the sequencing based approach revealed that conventionally identified bacteria are only a subset of many more bacteria that are present and that do not have to be casually linked to any disease. These problems become even more pressing when applying culturing techniques on materials with lower bacterial load such as blood, where NGS based data have been shown to be superior in identifying bacteria that cause sepsis (Grumaz et al., 2019). Further, culture based methods lack the capabilities to feasibly analyze entire microbiome compositions and assess metrics like the diversity of a sample.

NGS techniques, particularly 16S rRNA sequencing due to its low cost, have the potential to become highly relevant in the clinical practice and have already yielded some clinical applications (Akram et al., 2017). They provide culture-independent classifications of all bacterial DNA and are therefore not limited by a predefined set of target bacteria and do not rely on the easy culturability of specimens. Whole genome shotgun sequencing in particular can provide information on antibiotic resistance genes that are present in the microbiome irrespective of their current host. This way, reservoirs of antibiotic resistance genes and hazards of horizontal gene transfer can be analyzed. NGS techniques can however not easily assess if sequenced DNA stems from living or dead bacteria and would need further modifications (e.g. transcriptomics) to make a statement about the actual expression of antibiotic resistance genes.

When looking at the concept of dysbiosis it appears that culture based methods and NGS have some overlapping but also some principally different insights that can be gained: Conventional methods, especially in the context of KRINKO, are mostly concerned with something that could

be called "acute dysbiosis" or "short-term dysbiosis". The identification of specific pathogens and their respective antibiotic resistance potential that demand direct action. NGS approaches - albeit they have been shown to be of use in identifying "acute dysbiosis" (Grumaz et al., 2019) - produce results that might be used to model the complex interactions in the microbiome e.g. in context of diseases that manifest far later in life, like asthma (Hufnagl et al., 2020) in an attempt to identify states that could be termed "chronic dysbiosis" or "long-term dysbiosis".

In light of the increasing risk of spreading antibiotic resistance genes and the need for reliable identification of emerging health hazards by infectious agents such as multidrug resistant *Enterobacter cloacae* (Tindall et al., 2017) on the one hand and concerns about the possible long term health impact of changes in the early developing infant gut, it seems wisest to combine the strengths of conventional methods and NGS techniques in the future to generate the most reliable assessment of the microbial ecosystems in the gut.

4.3.2 What is a normal microbiota in infants?

A common understanding of the "normal" microbiota in infants is still largely lacking (Underwood et al., 2020). Since the development of the early microbiome is highly dynamic, it is reasonable to differentiate the analysis between meconium microbiome and the gut microbiome of infants after about four weeks.

Meconium microbiome

In the context of the first stool, the meconium, opinions about the microbiome composition vary greatly. The data gathered by some researches seem to suggest a bacterial colonization of the meconium *in utero*. Other researches, invoking data produced with rigorous controls strictly deny the colonization of meconium prior to birth (Kennedy et al., 2021). Overall, the information on meconium microbiomes is limited (Devi Oktaviyani et al., 2021). However, some studies have identified correlations between the composition of meconium microbiomes and diseases like necrotizing enterocolitis (Heida et al., 2016). In the current study, the meconium microbiome of 565 infants was analyzed to provide further baseline characterization of this understudied population. Robust negative and positive controls were used to preempt contamination from obscuring the microbiome. Meconium samples were found to possess a very high relative abundance of the genus *Staphylococcus* which is in line with previous reports in the literature in which *Staphylococcus aureus* and *Staphylococcus epidermidis* were ubiquitous in the gut microbiome of preterm infants (Rose et al., 2017). These bacteria are part of the typical adult skin microbiota and are likely transfered to the infant through mother-child contact (Ferretti et al., 2018).

Further tests with spike-in controls revealed that the bacterial load in the PRIMAL studies

was very low - a finding that is consistent with a rapid development from an unpopulated or sparsely populated gut. Overall, the data produced during the current study show that meconium has a distinct microbiome that likely interacts with the host immune system e.g. through produced metabolites (Petersen, Dai, et al., 2021) from the earliest stages of development.

Microbiome composition after one month - Comparison with reference datasets

To set the microbiome structures of the PRIMAL visit 2 samples at one month of life into context, reference studies from full and preterm infants were used to identify similarities and differences between the studied populations. The most dominant bacterial genus in all studied samples was *Bifidobacterium*. Given what is known about the early life microbiota in infants that was expected. *Bifidobacterium* is often described as a key taxon and numerous species of that genus like *Bifidobacterium longum*, *Bifidobacterium breve* and *Bifidobacterium bifidum* provide essential metabolites that cross talk with the developing infant immune system and perform other physiological tasks like metabolizing human milk oligosaccharides (Laursen et al., 2021). Interestingly, the PRIMAL samples showed a statistically significant difference in regards to alpha diversity compared to preterm reference samples but not to full term reference samples (see fig. 3.32). Based on the literature, one could have expected that the alpha diversity in preterm samples like PRIMAL and the preterm reference samples would be more similar to each other and lower than compared to full term references (Dahl et al., 2018). This discrepancy might be due to technical challenges like the use of different regions of the 16S rRNA gene, which is often cited as a cause for major complications in meta-analyses (Walker et al., 2015; Tremblay et al., 2015; Benjamin J. Callahan et al., 2017). Especially the use of the V4+V5 region of the 16S rRNA gene that was used in the full-term reference samples has been found to underrepresent taxa and alter the diversity profile in a systematic comparison study (Alcon-Giner, Caim, et al., 2017).

The relative abundance of *Bifidobacterium* was not significantly different in the preterm and PRIMAL samples but increased in the full term reference samples. Since the PRIMAL cohort and the preterm reference cohort were both exposed to a probiotic containing *Bifidobacterium*, the comparable detection levels of said genus might be evidence for a reproducible colonization in this population of preterm infants. A higher amount of *Bifidobacterium* in full term infants is extensively described in the literature (Milani et al., 2017) and forms the rationale for probiotic intervention studies, in order to bring preterm microbiomes closer to that of full term infants.

Reference and PRIMAL samples were compared on the basis of conventional notions of dysbiosis in the infant gut microbiome, i.e. low diversity and loss of beneficial taxa such as *Bifidobacterium* (Groer et al., 2020) (see also 1.1.4). The hypothesis was that a combination of low alpha diversity and a low relative abundance of *Bifidobacterium* would be indicative of a dysbiotic microbiome and would be primarily present in preterm infants and not in full term infants. When analyzing the bottom 10 % of samples with the lowest values for each of the two parameters in question, no clear trend was visible. Samples of all affiliations (preterm, full term and PRIMAL with both

treatments) were within the lowest ranking 10 % (see fig. 3.34). This lack of a clear clustering between full term and preterm infants when analyzed according to conventional metrics of dysbiosis rises doubts about the validity of these indicators as true markers of dysfunctional microbiome compositions.

4.3.3 Interactions in the early life microbiota and impacts of probiotic supplementation

Supplementing the diet of infants with probiotics during the earliest phase of life has been shown to produce remarkable health benefits (see 1.2.1). Most prominently it leads to a reduction in the occurrence of necrotizing enterocolitis (Sharif et al., 2020) and late-onset sepsis (Athalye-Jape and Patole, 2019) and, especially when *Lactobacillus acidophilus* and *Bifidobacterium infantis* are administered, to an increase in the growth rate of preterm infants (Härtel et al., 2017).

The co-occurrence networks calculated during the current study show a web of statistically significant shared appearances of specific taxa in the microbiome of infants that increases in complexity over the visits (see fig. 3.35). At visit 1 right after birth, the primary correlations are positive associations between *Bifidobacterium* and *Lactobacillus*, likely due to the supplementation with these taxa and not indicative of developed microbial ecological interactions. At visit 2 after one month of life *Bifidobacterium* has developed into the taxon with the most co-occurrences in the network. The presence of this genus was negatively correlated with many genera that include pathobionts such as *Clostridium*, *Klebsiella*, *Enterococcus* and *Enterobacter*. As statistical co-occurrence alone does not necessarily imply an interaction between taxa, examples of probiotic mechanisms in the literature can help to illuminate these observations further.

The probiotic mechanisms of specific probiotic strains like *Bifidobacterium animalis* subsp. *lactis* BB-12 have been studied extensively in an effort to gain a mechanistic explanation for the health benefits conveyed by taxa such as *Lactobacillus* and *Bifidobacterium* (Jungersen et al., 2014). Three major mechanisms that have been tested *in vitro* are hallmarks for probiotic action: First, an enhancement in the barrier function of the gut. Second, interactions with the host's immune system and third, the inhibition of pathogens. In regards to the results of the network analysis, especially the last point seems highly relevant. *In vitro* antagonism of *Bifidobacterium* against pathogens from diverse genera such as *Clostridium* and *Enterococcus* has been shown (Martins et al., 2009). Various implementations of this pathogen antagonism by probiotic bacteria have been reported in the literature (van Zyl et al., 2020):

1. Production of antimicrobial compounds (van Zyl et al., 2020)
    - bacteriocins (Cavera et al., 2015) like bifidocin A and b or nisin
    - organic acids (Peter A Vandenbergh, 1993)
    - bacteriocin-like inhibitory substances (Servin, 2004)

- hydrogen peroxide (Fayol-Messaoudi et al., 2005)
- siderophores (D. J. O'Sullivan, 2004)
- biosurfactants (Ciandrini et al., 2016)

2. Competitive exclusion of pathogens (Mathipa and Thantsha, 2017)
    - competition for nutrients
    - competition for adherence space

All these effects might have contributed to some extent to the observed negative correlation of *Bifidobacterium* and the other genera in the visit 2 (day 28) network. Strikingly, although *Lactobacillus* shares many of the features in regards to combatting pathogens e.g. the production of lacticin B against *Clostridium* (van Zyl et al., 2020), the genus was not significantly correlated to any other genus of the top 15 most abundant taxa in visit 2 samples (day 28). The overall low relative abundance of *Lactobacillus* in the samples might have been a contributing factor to the statistical non-significance. In the later stages of microbiome development (visit 3 after one year and mother samples) a further increase in complexity and a decrease of the importance of *Bifidobacterium* in the ecology of the gut microbiome can be observed, probably owing to less need of the genetic capabilities that this specific genus provides.

Data on microbiome changes during probiotic supplementation in the early life remain sparse. However, a steep increase in the relative abundance of *Bifidobacterium* has often been observed (Alcon-Giner, Dalby, et al., 2020; van Best et al., 2020), when probiotic combinations of *Bifidobacterium* and *Lactobacillus* have been administered. The relative abundance of *Lactobacillus* and its increase is substantially lower, even though the probiotics used often have the same amounts of *Bifidobacterium* and *Lactobacillus*. The same can be seen in the data of the current study were Probactiol®infantis consisting of equicellular amounts of *Bifidobacterium animalis* subsp. lactis (BB-12), *Bifidobacterium infantis* and *Lactobacillus acidophilus* (La-5) (each as $1.5*10^9$ CFU) was administered. Both treatment groups had considerable higher amounts of *Bifidobacterium* than *Lactobacillus* (although both genera were increased in treatment group A compared to B). The quality control measures during the current study confirmed the presence of *Lactobacillus* in the verum capsules. Thus, in turn a biological reason for the lower abundance of *Lactobacillus* can be suspected.

Bifidobacterium and *Lactobacillus* are both lactic acid bacteria (LAB) (Tabasco et al., 2014). As such they occupy similar niches and might be in a competitive struggle for the same resources (Tabasco et al., 2014). Thus, while both genera might be beneficial to the host, they might not necessarily be to each other and *Bifidobacterium* might be more potent in out-competing *Lactobacillus* than the other way around leading to the unequal relative abundance data in the current and other studies. In a study by Maldonado-Gomez et al. the long-term (6 months) persistence of *Bifidobacterium longum* could be predicted by the prior underrepresentation of genes for the utilization of complex carbohydrates, indicating a competition driven colonization

of the gut ecosystem that depends on individualized factors in the host resident microbiome (Maldonado-Gómez et al., 2016)

Overall, the effect of the treatment on the microbiome seemed limited to changes in the relative abundance of *Bifidobacterium* and *Lactobacillus*, as no significant alterations in any other taxa or alpha diversity was registered (see fig. 3.23 and 3.24). These findings are consistent with reports from the literature that detected increased levels of *Bifidobacterium* yet no significant changes in the alpha diversity when administering probiotics to infants (Quin et al., 2018).

The changes in the relative abundance of *Bifidobacterium* and *Lactobacillus* according to treatment group do not seem to persist after one year, since no significant differences in genus abundance could be seen in the visit 3 sample.

4.4 Conclusion

In the current study, technical and conceptual challenges during the analysis of microbiomes using conventional and sequencing based methods have been highlighted and discussed. The traditional concept of dysbiosis has been investigated in terms of possible interactions between different taxa, the presence of antibiotic resistance genes and though comparisons with "normal" microbiomes based on deferentially abundant taxa and diversity. In the following sections, lessons learned during the current study will be compiled and conclusions on the nature of the concepts eu- and dysbiosis will be proposed.

4.4.1 Dysbiosis - causality versus correlation

It is notoriously difficult to differentiate between true causality and mere correlation in science (Rohrer, 2018). This applies in the same way to the highly complex systems of microbiomes with their myriad of influencing parameters and is especially true for the study of dysbiotic states (Walter et al., 2020). Even double-blinded, placebo controlled studies are subjected to several confounding factors (Vujkovic-Cvijin et al., 2020). This can also be seen in the current dataset from PRIMAL where factors impacting the microbiome like the hospital site of collection, complicate the statistical analysis. Dimensionality reduction algorithms further introduce imprecisions by aggregating data to allow for potential generalizable conceptualization based on observed data.

In an attempt to better understand mechanisms of action of host-bacteria interactions at a higher resolution, there are two major approaches besides *in vivo* studies in humans. One is to create *in vitro* models of single bacterial species or entire microbial ecosystems to emulate physiological processes (McDonald, 2017). As discussed earlier, this has been successfully done for the mode of action of many probiotic bacteria. The other option is the development of humanized rodent models based on gnotobiotic animals (Samuel and Gordon, 2006). Germ-free

mice are colonized with bacteria originating from the human gut and concrete changes in the metabolome or the presence of specific genes can then be exploited to propose mechanistic explanations for e.g. observed protection against food allergies (Feehley et al., 2019). Both methods aim to provide a functional understanding of how commensal or pathogenic bacteria interact with a host organism and with each other during health and disease to form eu- and dysbiotic states. However, often times the approaches are hindered in doing so by the sheer complexity of the system in question (Walter et al., 2020).

In light of these challenges and uncertainties, it is hardly surprising that the translational microbiome research is still in its infancy (Gambardella et al., 2021). Most interventional studies with e.g. probiotics, that show a positive effect on the host health still operate in a "black box" modus operandi with no conceptual understanding of why the intervention works or where its limitations lie (Suez, 2019). Current microbiome research like a recent paper from Henrick et al. has increased its focus on functional descriptions of dysbioitc states. In their study, the authors described events of early life immune system imprinting that were dependent on the presence of *Bifidobacterium* and especially certain genes (HMO metabolizing genes) that are provided by this genus. They further propose mechanistic explanations for how and why the presence or absence of said genes and their injection into the system through the administration of a probiotic could mediate intestinal cytokine levels by upregulating the immunoregulatory factor galectin-1 (Henrick et al., 2021). However, large scale studies like these are very resource consuming and heavily rely on a prior understanding of the composition of the microbiome in question. Due to the huge variety of microbiomes especially in infants, both temporal and individual, studies like PRIMAL that rigorously describe the taxonomic and genetic makeup of a huge number of samples will further on be the basis to provide frameworks for any fruitful microbiome research.

4.4.2 Why dysbiosis itself is not an answer - dysbiosis as a defined dysfunctional state

A pubmed search with the term "dysbiosis" yields 12,105 results (as of November 2021). Few other topics dominate the realm of microbiome research as much as dysbiosis does. The term is used in the context of nearly every microbiome research, from agrarian studies (Balbín-Suárez et al., 2021) to the study of microbial changes during a stay in outer space (Cervantes and Hong, 2015; S. Turroni et al., 2020). Most often however it is used to discuss changes in the human gut microbiome compared to a proposed eubiosis in order to explain the emergence of specific disease (Iebba et al., 2016). As summarized in 1.1.4, the common definitions of dysbiosis entail 1. a loss of overall diversity, 2. a loss of beneficial taxa, 3. an overrepresentation of pathogens and in some cases 4. shifts in metabolic activity (see tab. 1.2). While, this consensus definition might be able to capture commonalities between most states described as dysbiotic, it does not provide actionable advice on the specifics of dysbiosis.

This context dependency of dysbiosis has nicely been illustrated by Fettweis et al. in a study performed during the integrative human microbiome project were an increase in diversity of vaginal microbiomes was clearly associated with detrimental health consequences such as compromised birth outcomes (Fettweis et al., 2019). Intervening in complex microbiome structures only having a very vague concept of dysbiosis such as described above can lead to various outcomes. In a best case scenario, interventions like probiotics can have a very positive impact on the health of the patients. These health promoting benefits have been extensively shown in preterm and caesarean-born infants when administering *Bifidobacterium* and *Lactobacillus* (Esaiassen, Hjerde, et al., 2018) and have also been shown for the exact supplementation used in the current study (Fortmann et al., 2020). They can help to increase the amount of beneficial *Bifidobacterium* and move their relative abundance levels closer to that of full term infants (Korpela et al., 2018). However, the occurrence of inadvertent effects is also possible: In a retrospective clinical comparison, the probiotic exposure during infancy had limited effects on gut microbial composition yet was associated with increased infections later in life (Quin et al., 2018). Other reports describe a hindrance rather than an aid through probiotic after antimicrobial therapy when trying to regain the gut microbiome baseline (Suez et al., 2018). Further problems might arise when the exact genetic makeup of the probiotic is not sufficiently assessed and the probiotic might serve as a reservoir of antibiotic resistance or other unintended genes. Especially when administered to highly vulnerable or immunocompromised patients, some probiotic strains might exploit the weakened host immune system and act as opportunistic pathogens (Kothari et al., 2019). Occasional reports of sepsis caused by *Bifidobacterium* can be found in the literature (Ha et al., 1999; Avcin et al., 2015). This lack of a real conceptual understanding further leads to confusions about what organizational level is import in a probiotic: Are all or most members of a higher taxonomic rank like genus (i.e. *Bifidobacterium*) beneficial or do only specific sub strains show probiotic activity? Studies that report colonization increases of *Bifidobacterium* in neonates even when administered dead cells (Tanaka et al., 2019) rise further questions and shift the focus of dysbiosis from bacteria to genes.

Bacterial taxonomy - what is a bacterial species?

In his article "The mind of the species problem", evolutionary biologist Jody Hey nicely illustrates the continual failure of biologists to unambiguously define what a specific species (or any taxonomic rank for that matter) is. (Hey, 2001) He lists over 20 common species concepts that are in part mutually exclusive. The lack of a coherent concept, he argues, stems primarily from the circumstance that evolution is a continuous process without sharp boundaries and the attempt to classify and categorize organisms into discrete taxa is thus fundamentally at odds with the fact of evolution. The idea of a species being members of a group that can sexually reproduce, is a viable approximation for macroscopic organisms like animals, fungi or plants. Sexual reproduction inherently selects members of the same reproductive group to have compatible sets of genes. A macroscopic phenotype of eukaryotes is thus a workable proxy for the representative core genome of most of its species members. For example, McCarthy et al. found between 80 % and

90 % of the genes from 4 fungal species to be core genes that were highly conserved (McCarthy and Fitzpatrick, 2019). In humans, only between 0.63 % and 1.33 % of the genes are not part of the core genome (Li et al., 2010)

The same does not hold true for bacteria. In bacteria, large fractions of the genome are considered to be part of a pan-genome. The pan-genome is defined as all the genes that are found within members of the same taxonomic clade e.g. a species (Vernikos et al., 2015) but not necessarily in every individual. In extreme cases like *Escherichia coli* the core genome present that can be found in all individuals is only around 3100 genes while the entirety of the pan-genome is around 89000 different genes (Land et al., 2015). This, in turn, means that only about 3.5 % of genes from the pan-genome are responsible for a phenotype that leads to a classification as *Escherichia coli*, strongly restricting the information on genetic capacity that can be derived from this taxonomic classification. Moreover, horizontal gene transfer is omnipresent in the bacterial world as is exemplary shown through the extensive transfer of antibiotic resistance genes in the human gut (Shoemaker et al., 2001), further obscuring the specific genetic makeup of a given species. Findings like these underscore the notion of some researches that species concepts used in the phylogeny of multicellular organisms do not apply in the world of asexually reproducing microbes (Doolittle, 2012).
A different approach of looking at genetic variation in bacteria, based on next-generation sequencing techniques, is summarized in a recent review by the EMBL researches Thea van Rossum, Pamela Ferretti, Oleksandr M. Maistrenko and Peer Bork (Van Rossum et al., 2020): Instead of conventional top-down species definitions that lack e.g. significance in establishing the presence or absence of specific genes, operationally defined groups of genetic variation can be established. These groups have no predefined size or phenotypic similarities and are dependent on the biological question that is being asked. This analysis of sequencing data at the level of individual genomes has several applications like source tracking of bacterial specimens or exact ecological niche inference based on genetic makeup, that are simply not possible using broader species definitions.

Dysbiosis should be interpreted as a concrete dysfunctional state with defined genetic deficits rather than a vague state of bacterial imbalance

Following the discussion above, it is abundantly clear that a dysbiosis definition based on vague criteria like the abundance of beneficial bacteria, bloom of pathogens and overall diversity is often too coarse-grained to be a workable supposition for most hypotheses driven research. However, the general term might still be useful to indicate the presence of some kind of atypical state in a microbiome. Much like the words "sickness" or "illness" are used by physicians to indicate a deviation of a health status from a theoretical status of "healthy" that is not defined in detail, dysbiosis might be used to indicate a deviation of a microbiome composition from a "normal" state without reference to any specific disease. While it is tempting to assume that a long co-evolution of hosts and their respective microbiomes has led to a close mutualistic

relationship for the benefit of all parties and that changes in the microbiome therefore represent a negative development per se, some researches have argued that this hologenome concept is not the most plausible null hypotheses (Moran and Sloan, 2015). Instead the scientist argue that in the evolutionary race, both the host and members of the microbiome do not necessarily have aligned priorities in terms of resource exploitation and reproduction, reducing the impact of selection at the holobiont level. Further, competition between different species within the gut could lead to drastic changes in the taxonomic composition of the microbiome without any measurable effect on the health of the host. Not unlike the above mentioned terms "sickness" and "illness" do for health in general, dysbiosis used in this way would not necessarily entail a reference to a specific disease or an inherent remedy suggestion to alleviate the abnormal status but instead solely indicate it. An example of this usage can be found in a study by Cardoso et al. where the term dysbiosis was applied to describe different microbiome compositions that individualized the fitness cost of antibiotic resistance genes for bacteria without reference to a specific host disease (Leónidas Cardoso et al., 2020).

There are at least two possible ways of how the dysbiosis concept can be moved into the realm of actionable hypotheses. First, a general definition of a clinical relevant dysbiosis might be proposed that entails concrete functional impairments of the microbiome that are statistically negative for the host health. These include a bacterial milieu that provides a weakened general colonization resistance or a diminished capacity to perform digestive tasks, immune education, or other beneficial physiological services (Tiffany and Bäumler, 2019).
Second, the dysbiosis concept can be used in the context of defined diseases. It will be valuable to differentiate between short-term (acute) and long-term (chronic) dysbiosis as proposed in 4.3.1. Analog to acute and chronic illnesses (E J Murrow and F M Oglesby, 1996), short-term dysbiosis would deal with the occurrence of an acute perturbation of the microbiome while long-term dysbiosis would address chronic ecological disturbances in the host-microbiome system. An example for acute dysbiosis might be instances of infections with pathogens that have a well characterized pathology like *Clostridium difficile* infections (Burke and Lamont, 2014) and would respond to an effective treatment immediately. Long-term dysbiosis would be more concerned with complex pathologies that develop over long time horizons and have been linked to interactions between the microbiome and host like the emergence of asthma (Hufnagl et al., 2020).

In either case, appropriate interventions, like rationally designed probiotics, will mandatorily need to operate within a framework of a deep mechanistic and microecological understanding of how the microbiota interacts to create and prevent diseases. While results gained by both, conventional microbiological techniques and metagenomic sequencing approaches, provide invaluable information to a basic understanding of these dysbiotic states (as has been shown during the current study), they will likely not be sufficient either in resolution or scope for the necessary deep understanding in the long run.

4.5 Future perspectives for microbiome and dysbiosis research - Rational probiotic engineering

There is no doubt that the complex interactions within and between microbiomes and their hosts, will offer plenty of opportunities for future research. Despite enormous progress in the field of microbiome research in the last 15 years, there are still some major knowledge gabs and challenges. These include the interplay between microbes and host tissues and host genetics, as well as interactions within the microbiome (NIH Human Microbiome Portfolio Analysis Team, 2019).

One of the major translational goals is - and will continue to be - the rational design of probiotics and other interventions that will allow the targeted control of enteric pathogens and help to alleviate microbiome related diseases (Mathipa and Thantsha, 2017). In order to achieve this, positive health outcomes that are empirically observed e.g. during the administration of probiotics in preterm infants (Härtel et al., 2017), need to be accompanied by a mechanistic understanding. This will reveal the scope and limitations of these interventions and will likely unveil options of rational improvement to the treatments. There is clearly an urge from companies and investors to bring microbiome based measurement methods (e.g. based on 16S rRNA gene sequencing (Liang et al., 2017)) and therapies to the market, yet a clear scientific consensus on the general effectiveness and the best modus operandi is still pending (Kothari et al., 2019). Some scientists even go so far as to state that the entire field of microbiome research should still largely be performed within a basic research setting and is far from ready for hypothesis driven studies yet alone rational interventions (A. Tripathi et al., 2018).

Besides the use of culture based microbiological methods and metagenomic sequencing based methods there is a number of promising techniques that could provide deeper insights and further the understanding of the complex world of the microbiota.
Examples for these include:

- 3-D microbiomics, in which the critical spatial information of microbial ecosystems that reveal local interactions of microbes are conserved (Garg et al., 2017).
- Multi-omics approaches combining meta-genomics, meta-transcriptomics, meta-bolomics and meta-proteomics to identify microbe-drug interactions (Zhang et al., 2019).
- Third-generation long read sequencing technologies like nanopore or pacific bioscience sequencing to boost the resolution of taxonomic classifications without the need for genome assembly (Earl et al., 2018).
- single cell microbiomics to identify the genetic features of individual members of the microbiome whether they are bacterial (Chijiiwa et al., 2020), protozoan (Rossi et al., 2019) or even viral (Garcia-Heredia et al., 2021)

Hopefully, the combination of all these methods will provide a big-data pool in the future that can be successfully mined to gain an in-depth understanding of the inner workings of microbiomes (Cheng et al., 2019). In line with the overarching meta trend of precision medicine (König et al., 2017), it is probably safe to assume that the microbiome research of the future will become highly individualized and every host-microbiome interaction will be analyzed as a unique ecosystem with unique challenges and opportunities for intervention.

Abstract

5

Dysbiosis, generally defined as a reduction of diversity and beneficial bacteria and a bloom of pathobionts in a given microbiome, has been proposed as a concept to explain why changes in the (gut) microbiome composition can lead to negative health outcomes, especially in the vulnerable population of preterm infants. This microbiome state is in contrast to an implicit state of balanced eubiosis.

The main aims of this thesis were to analyze and comprehensively describe the gut microbiome composition of preterm neonates and mothers enrolled in the PRIMAL study cohort and to evaluate the overall utility of the dysbiosis concept in the context of the preterm infant gut microbiome and the use of probiotics.

For 1353 infant fecal samples comprised of three visits (day 0, day 28 and day 365 of life) and 290 associated mother samples, taxonomic profiles were compiled through 16S rRNA gene sequencing to establish a baseline characterization of their microbiome composition. Right after birth, infant gut microbiomes were dominated by the genera *Staphylococcus* and *Bifidobacterium*. During the further development over the course of the first month of life, *Bifidobacterium* gained dominance in the microbiomes. *Bifidobacterium* remained the most abundant genus after one year of life during which the infant microbiome slowly matured to more closely resemble the adult mother microbiome which was primarily dominated by the genera *Bacterioides*, *Blautia* and *Faecalibacterium*.

To evaluate the usefulness and validity of the dysbiosis concept, results from conventional, culture based microbiological analyses and next-generation sequencing were compared with a focus on taxonomic classifications and antibiotic resistance capacity. Further, possible interactions in the microbiomes from the PRIMAL study cohort were analyzed through co-occurrence network analyses and the PRIMAL microbiomes were set into the context of publicly available preterm and full term reference microbiome datasets. Conceptual challenges with the generic term dysbiosis were resolved by proposing that dysbiosis should be used as a description of a dysfunctional state with defined genetic deficits within the microbiome. A distinction can be made between dysbiosis as an umbrella term for a general deviation from the genetic capabilities of a normal reference microbiome on the one hand and a two-part actionable concept on the other hand that can drive hypothesis driven research. This two-part concept differentiates between a short-term (acute) dysbiosis characterized by the presence of a concrete pathogen and a long-term (chronic) dysbiosis characterized by complex microecological interactions with the host that cause chronic diseases. Overall, this study demonstrates the complexity within the human gut microbiome even at the earliest stages of life and stresses the importance of a deep mechanistic understanding of host-microbe interactions to develop rationally designed therapeutic interventions.

Abstract (Deutsch)

Das Konzept der Dysbiose, gemeinhin definiert als ein Verlust von Diversität und nützlichen Bakterien sowie einer Zunahme von Pathobionten, wurde als mögliches Modell vorgeschlagen, um zu erklären, wie Änderungen in der Zusammensetzung des Mikrobioms - vor allem in besonders vulnerablen Populationen wie Frühgeborenen - zu Krankheiten führen können. Dabei wird der dysbiotische Zustand einem impliziten eubiotischen Zustand gegenübergestellt.
Die primären Ziele dieser Arbeit waren die Mikrobiom Zusammensetzung von frühgeborenen Neonaten und Müttern der PRIMAL Studie umfassend zu beschreiben und im Weiteren die Nützlichkeit des Dysbiose Konzepts im Kontext Frühgeborener und probiotischer Intervention zu analysieren. Für 1353 Säuglingsproben bestehend aus drei Untersuchungszeitpunkten (Tag 0, Tag 28 und Tag 365) und für 290 assoziierte Mutterproben wurden mittels 16S rRNA Gen Sequenzierung taxonomische Profile erstellt.
Direkt nach der Geburt waren die Mikrobiome dominiert durch Bakterien der Gattungen *Staphylococcus* und *Bifidobacterium*. Im Laufe der weiteren Entwicklung innerhalb des ersten Lebensmonats wurde *Bifidobacterium* zum dominantesten Keim. Nach der Maturierung im ersten Lebensjahr war noch immer *Bifidobacterium* die häufigste Gattung, während das restliche kindliche Mikrobiom in seiner Zusammensetzung zusehends dem erwachsenen Mikrobiom entsprach, das primär aus Bakterien der Gattung *Bacterioides*, *Blautia* und *Faecalibacterium* zusammengesetzt war.
Um die Nützlichkeit des Dysbiose Konzepts zu untersuchen wurden Ergebnisse aus der konventionellen Mikrobiologie mit Daten aus dem Next-Generation-Sequencing besonders im Hinblick auf taxonomische Zusammensetzung und Antibiotika Resistenz verglichen. Zudem wurden mögliche Interaktionen innerhalb der Mikrobiome durch Netzwerk-Analysen untersucht und die PRIMAL Daten wurden mit öffentlich zugänglichen Referenzdatensätzen von Früh- und Reifgeborenen verglichen. Als Vorschlag zur Lösung konzeptioneller Probleme mit der generischen Definition der Dysbiose wurde diese als spezifischer dysfunktionaler Zustand mit charakteristischen genetischen Defiziten innerhalb des Mikrobioms umdefiniert. Es kann eine Unterscheidung zwischen Dysbiose als generelle Abweichung vom genetischen Normalzustand innerhalb des Mikrobioms auf der einen Seite und einer krankheitsspezifischen Dysbiose auf der anderen Seite getroffen werden. Die spezifische Definition kann sich dabei auf akute Dysbiose wie die Anwesenheit eines konkreten Krankheitserregers im Mikrobiom oder auf einen chronischen Zustand, wie komplexe mikroökologischen Interaktionen, die zu Langzeitschäden in der Gesundheit des Wirtes führen, beziehen. Zusammenfassend zeigt diese Arbeit die enorme Komplexität im menschlichen und sogar im frühkindlichen Darmmikrobiom sowie die Notwendigkeit für weitere Forschung zum mechanistischen Verständnis der Wirt-Mikrobiom Interaktionen um rationale therapeutische Interventionen entwickeln zu können.

Acknowledgment

First, I would like to thank Prof. Thomas Hankeln and Prof. Stephan Gehring for supervising and reviewing this thesis.

Further, I want to thank and acknowledge all the people that provided invaluable technical support over the course of this thesis: Dr. Claudius Meyer for organizing and operating the Fecal Core Center Mainz (FCCM) with the PRIMAL main database for providing data critical for the success of this study. Katja Hilbert for her wet lab expertise in processing the PRIMAL samples and her organizational skills in sample handling during the entire course of the study, thus providing the main body of raw data (16S rRNA gene data) for the analyses in this thesis. Dr. Ekkehard Siegel and his entire team at the institute for "Medizinische Mikrobiologie und Hygiene" of the university medicine Mainz for performing the culture based KRINKO analyses. Dr. Vladimir Benes and his entire team at the Genomics Core Facility at the European Molecular Biology Laboratory (EMBL), Heidelberg, for performing the whole-genome shotgun sequencing. Thanks also goes to Prof. Peer Bork and his bioinformatics team at the EMBL, Heidelberg, especially Dr. Thea Van Rossum for performing and providing data on the presence of antibiotic resistance genes in the PRIMAL sub-cohort.

I would like to thank the whole PRIMAL Consortium, especially Prof. Christoph Härtel and Prof. Phillip Henneke, for the inception of the PRIMAL study. For collecting and coordinating the transport of numerous stool samples I would like to thank all the nurses and physicians at the participating PRIMAL centers, specifically: Bastian Siller (Lübeck), Thorsten Körner and Birte Tröger (Bremen), Anja Stein (Essen), Hans Fuchs (Freiburg), Lutz Koch (Hamburg), Hannover, Johannes Pöschl (Heidelberg) , Christoph Derouet (Homburg), Hans Proquitté and Libor Swoboda (Jena), Friedrich Ebinger and Frank Dohle (Paderborn), Dirk Olbertz (Rostock), Markus Knuf (Wiesbaden), Esther Schmidt (Schwerin), Christian Wieg (Aschaffenburg), Andreas Müller and Soyhan Bagci (Bonn). A Special thanks to all the participating parents and infants without whom the study could not have taken place in the first place.

Further thanks goes to Flores Kneilmann and all the other members of the "AG Pädiatrische Immunologie und Infektiologie" of the University Medical Center Mainz for executive assistance in the laboratory.

For their professional input at all stages of this thesis I would especially like to thank Prof. Stephan Gehring, Dr. Claudius Meyer, Katja Hilbert, Dr. Thea Van Rossum and Dr. Rebecca Knoll.

I would like to thank all members of the "AG Pädiatrische Immunologie und Infektiologie" of the University Medical Center Mainz for the great time I had during my doctoral studies.

A special thanks goes to my former biology teacher at the Max-Planck-Gymnasium Trier, Guido Börner, whose enthusiasm for molecular biology was contagious and made me want to pursue the subject professionally.

For their continuous support on all levels, I want to deeply thank my parents, Bettina Schömer-Klopp and Heinz-Peter Klopp, and the best brother one could ever wish for, David Klopp.

Bibliography

Aceti, Arianna, Isadora Beghetti, Luca Maggio, et al. (Oct. 10, 2018). “Filling the Gaps: Current Research Directions for a Rational Use of Probiotics in Preterm Infants”. In: *Nutrients* 10.10, p. 1472 (cit. on p. 15).

Akram, A., M. Maley, I. Gosbell, T. Nguyen, and R. Chavada (2017). “Utility of 16S rRNA PCR Performed on Clinical Specimens in Patient Management”. In: *International Journal of Infectious Diseases* 57, pp. 144–149 (cit. on p. 83).

Alcon-Giner, Cristina, Shabhonam Caim, Suparna Mitra, et al. (Dec. 2017). “Optimisation of 16S rRNA Gut Microbiota Profiling of Extremely Low Birth Weight Infants”. In: *BMC Genomics* 18.1, p. 841 (cit. on p. 85).

Alcon-Giner, Cristina, Matthew J. Dalby, Shabhonam Caim, et al. (Aug. 2020). “Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study”. In: *Cell Reports Medicine* 1.5, p. 100077 (cit. on pp. 18, 32, 69, 87).

Allen, Heather K, Luke A Moe, Jitsupang Rodbumrer, Andra Gaarder, and Jo Handelsman (2009). “Functional Metagenomics Reveals Diverse B-Lactamases in a Remote Alaskan Soil”. In: *The ISME Journal*, p. 9 (cit. on p. 82).

Altschul, S F, W Gish, W Miller, E W Myers, and D J Lipman (1990). “Basic Local Alignment Search Tool.” In: *Journal of molecular biology* 215.3, pp. 403–10 (cit. on p. 31).

Amon, Protima and Ian Sanderson (Oct. 2017). “What Is the Microbiome?” In: *Archives of disease in childhood - Education & practice edition* 102.5, pp. 257–260 (cit. on p. 1).

Anaconda Inc. (2021). *Anaconda Software Distribution*. Version 4.9.2 (cit. on p. 24).

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data* (cit. on p. 24).

Arnold, Carrie (July 2014). “Rethinking Sterile: The Hospital Microbiome”. In: *Environmental Health Perspectives* 122.7 (cit. on p. 78).

Arrieta, Marie-Claire, Leah T. Stiemsma, Nelly Amenyogbe, Eric M. Brown, and Brett Finlay (Sept. 5, 2014). “The Intestinal Microbiome in Early Life: Health and Disease”. In: *Frontiers in Immunology* 5 (cit. on p. 78).

Athalye-Jape, Gayatri, Girish Deshpande, Shripada Rao, and Sanjay Patole (Dec. 1, 2014). “Benefits of Probiotics on Enteral Nutrition in Preterm Neonates: A Systematic Review”. In: *The American Journal of Clinical Nutrition* 100.6, pp. 1508–1519 (cit. on p. 14).

Athalye-Jape, Gayatri and Sanjay Patole (2019). “Probiotics for Preterm Infants – Time to End All Controversies”. In: *Microbial Biotechnology* 0, pp. 1–5 (cit. on pp. 14, 86).

Avcin, Simona Lucija, Marko Pokorn, Lidija Kitanovski, Manica Mueller Premru, and Janez Jazbec (Sept. 2015). “*Bifidobacterium Breve* Sepsis in Child with High-Risk Acute Lymphoblastic Leukemia”. In: *Emerging Infectious Diseases* 21.9, pp. 1674–1675 (cit. on p. 90).

Babakhani, Sajad and Mana Oloomi (Nov. 2018). “Transposons: The Agents of Antibiotic Resistance in Bacteria”. In: *Journal of Basic Microbiology* 58.11, pp. 905–917 (cit. on p. 82).

Backhed, F. (Mar. 25, 2005). “Host-Bacterial Mutualism in the Human Intestine”. In: *Science* 307.5717, pp. 1915–1920 (cit. on p. 1).

Bäckhed, Fredrik, Josefine Roswall, Yangqing Peng, et al. (May 2015). “Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life”. In: *Cell Host & Microbe* 17.5, pp. 690–703 (cit. on p. 5).

Balbín-Suárez, Alicia, Samuel Jacquiod, Annmarie-Deetja Rohr, et al. (Mar. 31, 2021). “Root Exposure to Apple Replant Disease Soil Triggers Local Defense Response and Rhizoplane Microbiome Dysbiosis”. In: *FEMS Microbiology Ecology* 97.4, fiab031 (cit. on p. 89).

Baquero, Fernando, Ana P. Tedim, and Teresa M. Coque (2013). “Antibiotic Resistance Shaping Multi-Level Population Biology of Bacteria”. In: *Frontiers in Microbiology* 4 (cit. on p. 82).

Berger, Richard, Werner Rath, Harald Abele, et al. (Dec. 13, 2019). “Reducing the Risk of Preterm Birth by Ambulatory Risk Factor Management”. In: *Deutsches Aerzteblatt Online* (cit. on p. 7).

Bertelli, Claire, Trestan Pillonel, Anaïs Torregrossa, et al. (Mar. 15, 2015). “Bifidobacterium Longum Bacteremia in Preterm Infants Receiving Probiotics”. In: *Clinical Infectious Diseases* 60.6, pp. 924–927 (cit. on p. 14).

Bhullar, Kirandeep, Nicholas Waglechner, Andrew Pawlowski, et al. (Apr. 11, 2012). “Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome”. In: *PLoS ONE* 7.4. Ed. by Ramy K. Aziz, e34953 (cit. on p. 82).

Blaser, Martin J., Suzanne Devkota, Kathy D. McCoy, et al. (Dec. 2021). “Lessons Learned from the Prenatal Microbiome Controversy”. In: *Microbiome* 9.1, p. 8 (cit. on p. 4).

Boers, Stefan A., Ruud Jansen, and John P. Hays (June 2019). “Understanding and Overcoming the Pitfalls and Biases of Next-Generation Sequencing (NGS) Methods for Use in the Routine Clinical Microbiological Diagnostic Laboratory”. In: *European Journal of Clinical Microbiology & Infectious Diseases* 38.6, pp. 1059–1070 (cit. on p. 80).

Bokeh Development Team (2018). *Bokeh: Python Library for Interactive Visualization* (cit. on p. 24).

Brüssow, Harald (Mar. 2020). “Problems with the Concept of Gut Microbiota Dysbiosis”. In: *Microbial Biotechnology* 13.2, pp. 423–434 (cit. on p. 8).

Burke, Kristin E. and J. Thomas Lamont (Jan. 15, 2014). “*Clostridium Difficile* Infection: A Worldwide Disease”. In: *Gut and Liver* 8.1, pp. 1–6 (cit. on p. 92).

Bushnell, Brian, Jonathan Rood, and Esther Singer (Oct. 26, 2017). “BBMerge – Accurate Paired Shotgun Read Merging via Overlap”. In: *PLOS ONE* 12.10. Ed. by Patrick Jon Biggs, e0185056 (cit. on p. 31).

Callahan, Ben J., Kris Sankaran, Julia A. Fukuyama, Paul J. McMurdie, and Susan P. Holmes (2016). “Bioconductor Workflow for Microbiome Data Analysis: From Raw Reads to Community Analyses”. In: *F1000Research* 5.3, pp. 1492–1492 (cit. on p. 30).

Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, et al. (July 2016). “DADA2: High-resolution Sample Inference from Illumina Amplicon Data”. In: *Nature Methods* 13.7, pp. 581–583 (cit. on pp. 24, 25).

Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes (2017). “Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis”. In: *ISME Journal* 11.12, pp. 2639–2643 (cit. on pp. 18, 19, 85).

Cavera, Veronica L., Timothy D. Arthur, Dimitri Kashtanov, and Michael L. Chikindas (Nov. 2015). "Bacteriocins and Their Position in the next Wave of Conventional Antibiotics". In: *International Journal of Antimicrobial Agents* 46.5, pp. 494–501 (cit. on p. 86).

Cervantes, Jorge L. and Bo-Young Hong (May 13, 2015). "Dysbiosis and Immune Dysregulation in Outer Space". In: *International Reviews of Immunology*, pp. 1–16 (cit. on p. 89).

Cheng, Mingyue, Le Cao, and Kang Ning (Oct. 9, 2019). "Microbiome Big-Data Mining and Applications Using Single-Cell Technologies and Metagenomics Approaches Toward Precision Medicine". In: *Frontiers in Genetics* 10, p. 972 (cit. on p. 94).

Chijiiwa, Rieka, Masahito Hosokawa, Masato Kogawa, et al. (Dec. 2020). "Single-Cell Genomics of Uncultured Bacteria Reveals Dietary Fiber Responders in the Mouse Gut Microbiota". In: *Microbiome* 8.1, p. 5 (cit. on p. 93).

Chu, Derrick M., Maxim Seferovic, Ryan M. Pace, and Kjersti M. Aagaard (Oct. 2018). "The Microbiome in Preterm Birth". In: *Best Practice & Research Clinical Obstetrics & Gynaecology* 52, pp. 103–113 (cit. on p. 7).

Ciandrini, Eleonora, Raffaella Campana, Luca Casettari, et al. (Aug. 2016). "Characterization of Biosurfactants Produced by Lactobacillus Spp. and Their Activity against Oral Streptococci Biofilm". In: *Applied Microbiology and Biotechnology* 100.15, pp. 6767–6777 (cit. on p. 87).

Cock, P. J. A., T. Antao, J. T. Chang, et al. (June 1, 2009). "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics". In: *Bioinformatics* 25.11, pp. 1422–1423 (cit. on p. 24).

Cortes, Laetitia, Harm Wopereis, Aude Tartiere, et al. (Mar. 21, 2019). "Metaproteomic and 16S rRNA Gene Sequencing Analysis of the Infant Fecal Microbiome". In: *International Journal of Molecular Sciences* 20.6, p. 1430 (cit. on p. 18).

Costea, Paul I, Falk Hildebrand, Manimozhiyan Arumugam, et al. (2018). "Enterotypes in the Landscape of Gut Microbial Community Composition". In: p. 21 (cit. on p. 9).

Costello, E. K., K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, and D. A. Relman (June 8, 2012). "The Application of Ecological Theory Toward an Understanding of the Human Microbiome". In: *Science* 336.6086, pp. 1255–1262 (cit. on p. 5).

Costeloe, Kate, Pollyanna Hardy, Edmund Juszczak, Mark Wilks, and Michael R Millar (Feb. 2016). "Bifidobacterium Breve BBG-001 in Very Preterm Infants: A Randomised Controlled Phase 3 Trial". In: *The Lancet* 387.10019, pp. 649–660 (cit. on p. 14).

Cunnington, Aubrey J, Kathleen Sim, Aniko Deierl, et al. (Feb. 23, 2016). ""Vaginal Seeding" of Infants Born by Caesarean Section". In: *BMJ*, p. i227 (cit. on p. 6).

D'Costa, Vanessa M., Christine E. King, Lindsay Kalan, et al. (Sept. 2011). "Antibiotic Resistance Is Ancient". In: *Nature* 477.7365, pp. 457–461 (cit. on p. 82).

Da Costa-Luis, Casper and Nikolay Nechaev (May 24, 2021). *Tqdm: A Fast, Extensible Progress Bar for Python and CLI* (cit. on p. 24).

Dahl, Cecilie, Hein Stigum, Jørgen Valeur, et al. (Oct. 1, 2018). "Preterm Infants Have Distinct Microbiomes Not Explained by Mode of Delivery, Breastfeeding Duration or Antibiotic Exposure". In: *International Journal of Epidemiology* 47.5, pp. 1658–1669 (cit. on pp. 8, 85).

Davis, Nicole M., Diana M. Proctor, Susan P. Holmes, David A. Relman, and Benjamin J. Callahan (Dec. 2018). "Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data". In: *Microbiome* 6.1, p. 226 (cit. on p. 29).

De Goffau, Marcus C., Susanne Lager, Ulla Sovio, et al. (Aug. 15, 2019). “Human Placenta Has No Microbiome but Can Contain Potential Pathogens”. In: *Nature* 572.7769, pp. 329–334 (cit. on p. 4).

DeGruttola, Arianna K., Daren Low, Atsushi Mizoguchi, and Emiko Mizoguchi (May 2016). “Current Understanding of Dysbiosis in Disease in Human and Animal Models:” in: *Inflammatory Bowel Diseases* 22.5, pp. 1137–1150 (cit. on p. 8).

Deng, Huimin, Siqi Yang, Yucheng Zhang, et al. (Dec. 21, 2018). “Bacteroides Fragilis Prevents Clostridium Difficile Infection in a Mouse Model by Restoring Gut Barrier and Microbiome Regulation”. In: *Frontiers in Microbiology* 9, p. 2976 (cit. on p. 2).

Dermyshi, Elda, Yizhong Wang, Chongbin Yan, et al. (2017). “The “Golden Age” of Probiotics: A Systematic Review and Meta-Analysis of Randomized and Observational Studies in Preterm Infants”. In: *Neonatology* 112.1, pp. 9–23 (cit. on p. 14).

Deshpande, Girish, Gayatri Jape, Shripada Rao, and Sanjay Patole (Dec. 2017). “Benefits of Probiotics in Preterm Neonates in Low-Income and Medium-Income Countries: A Systematic Review of Randomised Controlled Trials”. In: *BMJ Open* 7.12, e017638 (cit. on p. 14).

Devi Oktaviyani, Raden Zulfa Alawiyyah, Putri Nusaiba, and Amarila Malik (Apr. 20, 2021). “A Review: Composition of Neonatal Meconium Microbiota and Its Role for Potential Probiotic”. In: *Pharmaceutical Sciences and Research* 8.1 (cit. on p. 84).

Dingle, Tanis C. and Susan M. Butler-Wu (Sept. 2013). “MALDI-TOF Mass Spectrometry for Microorganism Identification”. In: *Clinics in Laboratory Medicine* 33.3, pp. 589–609 (cit. on pp. 16, 80).

Dixon, Philip (Dec. 2003). “VEGAN, a Package of R Functions for Community Ecology”. In: *Journal of Vegetation Science* 14.6, pp. 927–930 (cit. on p. 31).

Dominguez-Bello, M. G., E. K. Costello, M. Contreras, et al. (June 29, 2010). “Delivery Mode Shapes the Acquisition and Structure of the Initial Microbiota across Multiple Body Habitats in Newborns”. In: *Proceedings of the National Academy of Sciences* 107.26, pp. 11971–11975 (cit. on p. 5).

Donia, Mohamed S., Peter Cimermancic, Christopher J. Schulze, et al. (Sept. 2014). “A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics”. In: *Cell* 158.6, pp. 1402–1414 (cit. on p. 19).

Doolittle, W Ford (2012). “Population Genomics: How Bacterial Species Form and Why They Don’t Exist”. In: p. 3 (cit. on p. 91).

Douglas, Chloe A., Kerry L. Ivey, Lito E. Papanicolas, et al. (Dec. 2020). “DNA Extraction Approaches Substantially Influence the Assessment of the Human Breast Milk Microbiome”. In: *Scientific Reports* 10.1, p. 123 (cit. on pp. 40, 78).

Draper, K., C. Ley, and J. Parsonnet (Aug. 24, 2017). “Probiotic Guidelines and Physician Practice: A Cross-Sectional Survey and Overview of the Literature”. In: *Beneficial Microbes* 8.4, pp. 507–519 (cit. on p. 12).

Duranti, Sabrina, Gabriele Andrea Lugli, Christian Milani, et al. (Oct. 2019). “*Bifidobacterium Bifidum* and the Infant Gut Microbiota: An Intriguing Case of Microbe-host Co-evolution”. In: *Environmental Microbiology* 21.10, pp. 3683–3695 (cit. on p. 5).

Dyar, O.J., B. Huttner, J. Schouten, and C. Pulcini (Nov. 2017). “What Is Antimicrobial Stewardship?” In: *Clinical Microbiology and Infection* 23.11, pp. 793–798 (cit. on pp. 3, 10).

E J Murrow and F M Oglesby (1996). “Acute and Chronic Illness: Similarities, Differences and Challenges”. In: pp. 47–51 (cit. on p. 92).

Earl, Joshua P., Nithin D. Adappa, Jaroslaw Krol, et al. (Dec. 2018). “Species-Level Bacterial Community Profiling of the Healthy Sinonasal Microbiome Using Pacific Biosciences Sequencing of Full-Length 16S rRNA Genes”. In: *Microbiome* 6.1, p. 190 (cit. on p. 93).

Edgar, Robert C. (2004). “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput”. In: *Nucleic Acids Research* 32.5, pp. 1792–1797 (cit. on p. 31).

Esaiassen, Eirin, Pauline Cavanagh, Erik Hjerde, et al. (Sept. 2016). “*Bifidobacterium Longum* Subspecies *Infantis* Bacteremia in 3 Extremely Preterm Infants Receiving Probiotics”. In: *Emerging Infectious Diseases* 22.9, pp. 1664–1666 (cit. on p. 14).

Esaiassen, Eirin, Erik Hjerde, Jorunn Pauline Cavanagh, et al. (Nov. 16, 2018). “Effects of Probiotic Supplementation on the Gut Microbiota and Antibiotic Resistome Development in Preterm Infants”. In: *Frontiers in Pediatrics* 6, p. 347 (cit. on p. 90).

Eschenbach, Da (Apr. 2018). “Vaginal Seeding: More Questions than Answers”. In: *BJOG: An International Journal of Obstetrics & Gynaecology* 125.5, pp. 537–537 (cit. on p. 6).

Fajardo, Alicia, Nadia Martínez-Martín, María Mercadillo, et al. (Feb. 20, 2008). “The Neglected Intrinsic Resistome of Bacterial Pathogens”. In: *PLoS ONE* 3.2. Ed. by Matthew Falagas, e1619 (cit. on p. 82).

Fayol-Messaoudi, Domitille, Cédric N. Berger, Marie-Hélène Coconnier-Polter, Vanessa Liévin-Le Moal, and Alain L. Servin (Oct. 2005). “pH-, Lactic Acid-, and Non-Lactic Acid-Dependent Activities of Probiotic Lactobacilli against *Salmonella Enterica* Serovar Typhimurium”. In: *Applied and Environmental Microbiology* 71.10, pp. 6008–6013 (cit. on p. 87).

FDA (July 2013). “Guidance for Industry”. In: p. 4 (cit. on p. 13).

Feehley, Taylor, Catherine H. Plunkett, Riyue Bao, et al. (Mar. 2019). “Healthy Infants Harbor Intestinal Bacteria That Protect against Food Allergy”. In: *Nature Medicine* 25.3, pp. 448–453 (cit. on p. 89).

Ferretti, Pamela, Edoardo Pasolli, Adrian Tett, et al. (2018). “Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome”. In: *Cell Host and Microbe* 24.1, 133–145.e5 (cit. on pp. 78, 79, 84).

Fettweis, Jennifer M., Myrna G. Serrano, J. Paul Brooks, et al. (June 2019). “The Vaginal Microbiome and Preterm Birth”. In: *Nature Medicine* 25.6, pp. 1012–1021 (cit. on p. 90).

Fischbach, Michael A and Christopher T Walsh (2009). “Antibiotics for Emerging Pathogens”. In: 325, p. 6 (cit. on p. 10).

Forslund, Kristoffer, Shinichi Sunagawa, Luis P. Coelho, and Peer Bork (Mar. 2014). “Metagenomic Insights into the Human Gut Resistome and the Forces That Shape It: Prospects & Overviews”. In: *BioEssays* 36.3, pp. 316–329 (cit. on p. 11).

Fortmann, Ingmar, Janina Marißen, Bastian Siller, et al. (Mar. 22, 2020). “Lactobacillus Acidophilus/Bifidobacterium Infantis Probiotics Are Beneficial to Extremely Low Gestational Age Infants Fed Human Milk”. In: *Nutrients* 12.3, p. 850 (cit. on p. 90).

Frese, Steven A., Andra A. Hutton, Lindsey N. Contreras, et al. (Dec. 27, 2017). “Persistence of Supplemented Bifidobacterium Longum Subsp. Infantis EVC001 in Breastfed Infants”. In: *mSphere* 2.6. Ed. by Rosa Krajmalnik-Brown, mSphere.00501–17, e00501–17 (cit. on p. 14).

Fundora, Jennifer B., Pallabi Guha, Darla R. Shores, Mohan Pammi, and Akhil Maheshwari (Jan. 2020). “Intestinal Dysbiosis and Necrotizing Enterocolitis: Assessment for Causality Using Bradford Hill Criteria”. In: *Pediatric Research* 87.2, pp. 235–248 (cit. on p. 9).

Gambardella, Jessica, Vanessa Castellanos, and Gaetano Santulli (Feb. 22, 2021). “Standardizing Translational Microbiome Studies and Metagenomic Analyses”. In: *Cardiovascular Research* 117.3, pp. 640–642 (cit. on p. 89).

Garcia-Heredia, Inmaculada, Ananda S. Bhattacharjee, Oscar Fornas, et al. (Mar. 2021). “Benchmarking of s INGLE-VIRUS Genomics: A New Tool for Uncovering the Virosphere”. In: *Environmental Microbiology* 23.3, pp. 1584–1593 (cit. on p. 93).

Garg, Neha, Mingxun Wang, Embriette Hyde, et al. (Nov. 2017). “Three-Dimensional Microbiome and Metabolome Cartography of a Diseased Human Lung”. In: *Cell Host & Microbe* 22.5, 705–716.e4 (cit. on p. 93).

Garrido, Daniel, David C. Dallas, and David A. Mills (Apr. 1, 2013). “Consumption of Human Milk Glycoconjugates by Infant-Associated Bifidobacteria: Mechanisms and Implications”. In: *Microbiology* 159 (Pt_4), pp. 649–664 (cit. on p. 5).

Gasbarrini, Giovanni, Fiorenza Bonvicini, and Annagiulia Gramenzi (2016). “Probiotics History”. In: *Journal of Clinical Gastroenterology* 50, S116–S119 (cit. on pp. 2, 12).

Gasparrini, Andrew J., Bin Wang, Xiaoqing Sun, et al. (Dec. 2019). “Persistent Metagenomic Signatures of Early-Life Hospitalization and Antibiotic Treatment in the Infant Gut Microbiota and Resistome”. In: *Nature Microbiology* 4.12, pp. 2285–2297 (cit. on p. 11).

Gensollen, Thomas, Shankar S. Iyer, Dennis L. Kasper, and Richard S. Blumberg (Apr. 29, 2016). “How Colonization by Microbiota in Early Life Shapes the Immune System”. In: *Science* 352.6285, pp. 539–544 (cit. on p. 1).

German, J. Bruce, Samara L. Freeman, Carlito B. Lebrilla, and David A. Mills (2008). “Human Milk Oligosaccharides: Evolution, Structures and Bioselectivity as Substrates for Intestinal Bacteria”. In: *Nestlé Nutrition Workshop Series: Pediatric Program*. Ed. by D.M. Bier, J.B. German, and B. Lönnerdal. Vol. 62. Basel: KARGER, pp. 205–222 (cit. on p. 5).

Gibbs, P.A. (2009). “Pathogenic Clostridium Species”. In: *Foodborne Pathogens*. Elsevier, pp. 820–843 (cit. on p. 16).

Gibson, Molly K, Terence S Crofts, and Gautam Dantas (Oct. 2015). “Antibiotics and the Developing Infant Gut Microbiota and Resistome”. In: *Current Opinion in Microbiology* 27, pp. 51–56 (cit. on p. 12).

Gibson, Molly K., Bin Wang, Sara Ahmadi, et al. (Apr. 2016). “Developmental Dynamics of the Preterm Infant Gut Microbiota and Antibiotic Resistome”. In: *Nature Microbiology* 1.4, p. 16024 (cit. on p. 11).

Gill, S. R., M. Pop, R. T. DeBoy, et al. (June 2, 2006). “Metagenomic Analysis of the Human Distal Gut Microbiome”. In: *Science* 312.5778, pp. 1355–1359 (cit. on p. 4).

Gnu, P (2007). *Free Software Foundation. Bash (3.2. 48)[Unix Shell Program]* (cit. on p. 24).

Groer, Maureen W., Elizabeth M. Miller, Amy D’Agata, et al. (May 2020). “Contributors to Dysbiosis in Very-Low-Birth-Weight Infants”. In: *Journal of Obstetric, Gynecologic & Neonatal Nursing* 49.3, pp. 232–242 (cit. on p. 85).

Grumaz, Silke, Christian Grumaz, Yevhen Vainshtein, et al. (May 2019). “Enhanced Performance of Next-Generation Sequencing Diagnostics Compared With Standard of Care Microbiological Diagnostics in Patients Suffering From Septic Shock:” in: *Critical Care Medicine* 47.5, e394–e402 (cit. on pp. 83, 84).

Ha, Gyoung Yim, Chang Heon Yang, Heesoo Kim, and Yunsop Chong (Apr. 1999). “Case of Sepsis Caused by *Bifidobacterium Longum*”. In: *Journal of Clinical Microbiology* 37.4, pp. 1227–1228 (cit. on p. 90).

Haas, B. J., D. Gevers, A. M. Earl, et al. (Mar. 1, 2011). “Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons”. In: *Genome Research* 21.3, pp. 494–504 (cit. on p. 27).

Hagberg, Aric A, Daniel A Schult, and Pieter J Swart (2008). “Exploring Network Structure, Dynamics, and Function Using NetworkX”. In: p. 5 (cit. on p. 24).

Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, et al. (Sept. 17, 2020). “Array Programming with NumPy”. In: *Nature* 585.7825, pp. 357–362 (cit. on p. 24).

Härtel, Christoph, Julia Pagel, Juliane Spiegler, et al. (Dec. 2017). “Lactobacillus Acidophilus/Bifidobacterium Infantis Probiotics Are Associated with Increased Growth of VLBWI among Those Exposed to Antibiotics”. In: *Scientific Reports* 7.1, p. 5633 (cit. on pp. 79, 86, 93).

Heida, Fardou H., Anne G. J. F. van Zoonen, Jan B. F. Hulscher, et al. (Apr. 1, 2016). “A Necrotizing Enterocolitis-Associated Gut Microbiota Is Present in the Meconium: Results of a Prospective Study”. In: *Clinical Infectious Diseases* 62.7, pp. 863–870 (cit. on p. 84).

Henrick, Bethany M., Lucie Rodriguez, Tadepally Lakshmikanth, et al. (June 2021). “Bifidobacteria-Mediated Immune System Imprinting Early in Life”. In: *Cell*, S0092867421006607 (cit. on p. 89).

Hey, Jody (2001). “The Mind of the Species Problem”. In: p. 4 (cit. on p. 90).

Hill, Cian J., Denise B. Lynch, Kiera Murphy, et al. (2017). “Evolution of Gut Microbiota Composition from Birth to 24 Weeks in the INFANTMET Cohort”. In: *Microbiome* 5.1, pp. 1–18 (cit. on pp. 18, 32).

Hill, Colin, Francisco Guarner, Gregor Reid, et al. (Aug. 2014). “The International Scientific Association for Probiotics and Prebiotics Consensus Statement on the Scope and Appropriate Use of the Term Probiotic”. In: *Nature Reviews Gastroenterology & Hepatology* 11.8, pp. 506–514 (cit. on p. 12).

Ho, Thao T. B., Maureen W. Groer, Bradley Kane, et al. (Dec. 2018). “Dichotomous Development of the Gut Microbiome in Preterm Infants”. In: *Microbiome* 6.1, p. 157 (cit. on p. 8).

Hooks, Katarzyna B. and Maureen A. O’Malley (Nov. 8, 2017). “Dysbiosis and Its Discontents”. In: *mBio* 8.5. Ed. by Julian E. Davies, mBio.01492–17, e01492–17 (cit. on pp. 8, 9).

Hufnagl, Karin, Isabella Pali-Schöll, Franziska Roth-Walter, and Erika Jensen-Jarolim (Feb. 2020). “Dysbiosis of the Gut and Lung Microbiome Has a Role in Asthma”. In: *Seminars in Immunopathology* 42.1, pp. 75–93 (cit. on pp. 84, 92).

Hunter, John D. (2007). “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science & Engineering* 9.3, pp. 90–95 (cit. on p. 24).

Huttenhower, Curtis, Dirk Gevers, Rob Knight, et al. (2012). “Structure, Function and Diversity of the Healthy Human Microbiome”. In: *Nature* 486.7402, pp. 207–214 (cit. on p. 1).

“Hygienemaßnahmen bei Infektionen oder Besiedlung mit multiresistenten gramnegativen Stäbchen” (Oct. 2012). In: *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 55.10, pp. 1311–1354 (cit. on p. 79).

Iebba, Valerio, Valentina Totino, Antonella Gagliardi, et al. (2016). “Eubiosis and Dysbiosis: The Two Sides of the Microbiota”. In: p. 12 (cit. on p. 89).

International Human Genome Sequencing Consortium (Feb. 2001). “Initial Sequencing and Analysis of the Human Genome”. In: *Nature* 409.6822, pp. 860–921 (cit. on p. 3).

Jiménez, Esther, María L. Marín, Rocío Martín, et al. (Apr. 2008). “Is Meconium from Healthy Newborns Actually Sterile?” In: *Research in Microbiology* 159.3, pp. 187–193 (cit. on p. 4).

Johnson, Jethro S., Daniel J. Spakowicz, Bo-Young Hong, et al. (Dec. 2019). “Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis”. In: *Nature Communications* 10.1, p. 5029 (cit. on p. 80).

Jungersen, Mikkel, Anette Wind, Eric Johansen, et al. (Mar. 28, 2014). “The Science behind the Probiotic Strain Bifidobacterium Animalis Subsp. Lactis BB-12®”. In: *Microorganisms* 2.2, pp. 92–110 (cit. on p. 86).

Karami, N., A. Martner, V. I. Enne, et al. (Sept. 17, 2007). “Transfer of an Ampicillin Resistance Gene between Two Escherichia Coli Strains in the Bowel Microbiota of an Infant Treated with Antibiotics”. In: *Journal of Antimicrobial Chemotherapy* 60.5, pp. 1142–1145 (cit. on p. 11).

Karas, Michael and Ralf Krüger (Feb. 2003). “Ion Formation in MALDI: The Cluster Ionization Mechanism”. In: *Chemical Reviews* 103.2, pp. 427–440 (cit. on p. 16).

Katayama, Takane (Apr. 2, 2016). “Host-Derived Glycans Serve as Selected Nutrients for the Gut Microbe: Human Milk Oligosaccharides and Bifidobacteria”. In: *Bioscience, Biotechnology, and Biochemistry* 80.4, pp. 621–632 (cit. on p. 5).

Keelan, Jeff A and John P Newnham (July 18, 2017). “Recent Advances in the Prevention of Preterm Birth”. In: *F1000Research* 6, p. 1139 (cit. on p. 7).

Kendall, Arthur (1918). “Recent Developments in Intestinal Bacteriology”. In: 156, pp. 157–73 (cit. on p. 4).

Kennedy, Katherine M., Max J. Gerlach, Thomas Adam, et al. (May 10, 2021). “Fetal Meconium Does Not Have a Detectable Microbiota before Birth”. In: *Nature Microbiology* (cit. on p. 84).

Kervinen, Kaisa, Ilkka Kalliala, Sivan Glazer-Livson, et al. (Oct. 2019). “Vaginal Microbiota in Pregnancy: Role in Induction of Labor and Seeding the Neonate’s Microbiota?” In: *Journal of Biosciences* 44.5, p. 116 (cit. on p. 6).

Kim, Gyungcheon, Jaewoong Bae, Mi Jin Kim, et al. (Sept. 11, 2020). “Delayed Establishment of Gut Microbiota in Infants Delivered by Cesarean Section”. In: *Frontiers in Microbiology* 11, p. 2099 (cit. on p. 18).

Klopp, Jonas, Pamela Ferretti, Claudius U. Meyer, et al. (Jan. 12, 2022). “Meconium Microbiome of Very Preterm Infants across Germany”. In: *mSphere*. Ed. by Vincent B. Young, e00808–21 (cit. on pp. 34, 40).

Kluyver, Thomas, Benjamin Ragan-Kelley, Fernando Pérez, et al. (2016). “Jupyter Notebooks—a Publishing Format for Reproducible Computational Workflows”. In: p. 4 (cit. on p. 24).

Koch-Institut, Robert (2013). “Epidemiologisches Bulletin 42/2013”. In: . *Oktober*, p. 16 (cit. on p. 16).

König, Inke R., Oliver Fuchs, Gesine Hansen, Erika von Mutius, and Matthias V. Kopp (Oct. 2017). “What Is Precision Medicine?” In: *European Respiratory Journal* 50.4, p. 1700391 (cit. on p. 94).

Korpela, Katri, Anne Salonen, Outi Vepsäläinen, et al. (2018). “Probiotic Supplementation Restores Normal Microbiota Composition and Function in Antibiotic-Treated and in Caesarean-Born Infants”. In: *Microbiome* 6.1, pp. 182–182 (cit. on p. 90).

Kothari, Damini, Seema Patel, and Soo-Ki Kim (Mar. 2019). “Probiotic Supplements Might Not Be Universally-Effective and Safe: A Review”. In: *Biomedicine & Pharmacotherapy* 111, pp. 537–547 (cit. on pp. 90, 93).

Kozich, James J, Sarah L Westcott, Nielson T Baxter, Sarah K Highlander, and Patrick D Schloss (Sept. 2013). “Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing Platform”. In: *Applied and Environmental Microbiology* 79.17, pp. 5112–5120. PMID: 23793624 (cit. on p. 75).

Kruskal, William H and W Allen Wallis (1952). “Use of Ranks in One-Criterion Variance Analysis”. In: 47, pp. 583–621 (cit. on p. 31).

La Rosa, P. S., B. B. Warner, Y. Zhou, et al. (Aug. 26, 2014). “Patterned Progression of Bacterial Populations in the Premature Infant Gut”. In: *Proceedings of the National Academy of Sciences* 111.34, pp. 12522–12527 (cit. on p. 8).

Lackey, Kimberly A., Janet E. Williams, Courtney L. Meehan, et al. (Apr. 17, 2019). “What’s Normal? Microbiomes in Human Milk and Infant Feces Are Related to Each Other but Vary Geographically: The INSPIRE Study”. In: *Frontiers in Nutrition* 6, p. 45 (cit. on pp. 6, 7).

Land, Miriam, Loren Hauser, Se-Ran Jun, et al. (Mar. 2015). “Insights from 20 Years of Bacterial Genome Sequencing”. In: *Functional & Integrative Genomics* 15.2, pp. 141–161 (cit. on p. 91).

Lauder, Abigail P., Aoife M. Roche, Scott Sherrill-Mix, et al. (Dec. 2016). “Comparison of Placenta Samples with Contamination Controls Does Not Provide Evidence for a Distinct Placenta Microbiota”. In: *Microbiome* 4.1, p. 29 (cit. on p. 4).

Laursen, Martin F., Mikiyasu Sakanaka, Nicole von Burg, et al. (Oct. 21, 2021). “Bifidobacterium Species Associated with Breastfeeding Produce Aromatic Lactic Acids in the Infant Gut”. In: *Nature Microbiology* (cit. on p. 85).

Lederberg, Joshua and Alexa T McCray (2001). “’Ome Sweet ’Omics– A Genealogical Treasury of Words”. In: p. 2 (cit. on p. 1).

Leónidas Cardoso, Luís, Paulo Durão, Massimo Amicone, and Isabel Gordo (Sept. 2020). “Dysbiosis Individualizes the Fitness Effect of Antibiotic Resistance in the Mammalian Gut”. In: *Nature Ecology & Evolution* 4.9, pp. 1268–1278 (cit. on pp. 82, 92).

Levy, Maayan, Aleksandra A. Kolodziejczyk, Christoph A. Thaiss, and Eran Elinav (Apr. 2017). “Dysbiosis and the Immune System”. In: *Nature Reviews Immunology* 17.4, pp. 219–232 (cit. on pp. 10, 79).

Li, Ruiqiang, Yingrui Li, Hancheng Zheng, et al. (Jan. 2010). “Building the Sequence Map of the Human Pan-Genome”. In: *Nature Biotechnology* 28.1, pp. 57–63 (cit. on p. 91).

Liang, Bingbing, Wenyu Fan, Jing Ginger Han, Ning Chen, and Nan Zhao (Nov. 2017). “An Implementation of Infants’ Gut Microbiome Maturation Analyses by 16s rRNA from Stool Samples in Extraction Solution of Room Temperature”. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Kansas City, MO: IEEE, pp. 910–914 (cit. on p. 93).

Lindemann, Steve, Julien Tremblay, Joseph Henry Vineis, Luisa W Hugerth, and Anders F Andersson (2017). “Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing A BRIEF HISTORY OF METHODOLOGIES FOR DETERMINING MICROBIAL COMMUNITY COMPOSITION”. In: *Front. Microbiol* 8.8 (cit. on p. 19).

Liu, C.-I, G. Y. Liu, Y. Song, et al. (Mar. 7, 2008). “A Cholesterol Biosynthesis Inhibitor Blocks Staphylococcus Aureus Virulence”. In: *Science* 319.5868, pp. 1391–1394 (cit. on p. 10).

Lozupone, Catherine, Manuel E Lladser, Dan Knights, Jesse Stombaugh, and Rob Knight (Feb. 2011). “UniFrac: An Effective Distance Metric for Microbial Community Comparison”. In: *The ISME Journal* 5.2, pp. 169–172 (cit. on p. 31).

Maldonado-Gómez, María X., Inés Martínez, Francesca Bottacini, et al. (Oct. 2016). “Stable Engraftment of Bifidobacterium Longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome”. In: *Cell Host & Microbe* 20.4, pp. 515–526 (cit. on p. 88).

Mancabelli, Leonardo, Christian Milani, Gabriele Andrea Lugli, et al. (Jan. 17, 2020). “The Impact of Primer Design on Amplicon-Based Metagenomic Profiling Accuracy: Detailed Insights into Bifidobacterial Community Structure”. In: *Microorganisms* 8.1, p. 131 (cit. on p. 18).

Mardis, Elaine R. (Feb. 2011). “A Decade’s Perspective on DNA Sequencing Technology”. In: *Nature* 470.7333, pp. 198–203 (cit. on p. 3).

Marißen, Janina, Annette Haiß, Claudius Meyer, et al. (Nov. 2019). “Efficacy of *Bifidobacterium Longum, B. Infantis and Lactobacillus Acidophilus* Probiotics to Prevent Gut Dysbiosis in Preterm Infants of 28+0–32+6 Weeks of Gestation: A Randomised, Placebo-Controlled, Double-Blind, Multicentre Trial: The PRIMAL Clinical Study Protocol”. In: *BMJ Open* 9.11, e032617 (cit. on pp. 15, 21).

Martins, Flaviano S., Aparecida A. Silva, Angélica T. Vieira, et al. (Aug. 2009). “Comparative Study of Bifidobacterium Animalis, Escherichia Coli, Lactobacillus Casei and Saccharomyces Boulardii Probiotic Properties”. In: *Archives of Microbiology* 191.8, pp. 623–630 (cit. on p. 86).

Mathipa, Moloko Gloria and Mapitsi Silvester Thantsha (Dec. 2017). “Probiotic Engineering: Towards Development of Robust Probiotic Strains with Enhanced Functional Properties and for Targeted Control of Enteric Pathogens”. In: *Gut Pathogens* 9.1, p. 28 (cit. on pp. 87, 93).

Maurice, Corinne Ferrier, Henry Joseph Haiser, and Peter James Turnbaugh (Jan. 2013). “Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome”. In: *Cell* 152.1-2, pp. 39–50 (cit. on p. 2).

McCarthy, Charley G. P. and David A. Fitzpatrick (Feb. 1, 2019). “Pan-Genome Analyses of Model Fungal Species”. In: *Microbial Genomics* 5.2 (cit. on p. 91).

McDonald, Julie A.K. (Nov. 30, 2017). “In Vitro Models of the Human Microbiota and Microbiome”. In: *Emerging Topics in Life Sciences* 1.4. Ed. by Julian R. Marchesi, pp. 373–384 (cit. on p. 88).

McDonnell, Lucy, Alexander Gilkes, Mark Ashworth, et al. (Jan. 1, 2021). “Association between Antibiotics and Gut Microbiome Dysbiosis in Children: Systematic Review and Meta-Analysis”. In: *Gut Microbes* 13.1, p. 1870402 (cit. on p. 81).

McKinlay, Christopher J D, Celso Rebello, and William Tarnow-Mordi (Aug. 2016). “Probiotics in Very Preterm Infants: The PiPS Trial”. In: *The Lancet* 388.10045, p. 655 (cit. on p. 14).

McMurdie, Paul J. and Susan Holmes (Apr. 22, 2013). “Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data”. In: *PLoS ONE* 8.4. Ed. by Michael Watson, e61217 (cit. on pp. 28, 31).

MetaHIT Consortium (additional members), Manimozhiyan Arumugam, Jeroen Raes, et al. (May 12, 2011). “Enterotypes of the Human Gut Microbiome”. In: *Nature* 473.7346, pp. 174–180 (cit. on p. 9).

Milani, Christian, Sabrina Duranti, Francesca Bottacini, et al. (Dec. 2017). “The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota”. In: *Microbiology and Molecular Biology Reviews* 81.4, e00036–17, e00036–17 (cit. on pp. 18, 85).

Moore, Aimée M., Sanket Patel, Kevin J. Forsberg, et al. (Nov. 13, 2013). “Pediatric Fecal Microbiota Harbor Diverse and Novel Antibiotic Resistance Genes”. In: *PLoS ONE* 8.11. Ed. by Herman Tse, e78822 (cit. on p. 11).

Moore, G.E. (Jan. 1998). “Cramming More Components Onto Integrated Circuits”. In: *Proceedings of the IEEE* 86.1, pp. 82–85 (cit. on p. 3).

Moore, Rebecca E. and Steven D. Townsend (Sept. 27, 2019). “Temporal Development of the Infant Gut Microbiome”. In: *Open Biology* 9.9, p. 190128 (cit. on pp. 4, 6).

Moran, Nancy A. and Daniel B. Sloan (Dec. 4, 2015). “The Hologenome Concept: Helpful or Hollow?” In: *PLOS Biology* 13.12, e1002311 (cit. on p. 92).

Morowitz, Michael J., Erica M. Carlisle, and John C. Alverdy (Aug. 2011). “Contributions of Intestinal Bacteria to Nutrition and Metabolism in the Critically Ill”. In: *Surgical Clinics of North America* 91.4, pp. 771–785 (cit. on p. 2).

Mueller, Noel T., Suchitra K. Hourigan, Diane E. Hoffmann, et al. (2019). “Bacterial Baptism: Scientific, Medical, and Regulatory Issues Raised by Vaginal Seeding of C-Section-Born Babies”. In: *Journal of Law, Medicine & Ethics* 47.4, pp. 568–578 (cit. on p. 6).

Nicholson, J. K., E. Holmes, J. Kinross, et al. (June 8, 2012). “Host-Gut Microbiota Metabolic Interactions”. In: *Science* 336.6086, pp. 1262–1267 (cit. on p. 2).

NIH Human Microbiome Portfolio Analysis Team (Dec. 2019). “A Review of 10 Years of Human Microbiome Research Activities at the US National Institutes of Health, Fiscal Years 2007-2016”. In: *Microbiome* 7.1, p. 31 (cit. on pp. 4, 93).

O’Sullivan, Daniel J. (2004). “ISOLATED BIFIDOBACTERIA THAT PRODUCE SIDEROPHORES WHICH INHIBIT GROWTH OF LACTOCOCCUS LACTIS”. Pat. (cit. on p. 87).

O’Sullivan, Denise M., Ronan M. Doyle, Sasithon Temisak, et al. (Dec. 2021). “An Inter-Laboratory Study to Investigate the Impact of the Bioinformatics Component on Microbiome Analysis Using Mock Communities”. In: *Scientific Reports* 11.1, p. 10590 (cit. on p. 18).

Olesen, Scott W. and Eric J. Alm (Dec. 2016). “Dysbiosis Is Not an Answer”. In: *Nature Microbiology* 1.12, p. 16228 (cit. on p. 8).

Olsen, Rie, Gorm Greisen, Morten Schrøder, and Jesper Brok (2016). “Prophylactic Probiotics for Preterm Infants: A Systematic Review and Meta-Analysis of Observational Studies”. In: *Neonatology* 109.2, pp. 105–112 (cit. on p. 14).

Pammi, Mohan, Julia Cope, Phillip I. Tarr, et al. (Dec. 2017). “Intestinal Dysbiosis in Preterm Infants Preceding Necrotizing Enterocolitis: A Systematic Review and Meta-Analysis”. In: *Microbiome* 5.1, p. 31 (cit. on p. 9).

Panigrahi, Pinaki (2017). “A Randomized Synbiotic Trial to Prevent Sepsis among Infants in Rural India”. In: p. 13 (cit. on p. 9).

Pannaraj, Pia S., Fan Li, Chiara Cerini, et al. (July 1, 2017). “Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome”. In: *JAMA Pediatrics* 171.7, p. 647 (cit. on pp. 18, 32).

Partridge, Sally R., Stephen M. Kwong, Neville Firth, and Slade O. Jensen (Oct. 2018). “Mobile Genetic Elements Associated with Antimicrobial Resistance”. In: *Clinical Microbiology Reviews* 31.4 (cit. on p. 82).

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, et al. (n.d.). “Scikit-Learn: Machine Learning in Python”. In: *MACHINE LEARNING IN PYTHON* (), p. 6 (cit. on p. 24).

Perez-Muñoz, Maria Elisa, Marie-Claire Arrieta, Amanda E. Ramer-Tait, and Jens Walter (Dec. 2017). “A Critical Assessment of the “Sterile Womb” and “in Utero Colonization” Hypotheses: Implications for Research on the Pioneer Infant Microbiome”. In: *Microbiome* 5.1, p. 48 (cit. on p. 4).

Peter A Vandenbergh (1993). “Lactic Acid Bacteria, Their Metabolic Products and Interference with Microbial Growth”. In: pp. 221–238 (cit. on p. 86).

Petersen, Charisse, Darlene L.Y. Dai, Rozlyn C.T. Boutin, et al. (May 2021). “A Rich Meconium Metabolome in Human Infants Is Associated with Early-Life Gut Microbiota Composition and Reduced Allergic Sensitization”. In: *Cell Reports Medicine* 2.5, p. 100260 (cit. on p. 85).

Petersen, Charisse and June L. Round (July 2014). “Defining Dysbiosis and Its Influence on Host Immunity and Disease: How Changes in Microbiota Structure Influence Health”. In: *Cellular Microbiology* 16.7, pp. 1024–1033 (cit. on pp. 8, 10).

Plotly Technologies Inc. (2015). *Plotly - Collaborative Data Science*. Montreal, QC: Plotly Technologies Inc. (cit. on p. 24).

Podolsky, Scott H (Nov. 2012). “Metchnikoff and the Microbiome”. In: *The Lancet* 380.9856, pp. 1810–1811 (cit. on p. 2).

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, et al. (Nov. 27, 2012). “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools”. In: *Nucleic Acids Research* 41.D1, pp. D590–D596 (cit. on p. 28).

Quin, C., M. Estaki, D. M. Vollman, et al. (Dec. 2018). “Probiotic Supplementation and Associated Infant Gut Microbiome and Health: A Cautionary Retrospective Clinical Comparison”. In: *Scientific Reports* 8.1, p. 8283 (cit. on pp. 88, 90).

Quince, Christopher, Alan W. Walker, Jared T. Simpson, Nicholas J. Loman, and Nicola Segata (2017). “Shotgun Metagenomics, from Sampling to Analysis”. In: *Nature Biotechnology* 35.9, pp. 833–844 (cit. on p. 19).

Rao, Shripada C., Gayatri K. Athalye-Jape, Girish C. Deshpande, Karen N Simmer, and Sanjay K. Patole (Mar. 2016). “Probiotic Supplementation and Late-Onset Sepsis in Preterm Infants: A Meta-analysis”. In: *Pediatrics* 137.3, e20153684 (cit. on p. 14).

Ravi, Anuradha, Fenella D Halstead, Amy Bamford, et al. (Sept. 1, 2019). “Loss of Microbial Diversity and Pathogen Domination of the Gut Microbiota in Critically Ill Patients”. In: *Microbial Genomics* 5.9 (cit. on p. 78).

Rinninella, Emanuele, Pauline Raoul, Marco Cintoni, et al. (Jan. 10, 2019). “What Is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases”. In: *Microorganisms* 7.1, p. 14 (cit. on pp. 1, 47).

Rohrer, Julia M (2018). “Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data”. In: p. 16 (cit. on p. 88).

Rose, Graham, Alexander G. Shaw, Kathleen Sim, et al. (Jan. 25, 2017). “Antibiotic Resistance Potential of the Healthy Preterm Infant Gut Microbiome”. In: *PeerJ* 5, e2928 (cit. on pp. 11, 82–84).

Rossi, Alessia, Alessio Bellone, Sergei I. Fokin, Vittorio Boscaro, and Claudia Vannini (July 2019). “Detecting Associations Between Ciliated Protists and Prokaryotes with Culture-Independent Single-Cell Microbiomics: A Proof-of-Concept Study”. In: *Microbial Ecology* 78.1, pp. 232–242 (cit. on p. 93).

Rychert, Jenna (July 1, 2019). “Benefits and Limitations of MALDI-TOF Mass Spectrometry for the Identification of Microorganisms”. In: *Journal of Infectiology* 2.4, pp. 1–5 (cit. on p. 80).

Samuel, B. S. and J. I. Gordon (June 27, 2006). “A Humanized Gnotobiotic Mouse Model of Host-Archaeal-Bacterial Mutualism”. In: *Proceedings of the National Academy of Sciences* 103.26, pp. 10011–10016 (cit. on p. 88).

Sawh, Sonja C., Santosh Deshpande, Sandy Jansen, Christopher J. Reynaert, and Philip M. Jones (Oct. 5, 2016). “Prevention of Necrotizing Enterocolitis with Probiotics: A Systematic Review and Meta-Analysis”. In: *PeerJ* 4, e2429 (cit. on p. 14).

Schloss, Patrick D., Dirk Gevers, and Sarah L. Westcott (Dec. 14, 2011). “Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies”. In: *PLoS ONE* 6.12. Ed. by Jack Anthony Gilbert, e27310 (cit. on p. 75).

SciPy 1.0 Contributors, Pauli Virtanen, Ralf Gommers, et al. (Mar. 2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17.3, pp. 261–272 (cit. on p. 24).

Secundus, Gaius Plinius (77). *Naturalis Historia* (cit. on p. 2).

Servin, Alain L. (Oct. 2004). “Antagonistic Activities of Lactobacilli and Bifidobacteria against Microbial Pathogens”. In: *FEMS Microbiology Reviews* 28.4, pp. 405–440 (cit. on p. 86).

Shannon, C E (1948). “A Mathematical Theory of Communication”. In: p. 54 (cit. on p. 31).

Shannon, P. (Nov. 1, 2003). “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. In: *Genome Research* 13.11, pp. 2498–2504 (cit. on p. 31).

Sharif, Sahar, Nicholas Meader, Sam J Oddie, Maria Ximena Rojas-Reyes, and William McGuire (Oct. 15, 2020). “Probiotics to Prevent Necrotising Enterocolitis in Very Preterm or Very Low Birth Weight Infants”. In: *Cochrane Database of Systematic Reviews*. Ed. by Cochrane Neonatal Group (cit. on p. 86).

Sharma, Chandra Madhur (2013). ““Neonatal Sepsis”: Bacteria & Their Susceptibility Pattern towards Antibiotics in Neonatal Intensive Care Unit”. In: *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH* (cit. on p. 10).

Shoemaker, N. B., H. Vlamakis, K. Hayes, and A. A. Salyers (Feb. 1, 2001). “Evidence for Extensive Resistance Gene Transfer amongBacteroides Spp. and among Bacteroides and Other Genera in the Human Colon”. In: *Applied and Environmental Microbiology* 67.2, pp. 561–568 (cit. on p. 91).

Shulman, S. T., H. C. Friedmann, and R. H. Sims (Oct. 15, 2007). “Theodor Escherich: The First Pediatric Infectious Diseases Physician?” In: *Clinical Infectious Diseases* 45.8, pp. 1025–1029 (cit. on p. 2).

Simpson, Kathleen Rice (2018). “Vaginal Seeding of Babies Born via Cesarean:” in: *MCN, The American Journal of Maternal/Child Nursing* 43.1, p. 60 (cit. on p. 5).

Smilowitz, Jennifer T., Jackelyn Moya, Melissa A. Breck, et al. (Dec. 2017). “Safety and Tolerability of Bifidobacterium Longum Subspecies Infantis EVC001 Supplementation in Healthy Term Breastfed Infants: A Phase I Clinical Trial”. In: *BMC Pediatrics* 17.1, p. 133 (cit. on p. 14).

Sommer, M. O. A., G. Dantas, and G. M. Church (Aug. 28, 2009). “Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora”. In: *Science* 325.5944, pp. 1128–1131 (cit. on p. 11).

Sommer, Morten O.A., George M. Church, and Gautam Dantas (July 2010). “The Human Microbiome Harbors a Diverse Reservoir of Antibiotic Resistance Genes”. In: *Virulence* 1.4, pp. 299–303 (cit. on p. 82).

Sonnenburg, Erica D. and Justin L. Sonnenburg (Nov. 2014). “Starving Our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates”. In: *Cell Metabolism* 20.5, pp. 779–786 (cit. on p. 9).

Stecher, B., R. Denzler, L. Maier, et al. (Jan. 24, 2012). “Gut Inflammation Can Boost Horizontal Gene Transfer between Pathogenic and Commensal Enterobacteriaceae”. In: *Proceedings of the National Academy of Sciences* 109.4, pp. 1269–1274 (cit. on pp. 11, 82).

Stewart, Christopher J., Nadim J. Ajami, Jacqueline L. O’Brien, et al. (Oct. 2018). “Temporal Development of the Gut Microbiome in Early Childhood from the TEDDY Study”. In: *Nature* 562.7728, pp. 583–588 (cit. on pp. 5, 78).

Stinson, Lisa F., Mary C. Boyce, Matthew S. Payne, and Jeffrey A. Keelan (2019). “The Not-so-Sterile Womb: Evidence That the Human Fetus Is Exposed to Bacteria Prior to Birth”. In: *Frontiers in Microbiology* 10 (JUN), pp. 1–15 (cit. on p. 4).

Stinson, Lisa F., Matthew S. Payne, and Jeffrey A. Keelan (2018). “A Critical Review of the Bacterial Baptism Hypothesis and the Impact of Cesarean Delivery on the Infant Microbiome”. In: *Frontiers in Medicine* 5 (MAY) (cit. on p. 6).

Strejcek, Michal, Tereza Smrhova, Petra Junkova, and Ondrej Uhlik (June 19, 2018). “Whole-Cell MALDI-TOF MS Versus 16S rRNA Gene Analysis for Identification and Dereplication of Recurrent Bacterial Isolates”. In: *Frontiers in Microbiology* 9, p. 1294 (cit. on p. 80).

Subramanian, Sathish, Sayeeda Huq, Tanya Yatsunenko, et al. (June 2014). “Persistent Gut Microbiota Immaturity in Malnourished Bangladeshi Children”. In: *Nature* 510.7505, pp. 417–421 (cit. on p. 18).

Suez, Jotham (2019). “The Pros, Cons, and Many Unknowns of Probiotics”. In: *Nature Medicine* 25, p. 14 (cit. on pp. 12, 89).

Suez, Jotham, Niv Zmora, Gili Zilberman-Schapira, et al. (2018). “Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT”. In: *Cell* 174.6, 1406–1423.e16 (cit. on p. 90).

Tabasco, Raquel, Pilar Fernández de Palencia, Javier Fontecha, Carmen Peláez, and Teresa Requena (Mar. 2014). “Competition Mechanisms of Lactic Acid Bacteria and Bifidobacteria: Fermentative Metabolism and Colonization”. In: *LWT - Food Science and Technology* 55.2, pp. 680–684 (cit. on p. 87).

Tackmann, Janko (Sept. 2019). “Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data”. In: p. 20 (cit. on p. 31).

Tan, Xing and Stuart Johnson (Dec. 2019). “Fecal Microbiota Transplantation (FMT) for C. Difficile Infection, Just Say ‘No’”. In: *Anaerobe* 60, p. 102092 (cit. on p. 13).

Tanaka, Katsunori, Yoshitaka Nakamura, Masaki Terahara, et al. (Apr. 13, 2019). “Poor Bifidobacterial Colonization Is Associated with Late Provision of Colostrum and Improved with Probiotic Supplementation in Low Birth Weight Infants”. In: *Nutrients* 11.4, p. 839 (cit. on p. 90).

Tasse, L., J. Bercovici, S. Pizzut-Serin, et al. (Nov. 1, 2010). “Functional Metagenomics to Mine the Human Gut Microbiome for Dietary Fiber Catabolic Enzymes”. In: *Genome Research* 20.11, pp. 1605–1612 (cit. on p. 2).

Tauchi, H., K. Yahagi, T. Yamauchi, et al. (July 10, 2019). “Gut Microbiota Development of Preterm Infants Hospitalised in Intensive Care Units”. In: *Beneficial Microbes* 10.6, pp. 641–651 (cit. on p. 8).

Team, R Core (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (cit. on p. 24).

The ENCODE Project Consortium (Oct. 22, 2004). “The ENCODE (ENCyclopedia Of DNA Elements) Project”. In: *Science* 306.5696, pp. 636–640 (cit. on p. 3).

“The Neighbor-Joining Method” (July 1987). “The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.” In: *Molecular Biology and Evolution* (cit. on p. 30).

The NIH HMP Working Group, J. Peterson, S. Garges, et al. (Dec. 1, 2009). “The NIH Human Microbiome Project”. In: *Genome Research* 19.12, pp. 2317–2323 (cit. on pp. 4, 9, 77).

The pandas development team (Feb. 2020). *Pandas-Dev/Pandas: Pandas* (cit. on p. 24).

Tiffany, Connor R. and Andreas J. Bäumler (Nov. 1, 2019). “Dysbiosis: From Fiction to Function”. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 317.5, G602–G608 (cit. on p. 92).

Tindall, B. J., G. Sutton, and G. M. Garrity (Feb. 1, 2017). “Enterobacter Aerogenes Hormaeche and Edwards 1960 (Approved Lists 1980) and Klebsiella Mobilis Bascomb et al. 1971 (Approved Lists 1980) Share the Same Nomenclatural Type (ATCC 13048) on the Approved Lists and Are Homotypic Synonyms, with Consequences for the Name Klebsiella Mobilis Bascomb et al. 1971 (Approved Lists 1980)”. In: *International Journal of Systematic and Evolutionary Microbiology* 67.2, pp. 502–504 (cit. on pp. 80, 84).

Tremblay, Julien, Kanwar Singh, Alison Fern, et al. (Aug. 4, 2015). “Primer and Platform Effects on 16S rRNA Tag Sequencing”. In: *Frontiers in Microbiology* 6 (cit. on p. 85).

Tripathi, Anupriya, Clarisse Marotz, Antonio Gonzalez, et al. (2018). “Are Microbiome Studies Ready for Hypothesis-Driven Research?” In: *Current Opinion in Microbiology* 44, pp. 61–69 (cit. on pp. 77, 93).

Tripathi, Nidhi, C. Michael Cotten, and P. Brian Smith (Mar. 2012). “Antibiotic Use and Misuse in the Neonatal Intensive Care Unit”. In: *Clinics in Perinatology* 39.1, pp. 61–68 (cit. on p. 10).

Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, et al. (Oct. 2007). “The Human Microbiome Project”. In: *Nature* 449.7164, pp. 804–810 (cit. on pp. 1, 4).

Turroni, Francesca, Christian Milani, Sabrina Duranti, et al. (Jan. 2018). “Bifidobacteria and the Infant Gut: An Example of Co-Evolution and Natural Selection”. In: *Cellular and Molecular Life Sciences* 75.1, pp. 103–118 (cit. on p. 5).

Turroni, Silvia, Marciane Magnani, Pukar Kc, et al. (Sept. 8, 2020). “Gut Microbiome and Space Travelers’ Health: State of the Art and Possible Pro/Prebiotic Strategies for Long-Term Space Missions”. In: *Frontiers in Physiology* 11, p. 553929 (cit. on p. 89).

Underwood, Mark A., Sagori Mukhopadhyay, Satyan Lakshminrusimha, and Charles L. Bevins (Nov. 2020). “Neonatal Intestinal Dysbiosis”. In: *Journal of Perinatology* 40.11, pp. 1597–1608 (cit. on pp. 7, 84).

Van de Peer, Y (Sept. 1, 1996). “A Quantitative Map of Nucleotide Substitution Rates in Bacterial rRNA”. In: *Nucleic Acids Research* 24.17, pp. 3381–3391 (cit. on p. 17).

Van Rossum, Thea, Pamela Ferretti, Oleksandr M. Maistrenko, and Peer Bork (Sept. 2020). “Diversity within Species: Interpreting Strains in Microbiomes”. In: *Nature Reviews Microbiology* 18.9, pp. 491–506 (cit. on pp. 80, 91).

Van Best, Niels, Sonja Trepels-Kottek, Paul Savelkoul, et al. (Nov. 9, 2020). “Influence of Probiotic Supplementation on the Developing Microbiota in Human Preterm Neonates”. In: *Gut Microbes* 12.1, p. 1826747 (cit. on p. 87).

Vangay, Pajau, Tonya Ward, Jeffrey S. Gerber, and Dan Knights (May 2015). “Antibiotics, Pediatric Dysbiosis, and Disease”. In: *Cell Host & Microbe* 17.5, pp. 553–564 (cit. on pp. 9, 10).

Van Nood, Els, Anne Vrieze, Max Nieuwdorp, et al. (Jan. 31, 2013). “Duodenal Infusion of Donor Feces for Recurrent *Clostridium Difficile*”. In: *New England Journal of Medicine* 368.5, pp. 407–415 (cit. on p. 13).

Van Rossum, Guido (1995). *Python Programming Language*. Amsterdam (cit. on p. 24).

Van Zyl, Winschau F., Shelly M. Deane, and Leon M.T. Dicks (Nov. 9, 2020). “Molecular Insights into Probiotic Mechanisms of Action Employed against Intestinal Pathogenic Bacteria”. In: *Gut Microbes* 12.1, p. 1831339 (cit. on pp. 86, 87).

Vasala, Antti, Vesa P. Hytönen, and Olli H. Laitinen (July 15, 2020). “Modern Tools for Rapid Diagnostics of Antimicrobial Resistance”. In: *Frontiers in Cellular and Infection Microbiology* 10, p. 308 (cit. on p. 81).

Venter, J. C. (Apr. 2, 2004). “Environmental Genome Shotgun Sequencing of the Sargasso Sea”. In: *Science* 304.5667, pp. 66–74 (cit. on p. 19).

Vernikos, George, Duccio Medini, David R Riley, and Hervé Tettelin (Feb. 2015). “Ten Years of Pan-Genome Analyses”. In: *Current Opinion in Microbiology* 23, pp. 148–154 (cit. on p. 91).

Vindigni, Stephen M. and Christina M. Surawicz (Mar. 2017). “Fecal Microbiota Transplantation”. In: *Gastroenterology Clinics of North America* 46.1, pp. 171–185 (cit. on p. 13).

Vujkovic-Cvijin, Ivan, Jack Sklar, Lingjing Jiang, et al. (Nov. 19, 2020). “Host Variables Confound Gut Microbiota Studies of Human Disease”. In: *Nature* 587.7834, pp. 448–454 (cit. on p. 88).

Walker, Alan W., Jennifer C. Martin, Paul Scott, et al. (Dec. 2015). “16S rRNA Gene-Based Profiling of the Human Infant Gut Microbiota Is Strongly Influenced by Sample Processing and PCR Primer Choice”. In: *Microbiome* 3.1, p. 26 (cit. on p. 85).

Walsh, Christopher (Aug. 2000). “Molecular Mechanisms That Confer Antibacterial Drug Resistance”. In: *Nature* 406.6797, pp. 775–781 (cit. on p. 81).

Walter, Jens, Anissa M. Armet, B. Brett Finlay, and Fergus Shanahan (Jan. 2020). “Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents”. In: *Cell* 180.2, pp. 221–232 (cit. on pp. 88, 89).

Wang, Jiunn-Wei, Chao-Hung Kuo, Fu-Chen Kuo, et al. (Mar. 2019). “Fecal Microbiota Transplantation: Review and Update”. In: *Journal of the Formosan Medical Association* 118, S23–S31 (cit. on p. 13).

Wang, Qiong, George M. Garrity, James M. Tiedje, and James R. Cole (Aug. 15, 2007). “Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy”. In: *Applied and Environmental Microbiology* 73.16, pp. 5261–5267 (cit. on pp. 28, 75).

Waskom, Michael (Apr. 6, 2021). “Seaborn: Statistical Data Visualization”. In: *Journal of Open Source Software* 6.60, p. 3021 (cit. on p. 24).

Wei, Ze-Gang, Xiao-Dan Zhang, Ming Cao, et al. (Mar. 24, 2021). “Comparison of Methods for Picking the Operational Taxonomic Units From Amplicon Sequences”. In: *Frontiers in Microbiology* 12, p. 644012 (cit. on p. 18).

Westcott, Sarah L. and Patrick D. Schloss (Dec. 8, 2015). “De Novo Clustering Methods Outperform Reference-Based Methods for Assigning 16S rRNA Gene Sequences to Operational Taxonomic Units”. In: *PeerJ* 3, e1487 (cit. on p. 18).

Wieërs, Grégoire, Leila Belkhir, Raphaël Enaud, et al. (Jan. 15, 2020). “How Probiotics Affect the Microbiota”. In: *Frontiers in Cellular and Infection Microbiology* 9, p. 454 (cit. on p. 12).

Wilcoxon, Frank (1945). “Individual Comparisons by Ranking Methods”. In: 1, p. 5 (cit. on p. 31).

Wilkins, Lamont J., Manoj Monga, and Aaron W. Miller (Dec. 2019). “Defining Dysbiosis for a Cluster of Chronic Diseases”. In: *Scientific Reports* 9.1, p. 12918 (cit. on p. 8).

Williams, Michael D., Christina Y. Ha, and Matthew A. Ciorba (Mar. 8, 2010). “Probiotics as Therapy in Gastroenterology: A Study of Physician Opinions and Recommendations”. In: *Journal of Clinical Gastroenterology* Publish Ahead of Print (cit. on p. 12).

Woese, C. R. and G. E. Fox (Nov. 1, 1977). “Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms”. In: *Proceedings of the National Academy of Sciences* 74.11, pp. 5088–5090 (cit. on p. 17).

Woo, P.C.Y., S.K.P. Lau, J.L.L. Teng, H Tse, and K.-Y. Yuen (Oct. 2008). “Then and Now: Use of 16S rDNA Gene Sequencing for Bacterial Identification and Discovery of Novel Bacteria in Clinical Microbiology Laboratories”. In: *Clinical Microbiology and Infection* 14.10, pp. 908–934 (cit. on p. 16).

Wood, Ryan L., Tanner Jensen, Cindi Wadsworth, et al. (Feb. 20, 2020). “Analysis of Identification Method for Bacterial Species and Antibiotic Resistance Genes Using Optical Data From DNA Oligomers”. In: *Frontiers in Microbiology* 11, p. 257 (cit. on p. 81).

Yang, Bo, Yong Wang, and Pei-Yuan Qian (Dec. 2016). “Sensitivity and Correlation of Hypervariable Regions in 16S rRNA Genes in Phylogenetic Analysis”. In: *BMC Bioinformatics* 17.1, p. 135 (cit. on p. 18).

Yang, Hongyi (2019). “Factors Affecting the Composition of the Gut Microbiota, and Its Modulation”. In: p. 31 (cit. on p. 8).

Yatsunenko, Tanya, Federico E Rey, Mark J Manary, et al. (2012). “Human Gut Microbiome Viewed across Age and Geography”. In: *Nature* 486.7402, pp. 222–227 (cit. on p. 78).

Zhang, Xu, Leyuan Li, James Butcher, Alain Stintzi, and Daniel Figeys (Dec. 2019). “Advancing Functional and Translational Microbiome Research Using Meta-Omics Approaches”. In: *Microbiome* 7.1, p. 154 (cit. on p. 93).

Zion Market, Research (Jan. 31, 2020). *Probiotics Market: Size, Share & Trends Analysis Report by Ingredient Type (Bacteria and Yeast), by Form (Liquid Probiotic and Dry Probiotic), by Application (Food & Beverages, Dietary Supplements, and Animal Feed), By End User (Human Probiotics and Animal Probiotics): Global Industry Perspective, Comprehensive Analysis, and Forecast, 2019 – 2026*. Zion Market Research (cit. on p. 12).

Zmora, Niv, Gili Zilberman-Schapira, Jotham Suez, et al. (2018). “Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features”. In: *Cell* 174.6, 1388–1405.e21 (cit. on p. 13).

Zoumpopoulou, Georgia, Effie Tsakalidou, Joelle Dewulf, Bruno Pot, and Corinne Grangette (Apr. 2009). “Differential Crosstalk between Epithelial Cells, Dendritic Cells and Bacteria in a Co-Culture Model”. In: *International Journal of Food Microbiology* 131.1, pp. 40–51 (cit. on p. 2).

Websites

Fang, Yuan and Sanjeena Subedi (Nov. 12, 2020). *Clustering Microbiome Data Using Mixtures of Logistic Normal Multinomial Models*. arXiv: `2011.06682 [stat]`. URL: `http://arxiv.org/abs/2011.06682` (visited on Feb. 8, 2021) (cit. on p. 77).

Wetterstrand, Kris A. (2020). *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. URL: `www.genome.gov/sequencingcostsdata` (visited on Aug. 31, 2020) (cit. on p. 3).

Zymo, Research (2021). *ZymoBIOMICS Spike-in Control I (High Microbial Load) Protocol*. URL: `https://files.zymoresearch.com/protocols/_d6320_zymobiomics_spike-in_control_i.pdf` (visited on Apr. 13, 2021) (cit. on p. 76).

List of Figures

List of Tables

Appendix

Jonas Klopp

Curriculum vitae

Katharina-Pfahler-Str. 4
55128 Mainz
Germany
0176 70369779
jonas.klopp@web.de

Personal Details

Date of Birth	26.02.1993
Place of Birth	Trier
Nationality	German

Education

2018 – 2022 **PhD Bioinformatics**, *Medical Center of Johannes Gutenberg-University*, Mainz.
Analysis of the gut microbiome in neonates through high throughput metagenomic sequencing

2015 – 2017 **M.Sc. Biology**, *Johannes Gutenberg-University*, Mainz.
Highly relevant modules included:
Molecular genetics – genome studies and bioinformatical sequence analysis
Gene Regulation, Epigenetics and Genome Stability

2012 – 2015 **B.Sc. Molecular Biology**, *Johannes Gutenberg-University*, Mainz.
Highly relevant modules included:
Protein Biochemistry and Bioinformatics
Molecular genetics of Eucaryotes
Analysis of Eucaryotic Genes

2012 **Abitur**, *Max-Planck-Gymnasium*, Trier.

PhD thesis

Title	*Study of eu- and dysbiosis in the intestinal microbiome of very preterm infants in the PRIMAL study cohort*
Supervisors	Prof. Dr. Thomas Hankeln and Prof. Dr. Stephan Gehring

Publication

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE

Meconium Microbiome of Very Preterm Infants across Germany

Jonas Klopp,[a] Pamela Ferretti,[b] Claudius U. Meyer,[a] Katja Hilbert,[a] Annette Haiß,[c] Janina Marißen,[d] Philipp Henneke,[e,f,g] Hannes Hudalla,[h] Sabine Pirr,[i] Dorothee Viemann,[d,i] Michael Zemlin,[j] Sofia Kirke Forslund,[b,k,l,m,n] Christoph Härtel,[d] Peer Bork,[b,l,o,p] Stephan Gehring,[a] Thea Van Rossum,[b] and the PRIMAL Consortium

[a]University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany
[b]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
[c]Department of Pediatrics, University of Lübeck, Lübeck, Germany
[d]Department of Pediatrics, University of Würzburg, Würzburg, Germany
[e]Center for Pediatrics and Adolescent Medicine Pediatrics, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany
[f]Institute for Immunodeficiency, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany
[g]Center for Chronic Immunodeficiency (CCI) Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany
[h]Department of Neonatology, University of Heidelberg, Heidelberg, Germany
[i]Hanover Medical School, Department of Pediatric Pneumology, Allergology and Neonatology, Hanover, Germany
[j]Department of Pediatrics, Saar University Homburg, Homburg, Germany
[k]Experimental and Clinical Research Center, a cooperation of Charité-Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany
[l]Max Delbrück Centre for Molecular Medicine, Berlin, Germany
[m]Charité – Universitätsmedizin Berlin, Berlin, Germany
[n]German Centre for Cardiovascular Research (DZHK), partner site Berlin, Berlin, Germany
[o]Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea
[p]Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany

ABSTRACT Meconium constitutes infants' first bowel movements postnatally. The consistency and microbial load of meconium are different from infant and adult stool. While recent evidence suggests that meconium is sterile *in utero*, rapid colonization occurs after birth. The meconium microbiome has been associated with negative health outcomes, but its composition is not well described, especially in preterm infants. Here, we characterized the meconium microbiomes from 330 very preterm infants (gestational ages 28 to 32 weeks) from 15 hospitals in Germany and in fecal samples from a subset of their mothers (N = 217). Microbiome profiles were compiled using 16S rRNA gene sequencing with negative and positive controls. The meconium microbiome was dominated by *Bifidobacterium*, *Staphylococcus*, and *Enterococcus* spp. and was associated with gestational age at birth and age at sample collection. Bifidobacterial abundance was negatively correlated with potentially pathogenic genera. The amount of bacterial DNA in meconium samples varied greatly across samples and was associated with the time since birth but not with gestational age or hospital site. In samples with low bacterial load, human mitochondrial sequences were highly amplified using commonly used, bacterial-targeted 16S rRNA primers. Only half of the meconium samples contained sufficient bacterial material to study the microbiome using a standard approach. To facilitate future meconium studies, we present a five-level scoring system ("MecBac") that predicts the success of 16S rRNA bacterial sequencing for meconium samples. These findings provide a foundational characterization of an understudied portion of the human microbiome and will aid the design of future meconium microbiome studies.

IMPORTANCE Meconium is present in the intestines of infants before and after birth and constitutes their first bowel movements postnatally. The consistency, composition and microbial load of meconium is largely different from infant and adult stool. While recent evidence suggests that meconium is sterile *in utero*, rapid colonization occurs

Editor Vincent B. Young, University of Michigan-Ann Arbor

Address correspondence to Thea Van Rossum, thea.van.rossum@embl.de.

The authors declare no conflict of interest.

Received 30 September 2021
Accepted 19 December 2021
Published 12 January 2022

after birth. The meconium microbiome has been associated with short-term and long-term negative health outcomes, but its composition is not yet well described, especially in preterm infants. We provide a characterization of the microbiome structure and composition of infant meconium and maternal feces from a large study cohort and propose a method to evaluate meconium samples for bacterial sequencing suitability. These findings provide a foundational characterization of an understudied portion of the human microbiome and will aid the design of future meconium microbiome studies.

KEYWORDS meconium, microbiome, 16S rRNA gene sequencing, mitochondria, mother

The development and composition of the early gut microbiome has been shown to impact the well-being of infants (1). This has been well-studied for infants of 2 to 46 months of life (2). However, the earliest developmental stages of the gut microbiome are still a topic of ongoing investigation (3, 4). Recent studies suggest that the uterus and thus the fetal meconium is sterile in healthy women (5). Sterility of fetal meconium was confirmed by sequencing and culturing stool samples obtained from the infant colon during elective breech cesarean deliveries (6). While it was previously thought that the first pass stool of infants after birth does not contain viable bacteria, recent studies, including the aforementioned, have found evidence of bacterial DNA in first pass meconium with the help of next-generation molecular techniques (6–9). First pass meconium is likely colonized with bacteria postnatally from breast milk, maternal gut and skin, and other environmental sources (10). Other studies have suggested that the uterus is not sterile (11) and that the post-birth meconium microbiome might mirror the *in utero* microbial environment (12, 13). Irrespective of whether meconium is colonized before, during or after birth, the composition and structure of early-life gut communities may impact health later in life (14).

Further explorations of associations between meconium microbiome composition and infant health require baseline characterization of the meconium microbiome. To date, few studies have investigated the meconium microbiome of newborns in sufficiently powered studies and with adequate controls. One noteworthy case-control trial from China investigated the association between neonatal jaundice and the meconium microbiome in 301 newborns (15) and found a higher abundance of *Bifidobacterium pseudolongum* and higher alpha diversity to be associated with a lower risk of jaundice in cesarean-born infants. They also discovered that meconium was mainly composed of bacteria from the phyla *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Cyanobacteria* and *Bacteroidetes*. The same primary taxa were identified in other meconium studies albeit in different relative proportions (16–18). However, many of these studies lack suitable controls. Studies are furthermore hampered by low bacterial DNA quantities in meconium compared to stool from older infants (1). Working with meconium thus requires the rigorous use of negative and positive controls to ensure reliable results (19, 20).

In this study, we provide a characterization of the microbiome in meconium samples of a representative cohort of preterm infants and their respective mothers from 15 hospitals in Germany as part of the "Priming Immunity at the beginning of life" (PRIMAL) study (21). We further elaborate on the problem of co-amplification of human mitochondrial 12S rRNA genes (the homologue of the 16S rRNA gene in prokaryotes [22]) that can arise while dealing with meconium samples and present a scoring system to aid the design of future meconium microbiome studies.

RESULTS

Sequencing and quantification of bacterial and mitochondrial amplicons. A total of 547 samples were sequenced and analyzed in this study (N = 330 infant meconium and N = 217 mother fecal samples). For each infant, the hospital site of origin, sex and the gestational age category (28 to 30 weeks or >30 to 32 weeks) were recorded (Table S1). Sequencing data were generated for 665 samples (including 118

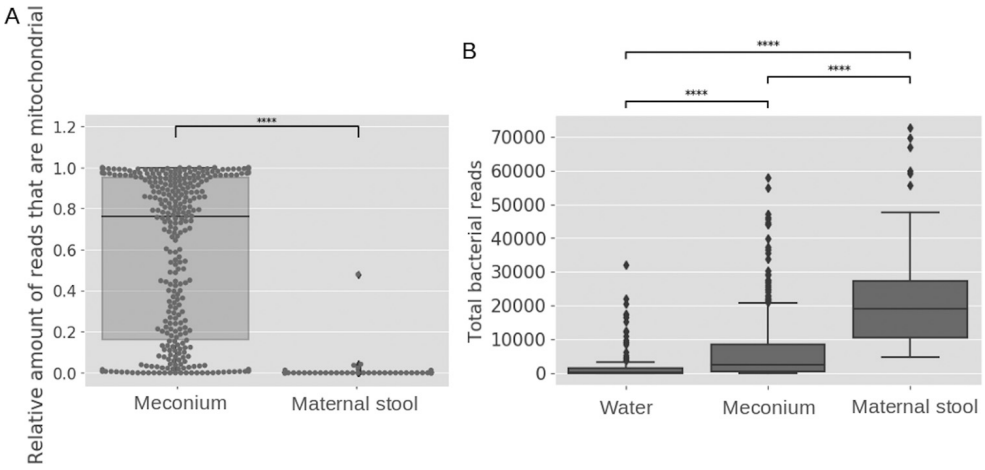


FIG 1 Differences in the yield of mitochondrial and bacterial sequences by sample type. (A) Proportion of total amplicon sequences that are mitochondrial sequences in infant meconium and maternal stool samples. *P* value annotation legend: ****: $P \leq$ 1.00e-04. (A) After removal of mitochondrial sequences, differences in the number of bacterial sequences in negative controls ("water," N = 118), infant meconium samples (N = 330) and maternal stool samples (N = 217). Two mother samples and two meconium samples with more than 80,000 total bacterial reads were omitted from the plot for legibility (mothers: 93,954 and 131,844 reads, infants: 91,079 and 121,745 reads). Kruskal-Wallis tests were performed between all samples (A & B). *P* value annotation legend: ns: 5.00e-02 $< P \leq$ 1.00e + 00, *, 1.00e-02 $< P \leq$ 5.00e-02, **, 1.00e-03 $< P \leq$ 1.00e-02, ***, 1.00e-04 $< P \leq$ 1.00e-03, ****, $P \leq$ 1.00e-04.

controls), with a mean of 23,792 reads per sample (median of 21,055 reads, 25th percentile of 9,586 reads, 75th percentile of 33,167 reads). The length distribution of ASVs was bi-modal, with peaks at 251 and at 200 nucleotides long. The longer length is the expected value for bacterial amplicons. The shorter ASVs were classified as human mitochondria using BLASTn and the NCBI nucleotide collection (nt).

The presence of bacterial genera known to be technical contaminants (*Delftia*, *Flavobacterium*, *Pseudomonas*, *Burkholderia*, *Sphingomonas*, *Corynebacterium*, and *Propionibacterium*) (23) was assessed and, except for *Corynebacterium*, all other known contaminants were found in less than 1% of samples and their 95th percentile abundance was less than 0.1% of the meconium microbiomes. *Corynebacterium* was present in 26% of samples, with a median abundance of 0.2% in those samples.

The relative number of mitochondrial sequences was strikingly high in infant meconium samples relative to the maternal stool samples; with a median of 76% in meconium compared to 0% in maternal stool (Fig. 1A, Kruskal-Wallis test, $P < 0.0001$). To determine whether the high abundance of mitochondrial sequences was due to low bacterial load in samples, a defined bacterial load was added to a meconium sample prior to extraction. This led to a drastic reduction in the relative number of mitochondrial sequences in the microbiome and a microbiome composed of only spike-in DNA (Fig. S1). The amount of spike-in used would be expected to make up at most 10% of the microbiome from a normal stool sample.

Before taxonomic analysis, quality control and removal of mitochondrial reads was performed. Based on negative controls and rarefaction curves (Fig. S1), samples were discarded if they contained less than 2500 bacterial reads after removal of mitochondrial sequences and filtration from the ASV identification. 165 meconium samples (50%) and 97 negative controls (82%) did not pass this threshold. Without mitochondrial sequences, the number of quality controlled bacterial reads in meconium samples was significantly higher than in the negative controls (Kruskal-Wallis test, $P < 0.0001$) and the number of bacterial reads in meconium samples was significantly lower than in maternal fecal samples (Kruskal-Wallis test, $P < 0.0001$) (Fig. 1B). Across all samples that passed all quality criteria (N = 165 infant meconium samples and N = 217 maternal samples), 6,267 ASVs were observed. Of these, 130 were identified as contamination with decontam, leaving 6137 ASVs. After decontamination, 469 ASVs had at least 1%

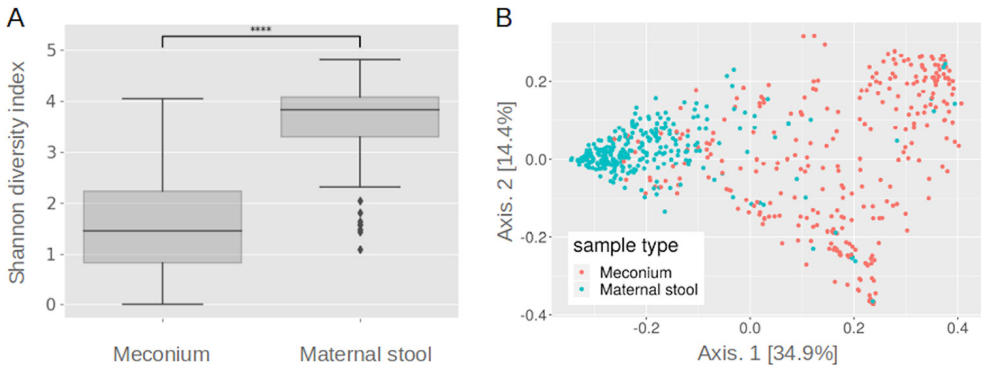


FIG 2 Diversity in meconium and maternal stool samples on the ASV level. (A) The difference in alpha diversity measured by the Shannon Index reveals that meconium samples have far less diversity than mother samples. *P* value annotation legend, ****, $P \leq$ 1.00e-04. (B) Beta diversity of infant meconium and maternal stool samples. Principal Coordinate Analysis (PCoA) based on the weighted UniFrac distances between samples is shown. Every point represents one sample and is colored based on the sample type.

abundance in at least two microbiome samples. In meconium samples, 2,312 ASVs were observed with 231 having at least 1% abundance in at least two samples. In mother samples, 4,538 ASVs were observed with 350 having at least 1% abundance in at least two samples.

Taxonomic composition of infant meconium and maternal stool. 16S rRNA gene-based microbiome taxonomic profiles were compiled from bacterial sequences for the samples that passed all quality criteria (N = 165 infant samples and N = 217 mother samples) (Tables S2, S3 and S4, https://seafile.rlp.net/d/37e022ec0ccb43018073/). As expected, significant differences in the microbiome composition of infant meconium and maternal stool samples were observed, (PERMANOVA $P < 0.001$). The alpha diversity was significantly higher in maternal stool samples than in meconium (Shannon index, Fig. 2A, Kruskal-Wallis test, $P < 0.0001$). If repeatedly subsampled to 2500 reads, meconium samples had a median of 12.5 ASVs per sample and maternal stool samples had a median of 112.5 ASVs per sample (meconium: 25th percentile = 6 ASVs, 75th percentile = 52 ASVs, ranging from 1 to 151 ASVs per sample, mean across subsampling; maternal stool: 25th percentile = 77 ASVs, 75th percentile = 145 ASVs, ranging from 5 to 253 ASVs). The diversity among mother samples was lower than the diversity among meconium samples, as seen in the spread of data in the PCoA (Fig. 2B, Permutation test for homogeneity of multivariate dispersions, *P* value < 0.001). There was no clustering of samples based on hospital sites (See Fig. S3).

The taxonomic composition of meconium was significantly associated with the gestational age at birth and infant age when the meconium was collected (Table 1). Hospital collection site was also associated with composition of meconium and maternal stool samples, however, permutation tests for homogeneity of multivariate dispersions indicated that these differences might be due to differences in internal variation within the hospital sites rather than differences in taxa (Table 1).

Maternal stool and meconium samples primarily consisted of bacteria that are members of the phyla *Firmicutes*, *Bacteriodota*, *Proteobacteria*, and *Actinobacteria* (Fig. 3A and B). At the genus level, the most common bacterial genera differed between mother and infant samples. The most abundant genus in meconium was *Bifidobacterium*, which was also commonly present in mothers. The other top genera in meconium (*Staphylococcus*, *Enterococcus*, *Streptococcus*, and *Escherichia-Shigella*) were also found in maternal stool samples but at much lower abundances (median: 0%, 0.04%, 0.72%, 0.12%, respectively). *Bacteroides* was the most common genus in maternal stool samples but did not constitute a major genus in the meconium samples (median: 0.04%) (Fig. 3C and D). A full list of genus and ASV abundances in meconium and maternal stool samples is depicted in Tables S6, S7, S8, and S9. Other top genera in

TABLE 1 Association of gestational age at birth, age at meconium sample collection, and hospital collection site on overall microbiome composition (ASV level) in meconium from preterm infants

	Gestational age	Hospital site	Age at meconium collection
Permanova *P* value	$P < 0.0001$	$P < 0.009$	$P = 0.028$
Effect size	0.018	0.12	0.02
Dispersion test	$P = 0.304$	$P = 0.001$	$P = 0.187$

maternal stool samples (*Blautia*, *Faecalibacterium*, and *Subdoligranulum*) were present in some of the meconium samples but at lower prevalence and abundance (prevalence: 38%, 37%, 28%, respectively, with median of 0% abudance and a 75th percentile $<$ 0.01% abundance in the meconium microbiome).

Meconium samples from infants and stool samples from their mothers were tested for shared taxa. Eighty-two mother-infant pairs were analyzed. A median of 28% of ASVs in the infant samples were also present in the corresponding mother samples, which was not significantly different from any other comparison between mothers and unrelated infants, either from the same or different hospitals (Kruskal-Wallis test, $P > 0.05$, with Bonferroni correction) (Fig. 4 and Table S5, https://seafile.rlp.net/d/37e022ec0ccb43018073/). Mother samples from the same hospital that shared 100% of ASVs were from the same mother that gave birth to twins, so was sampled twice. Only one meconium sample had complete presence of its ASVs in its corresponding mother sample and the infant in question only had 2 ASVs.

Taxa across meconium samples were tested for co-occurrence. Out of the 15 most abundant genera (selected by median and mean abundance), significant correlations in abundances were observed for six taxa (discretized mutual information test as implemented by FlashWeave, alpha = 0.01, Fig. 5). Two ASVs (taxa of approximately species-level resolution) that were classified as members of the genus *Bifidobacterium* showed a positive co-occurrence and one of these is also highly correlated with the

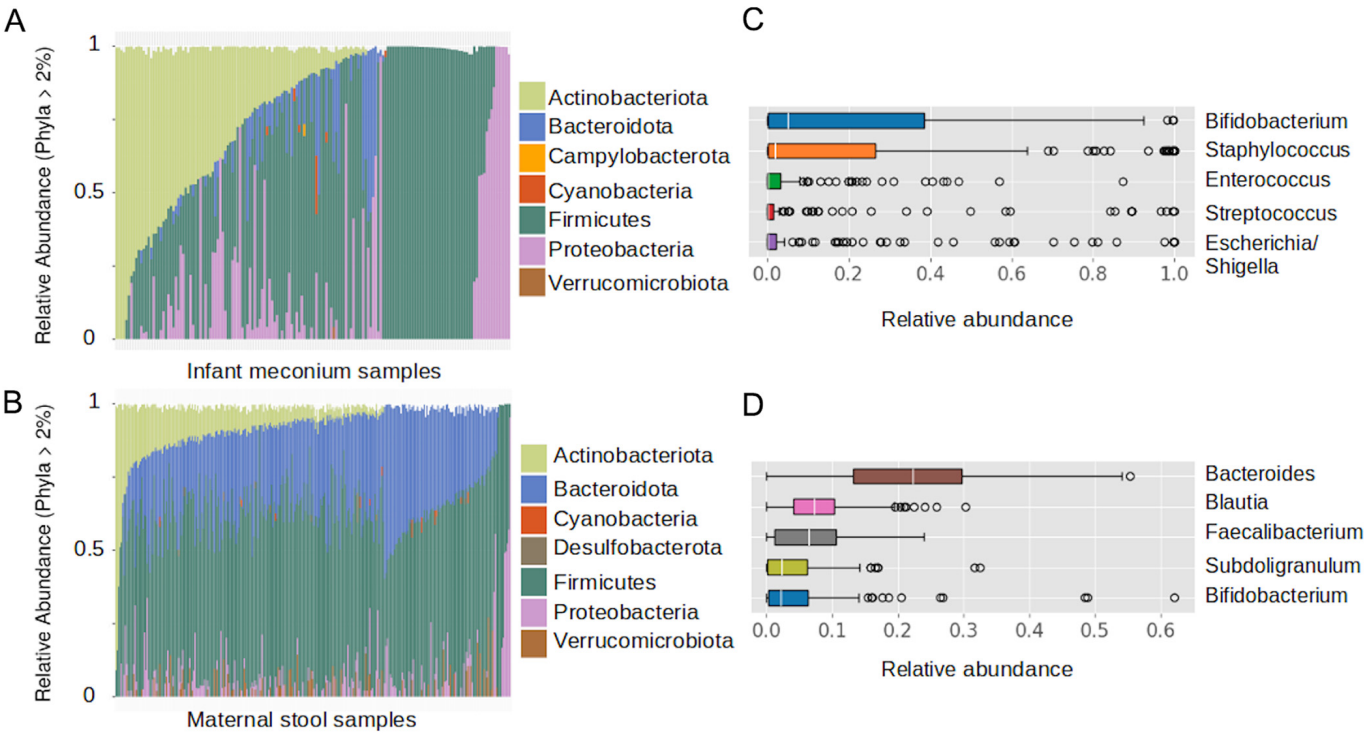


FIG 3 Taxonomic composition of the meconium and maternal stool samples. (A) Phylum composition of the infant meconium samples (N = 165). (B) Phylum composition of the maternal stool samples (N = 217). Only phyla that comprised more than 2% of total microbiome composition across samples are shown. (C) The five most abundant genera by median abundance across samples in meconium and (D) maternal stool microbiomes.

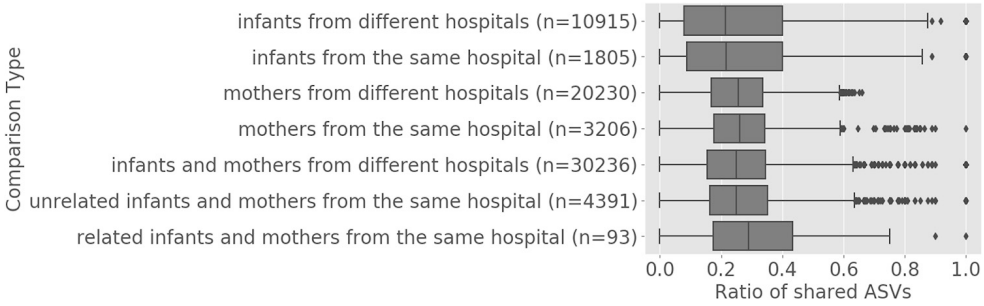


FIG 4 Proportion of shared ASVs in meconium and maternal stool samples. Proportion of shared ASVs was computed for different sample sets. Each boxplot shows one type of comparison. On the *y* axis the number (n) of pairwise comparisons for each comparison type is listed. No statistically significant differences between comparison types were found.

presence of *Lactobacillus*. *Staphylococcus*, and *Enterococcus* are each negatively correlated with a *Bifidobacterium* ASV, as well as to each other. The presence of a *Staphylococcus* ASV was negatively correlated with *Escherichia-Shigella*. Table 2 shows an overview of the prevalence and abundance of the correlated taxa. Nine of the 15 most abundant ASVs showed no significant correlation with each other. The ASV sequences of "*Bifidobacterium* 1" and "*Bifidobacterium* 2" (numbering is study specific) are 98% identical, with the latter likely belonging to *Bifidobacterium animalis*, based on nucleotide sequence similarity to reference sequences in NCBI nt database.

Scoring meconium suitability for bacterial sequencing. The taxonomic analysis described above revealed that the primers used in this study amplified a product from the human mitochondrial 12S rRNA gene. The sequencing data showed that this prod-

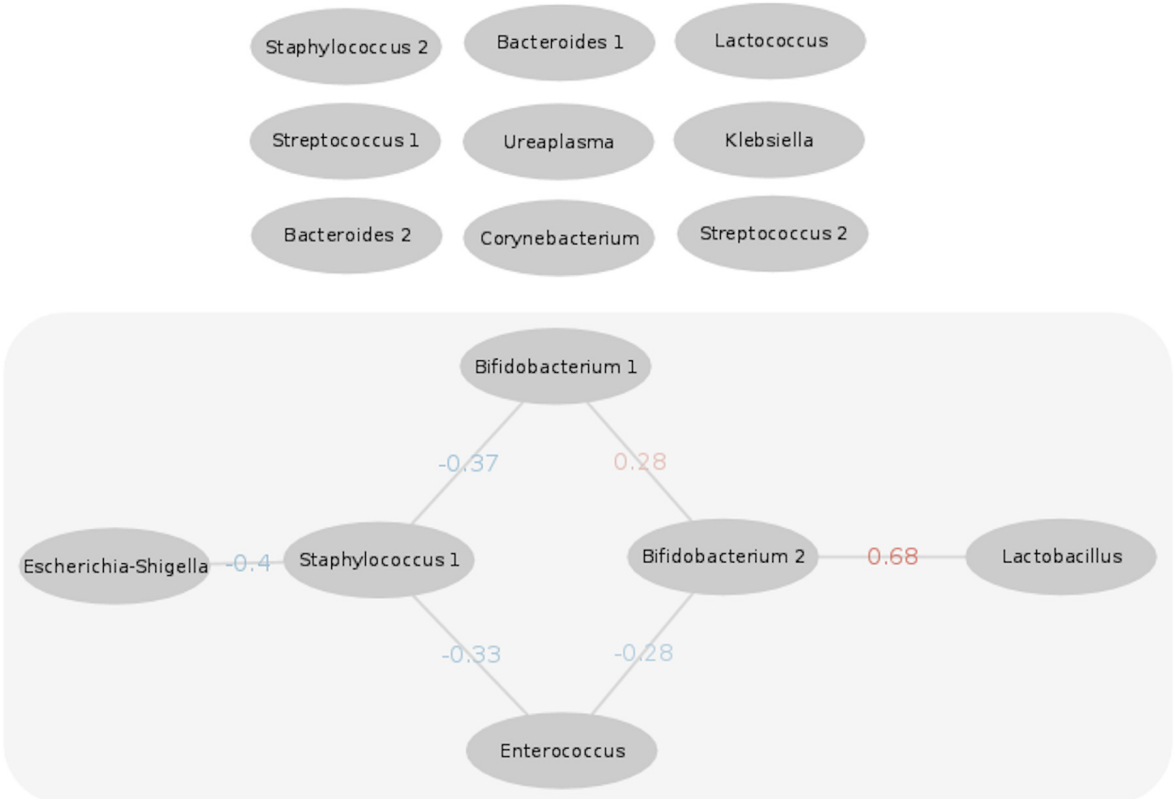


FIG 5 Co-occurrence network of the most abundant genera in infant meconium. Nodes are ASVs (taxa of approximately species-level resolution) with their respective genus classifications as labels. The 15 most abundant genera were selected according to median and mean abundance. Edges represent statistically significant correlations. The weight of the correlation is color coded from blue (negative correlation) to red (positive correlation). The significance level is alpha = 0.01.

TABLE 2 Prevalence and abundance statistics and taxonomic classifications of bacterial ASVs in meconium samples that showed significant correlations in the co-occurrence network analysis

AASV	Prevalence	Median abundance	25th percentile	75th percentile
Bifidobacterium 1	38 %	3.6 %	0.04 %	31.91 %
Bifidobacterium 2	14 %	0 %	0%	0.12 %
Lactobacillus	17 %	0 %	0 %	0.16 %
Staphylococcus	75 %	2 %	0.05 %	12 %
Escherichia-Shigella	63 %	0.48 %	0 %	7 %
Enterococcus	63 %	0.8 %	0 %	5.4 %

uct was consistently and distinctively shorter than the bacterial product (ca. 200 bp for mitochondrial amplicons versus ca. 250 bp for the bacterial amplicons). This difference can be leveraged to test for the presence of both products in samples prior to DNA sequencing. After extraction and amplification of DNA from meconium samples, five distinct patterns of amplicon length distributions per sample emerged when the DNA was subjected to gel electrophoresis. We classified these in a scoring system from one to five (Table 3), which we call a Meconium Bacterial Load ("MecBac") score. Samples with a score of one include only the amplicon product that has the expected size from amplification from bacterial DNA. Samples with scores of two and three have some bacterial product, but also contain mitochondrial product. Scores of four and five do not support the presence of any bacterial product in the sample. 47% of the meconium samples in this study had a score of 3 or better, 53% were unsuitable for typical microbiome sequencing.

The quantity of bacterial reads per sample, after quality control and removal of mitochondrial sequences, correlates strongly with the MecBac score ("total reads" analysis [see methods], Pearson correlation coefficient: -0.61, $P < 10^{-34}$, Spearman correlation coefficient: -0.78, $P < 10^{-67}$, N = 327). Samples that were classified with the best score (1) had the most bacterial reads but not the most total overall reads (Fig. 6).

MecBac scores were not correlated with the gestational age of infants at birth or the hospital site (Spearman correlation coefficient: -0.08, $P > 0.15$, 0.02, $P > 0.72$, N = 327) but showed an inverse correlation with the age of the infants when the meconium was collected (Spearman correlation coefficient -0.358, $P < 0.001$, N = 241). Samples collected later in life (after more time has passed for bacterial growth) had a better MecBac score. After removal of contaminants and mitochondrial sequences, the composition of meconium bacterial communities was significantly associated with a sample's MecBac score, after controlling for gestational age and hospital (PERMANOVA, $P < 0.001$, effect size = 5.8%). There were also significant differences in alpha diversity between samples with different MecBac scores (Kruskal-Wallis test, $P < 0.00001$). The samples with the best score tended to have lower alpha diversity. The higher diversity in samples with worse MecBac scores could be due to their lower bacterial loads, which increases the likelihood of contamination and random amplification of this DNA or it could be a real biological difference, whereby samples that have a lower bacterial load also tend to have a different community composition. Samples with the best MecBac score (1) had significantly fewer sequences identified as contamination during the decontamination step than other meconium samples (Kruskal-Wallis test, $P < 0.00001$).

DISCUSSION

The meconium microbiome has been discussed as a potential source of information about long term health and disease (13). Recent evidence demonstrates that the meconium microbiome may predict and impact health and disease both early in life, in terms of weight gain and growth (17), sepsis risk (24), and NEC propensity (25), as well as later in life, e.g., concerning the incidence of asthma, allergies, obesity and psychiatric illnesses (26). A comprehensive understanding of baseline bacterial structure and composition of

TABLE 3 Meconium bacterial load scoring system ("MecBac" Score) from 1 (best for microbiome sequencing) to 5 (worst for microbiome sequencing)

Meconium bacterial load score ("MecBac" score)	Gel band of size 200 bp (mitochondrial)	Gel band of size 250 bp (bacterial)	Suitability for bacterial microbiome sequencing	Percentage of meconium samples in this study
1	Absent	Present	Excellent	25 %
2	Present	Present	High	14 %
3	Present	Present (Lighter)	Reasonable	8 %
4	Present	Absent	Unsuitable	36%
5	Absent	Absent	Unsuitable	17 %

meconium may complement our knowledge in this respect. In this study, we provide a basic characterization of the meconium microbiome in 330 infants and comparisons with a subset of their mothers. These findings are representative of very preterm infants in Germany. Preterm infants may have a different meconium composition than full term infants as passage takes longer and is often delayed in preterm infants (27). As preterm infants are at increased risk of poor health outcomes, possibly linked to gut dysbiosis (28), gut microbiome studies are valuable in this population to understand mechanisms and find new targets for preventive and therapeutic strategies.

Since meconium samples contain very small amounts of bacterial biomass, it is especially important to assess possible contamination (19, 20) and to avoid spurious results by using negative controls and purpose-built statistical methods (29). To only consider high quality microbiome profiles in our analysis, we included negative and positive controls and used specific software to identify likely contamination. Samples that contained fewer bacterial reads than the mean in the negative controls were discarded. Any bacterial community in these samples was not considered to be well represented by the typical amplicon sequencing approach used here. Interestingly, about half of all meconium samples in this study did not pass this threshold and had a high abundance of mitochondrial sequences. This finding suggests that the bacterial load in meconium varies greatly across infants, with many having a smaller amount than is accessible to study using typical amplicon microbiome approaches. This variation appears to be correlated with the age of the infant, such that meconium passed later in life (e.g., at day 2) tends to have more bacterial load than meconium passed earlier in life (e.g., at day 1). This trend fits with the hypothesis that meconium is sterile at birth and then is colonized postnatally. This correlation is weak however, so there are likely other factors that also impact the bacterial load, such as prolonged rupture of membranes, mode of birth, diet and environmental exposure. This variation in bacterial presence has been previously reported in another much smaller study (N = 15) (30), and hints at a possible reason for the ongoing controversy (31) on the presence and source of bacteria in meconium.

In our study, the main driver of DNA load in many meconium samples was human mitochondrial DNA (comprising a median of 76% of amplicon reads) instead of bacterial

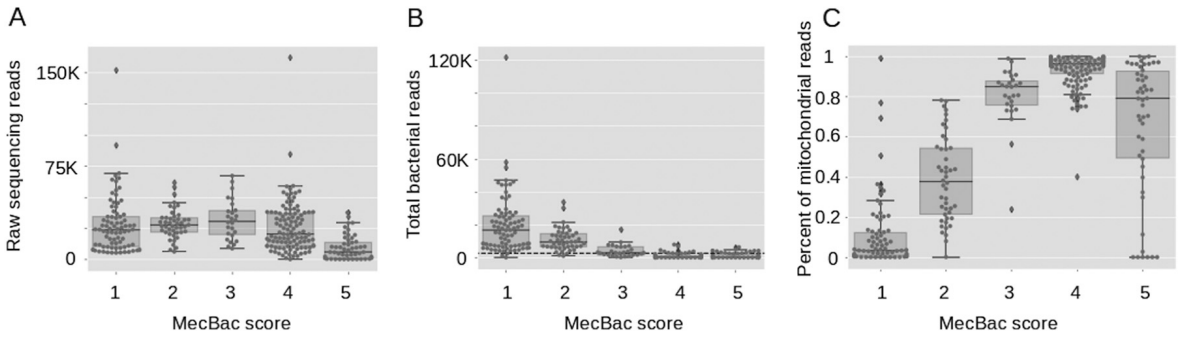


FIG 6 The quantities of raw, bacterial, and mitochondrial reads are associated with the meconium gel quality (MecBac) score. (A) Samples with the worst quality score tend to yield the fewest reads when sequenced. (B) Meconium with a better score yields higher numbers of high-quality bacterial reads. The dashed line shows the minimum threshold of 2500 reads. (C) A better (lower) MecBac score is correlated with a lower proportion of mitochondrial reads (Spearman correlation coefficient, 0.64, $P < 0.0001$, N = 327).

DNA. As the first pass stool, such abundance of mitochondrial DNA is not unexpected as meconium contains many human cells such as intestinal epithelial cells of the infant. However, a PCR using 16S rRNA gene specific primers should not result in the amplification of these mitochondrial sequences. We speculate that the widely used protocol of 16S primer pairs 515f and 806r, first published by Caporaso et al. (32) and later optimized (33) might have led to this spurious amplification. The forward primer has a sequence identity of 85.71% and the reverse primer of 89.47% to the 12s rRNA gene of the human mitochondrium spanning a 199 bp long segment, explaining the gel band at around 200 bp in our study. In fecal samples with a sufficient amount of bacterial DNA the competition is large enough to inhibit the primer binding to mitochondrial DNA (Fig. 1B). However, in a very low biomass setting like meconium, imperfect binding to the mitochondrial DNA might facilitate their amplification. This hypothesis is further strengthened by the observation that the relative amount of amplified mitochondrial DNA is drastically reduced when using spike-in controls of defined bacterial load (Fig. S2), suggesting very low or absent bacterial biomass in the original sample.

Interestingly, other studies investigating the meconium microbiome composition using the V4 region of the 16S rRNA gene such as one from China (18) and one from Brazil (17) found *Proteobacteria* to be a highly abundant member of the microbiome in most samples, while we found it only sporadically (Fig. 3A). Using the taxonomic classification provided by SILVA, human mitochondrial amplicon sequences are classified as Kingdom: *Bacteria*, Phylum: *Proteobacteria*, Class: *Alphaproteobacteria*, Order: *Rickettsiales*. It might thus be possible that mitochondrial sequences introduce bias to the structure of underlying microbial communities. However, it is well known that other covariates like geographical origin of the samples (34) and differences in extraction methods (especially when working with low biomass samples such as meconium [35]) might explain microbial differences found in the infant gut.

In the literature, the early ecological succession in the infant gut is described as being driven by niche-based competition after an initial seeding with bacteria whose taxonomic composition is in part subject to stochastics effects (36). Evidence for these stochastic elements can be seen in the large beta diversity between meconium samples in this study. Similarly to full term infants (37), the genus *Bifidobacterium* was one of the main genera in the meconium of premature infants. This highlights the pivotal role of this genus in the developing infant gut microbiome from birth. Microbiomes of preterm infants also tend to show higher relative abundances of *Staphylococcus* immediately after birth compared to full term infants, which is consistent with our data (38). In line with earlier studies (39, 40), the meconium composition in our study is characterized by lower alpha diversity compared to the later microbiomes in preterm infants.

Mitochondrial amplification from the V4 primers led to most meconium samples having sufficient DNA concentration to pass typical sequencing library preparation quality controls (based on Qubit measurements). However, after removing mitochondrial sequences, 50% of samples lacked sufficient bacterial data for analysis and therefore wasted significant resources. Typical quality control procedures such as NanoDrop or Qubit measurements would not differentiate between bacterial and mitochondrial amplicon products and thus not predict these wasted samples. It is possible that this issue would not be as severe if using other 16S amplicon primers that targeted another region on the 16S rRNA gene. However, the V4 region of the 16S rRNA gene used in this study has substantial advantages over using other or multiple 16S rRNA regions in microbiome studies, such as a very low error rate (41) and suitability for meta-analysis, due to its widespread popularity in the field. To support future use of V4 primers to study meconium despite the nonspecific amplification issue, we developed a five-point scoring system to assess the suitability of meconium samples for bacterial sequencing ("MecBac" score) based on a simple gel electrophoresis of amplicon products. This allows the selection of promising samples prior to sequencing, which, based on the score distribution in our study, would potentially save 50% of the sequencing costs in future studies. While developed for 16S rRNA sequencing, this scoring system can also

be used for sample selection for more expensive sequencing, such as metagenomic whole-genome shotgun sequencing.

For the infants in our study that had a high number of bacterial reads, the alpha diversity was sometimes especially low (Shannon index close to 0), indicating the dominance of only one bacterial genus or species. 35 infants were dominated by less than 5 ASVs. The main genera present in these microbiome samples were *Staphylococcus*, *Escherichia-Shigella*, and *Streptococcus*. All of these genera have members that are well known human pathogens, such as methicillin-resistant *Staphylococcus aureus* (MSRA), certain serotypes of *Escherichia coli* and many *Shigella* species that are the cause of shigellosis. The microbiome of the infants concerned might thus be in a state of extreme dysbiosis. However, two limitations of the 16S rRNA gene sequencing approach make it hard to draw definite conclusions. First, the lack of sufficient taxonomic resolution to identify bacteria below the genus level does not allow the discrimination of pathogenic and nonpathogenic strains in the aforementioned genera. Second, it is possible that the apparent abundance of bacterial species was artificially inflated through PCR kinetics early in the amplicon methodology. To validate these findings in future studies, other methods, such as targeted qPCR or amplification-free metagenomics, would be required.

The statistically significant co-occurrence correlations among abundant taxa in the meconium microbiome may provide insight into microbial interactions. Two ASVs belonging to the *Bifidobacterium* genus, which is well known to be a beneficial member of the early microbiome in infants (42), were negatively correlated with the potentially pathogenic genera *Staphylococcus* and *Enterococcus*. These negative correlations might be further evidence for the competition-based mechanism hypothesis of *Bifidobacterium* in the clinical effectiveness of probiotics (43). Several molecular mechanisms are known by which *Bifidobacterium* and other probiotic bacteria can have suppressing effects on intestinal pathogenic bacteria (reviewed in van Zyl et al. [44]). Certain species of the genus *Bifidobacterium*, for example, are able to produce proteins that actively prevent pathogenic bacteria from attaching to the mucosal cells of the host, produce unique bacteriocins, called bifidocins, or simply competitively outperform pathogens in the struggle for resources (44). The strong positive correlation between *Bifidobacterium* and *Lactobacillus* is in line with network analyses of the gut microbiome from other groups (45) and these two genera are also known to be abundant in the gut microbiomes of term infants (46) especially after vaginal delivery (47). In our study, these taxa were included in a probiotic given to half of the infants. Since the study is still blinded, we could not assess in detail to what extent the probiotic could be strengthening or even driving the association. Unfortunately, the resolution of 16S rRNA gene sequencing does not always allow the differentiation of different species within a genus. Thus, the identity of the *Bifidobacterium* and *Lactobacillus* cannot be unambiguously assessed for all ASVs at a level lower than genus. However, since the genus *Bifidobacterium* is present in 76.8% of samples while only 50% of the infants received a probiotic, it is reasonable to assume that naturally occurring *Bifidobacterium* are highly relevant to the detected associations in the meconium samples. In future studies aimed at understanding dysbiosis in infants, the interactions between different taxa in the normal microbiome between presumably beneficial bacteria (*Bifidobacterium*, *Lactobacillus*) and potentially pathogenic bacteria should thus be given a pivotal role.

Earlier studies on the transfer of bacteria from mothers to full term, nonhospitalised infants found that the gut microbiome is the major source of shared strains (48) and that up to 63% of the infant's microbiome can be explained by bacteria in the maternal gut (49). Mother-infant pairs were characterized by an increased number of shared ASVs, compared to unrelated pairs. However, probably due the sample size of this study, this difference was not significant. This indicates that while a transfer might have occurred, we do not have evidence that it was the main source for bacteria in the meconium. However, insufficient taxonomic resolution prohibits final conclusions on the exact transfer of bacteria from mothers to infants. Future studies would require strain level information to make these conclusions, such as from metagenomic whole-genome shotgun sequencing.

Summary and conclusion. Analysis of meconium, the first pass stool of newborns, offers insights into the very early development of the human gut microbiome. Characterization of the microbiome composition of very preterm infants from 15 geographically dispersed German hospitals revealed a large variability in the level of detectable bacteria across meconium samples, ranging from high abundance of bacteria to undetectable bacterial loads. Bacterial load was weakly correlated with the time span between birth and sample collection. This is in line with the hypothesis that meconium is sterile in the womb and is colonized postnatally. In low bacterial mass samples, standard 16S rRNA methodology amplified mitochondrial DNA instead, which can be misclassified as Proteobacteria using standard analytical approaches. Through exclusion of these mitochondrial sequences and *in silico* removal of likely contaminants, identified by using negative controls, we produced robust bacterial profiles of meconium. These profiles revealed that meconium shows lower alpha diversity than adult stool and is often dominated by just a few species. Common taxa in meconium include known infant associated but also pathogen-containing genera (*Bifidobacterium*, *Staphylococcus*, *Enterococcus* and *Streptococcus*, respectively). *Bifidobacterium* was negatively correlated with pathogen-containing genera, providing further support for their role as beneficial bacteria. Due to the large range of bacterial load in meconium samples, from essentially absent to abundant, bacterial sequencing effort is easily wasted on many samples. To make progress in this respect, we developed a five-level scoring system based on simple gel electrophoresis ("MecBac" score) that allows identification of promising samples prior to sequencing. This system, as well as the foundational meconium characterization provided in this study, should be helpful for future studies of the meconium microbiome, in particular for higher resolution techniques, such as metagenomic sequencing.

MATERIALS AND METHODS

Sample collection. The samples collected in this study are part of the PRIMAL study. The study protocol describing in detail the infant population, sample collection and storage, exclusion and inclusion criteria and implementation has been published (21). In short, very preterm infants (VPI, gestational age 28–32 weeks) were recruited from 15 German pediatric hospitals. After written consent was obtained from the parents, meconium samples from the infants and fecal samples from the mothers were collected, transported on dry ice in an anaerobic environment, and stored at −80 Celsius until further processing. Maternal stool samples were collected in 4 hospitals (Freiburg, Heidelberg, Homburg, and Lübeck) within the first 7 days after delivery. Meconium samples were collected with a median time of 2 days after birth (25th percentile: 1 day, 75th percentile: 3 days). 57.5% of the infants were male and 42.5% were female. Exactly 50% of infants were born in the gestational weeks 28 to 30 and in >30 to 32, respectively. Meconium samples were analyzed for color and texture and only samples that had typical meconium characteristics (dark color, sticky and tar-like) were considered for this study. Half of the infants received a probiotic supplement (Probactiol infantis) prior to meconium collection, consisting of *Bifidobacterium animalis subsp. lactis*, (BB[1]12), *B. infantis* and *Lactobacillus acidophilus* (La-5) each at $1.5*10^9$colony forming units. The present study represents a pilot analysis of the wider PRIMAL cohort using only the early life (meconium) time points and is still blinded in regard to probiotic supplementation. By the preregistered protocol of the PRIMAL RCT, unblinding the data are scheduled to take place only once endpoints are reached, ensuring methodological rigor and adherence to regulatory requirements. Institutional review board approvals have been obtained at all participating sites.

DNA extraction and 16S rRNA gene sequencing. Positive (N = 5) and negative (N = 118) control samples were used to monitor sample contamination. ZymoBIOMICS Microbial Community DNA Standard was used as a positive control. Every step described below was accompanied by sterile, nuclease-free, DEPC-treated water as negative controls.

DNA was extracted from meconium and maternal stool samples using the DNeasy PowerSoil Pro kit (Qiagen, Germany) including a bead-beating step at 30 Hertz for 2*7 min on a TissueLyser (Qiagen, Germany). The V4 region of the 16S rRNA bacterial genes was targeted for amplification with primers (forward: 5′-GTGCCAGCMGCCGCGGTAA-3′ 515F, reverse: 5′-GACTACHVGGGTWTCTAATCC-3′ 806R) (32). Gel electrophoresis was used to assess the approximate quantity and length of the amplicons. Gel bands of the shorter than expected amplification products (200 bp) were cut out, sequenced and classified as mitochondrial sequences using the BLASTn algorithm (50) with the NCBI nucleotide collection (nt) database.

All products were sequenced on a MiSeq (Illumina, Inc., USA) using V2 chemistry producing 2 × 250 bp paired-end reads. Samples were sequenced over 5 sequencing runs, each with positive and negative controls. To approximate the absolute cell count in one sample, 2 μl and 10 μl ZymoBIOMICS Spike-in Control I (High Microbial Load) were added to two aliquots of the same sample, respectively, and sequenced as separate samples. The spike-in control is composed of a known quantity of DNA from *Imtechella halotolerans* and *Allobacillus halotolerans*. For normal fecal samples, Zymo Research suggests using 20 μl of ZymoBIOMICS™ Spike-in Control I to allow the quantification of bacterial cells in the

sample. This added DNA, usually accounts for less than 10% of the resulting microbiome after analysis (51). Bacterial load estimates were based on PCR band results.

Bioinformatic analysis. Resulting sequencing data were processed using the R statistical environment (version 3.4.4) (52). Two different analyses were performed using the pipeline described below. One analysis in which all sequences were kept, including mitochondrial reads ("total reads"), and one in which all sequences much shorter than the expected amplicon length of 249–254 bp (i.e., mitochondrial sequences) were discarded ("bacterial reads"). Samples with 0 reads were not considered for the following analysis.

The DADA2 package (53) was used to quality filter the reads, trim off the primer sequences and truncate reads to 240 bp. This truncation step was skipped in the "total reads" analysis. Unpaired reads and reads that did not pass quality control were discarded. One error model per sequencing run was trained to infer exact amplicon sequence variants (ASVs). Forward and reverse reads were merged and chimeric sequences removed. ASVs were taxonomically classified using the RDP Naive Bayesian Classifier (54) and the SILVA high quality rRNA database version 138 (55). The ASVs were then decontaminated using the decontam package version 1.1.2 (29). This algorithm statistically assesses if an ASV is likely to be technical contamination considering its presence in negative-control samples and the DNA concentration in the samples, assuming that contaminating sequences have a higher frequency in low-concentration samples. Only ASVs that had a less than 5% chance of being contaminants were kept for further analysis.

Subsequent data handling, processing and visualization was performed in the python programming language (version 3.7.3) (56) with a suite of data science packages (pandas (57), SciPy (58), numpy (59), seaborn (60), matplotlib (61), networkx (62)) provided in the anaconda (version 4.9.2) environment (63). Statistical analysis, including alpha and beta diversity calculations (Shannon diversity index and weighted Unifrac distances) and Principal Coordinate Analysis (PCoA) were performed in R using the phyloseq package version 1.22.3 (64). Permutational multivariate analysis of variance (PERMANOVA) was performed with the vegan package (version 2.5.6) (65) in R using the adonis2 function. Collection dates of meconium that were unusually high ($>$7 days) or not recorded were not used for the multivariate analysis of the influence of age at collection on the microbiome. Differential abundance analysis between groups was performed with the Mann-Whitney U test implemented in SciPy and multiple testing correction was performed using the Benjamini-Hochberg procedure. The co-occurrence network analysis was performed with FlashWeave (66) in the Julia programming language (version 1.6.1).

Data availability. The data sets generated and analyzed during the current study are available in the ENA repository under accession no. PRJEB47767. Metadata, including demographic and clinical data, are available in this article's supplemental material files.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.1 MB.
FIG S2, TIF file, 0.03 MB.
FIG S3, TIF file, 0.3 MB.
TABLE S1, CSV file, 0.1 MB.
TABLE S6, CSV file, 0.4 MB.
TABLE S7, CSV file, 1 MB.
TABLE S8, CSV file, 0.02 MB.
TABLE S9, CSV file, 0.01 MB.

ACKNOWLEDGMENTS

We thank all participating parents and their infants for their cooperation and for providing the necessary biological material. We also thank the clinicians and nurses at the study sites for their support in collecting and coordinating the transport of fecal and meconium samples and all others involved for their support, specifically, Bastian Siller (Lübeck), Thorsten Körner and Birte Tröger (Bremen), Anja Stein (Essen), Hans Fuchs (Freiburg), Lutz Koch (Hamburg), Hannover, Johannes Pöschl (Heidelberg), Christoph Derouet (Homburg), Hans Proquitté and Libor Swoboda (Jena), Friedrich Ebinger and Frank Dohle (Paderborn), Dirk Olbertz (Rostock), Markus Knuf (Wiesbaden), Esther Schmidt (Schwerin), Christian Wieg (Aschaffenburg), Andreas Müller, and Soyhan Bagci (Bonn). This work comprises results from Jonas Klopp's doctoral thesis.

JK, TVR, and SG conceived the study. KH and CM organized and performed the processing of samples. AH, JM, PH, HH, SB, DV, MZ, and CH planned and executed the sample collection at their respective sites. JK performed the bioinformatic analyses with the help of TVR, prepared the figures, and wrote the manuscript. All authors provided critical intellectual input and approved the final manuscript.

This study is supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) grant (01GL1746B). TVR, PF, and PB were supported by the European Molecular Biology Laboratory (EMBL). SKF was supported by the German Research Foundation (DFG; KFO338: "food@") and the Joint Programming Initiative (JPI-HDHL: FATMAL).

REFERENCES

1. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. 2019. The human microbiome and child growth–first 1000 days and beyond. Trends Microbiol 27:131–147. https://doi.org/10.1016/j.tim.2018.09.008.
2. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni H, Metcalf GA, Muzny D, Gibbs RA, Vatanen T, Huttenhower C, Xavier RJ, Rewers M, Hagopian W, Toppari J, Ziegler A-G, She J-X, Akolkar B, Lernmark A, Hyoty H, Vehik K, Krischer JP, Petrosino JF. 2018. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature 562:583–588. https://doi.org/10.1038/s41586-018-0617-x.
3. Blaser MJ, Devkota S, McCoy KD, Relman DA, Yassour M, Young VB. 2021. Lessons learned from the prenatal microbiome controversy. Microbiome 9:8. https://doi.org/10.1186/s40168-020-00946-2.
4. Perez-Muñoz ME, Arrieta M-C, Ramer-Tait AE, Walter J. 2017. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. Microbiome 5:48. https://doi.org/10.1186/s40168-017-0268-4.
5. de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS, Smith GCS. 2019. Human placenta has no microbiome but can contain potential pathogens. Nature 572:329–334. https://doi.org/10.1038/s41586-019-1451-5.
6. Kennedy KM, Gerlach MJ, Adam T, Heimesaat MM, Rossi L, Surette MG, Sloboda DM, Braun T. 2021. Fetal meconium does not have a detectable microbiota before birth. Nat Microbiol 6:865–873. https://doi.org/10.1038/s41564-021-00904-0.
7. Liu C-J, Liang X, Niu Z-Y, Jin Q, Zeng X-Q, Wang W-X, Li M-Y, Chen X-R, Meng H-Y, Shen R, Sun S-Y, Luo Y-Y, Yang E, Geng J-W, Li X-R. 2019. Is the delivery mode a critical factor for the microbial communities in the meconium? EBioMedicine 49:354–363. https://doi.org/10.1016/j.ebiom.2019.10.045.
8. Shi Y-C, Guo H, Chen J, Sun G, Ren R-R, Guo M-Z, Peng L-H, Yang Y-S. 2018. Initial meconium microbiome in Chinese neonates delivered naturally or by cesarean section. Sci Rep 8:3255. https://doi.org/10.1038/s41598-018-21657-7.
9. Wong WSW, Sabu P, Deopujari V, Levy S, Shah AA, Clemency N, Provenzano M, Saadoon R, Munagala A, Baker R, Baveja R, Mueller NT, Dominguez-Bello MG, Huddleston K, Niederhuber JE, Hourigan SK. 2020 Prenatal and peripartum exposure to antibiotics and cesarean section delivery are associated with differences in diversity and composition of the infant meconium microbiome. *Microorganisms 8: 179.*
10. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S, Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2017. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. Microbiol Mol Biol Rev 81:e00036-17. https://doi.org/10.1128/MMBR.00036-17.
11. Stinson LF, Boyce MC, Payne MS, Keelan JA. 2019. The not-so-sterile womb: evidence that the human fetus is exposed to bacteria prior to birth. Front Microbiol 10:1–15.
12. He Q, Kwok L-Y, Xi X, Zhong Z, Ma T, Xu H, Meng H, Zhao F, Zhang H. 2020. The meconium microbiota shares more features with the amniotic fluid microbiota than the maternal fecal and vaginal microbiota. Gut Microbes 12:1794266. https://doi.org/10.1080/19490976.2020.1794266.
13. Wilczyńska P, Skarżyńska E, Lisowska-Myjak B. 2018. Meconium microbiome as a new source of information about long-term health and disease: questions and answers. J Matern Fetal Neonatal Med :1–151.
14. Petersen C, Dai DLY, Boutin RCT, Sbihi H, Sears MR, Moraes TJ, Becker AB, Azad MB, Mandhane PJ, Subbarao P, Turvey SE, Finlay BB. 2021. A rich meconium metabolome in human infants is associated with early-life gut microbiota composition and reduced allergic sensitization. Cell Rep Med 2:100260. https://doi.org/10.1016/j.xcrm.2021.100260.
15. Dong T, Chen T, White RA, Wang X, Hu W, Liang Y, Zhang Y, Lu C, Chen M, Aase H, Xia Y. 2018. Meconium microbiome associates with the development of neonatal jaundice. Clin Transl Gastroenterol 9:e182. https://doi.org/10.1038/s41424-018-0048-x.
16. Morais J, Marques C, Teixeira D, Durão C, Faria A, Brito S, Cardoso M, Macedo I, Pereira E, Tomé T, Calhau C. 2020. Extremely preterm neonates have more *Lactobacillus* in meconium than very preterm neonates – the *in utero* microbial colonization hypothesis. Gut Microbes 12:1785804. https://doi.org/10.1080/19490976.2020.1785804.
17. Terrazzan Nutricionist AC, Procianoy RS, Roesch LFW, Corso AL, Dobbler PT, Silveira RC. 2020. Meconium microbiome and its relation to neonatal growth and head circumference catch-up in preterm infants. PLoS ONE-Plos ONE 15:e0238632. https://doi.org/10.1371/journal.pone.0238632.
18. Yang J, Yao S, Cheng K, Xu L, Hou L, Wei Y, Feng H, Yu X, Zhang X, Tong X, Li Z, Zhao Y. 2020. Comparison of meconium microbiome in dizygotic and monozygotic twins born by caesarean section (CS). Front Microbiol 11:1139. https://doi.org/10.3389/fmicb.2020.01139.
19. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol 27:105–117. https://doi.org/10.1016/j.tim.2018.11.003.
20. Hornung BVH, Zwittink RD, Kuijper EJ. 2019. Issues and current standards of controls in microbiome research. FEMS Microbiol Ecol 95. https://doi.org/10.1093/femsec/fiz045.
21. Marißen J, Haiß A, Meyer C, Van Rossum T, Bünte LM, Frommhold D, Gille C, Goedicke-Fritz S, Göpel W, Hudalla H, Pagel J, Pirr S, Siller B, Viemann D, Vens M, König I, Herting E, Zemlin M, Gehring S, Bork P, Henneke P, Härtel C, PRIMAL consortium. 2019. Efficacy of *Bifidobacterium longum, B. infantis and Lactobacillus acidophilus* probiotics to prevent gut dysbiosis in preterm infants of 28+0–32+6 weeks of gestation: a randomised, placebo-controlled, double-blind, multicentre trial: the PRIMAL Clinical Study protocol. BMJ Open 9:e032617. https://doi.org/10.1136/bmjopen-2019-032617.
22. Eperon IC, Anderson S, Nierlich DP. 1980. Distinctive sequence of human mitochondrial ribosomal RNA genes. Nature 286:460–467. https://doi.org/10.1038/286460a0.
23. Douglas CA, Ivey KL, Papanicolas LE, Best KP, Muhlhausler BS, Rogers GB. 2020. DNA extraction approaches substantially influence the assessment of the human breast milk microbiome. Sci Rep 10:123. https://doi.org/10.1038/s41598-019-55568-y.
24. Dornelles LV, Procianoy RS, Roesch LFW, Corso AL, Dobbler PT, Mai V, Silveira RC. 2020. Meconium microbiota predicts clinical early-onset neonatal sepsis in preterm neonates. J Matern Fetal Neonatal Med :1–9. https://doi.org/10.1080/14767058.2020.1774870.
25. Heida FH, van Zoonen AGJF, Hulscher JBF, Te Kiefte BJC, Wessels R, Kooi EMW, Bos AF, Harmsen HJM, de Goffau MC. 2016. A necrotizing enterocolitis-associated gut microbiota is present in the meconium: results of a prospective study. Clin Infect Dis 62:863–870. https://doi.org/10.1093/cid/ciw016.
26. Cuna A, Morowitz MJ, Ahmed I, Umar S, Sampath V. 2021. Dynamics of the preterm gut microbiome in health and disease. Am J Physiol Gastrointest Liver Physiol 320:G411–G419. https://doi.org/10.1152/ajpgi.00399.2020.
27. Bekkali N, Hamers SL, Schipperus MR, Reitsma JB, Valerio PG, Van Toledo L, Benninga MA. 2008. Duration of meconium passage in preterm and term infants. Arch Dis Child - Fetal Neonatal Ed 93:F376–F379. https://doi.org/10.1136/adc.2008.138024.
28. Underwood MA, Mukhopadhyay S, Lakshminrusimha S, Bevins CL. 2020. Neonatal intestinal dysbiosis. J Perinatol 40:1597–1608. https://doi.org/10.1038/s41372-020-00829-2.
29. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6:226. https://doi.org/10.1186/s40168-018-0605-2.

30. Hansen R, Scott KP, Khan S, Martin JC, Berry SH, Stevenson M, Okpapi A, Munro MJ, Hold GL. 2015. First-pass meconium samples from healthy term vaginally-delivered neonates: an analysis of the microbiota. PLoS One PLoS One 10:e0133320. https://doi.org/10.1371/journal.pone.0133320.
31. Siena MD, Laterza L, Matteo MV, Mignini I, Schepis T, Rizzatti G, Ianiro G, Rinninella E, Cintoni M, Gasbarrini A. 2021. Gut and reproductive tract microbiota adaptation during pregnancy: new insights for pregnancy-related complications and therapy 15.
32. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108:4516–4522. https://doi.org/10.1073/pnas.1000080107.
33. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R. 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. msystems 1:sys0029. https://doi.org/10.1128/mSystems.00009-15.
34. Lackey KA, Williams JE, Meehan CL, Zachek JA, Benda ED, Price WJ, Foster JA, Sellen DW, Kamau-Mbuthia EW, Kamundia EW, Mbugua S, Moore SE, Prentice AM, K DG, Kvist LJ, Otoo GE, García-Carral C, Jiménez E, Ruiz L, Rodríguez JM, Pareja RG, Bode L, McGuire MA, McGuire MK. 2019. What's normal? microbiomes in human milk and infant feces are related to each other but vary geographically: the INSPIRE study. Front Nutr 6:45. https://doi.org/10.3389/fnut.2019.00045.
35. Stinson LF, Keelan JA, Payne MS. 2018. Comparison of meconium DNA extraction methods for use in microbiome studies. Front Microbiol 9:1–14.
36. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. 2012. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. Science 336:1255–1262. https://doi.org/10.1126/science.1224203.
37. Nagpal R, Tsuji H, Takahashi T, Kawashima K, Nagata S, Nomoto K, Yamashiro Y. 2016. Sensitive quantitative analysis of the meconium bacterial microbiota in healthy term infants born vaginally or by Cesarean section. Front Microbiol 7. https://doi.org/10.3389/fmicb.2016.01997.
38. Tauchi H, Yahagi K, Yamauchi T, Hara T, Yamaoka R, Tsukuda N, Watanabe Y, Tajima S, Ochi F, Iwata H, Ohta M, Ishii E, Matsumoto S, Matsuki T. 2019. Gut microbiota development of preterm infants hospitalised in intensive care units. Benef Microbes 10:641–651. https://doi.org/10.3920/BM2019.0003.
39. Chernikova DA, Madan JC, Housman ML, Zain-Ul-Abideen M, Lundgren SN, Morrison HG, Sogin ML, Williams SM, Moore JH, Karagas MR, Hoen AG. 2018. The premature infant gut microbiome during the first 6 weeks of life differs based on gestational maturity at birth. Pediatr Res 84:71–79. https://doi.org/10.1038/s41390-018-0022-z.
40. Moles L, Gómez M, Heilig H, Bustos G, Fuentes S, de Vos W, Fernández L, Rodríguez JM, Jiménez E. 2013. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. PLoS One 8:e66986. https://doi.org/10.1371/journal.pone.0066986.
41. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol 79:5112–5120. https://doi.org/10.1128/AEM.01043-13.
42. Turroni F, Milani C, Duranti S, Ferrario C, Lugli GA, Mancabelli L, van Sinderen D, Ventura M. 2018. Bifidobacteria and the infant gut: an example of co-evolution and natural selection. Cell Mol Life Sci 75:103–118. https://doi.org/10.1007/s00018-017-2672-0.
43. Dermyshi E, Wang Y, Yan C, Hong W, Qiu G, Gong X, Zhang T. 2017. The "golden age" of probiotics: a systematic review and meta-analysis of randomized and observational studies in preterm infants. Neonatology 112:9–23. https://doi.org/10.1159/000454668.
44. van Zyl WF, Deane SM, Dicks LMT. 2020. Molecular insights into probiotic mechanisms of action employed against intestinal pathogenic bacteria. Gut Microbes 12:1831339. https://doi.org/10.1080/19490976.2020.1831339.
45. Luo J, Li Y, Xie J, Gao L, Liu L, Ou S, Chen L, Peng X. 2018. The primary biological network of Bifidobacterium in the gut. FEMS Microbiol Lett 365. https://doi.org/10.1093/femsle/fny057.
46. Korpela K, Renko M, Paalanne N, Vänni P, Salo J, Tejesvi M, Koivusaari P, Pokka T, Kaukola T, Pirttilä AM, Tapiainen T. 2020. Microbiome of the first stool after birth and infantile colic. Pediatr Res 88:776–783. https://doi.org/10.1038/s41390-020-0804-y.
47. Moore RE, Townsend SD. 2019. Temporal development of the infant gut microbiome. Open Biol 9:190128. https://doi.org/10.1098/rsob.190128.
48. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo M, Beghini F, Bertorelli R, De Sanctis V, Bariletti I, Canto R, Clementi R, Cologna M, Crifò T, Cusumano G, Gottardi S, Innamorati C, Masè C, Postai D, Savoi D, Duranti S, Lugli GA, Mancabelli L, Turroni F, Ferrario C, Milani C, Mangifesta M, Anzalone R, Viappiani A, Yassour M, Vlamakis H, Xavier R, Collado CM, Koren O, Tateo S, Soffiati M, Pedrotti A, Ventura M, Huttenhower C, Bork P, Segata N. 2018. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. Cell Host Microbe 24:133–145. https://doi.org/10.1016/j.chom.2018.06.005.
49. Maqsood R, Rodgers R, Rodriguez C, Handley SA, Ndao IM, Tarr PI, Warner BB, Lim ES, Holtz LR. 2019. Discordant transmission of bacteria and viruses from mothers to babies at birth. Microbiome 7:156. https://doi.org/10.1186/s40168-019-0766-7.
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
51. Zymo Research. 2020. ZymoBIOMICS spike-in control i (high microbial load) protocol, v1.1.5. Zymo Research Group, Irvine, CA. https://files.zymoresearch.com/protocols/_d6320_zymobiomics_spike-in_control_i.pdf.
52. R Core Team. 2018. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
53. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
54. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267. https://doi.org/10.1128/AEM.00062-07.
55. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.
56. van Rossum G. 1995. Python Programming language. Python, Amsterdam.
57. The pandas development team. 2020. pandas-dev/pandas: pandas. Zenodo.
58. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0 Contributors. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2.
59. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020. Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2.
60. Waskom M. 2021. seaborn: statistical data visualization. JOSS 6:3021. https://doi.org/10.21105/joss.03021.
61. Hunter JD. 2007. Matplotlib: a 2D graphics environment. Comput Sci Eng Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55.
62. Hagberg AA, Schult DA, Swart PJ. 2008. Exploring network structure. Dynamics, and Function uUsing NetworkX 5
63. Anaconda Inc. 2021 Anaconda software distribution. Anaconda, Inc., Austin, TX. https://anaconda.org/.
64. McMurdie PJ, Holmes S. 2013. phyloseq: an R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.
65. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.
66. Tackmann J. 2019. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data 20.

Electronic supplement

- **Digital version of this thesis**
- **Algorithms and scripts**
  - ***Data processing***
    - check_metadata_sampleIds_for_redundancy.py
    - shared_ASVs_parallel.py
    - combine_runs_comparison.py
    - subset_asv_tables_positive_control.py
    - add_readcounts_and_run_number_to_metadata.py
    - subset_asv_tables_for_Flashweave_args.py
    - get_mito_percent_in_meconium_samples.py
    - visulalize_network.py
    - annotate_gml_FlashWeave_network.py
    - FlashWeave_pipeline_args.jl
    - unzip_all_files_in_dir.py
    - phylogenetic_assingment_of_ASV_tables_positive_control.py
    - DADA2_without_tree_argu.R
    - FlashWeave_pipeline.sh
    - total_reads_gz_in_Miseq_run
    - Phylogenetic_Tree_PRIMAL_References.R
    - add_KRINKO_to_metadata.py
    - rename_numbers_to_samplenames.py
    - readcount_check_for_samples_in_run.py
    - excel_metadata_to_complemented_csv.py
    - subset_asv_tables_for_Flashweave_args (copy).py
    - ARG_taxa_correlation.py
  - ***Data visualization***
    - Mikrobiologie_vs_16s_visualisierung.py

- Beta_diversity_plots.R
- Mitochondria_relative_amount_across_sampletypes.py
- sequencing_run_statistics.py
- beta_diversity_between_samples.py
- Most_abundant_genera_across_hospitals_legend.py
- phylogenetic_origin_of_ARGs.py
- Most_abundant_genera_across_sample_types statistics.py
- antibiotic_restistance_analysis.py
- antibiotic_restistance_for_phylogenetic_level_analysis.py
- dysbioitc_statistics.py
- stacked_barplot_spike-in.py
- percent_real_meconium.py
- db_overview.py
- Resistome_Machine_Learning_Random_Forest.py
- antibiotic_phylogeny.R
- dada_stats.py
- correlation_pc1_and_bifidobacterium_in_visit2.py
- Most_abundant_phyla_across_sample_types statistics.py
- stacked_barplot_positive_control.py
- ASV_rarefied_distribution_statistics.py
- Most_abundant_genera_across_hospitals.py
- Probiotic_taxa_distribution.py
- alpha_diversity_by_KRINKO.py
- PRIMAL_treatment_differences.py
- Most_abundant_genera_across_sample_types_legend.py
- Most_abundant_genera_across_sample_types.py
- upset_plot.py
- Beta_diversity_plots_reference_studies.R

- alphadiversity_per_sample_type.py
- Taxonomic_plots_phylum_level.R
- PERMANOVA.R
- relative_abundance_color_scheme.txt

Declaration

I hereby declare that I have written the present thesis independently and without use of other than the indicated means. I also declare that to the best of my knowledge all passages taken from published and unpublished sources have been referenced. The thesis has not been submitted for evaluation to any other examining authority.

Mainz, Janurary 13, 2022

Jonas Klopp