Norwegian University of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Associations of gut microbiota- and short-chain fatty acid composition during the first year of life with immune cells in one-year children

Unni Lise Albtersdóttir Jonsmoen
MSc Biotechnology

Associations of gut microbiota- and short-chain fatty acid composition during the first year of life with immune cells in one-year children

Norwegian University of Life Sciences (NMBU),
Faculty of Chemistry, Biotechnology and Food Science

Acknowledgments

This thesis was performed at the Faculty of Chemistry, Biotechnology, and Food Sciences, at the Norwegian University of Life Sciences, under the supervision of Professor Knut Rudi and Ph.D. Student Morten Nilsen.

First of all, I would like to thank my supervisor Knut Rudi for taking me on as a student, introducing me to the project, and for sharing his knowledge and good advice whenever I ran into challenges. I also would like to thank my secondary supervisor Morten Nilsen, his joy for the field, and willingness to help others is admirable. Without all the hours of Nilsen's help troubleshooting the gas chromatograph, this thesis would not be possible. Rudi and Nilsen have both my deepest gratitude.

I would like to thank the entire PreventADALL project, under the lead of Karin C. Lødrup. Especially thanks to Alex Olin and Petter Brodin, at the Karolinska Institute, for sharing the immune cell profile data used in this thesis.

Thanks to the laboratory engineers Ida Ormaasen and Inga Lena Angell for all the helping hands and many laughs. Thanks to the master students Cecilie Fredheim, Fredrik Johansen, Mari Raudstein, and Regina Sørensen for the good talks and support throughout this year. I wish you all continued success. You and the rest of MiDiv have made this year truly memorable.

Lastly, thanks to my family and friends for encouragement and support, not only through this year but throughout my studies.

Ås, 2020
Unni Lise Albertsdóttir Jonsmoen

Sammendrag

Mikroorganismene i tarmen etablerer komplekse mutualistiske forhold til verten og spiller en viktig rolle i modningen av vertens immunsystem. Modningsprosessen påvirkes enten direkte av bakterielle komponenter eller indirekte gjennom bakterienes fermenteringsprodukter slik som korkjedede fettsyrer. Med dette som grunnlag utforsker denne studien den langsgående utviklingen av tarmmikrobiotaen og nivået av kortkjedede fettsyrer hos barn gjennom deres første leveår, og ser det i sammenheng med barnas immuncelleprofiler ved ett års stadiet.

Fekale prøver fra 180 12 måneder gamle barn ble hentet fra studien Prevent Atopic Dermatitis and Allergies (PreventADALL). Immuncellekomposisjonsdata ble mottatt for 67 av barna og disse barna ble studert langsgående ved analyse av fekale prøver hentet fra barnas mødre i svangerskapets 18. uke, og fra barna ved 0-, 3-, 6- og 12 måneder. Bakterie- og kortkjedet fettsyresammensetning ble bestemt ved henholdsvis nestegenerasjonssekvensering og gasskromatografi. Informasjonen om immuncellesammensettingen ved 12 måneder ble brukt videre i den statistiske analysen. Bakteriedataene ble behandlet gjennom QIIME-pipeline. Videre dataanalyse ble utført ved paret t-tester og Spearman korrelasjonsanalyse justert for multippel testing.

Studien avdekket sammenhenger mellom immuncelleprofiler og miljøassosierte bakterier. Funnene viste at både Methylophiliales og Methylococcales påvirket vertens immuncellekomposisjon, på en henholdsvis pro- og anti-inflammatorisk måte. Det ble ikke avdekket korrelasjoner mellom kortkjedede fettsyrer og immunceller, noe som hadde vært forventet utfra dyrestudier. Som i tidligere studier, viste den mikrobielle sammensetningen at Enterobacteriales dominerte i barnas første avføring, mens Clostridiales dominerte i mødrenes tarmer. Mengden eddiksyre endret seg i samsvar med mengde eddiksyre-produserende Bifidobactierales ved alle aldere. Positive korrelasjoner ble påvist mellom Lactobacillales og smørsyre ved 3-, 6- og 12 måneder, noe som kan skyldes bakterielle kryssfôringsmekanismer.

Oppsummert avdekket denne eksplorative studien korrelasjoner mellom enkelte immunceller og tarmbakterier. Miljøassosierte bakterier påvirket vertens immunsystem på ulike pro- og anti-inflammatoriske måter. Det er et behov for videre forskning på forbindelsen mellom tarmmikrobiotaen og immunsystemet i friske mennesker, og denne studien taler for at miljømikroorganismer i større grad bør vektlegges i slik forskning.

Abstract

The microorganisms in the gut establish complex mutualistic relationships with their human host and serve important functions in the maturation of the host's immune system. The immune system maturation process is affected both by bacterial components directly and through their fermentative products, such as the short-chain fatty acids. This thesis aimed to explore the longitudinal development of the infant gut microbiota and short-chain fatty acid (SCFAs) concentrations in relation to the immune cell status at 12 months.

Fecal samples from 180 12 months-old infants were retrieved from the Prevent Atopic Dermatitis and Allergies (PreventADALL) study cohort. Data on immune cell profiles were obtained for 67 of the infants, and these infants were studied longitudinally with fecal samples from the mothers 18-weeks pregnant, and the infants at 0-, 3-,6- and 12 months were analyzed. The bacterial and SCFA composition was determined by next-generation sequencing, and gas chromatography, respectively. The immune cell data was used in statistical analysis. The bacterial data was processed using the QIIME pipeline, and further statistical analysis was conducted using paired t-tests and Spearman rank's correlation adjusted for false discoveries.

This study revealed interesting correlations between immune cell profiles and environmental associated bacteria. The orders of Methylophiliales and Methylococcales both affected the immune cell profiles in pro- and anti-inflammatory manner, respectively. No correlations between SCFAs and immune cells were detected, which would have been expected based on animal studies. With respect to the general age-related development of the microbiota did the findings of this study correspond to previous studies. Enterobacteriales dominated the meconium samples and Clostridiales dominated in the mothers. Acetate levels coincided with the abundance of the acetogenic Bifodobacteriales in all age groups. Positive correlations were detected between Lactobacilliales and butyric acid at 3-, 6- and 12 months, which was likely caused by bacterial cross-feeding mechanisms.

This exploratory study reviled correlations between immune cells and gut bacteria. Environmental associated bacteria affected the immune system and influenced both pro- and anti-inflammatory associated responses. There is a need for further research on the connection between the gut microbiota and the host's immune system in healthy humans, and this study notes the importance of including environmental associated bacteria in this research.

Table of contents

List of Figures and Tables

Abbreviations

BH – Benjamini-Hockberg
CyTOF – Cytometry by Time-Of-Flight
DNA – Deoxyribonucleic Acid
GALT – Gut Associated Lymphoid Tissues
GC – Gas Chromatography
IBD – Inflammatory Bowel Diseases
Ig – Immunoglobulin
NGS – Next-Generation Sequencing
PCoA - Principal Coordinates Analysis
PCR – Polymerase Chain Reaction
SCFAs – Short-Chain Fatty Acids.
dNTP – Deoxyribonucleotide Triphosphate
ddPCR – Digital Droplet PCR
qPCR – Quantitative PCR
rRNA – Ribosomal Ribonucleic Acid
ssDNA – Single-Stranded

1. Introduction

1.1 Human Gut Microbiota

The gut harbors the largest collection of microorganisms that lives inside or on humans with 10^{11} – 10^{12} microbes/ml luminal content (Palmer, Bik, DiGiulio, Relman, & Brown, 2007; Riviere, Selak, Lantin, Leroy, & De Vuyst, 2016). The microorganisms establish complex relationships with their human host as well as with each other, and the relationships are ranging from symbiotic to parasitic (Milani et al., 2017). The intestinal habitat of an individual is estimated to contain at least 150 bacterial species (Qin et al., 2010). However, representatives from all three domains of life are found in the lumen, not to mention viruses (Milani et al., 2017). Collectively, all these microorganisms are defined as the human gut microbiota. The microbiota plays a significant role in human metabolism, contributing to 100-fold more genes than the human genome (Rook, Raison, & Lowry, 2014). The early colonized microbiota plays a vital role in the maturation of metabolic and development pathways, and emerging evidence states that disruptions in obtaining an optimal microbiota may lead to an increased lifelong risk of various diseases and disorders (S. Wang et al., 2020).

1.1.1 Developmental stages of the gut microbiota

The gut microbiome in early life is characterized by massive changes in microbial richness, diversity, and functions. However, there is no current general standard for the development patterns in the gut microbiota, despite the development being highly stage-specific (de Muinck & Trosvik, 2018; S. Wang et al., 2020). Historically, the gut microbiota has been studied and described based on culture-depended methods (Cooperstock, 1983), and in 1983 four stages of the infant gut microbiota development were described. The first phase begins with the initial colonization, followed by the second stage – the period of exclusive breastfeeding. The third stage occurs as the infant starts receiving dietary supplements, called weaning, and lasts until the cessation of breastfeeding. The fourth and final stage extends from the completion of weaning until the conversion of an adult-like microbiota community (Cooperstock, 1983; S. Wang et al., 2020), which is generally around 2.5 to 3 years of age (Milani et al., 2017). In later years, with the accessibility of culture-independent methods, next-generation sequencing, and more extensive study cohorts, predictable patterns are beginning to emerge. A study by Stewart et al. observed three distinct phases in the microbiome development in children sampled from

the age of 3 months to 46 months (Stewart et al., 2018). The first phase (months 3 to 14), called the developmental phase, was characterized by five phyla, Actinobacterium, Firmicutes, Verrucomicrobia, Proteobacteria, and Bacteroidetes, significantly changing. During the transitional phase (months 15 to 30), experienced the phyla of Proteobacteria and Bacteroidetes significant changes. From 31 months onward, reached the microbiota a stable phase where all phyla were unchanged. Most microbiome profiles at this stage were characterized by high alpha diversity and dominated by the Firmicutes phyla (Stewart et al., 2018).

1.1.2 **The initial colonization of the neonatal gut**

The exact timing when the colonization process starts remains a controversial topic. The consensus has been that the gut is sterile until the rupture of the amniotic membrane (Perez-Munoz, Arrieta, Ramer-Tait, & Walter, 2017; Rehbinder et al., 2018; Rodriguez et al., 2015). However, in recent years this consensus has been challenged by studies suggesting that the placenta (Aagaard et al., 2014) and amniotic fluid have a microbiota of their own (Collado, Rautava, Aakko, Isolauri, & Salminen, 2016). Most evidence supporting the sterile womb hypothesis was generated with traditional methods like microscopy and culture-based techniques, which today are not considered efficient for assessing a microbiome (Perez-Munoz et al., 2017). There are also few studies conducted with the sole purpose of assessing the microbiology of the amniotic fluid from healthy pregnancies delivered at term. However, the findings of *in utero* microbiota using polymerase chain reactions and next-generation sequencing methods have been criticized as those approaches do not have the detection limits necessary to study low-density bacterial populations, and the methods are susceptible to contamination (Perez-Munoz et al., 2017). There is currently no clear consensus on when the colonization process starts, and studies with larger sample sizes are needed (S. Wang et al., 2020).

The discussion of the first colonization aside, the neonates' first major exposure to microbes happens as the amniotic membrane ruptures. The complexity of the microbial community in the gut is low and typically dominated by one phylum. Because the gut is at first an aerobic environment, it is inhabited by facultative anaerobic bacteria such as *Escherichia, Streptococcus, Enterococcus,* and *Klebsiella.* These facultative anaerobic microbes are responsible for creating a reduced and oxygen-depleted environment that is favorable for the strict anaerobic microbes (Backhed et al., 2015; S. Wang et al., 2020). During a normal vaginal birth, the infant is in contact with the maternal vaginal and fecal microbiota, and the first

colonizers derive from vagina-associated microbes, e.g., *Lactobacillus* and *Prevotella* (Milani et al., 2017; Rodriguez et al., 2015). The maternal vaginal species are only detected during the first few days before they eventually are lost or undetectable by one-week post-birth (S. Wang et al., 2020).

By the time the gut lumen is depleted of oxygen, the obligate anaerobic species start their colonization process, leading to a decrease in facultative anaerobes (Avershina et al., 2016). How these late obligate anaerobic colonizers are recruited is a debated subject, but there are two main hypotheses. The first one states that bacteria are vertically transmitted from mother to child during early life and kept in low abundant populations waiting for the right condition (Avershina et al., 2016). The second recruitment hypothesis states that the obligate anaerobic bacteria are acquired later in life through spores or other forms that survive in the aerobic environment outside the gut lumen (Avershina et al., 2016).

1.1.3 **Stabilization of the fluctuating infant microbiota**

Following the recruitment of the first anaerobic species, the fluctuating gut microbiota goes through bacterial shifts before stabilizing with an adult-like microbiome at 2 to 3.5 years of age (Milani et al., 2017). By one month of life, the strict anaerobic species are fully emerged, especially *Bifidobacterium* and *Bacteroides.* The phylum Actinobacteria continues to be dominating during the next few months, together with representatives from Firmicutes phylum (Azad et al., 2013; Backhed et al., 2015; S. Wang et al., 2020), while *Bacteroides* loses some of its prevalence. The bacterial shift is due to the pro- and antimicrobial agents in breast milk that favors a so-called milk-oriented microbiota. Particularly, the human milk oligosaccharides (HMOs) have been found to selectively shape the growth and function of the gut microbiota, as well as to modulate the intestinal epithelial cells' immune response. *Bifidobacterium* and *Lactobacillus* are especially associated with the milk-oriented microbiota as they utilize the complex oligosaccharides (Milani et al., 2017; S. Wang et al., 2020). There is a reduction in the abundance of *Bifidobacterium*, *Clostridium*, and *Enterobacter* after 6 months, corresponding to the cessation of breastfeeding and the increase of solid food in the infant's diet. At the same time starts representatives from the Bacteroides phylum to reemerge. With the introduction of solid food, access to novel substrates leads to a shift of the dominating phyla in the gut from Proteobacteria and Actinobacteria to Firmicutes and Bacteroidetes, as well as an increase in alpha diversity (Milani et al., 2017). The observed increase in alpha diversity and a simultaneous decrease of beta diversity suggest that the microbiota becomes more complex and

less dissimilar with the microbiota of other infants at the same age (Backhed et al., 2015). The increase of microbial diversity contributes to an increase in functional capacity as the gut microbiota is able to degrade polysaccharides, and an increase in production of short-chain fatty acids (SCFAs) is observed (Backhed et al., 2015; Milani et al., 2017). However, the functional changes are not very apparent until the infant stops breastfeeding, coinciding with the reduction of the heavily dominating Bifidobacteria class (Avershina et al., 2016). At 12 months, species belonging to the phyla Firmicutes and the class of Clostridia, like the *Ruminococcus* and *Roseburia* genera, are dominating (Backhed et al., 2015; S. Wang et al., 2020).

1.1.4 **Factors influencing the microbiome development**

Several factors interfere with the development of the gut microbiota, including gestational age at birth, mode of delivery, infant's diet, and medical interventions. Premature infants, born prior to 37 full weeks of gestation, tend to exhibit delayed gut colonization of commensal anaerobic microbes. Their feces contain higher levels of opportunistic pathogenic such as *Enterobacteriaceae* and *Enterococcus*, than fecal samples from full-term neonates (Milani et al., 2017). Premature infants are often born with immune, respiratory and neurological issues that need medical treatment (Milani et al., 2017). For cesarean section (C-section) delivered infants may deviations in the microbiome development occur, as they are not directly presented the same maternal microbial niches as vaginally delivered infants. Instead, they are more likely to be colonized by microorganisms from the surrounding environment, such as microbes from the maternal skin and hospital environment. While the diversity of the microbiota increases over the first month of a vaginally delivered infants life, the changes in the microbiota of a C-section delivered infant are limited. Infants delivered through C-section hold a microbial signature characterized the phyla Firmicutes, Proteobacteria, and Actinobacteria (Milani et al., 2017; Rodriguez et al., 2015), as well as delayed Bacteroides levels and low intra-individual diversity when compared to vaginally delivered infants (S. Wang et al., 2020). C-section delivery has been associated with an increased risk of immune disorders such as asthma, allergy, and type 1 diabetes (Milani et al., 2017).

Formula-fed infants are exposed to different nutrients and carbohydrates than breastfed infants, which causes different colonization patterns (Milani et al., 2017). Formula-fed infants have a more diverse microbiota dominated by *Staphylococcus, Bacteroides,* Clostridia, and *Enterobacteria,* compared to infants receiving breast milk. Formula-fed infants also have a microbiota with higher potential pathogens (Milani et al., 2017).

Lastly, medical interventions may interfere with the natural development and cause deviations in the establishment of the infant gut microbiota (Milani et al., 2017). Infants born preterm, by C-section, or with a low birth weight are at high risk for antibiotic-associated perturbations, as they receive antibiotic therapy at birth (Milani et al., 2017). Antibiotics are widely used to remove or prevent bacterial colonization in the human body, without targeting specific types of bacteria, and board-spectrum antibiotics may have substantial effects on the gut microbiota composition (Angelucci, Cechova, Amlerova, & Hort, 2019). This bacterial removal reduces biodiversity, delays the colonization process for a long time after administration, and disrupts the infant's metabolic and immune development (Angelucci et al., 2019; Gibson, Crofts, & Dantas, 2015). In addition to diversity decrease, there is a decline of Bifidobacterial and *Bacteroidetes sp.,* increasing the relative abundance of *Enterobacteria* and *Enterococci.* The use of antibiotics further leads to a potential enrichment of the reservoir of antibiotic resistance genes available for pathogens, the so-called resistome (Gibson et al., 2015).

1.1.5 **The functional adult gut microbiota**

The microbiota in the gut provides crucial functions in human health and affects the host via different host-microbiota pathways (von Martels et al., 2017). It is possible to divide the beneficial effects of the gut microbiome into three functional categories; metabolic, trophic, and protective (Guarner & Malagelada, 2003). The bacterial diversity provides a diverse set of genes, enriching the gut with bacterial pathways distinct from the hosts. The microbes take advantage of these pathways and can ferment complex non-digestible carbohydrates or other dietary residues that escape digestion and absorption in the small intestine. The anaerobic fermentation of carbohydrates leads to the production of host-beneficial compounds such as short-chain fatty acids and vitamins B and K (LeBlanc et al., 2017; von Martels et al., 2017). With the depletion of carbohydrates in the lumen, the gut microbiota switches to other substrates, notably proteins and amino acids, which in addition to the production of SCFA, also generates possibly toxic substrates as ammonia, amine, phenols, thiols, and indols (Marchesi et al., 2016). These toxic compounds play a role in diseases such as inflammatory bowel diseases (IBD) and colon cancer (Marchesi et al., 2016). The trophic effect of the microbes asserts itself in the stimulation factors that the bacterial fermentative products have on epithelial cell growth and differentiation (Guarner & Malagelada, 2003). Through the bacterial products, the healthy gut microbiome plays an essential role in the development of a balanced and competent immune system (von Martels et al., 2017). Lastly, the resident bacteria in the gut lumen provides a

protective line of resistance to exogenous colonizers by means of direct competition and competition over nutrients (Guarner & Malagelada, 2003; Modi, Collins, & Relman, 2014).

The adult gut microbiota is seen as a two phyla system with a high abundance of Firmicutes and Bacteroidetes (Marchesi et al., 2016), followed by Actinobacteria (Rodriguez et al., 2015). Though, members of over 10 different phyla are found to hold important functional contributions (Marchesi et al., 2016). The gut microbiota is dynamic, nonetheless stable around an equilibrium state. The community is resilient to small changes in the environment, however larger environmental changes will lead to equilibrium shifts (Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012). In many cases, are shifts in the equilibrium state associated with dysbiosis, a term used to describe a disturbance in the balance between beneficial bacteria with anti-inflammatory properties and bacteria with proinflammatory properties. Alteration in the gut microbiota and dysbiosis is linked to gut-related disorders, such as IBD, Crohn's disease, and ulcerative colitis (von Martels et al., 2017). Dysbiosis and a lower degree of diversity have been reported in individuals struggling with obesity (Turnbaugh et al., 2009).

1.2 Short-chain fatty acids

Short-chain fatty acids (SCFAs) are branched or unbranched organic fatty acids with backbones between 1 to 6 carbon atoms in length. As previously mentioned, the SCFAs are the principal anions that arise as byproducts from bacterial fermentation of carbohydrates, proteins, and glycoproteins that escapes digestion and absorption in the small intestine (Y. Sun & O'Riordan, 2013; Wong, de Souza, Kendall, Emam, & Jenkins, 2006). The SCFAs are then absorbed by colonic epithelial cells, where they are oxidized for energy, and the SCFAs that escapes oxidation enter the hepatic portal blood (Barrett, 2014; Henningsson, Björck, & Nyman, 2001; Wong et al., 2006).

1.2.1 Production of SCFAs

Multiple pathways lead to the formation of SCFAs performed by the diverse members of the gut microbiota. Most of the microorganisms in the gut are saccharolytic, meaning they utilize carbohydrates as substrate, forming SCFAs and H_2-, CO_2- and CH_4-gas as byproducts of their bacterial fermentation (Henningsson et al., 2001; Wong et al., 2006). Firstly, primary fermenters, e.g., *Bacteroides,* ferment mono- and oligosaccharides and generate SCFAs, such

as lactic acid and acetic acid, which are then released into the environment. These SCFAs are subsequently utilized by secondary fermenters to generate additional SCFAs. Besides, acetogenes deplete the hydrogen that is released from the reactions together with carbon dioxide to form acetate and thereby contributing to the intestinal SCFA content (Y. Sun & O'Riordan, 2013). The most abundant SCFAs in the gut lumen are acetic acid, propionic acid, and butyric acid, and are typically found in the ratio 60:20:20 (Cummings, 1981). Acetate, being the most abundant, is produced by most enteric bacteria from acetyl-CoA derived from glycolysis (Parada Venegas et al., 2019). Bacteria belonging to the Bacteroidetes phyla are responsible for most intestinal propionate production, which is mainly produced through the carbohydrate metabolism in glycolysis, but also from the organic acid- and amino acid metabolism. Butyrate is predominantly produced by bacteria belonging to the phylum of Firmicutes. However, sugar and lactate-utilizing bacteria may produce butyrate from lactate and acetate, such as *Eubacterium hallii* and *Anaerostipes spp*. (Parada Venegas et al., 2019).

1.2.2 **Absorption and function of SCFAs**

About 90% of the SCFAs are absorbed in the colon, and the rest is excreted in the feces (Garcia-Villalba et al., 2012; Henningsson et al., 2001). Two mechanisms for the absorption of the SCFAs have been proposed. The first is by passive diffusion of protonated SCFAs over the cell membrane, and the second is absorption by anion exchange, through the proton-coupled monocarboxylate-transporter 1 and sodium-coupled monocarboxylate-transporter 1 (Parada Venegas et al., 2019; Wong et al., 2006). Once absorbed, the SCFAs are sent to three different body sites for metabolization; the colonic epithelial cells, the liver cells, and periphery cells (Wong et al., 2006). Propionate and acetate found in peripheral blood are metabolized by the cells in the liver or by peripheral tissues, such as muscle tissues. They also play a role as modulators of glucose metabolism. In rats, propionate was found to inhibit gluconeogenesis and increase glycolysis in hepatocytes (Henningsson et al., 2001). Acetic acid is transferred into acetyl-CoA in the liver, further to act as a precursor for lipogenesis, and stimulates gluconeogenesis (Henningsson et al., 2001). Butyrate is, however, the preferred source of energy for the colonocytes (Pryde, Duncan, Hold, Stewart, & Flint, 2002), even favored over the glucose and glutamine supplied by the blood (Henningsson et al., 2001; Wong et al., 2006). In addition to fuel the colonocytes, butyrate has an effect on cell proliferation and differentiation, and has anti-inflammatory effects by inhibition of NFκB and reducing the formation of proinflammatory cytokines (Segain et al., 2000; Wong et al., 2006).

1.3 **Immunology in the intestine**

The immune system is the body resistance mechanism against infection and toxins, and without a working immune system, even minor infections could be fatal. The immune system is divided into the innate and the adaptive immune system (Parham, 2014, pp. 1-25). The innate is a non-specific defense mechanism and is the first to act when an intruder appears in the body. If the innate immune system cannot handle the infection by itself, it calls in the adaptive immune system. The adaptive immune system is antigen-specific and provides the immunological memory of the infectious agent (Parham, 2014, pp. 1-25).

1.3.1 **Mucosal immunity**

With an area between 250 and 400 m^2, the gastrointestinal tract is the second largest surface area on the human body, following the skin (Bengmark, 1998). Being the most densely colonized site on the body, the gut is in constant contact with luminal microorganisms (Turnbaugh et al., 2007) as well as food particles, and must deal with the constant threat of infection. The immune system in the gut is highly specialized, and it separates what is harmless from harmful. Immunotolerance is key, a wrongly induced inflammatory response could do more harm than good, as an infection could utmostly disrupt the protective barriers and opportunistic bacteria would gain entrance into the body tissues (Parham, 2014, pp. 1-25).

The adaptive immune system plays a vital role in maintaining tolerogenic responses towards symbionts and sustain barrier integrity, while the innate immune system regulates the adaptive immune response to the commensal bacteria in the gut (L. Wang, Zhu, & Qin, 2019). The lower part of the gastrointestinal tract is surrounded by the gut-associated lymphoid tissue (GALT). GALT is the largest organ of the immune system. It consists of both scattered and aggregated B-cell follicles, which all share common characteristics of being follicle-associated epithelium with membrane cells that samples luminal antigens into the lymphoid tissue bellow (Brandtzaeg, 1998). The sampling of luminal antigens enables immune cell differentiation towards the luminal content, even before there is a threat of invasion.

The mucosal adaptive immune system has two main strategies; immune exclusion and immunosuppression. Immune exclusion is the strategic act of preventing pathogenic bacteria initial access to the body tissues. The first line of defense is the secretory immunoglobulin (Ig)

A, together with several non-specific protective innate factors. Secretory IgA is a class of antibodies secreted into the lumen in large amounts, where they bind to their antigen and neutralizes them without giving rise to an inflammatory response. The neutralization inhibits foreign materials to penetrate the epithelial barrier (Brandtzaeg, 1998; L. Wang et al., 2019). An important innate factor contributing to the exclusion of foreign material is mucus, which is produced by specialized epithelial cells named Goblet cells (Parham, 2014, pp. 1-25). In the large intestine, the mucus is found to be layered into a dense layer closer to the epithelial cells, followed by a looser layer further out in the luminal space. The dense layer creates a barrier for penetration, while the looser layer works as a habitat for luminal bacteria that metabolize the mucosal substrates (Parham, 2014, pp. 1-25). The mucus does, however, also contain antimicrobial compounds such as antimicrobial peptides and lysozymes, produced by other specialized epithelial cells (Parham, 2014, pp. 1-25).

Oral tolerance is the act of avoiding or suppressing an inflammatory response that might come to harm the epithelial barrier. When soluble dietary antigens or the commensal bacteria penetrate the mucosal barriers, the immune system acts with several different mechanisms. Which mechanism that occurs depends on the amount of antigen introduced (Weiner, da Cunha, Quintana, & Wu, 2011). Low antigen concentrations call for immunoregulation, where regulatory T-cells in the GALT suppress the proinflammatory responses (Brandtzaeg, 1998). Higher doses of antigens favor the induction of clonal anergy or clonal deletion of the reactive immune cells (Weiner et al., 2011). Intestinal homeostasis is dependent on a balance between the gut microbiota and the adaptive immune cells, where both sides contribute towards a mutualistic relationship. Disturbances to this equilibrium might result in the host developing inflammatory or autoimmune disorders.

1.3.2 **The neonate immune system**

The gastrointestinal age at birth is important for the neonate immune competence level (Gleeson & Cripps, 2004). During gestation, the mother and fetus are interacting and there is evidence of transferal of maternal molecular components to the fetus. Most studied is the transfer of maternal antibodies to the newborn, called maternal immunization. The maternal immunization provides passive immunity against pathogens prevalent in the environment. Transfer of maternal antibodies starts *in utero*, mainly through the placenta, but also, to some degree through the amniotic fluid, and breast milk post-birth (Jennewein, Abu-Raya, Jiang, Alter, & Marchant, 2017). At birth, the infant's immune system is biased towards tolerance and consists

mostly of cells with naïve phenotypes (Brodin & Davis, 2017). IgG dominates at birth, alongside low levels of IgM, IgE, or IgA found in serum samples (Jennewein et al., 2017). Lymphocytes in GALT migrate to the mammary glands during lactation, here they secrete primary IgA and IgM into the breast milk. The maternal antibodies hold an immunosuppressive role in the infant's gut and prevent inflammation by the binding of toxins, bacteria, and macromolecules in the gut and aids in the shaping of the infant gut microbial community (Jennewein et al., 2017). The protective maternal antibodies, together with the innate immune factors of the neonate, lead to the rapid development of the mucosal immune system, which has been estimated fully mature during the first year of life, although inter-individual variations occur (Gleeson & Cripps, 2004).

The gut microbiota provides specific signals for immune stimulation and development (Pennock et al., 2013). At the onset of the immune development, the immune system has muted proinflammatory responses, enabling a regulatory profile that favors microbiota establishment. The establishment of the early microbial colonizers gives repeated exposure to foreign microbial antigens and different bacterial products (Pennock et al., 2013). Further, the first colonizers influence the development onward by facilitating a favorable environment for further colonization, e.g., by depletion of oxygen. Since the immune system development is interconnected with gut development, will prenatal factors influencing microbial development, such as delivery mode and use of antibiotics and other medical interventions, affect the development of the immune system. However, how the infant gut tissue adapts to the continuous microbial exposure is not yet fully understood (Pennock et al., 2013).

1.4 Methodology

1.4.1 DNA markers for species identification

In modern microbial and environmental research have amplicon sequencing allowed the profiling of entire bacterial communities and is a culture-free method for simultaneous multiple species detection (Knetsch, van der Veer, Henkel, & Taschner, 2019). The use of molecular target markers for amplicon sequencing is widespread due to their ease of use, stability, and broad application (Grover & Sharma, 2016). The molecular markers are retrieved from either a short sequence of DNA or protein (Chakraborty, Doss, Patra, & Bandyopadhyay, 2014). Different genes mutate at different rates, providing various molecular markers, yet which gene to use depends on the purpose of the study. Studies of closely related species require more

diverse genes to gain good taxonomic depth. Metabarcoding studies for simultaneous identification of many taxa, on the other hand, requires genes that are conserved across most species. Nonetheless, the sequence needs to be diverse enough for species separation. The genes used are often the highly conserved housekeeping genes that are important for basic cellular function (Chakraborty et al., 2014). Multiple genes are fulfilling the requirements for bacterial metabarcoding. To this day, the 16S ribosomal RNA (16S rRNA) genes are the most commonly used genomic fragment for taxonomic determination (Clarridge, 2004). However, the *cpn60* gene encoding the universal 60-kDa chaperonin protein, known as GroEl or Hsp60, (Hill et al., 2010), and the *COI* gene encoding cytochrome c oxidase subunit I (Naseem & Tahir, 2018), have also been used as markers for species identification.

Using ribosomal RNA genes as molecular markers for classification was proposed in 1977 by Carl Woese (Escobar-Zepeda, Vera-Ponce de Leon, & Sanchez-Flores, 2015). The fragment is useful as it makes a good model of the overall evolutionary rate as the sequence occurs in all organisms, and therefore can be compared not only among bacteria but also with the *16S rRNA* gene of archaea and the 18S rRNA gene sequence of eukaryotes (Clarridge, 2004). The gene is highly conserved due to the fragment's importance for cell function, as some regions' specificity is essential in the interaction between mRNA and the tRNA/amino acid complex during translation. Different parts of the sequence mutate at different rates resulting in some regions being more variable than others. The more variable regions are used for taxonomic purposes, while the conserved regions make ideal binding sites for universal primers needed for fragment amplification. The fragment length of the *16S rRNA* gene is about 1 550 bp (Clarridge, 2004).

Different approaches are used for describing and comparing metagenomic data from complex communities, like the gut. One of the approaches is to look at bacterial diversity. The most used diversity metrics are termed alpha, gamma, and beta, describing the local community diversity, total regional diversity, and the differences in diversity between niches, respectively (Escobar-Zepeda et al., 2015). Alpha diversity indices illustrate species richness and evenness in a given niche, where richness refers to the number of species, and evenness is the measure of relative abundance of different species. The Observed species index, the Shannon-Weaver index (Shannon & Weaver, 1949), and the inverted Simpson index (Simpson, 1949) are all used for alpha diversity assessment. Observed species consider the number of species in a given niche. The Shannon-Weaver index emphasizes rare species, while the Simpson index gives higher

weight to species that are more frequent and dominant in the niche (Escobar-Zepeda et al., 2015). An increase in Shannon-Weaner and inverted Simpson index values are equal to an increase of sample diversity and evenness. Differences between bacterial niches are determined by beta diversity indices and are based on dissimilarity. Among the dissimilatory indices is the binary Jaccard index (Jaccard, 1901). This qualitative index recognizes the presence and absence of species in different niches and uses that information to calculate a distance matrix. Another metric for dissimilarity is the Bray-Curtis method (Bray & Curtis, 1957), which applies a quantitative measure of community dissimilarity by using species abundance in each niche.

1.4.2 **Sequencing technologies**

DNA sequencing is the process of determining the primary structure of DNA by resolving the order of the four bases, adenine, guanine, cytosine, and thymine. The knowledge about DNA sequences has become invaluable for basic biological research as it may be applied to individual genes, larger genetic regions, full chromosomes, or entire genomes of any organism. Dideoxynucleotide sequencing is an enzymatic sequencing method and was one of the first methods to sequence fragments of DNA (McCombie, McPherson, & Mardis, 2019). The method was introduced in 1977 by Sanger and is now better known as Sanger sequencing (Sanger, Nicklen, & Coulson, 1977). Sanger sequencing uses a mixture of template DNA, DNA polymerase, primer, native dNTPs, and one of the four 2'3'-dideoxynucleotide (ddNTP). The incorporation of ddNTP terminates elongation by preventing the addition of further nucleotides (McCombie et al., 2019). After sequencing, the sample contains a variety of fragments of different lengths. A subsequent application on a denaturing polyacrylamide gel produces a ladder of fragments across four lanes, one for each ddNTP. The nucleotide sequence is read from the smallest fragments to the largest, from the bottom to the top of the gel (McCombie et al., 2019).

At the same time as the Sanger sequencing method arrived, other enzymatic sequencing methods were published. For instance, a couple of years prior, Sanger came out with the plus and minus method (1975), Maxam and Gilbert published the chemical cleavage method in 1977, and in 1978, Barnes introduced the partial ribosubstitution method (McCombie et al., 2019). However, these methods never became as popular as Sanger sequencing, and today is Sanger sequencing seen as the primary technology from the first-generation of sequencing (Liu et al., 2012).

The next-generation sequencing (NGS), or second-generation sequencing differs from Sanger sequencing as both enzymology and data acquisition are orchestrated in a stepwise fashion, enabling massive parallel sequencing. Massive parallel sequencing generates data from large amounts of templates simultaneously, ranging from tens of thousands to billions of templates (McCombie et al., 2019). NGS technologies have higher throughput and a reduced cost compared to Sanger sequencing and the other first-generation methods (Liu et al., 2012). The first commercially available NGS system was the Roche 454, using pyrosequencing technology. This technology relies on the detection of light emission from pyrophosphate that is released during the incorporation of nucleotides. Nevertheless, in later years the Illumina sequencing method has been almost synonymous with next-generation sequencing (Knetsch et al., 2019), as it has dominated the market after buying Solexa and their technology in 2006 (Liu et al., 2012).

With the increased use of second-generation sequencing technologies and the addition of new modifications, the third generation of sequencing technologies has arisen. The new generation of sequencing brings new insights by removing the need for polymerase chain reaction (PCR) amplification, thereby removing the bias associated with amplification, as well as shortening preparation time (Knetsch et al., 2019; Liu et al., 2012). The Oxford Nanopore MinION sequencer is an example of this generation sequencing methods. Another example is the Pac Bio RS II sequencer by Pacific Biosciences. The first one measures change electrical conductance during nucleotide translocation of template DNA through a tiny biopore (McCombie et al., 2019), while the latter measures enzymatic incorporation of fluorescence-marked nucleotides in real-time (Knetsch et al., 2019).

Today Illumina is the dominating sequencing technology and accepts input generated by any method that gives adaptor-flanked fragments up to several hundred base-pairs in length (Knetsch et al., 2019; Shendure & Ji, 2008). Illumina sequencing by synthesis has two main factors, that is its flow cell technology and the reversible dye terminator sequencing chemistry (Knetsch et al., 2019). The sequencing procedure begins with the template DNA being separated into single strands. Adaptor sequences are ligated to both ends of the single-strand DNA (ssDNA) and hold a site complementary to the oligonucleotides (oligos) on the flow cell, allowing the strands to attach. The adapter is also equipped with a site for primer binding, as well as a barcode for recognition. Through PCR-based methods are the templates amplified, giving discrete clusters of copies of the same template DNA. The clusters are generated by

bridge amplification where the ssDNA molecules, attached with one end to the flow cell, are forced to bind with their free end to a complementary nearby, forming a bridge. While in bridge formation, the DNA molecules are synthesized forming a double-stranded bridge, which then denatured into two single DNA strands anchored to the flow cell. This results in local clusters of around one thousand clonal molecules close to the original template-strand (Knetsch et al., 2019).

Illumina sequencing by synthesis starts with binding of a primer to the adapter sequence, followed by incorporation of fluorescent nucleoside triphosphate (dNTPs) that are blocked at the 3'-OH-end (Knetsch et al., 2019). The incorporation is done in several cycles, for every cycle all four dNTPs are pumped through the lanes and incorporated where possible. As the end is blocked, only a single nucleotide is added to the growing complementary DNA strand per cycle. The different dNTPs are labeled with a different fluorochrome, and the incorporation is monitored. After the incorporation of a dNTP the entire surface is imaged, illuminating the fluorescent coloring of the newly incorporated nucleotide. As the last step, the blocking of the strand is removed, and the clusters are ready for the next dNTP. The fluorescent signal of millions of individual clusters are captured by sensitive optics, and all signals are translated into nucleotide assignments (Knetsch et al., 2019).

1.4.3 **Gas Chromatography**

Analysis of SCFAs is predominantly performed using gas chromatography (GC) (Primec, Micetic-Turk, & Langerholc, 2017). Chromatography is a technique for chemical separation of molecular components in complex samples. Following separation, the compounds can be measured quantitatively or qualitatively (Poole, 2012, pp. 19-75).

The injected sample vaporizes in contact with a hot glass liner in the injector site of the chromatograph. Further, the samples are led through a column that separates the sample components and ends in the detector site, where composition and concentrations of substances are determined (Poole, 2012, pp. 19-75). The method of gas chromatography separates the compounds by exploiting the chemical and physical properties of the molecules. The column contains a stationary phase and a mobile phase. The mobile phase flows through the column, while the stationary phase is fixed. The mobile phase in gas chromatography is usually an inert gas, e.g., helium or nitrogen, and is called the carrier gas. The carrier gas carries the components of the sample mixture along the column, while the stationary phase interacts with the sample

components retaining them in the column. Stronger interactions lead to longer retention time than weaker interactions, thus separating the sample components (Poole, 2012, pp. 19-75). As each compound uses a different amount of time to reach the detector site, the time becomes specific for each component under given pressure and temperature and is termed the retention time of the compound. After a chromatographic run, the data is presented as a chromatogram, a graph of detector response against retention time. The retention time is used for qualitative determination. The area under the peaks of the chromatogram is used for quantitative measurement of the compound, as the peak is assumed proportional to the amount of analyte present in the sample (Primec et al., 2017).

SCFAs have been analyzed in various biological materials, such as blood plasma and serum, brain, and feces (Primec et al., 2017). The most popular biological material for SCFA analysis is fecal material, as it is accessible and non-invasive to collect. It should, however, be noted that 90-95% of the SCFAs are taken up in the gastrointestinal tract leaving only about 5% of the microbially produced SCFAs to the feces (Primec et al., 2017). The fecal material is complex, and as SCFAs are volatile, which poses a challenge in sample preparation and for fast and reliable determination of SCFAs content.

To handle the complex fecal samples, a number of pretreatment methods have been proposed. The fastest and simplest are pretreatment methods that avoid extraction of the SCFAs from the samples. These are treatments involving dilution, filtration, ultrafiltration, or centrifugation (Primec et al., 2017). However, fast and simple, they have the problem of impurities overloading the column. Other preparation methods include simple acidification using several acidification agents (Weir et al., 2013), including formic acid (Primec et al., 2017).

Separation of the SCFAs occurs in the column, where the sample components are separated based on interactions between analytes and the stationary phase. In GC, capillary columns are the most effective, and many contain silica coated with polyethylene glycol (PEG) as a stationary phase (Primec et al., 2017). As the stationary phase is highly polar, they are widely used for analyzing compounds with polar functional groups, like SCFAs, by hydrogen bonding and acid-base interaction (Hayward, Hua, Gras, & Luong, 2017).

The most conventional detector used in SCFA analysis is the flame ionizing detector (FID). This detector consists of a hydrogen flame that ionizes the sample molecules and holds a negatively biased collector that collects all the positively charged ions. The response is proportional to the mass of carbon that passes through in a unit of time (Primec et al., 2017).

1.4.4 **Cytometry by Time-Of-Flight (CyTOF)**

Cytometry by Time-Of-Flight (CyTOF) or mass cytometry is a novel technology for detailed phenotypic and functional analysis of single cells (Kay, Strauss-Albee, & Blish, 2016). The method was developed by Tanner and colleagues at the University of Toronto, with the aim of increasing the number of simultaneous protein measurements in individual cells, compared to what was possible using traditional flow cytometry (Lakshmikanth & Brodin, 2019). Mass cytometry has a variety of applications and has been used for describing human immune system variations (Brodin et al., 2015). CyTOF relies on antibodies conjugated with heavy metal isotopes, which are used to stain cells from biological samples before they are applied to the system (Kay et al., 2016). The antibodies bind to the target of interest, either on or within the cells. The metal isotopes act as reporters for expression levels of the targets, and more than 42 isotopes can be applied simultaneously (Kay et al., 2016). The cells in the sample pass in a single-cell suspension trough a nebulizer, forming droplets for introduction to the mass cytometer. As the cells enter the instrument, they are exposed to argon plasma, which induces the covalent bonds to break, forming free, charged atoms. The cloud of ions is filtered to discard common biological elements, thereby enriching the heavy metal ions, which are then separated by their mass-to-charge ratio in a time-of-flight mass spectrometer. Electrical signals are obtained from the counted ions, integrated on a single-cell basis into single-cell events for downstream analysis (Bandura et al., 2009).

1.5 **PreventADALL cohort**

Allergic diseases and other immune-related noncommunicable diseases have become common in the Western world. Atopic dermatitis and food allergies are two of the earliest manifestations of this trend, affecting 20-30%, and 5-10% of infants, respectively (Lowe, Leung, Tang, Su, & Allen, 2018). The Preventing Atopic Dermatitis and ALLergies (PreventADALL) in Children cohort is a general population-based mother-child birth cohort. The main objectives of the cohort study are to determine whether primary prevention of allergic diseases is possible by

simple, low-cost strategies and to assess early life factors and exposures, including intrauterine environment, microbiota, and xenobiotics, involved in the development of asthma and allergic diseases or other noncommunicable diseases including cardiovascular diseases, obesity, and diabetes (Lodrup Carlsen et al., 2018).

The cohort recruited mothers at 18-weeks of pregnancy, not discriminating based on the mother's age, race, or medical history. The three main sites for recruitment were the Oslo University Hospital (Norway), Østfold Hospital Trust (Norway), and Karolinska Institute (Sweden). The first pregnant woman was enrolled in the study in 2014, and the last one was included in 2017, making it a total of 2697 mothers. The study would thereafter include their children from the day of birth until 36 months, with periodic controls at 3-, 6-, 12-, and 24 months. The biological sampling included, among others, samples from urine, blood, saliva, breast milk, skin swabs, fecal samples (Lodrup Carlsen et al., 2018).

1.6 **Study aim**

The relationship between the gut microbiota and the human host is of importance for the host health status, and the bacterial products, such as SCFAs, have been shown to play a vital role in the development of our immune system. Most of the established associations between the gut microbiota and the host's immune system are derived from studies of model organisms, and there is a lack of studies looking at these effects in healthy humans. The present study aims to explore the longitudinal development of the infant gut microbiota and short-chain fatty acid in relation to the immune cell status at 12 months. This was achieved by the following sub-goals:

- Bacterial composition determination using the *16S rRNA* gene and Illumina sequencing.
- Short-chain fatty acid analysis through gas chromatography.
- Correlation analysis of the bacterial composition, SCFAs, and immune cell profiles at 12 months.

2. **Materials and methods**

This study included the determination of bacterial composition by DNA extraction of the *16S rRNA* gene from fecal samples, and next-generation sequencing using the Illumina MiSeq system. The short-chain fatty acid composition of the fecal samples was determined by gas chromatography on a Trace 1310 equipped with an autosampler (ThermoFisher Scientific). Immune cell composition determination using mass cytometry on CyTOF2 of serum samples was conducted by Alex Olin and Petter Brodin, at Karolinska Institute. The project setup is illustrated in Figure 2.1.

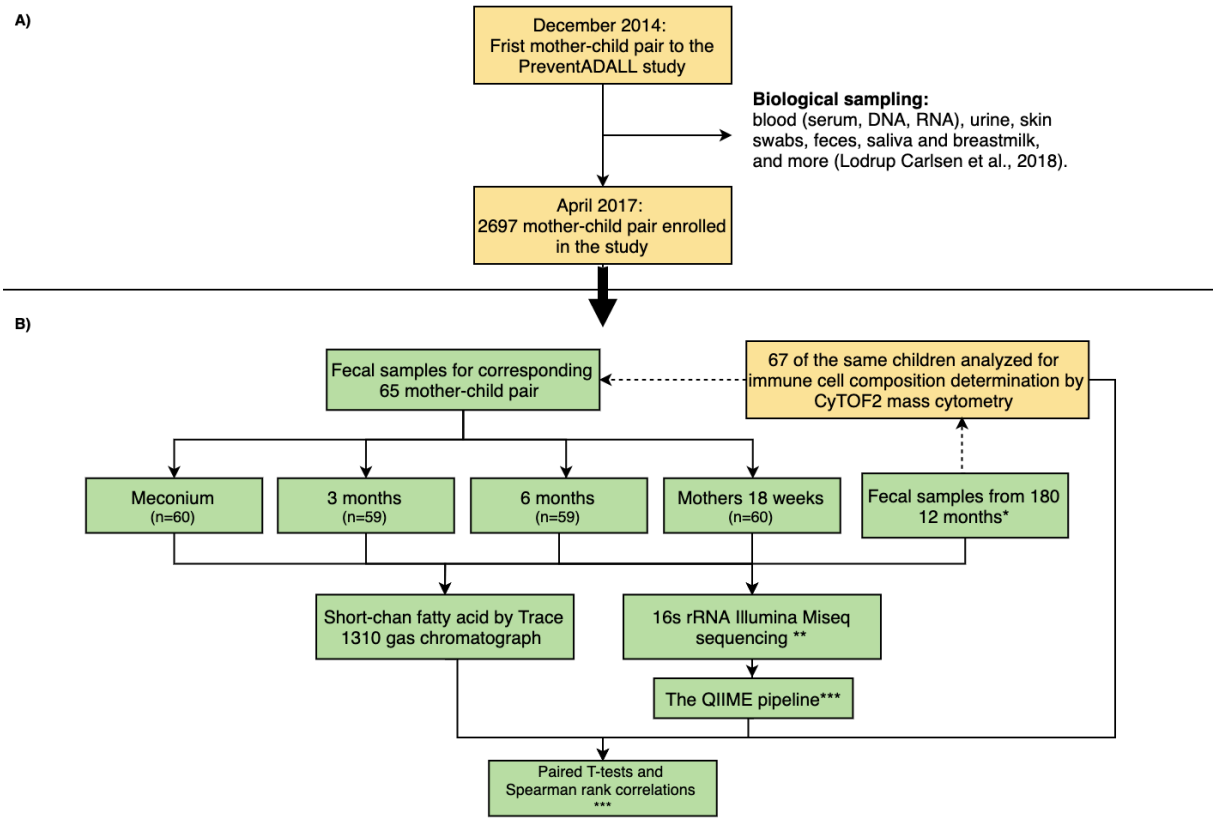


Figure 2.1: Flowchart illustrating the methodical setup. Analysis conducted in this thesis is marked in green, while yellow is done by other groups working within PreventADALL. A) represent the biological sampling which was done by the PreventADALL project. B) represents the analysis of fecal material analysis of bacterial and SCFA compositions conducted in the present sample. The fecal samples were prepared for DNA sequencing, sequenced on Illumina MiSeq, and analyzed for SCFAs on a Trace1310 gas chromatograph. The manuscript for immune cell analysis by CyTOF2 mass cytometry is listed in Appendix A. The bacterial data were processed using the QIIME pipeline for sequencing results, while the statistical analysis was done using RStudio and Microsoft Excel.

* In total, fecal samples were obtained from 180 infants at 12 months of age, and as immune cell composition data was acquired for 67 of the 180 infants, fecal samples collected at 0(meconium), 3 and 6 months, and from their mothers (18-weeks pregnant) were obtained as well, making 65 mother-child pairs.

** Most of the 16S rRNA Illumina library preparation was conducted in this thesis, except for library quantification using qPCR and ddPCR, and loading onto the MiSeq v3 reagent cartridge, which was done by laboratory personnel.

*** The QIIME run was performed by Ph.D. student Morten Nilsen. Nilsen helped in computing the RStudio scripts for Spearman rank correlations and false discovery rate adjustments.

2.1 Sample collection and initial handling

2.1.1 Sample collection

The fecal samples analyzed in this study were collected at the infant and mother's periodic controls and diluted 1:10 with a DNA/RNA shield buffer (Nordic Biolabs R1100-250). Before being immediately frozen at -80°C to prevent any degradation or bacterial growth.

2.1.2 Initial handling

The samples were transported from the University of Oslo Biobank to Norwegian University of Life Sciences. During transportation, the samples were kept cold using a cooler box with cooling elements. Initial outtakes from the samples were 1.2mL of fecal material, enough for DNA extraction and short-chain fatty acid analysis.

2.2 DNA extraction

The DNA extraction involved lysis of the cells, solubilizing DNA, followed by chemical and enzymatic treatment for removal of other cellular components, such as macromolecules, lipids, RNA, and proteins (Gupta, 2019).

2.2.1 Mechanical lysis

To isolate the DNA from fecal samples the cells were first disrupted by mechanical lysis using 0.2g acid-washed glass beads with the size <106µm, 450-600µm, and 2 beads of 2.5-3.5µm (Sigma Aldrich). Different sized beads were used as it resulted in better cell lysis compared to the use of one bead size only. The combination ensures higher diversity and less bias in mechanical disruption of the bacterial cells (Bakken & Frostegård, 2006). The glass beads and the samples were processed twice in FastPrep96 (MP Biomedicals, USA) at 1800 rpm for 40 seconds. After mechanical disruption, the samples were centrifuged at 1300 rpm for 5 minutes at 4°C.

2.2.2 Proteinase treatment and chemical lysis

Chemical lysis treatment with elevated temperatures ensured the disruption of all cells before DNA extraction. From the centrifuge step, 50µL of the supernatant was transferred to a KingFisher96 well (KF96) plate, together with lysis buffer (Thermo Fisher Scientific, USA) and Proteinase K (Thermo Fisher Scientific, USA) for chemical lysis. Using a KingFisher Flex

Nucleic Acid Purification system (Thermo Fisher Scientific, USA), the samples were incubated for 10 minutes at 55°C. The lysis buffer uses high salt concentrations to lyse the cells, while the unspecific Proteinase K digest proteins in the sample (Ebeling et al., 1974).

2.2.3 **DNA purification**

The purification of DNA was conducted using paramagnetic beads, which in high saltine solutions, forms a bridge between the DNA and the bead particles (Boom et al., 1990). The particle-DNA complexes were precipitated by the addition of ethanol, and subsequently, in multiple wash steps, were alcohol- and salt-based buffers used to remove contaminations. To elute the extracted DNA, water was added leading to disruption of the salt bridges, and the DNA was extracted from the eluate. The extraction was done on the KingFisher Flex Robot using a MagMidi LGC kit (LGC Biosearch Technologies, United Kingdom), following the kit recommendations. The extracted DNA was later tested quantitatively by Cambrex-FLX 800 CSE (Thermo Fischer Scientific, USA), Qubit Fluorometer (Life Technologies, USA), and qPCR (LightCycler 480 II, Roche, Germany).

2.3 **Polymerase chain reactions**

Polymerase chain reactions is a widely used method for amplification and detection of small amounts of specific nucleotide sequences. The principle behind PCR is a three-step process that is repeated in cycles. The first step is denaturation, during which an increase in temperature leads to the breakage of the hydrogen bonds that hold the complementary strands of the DNA together (Schochetman, Ou, & Jones, 1988). As the mixture cools down, the single strands associate with their complementary sequences. However, as there is an excess of primers in the solution, the ssDNA is more likely to anneal to the primers and not to their complimentary ssDNA sequences. Once annealed, the DNA polymerase starts extending the primers by adding free nucleotides on the 3'OH-end using the DNA strand as a template. These three steps make up a cycle, which is repeated between twenty to fifty times (Kubista et al., 2006; Schochetman et al., 1988).

2.3.1 **Qualitative PCR**

Amplification of the *16S rRNA* gene fragment was done using the primer combination of PRK341F and PRK806R, which permit amplification of the V3 -V4 region of the gene. More information about the primers is given in Table 2.1. The PCR was performed using a reaction cocktail with a final concentration of 1x HOT FIREPol® Blend Master Mix Ready to Load (Soils BioDyne, Estonia), 0.2µM of each primer and nuclease-free water (VWR, USA). The PCR master mix was distributed into a 96-well plate, and 2µL template DNA was added.

The PCR cycle for amplification began with an initial denaturation for 15 minutes at 95ºC, followed by 25-30 cycles of denaturation at 95ºC for 30 seconds, annealing at 55ºC for 30 seconds and elongation at 72ºC for 45 seconds. Final elongation lasted 7 minutes at 72ºC, and the samples were subsequently stored at 10°C. The temperature cycles were conducted using the 2720 Thermal Cycler (Applied Biosystems, USA).

Table 2.1: Description of primer sequences used in different PCR reaction.

	Primer		**Sequence (5'-3')**	**Annealing temperature (ºC)**	**Fragment size (bp)**	**Reference**
Amplicon PCR and qPCR	PRK341F	Forward	5'-CCRACGGGRBGCASCAG-3'	55	466	(Yu, Lee, Kim, & Hwang, 2005)
	PRK806R	Reverse	5'-GGACTACYVGGGTATCTAAT-3'			
Index PCR	1-16	Forward	Table B.1 in Appendix B	55	594	
	13-24	Reverse				

2.3.2 **Index PCR**

Index PCR was performed to attach Illumina adapters to the ends of the *16S rRNA* gene fragments. The flanking adaptors are essential for Illumina flow cell attachment, polymerase binding, and determination of sample origin after sequencing. A combination of 16 different forward primers and 24 reverse primers were used, resulting in 384 index combinations. Primer information is found in Table 2.1 and Table B.1 in Appendix B.

The PCR reaction cocktail was prepared to be 1x HOT FIREPol® Blend Master Mix Ready to Load with nuclease-free water and dispensed into each sample well of a 96-well plate. Eppendorf epMotion5070 (Eppendorf, Germany) robot distributed the specific combination of primers, to make a 0.2µM primer concentration. To this reaction mixture, 2µL of cleaned DNA was added.

The PCR amplification was done by an initial denaturation at 95ºC for 5 minutes. Then 10 cycles of denaturation at 95ºC for 30 seconds, annealing for 1 minute, and elongation at 72ºC for 45 seconds follows. The final elongation lasted for 7 minutes at 72ºC. The PCR was conducted using a 2720 Thermal Cycler.

Samples which did not show satisfactory results after indexing were treated a second time to ensure that most samples had sufficient DNA quantity and quality for sequencing.

2.3.3 **Quantitative PCR**

Quantitative PCR (qPCR), also known as real-time PCR, is based on the same principles as conventional PCR. However, probes for fluorescent monitoring of PCR products are included (Kubista et al., 2006). qPCR was conducted by adding a master mix of 1x HOT FIREPol® EvaGreen® qPCR supermix (Solis BioDyne, Estonia), primers (PRK341F and PRK806R) to a concentration of 0.2µM, and nuclease-free water to 1µL of template DNA. The products were amplified using LightCycler480 II (Roche, Germany). Based on the fluorescence and the amplification efficiency, it was possible to calculate the original number of DNA molecules in each sample. The sample DNA was quantified by comparing the number of cycles needed to reach a set threshold of fluorescence, called the Ct-value (Kubista et al., 2006).

The initial denaturation was done at 95°C for 15 minutes, followed by 40 cycles of denaturation on 95°C for 30 seconds, annealing at 55°C for 30 seconds and elongation 72°C for 45 seconds.

2.3.4 **PCR-product clean up**

Before further treatment of the DNA, the PCR products were cleaned to remove any nucleotides, polymerases, and primers. The clean-up was done using Sera-Mag beads (Thermo Fisher Scientific, USA). Sera-Mag beads are magnetic particles that bind to the negatively charged DNA. When put near magnets, the beads and the bound DNA form a pellet, separating

from the supernatant. The following wash steps use the beads' magnetic properties to further clean the sample DNA from the rest of the sample contents. Lastly, in the elution step, the magnetic particles are suspended in the elution buffer, and the DNA will be freed from the magnetic particles. When put near magnets the DNA will stay suspended in the elution, but the magnetic particles will pelletize. The eluate containing the amplicons was transferred and stored in a 96-well storage plate.

Clean-up of the qualitative PCR products was done using 0.1% Sera-Mag Speed beads solution, in a 1:1 ratio to DNA products using Beckman Biomek 3000 robot (Beckman Coulter, USA). In the case of the pooled library after PCR indexing, a 0.1% Sera-Mag Speed Beads solution, with a ratio of 0.8:1 Sera-Mag beads to sample was used. The Biomek 3000 followed a protocol for clean-up that included two washing steps with 80% EtOH (Antibac, Norway), and elution in nuclease-free water.

All the cleaned PCR products and pooled libraries were measured qualitative using gel electrophoresis on 1.5% agarose (Invitrogen, USA).

The cleaned pooled samples were frozen and saved for quantification using qPCR and ddPCR and loading onto the MiSeq v3 reagent cartridge (Illumina, USA), which was done by laboratory personnel according to the manufacturer's recommendations.

2.4 Quantitative and qualitative measurements of DNA

2.4.1 Cambrex-FLX 800 CSE and Qubit fluorometer

The DNA concentrations were measured using Cambrex-FLX 800 CSE. A Qubit working solution was prepared by diluting Qubit dsDNA HS Reagent (Invitrogen, USA) in Qubit dsDNA HS Buffer (Invitrogen, USA) in a ratio of 1:200. This solution was added to each well in a Nunc 96 well Nontreated Black Microwell plate (Thermo Fisher Scientific, USA), with 2µL sample DNA, before the measured on Cambrex-FLX 800 CSE.

From the concentrations measured on Cambrex, some samples at different parts of the concentration range were chosen to be used to make a standard curve. These samples were measured using the Qubit fluorometric quantification system, following the kit protocol, mixing 198µL working solution (Quant-iT reagent diluted 1:200 in Quant-iT buffer) with 2µL sample

DNA. In addition, two Qubit standards were made to calibrate the fluorometer. A regression line was made by connecting the Qubit- and the Cambrex measurements of the samples. The regression line was used to estimate the Qubit value of all the other samples based on the Cambrex result.

2.4.2 **Gel electrophoresis**

The size and quality of PCR- and clean-up products were determined using gel electrophoresis on a 1.5% agarose with 4 % PeqGreen (PeqLab, Germany).

To the clean-up products loading dye (NewEngland Biolabs, USA) was added in a ratio of 1:6. The quality assessment was determined based on fragment size, band density, and if there were other bands, due to either formation of by-products, mispriming, or damaged products (van Pelt-Verkuil & te Witt, 2019). The electrophoresis ran for 30 minutes, at 80V and 400A. A 100-base pair DNA ladder (Soils BioDyne, Estonia) was used as a size marker for fragment length. The expected fragment size after each PCR run is listed in Table 2.1. The fragments were visualized using the Molecular Imager Gel Doc XR Imaging system with Quantity One 1-D analysis software v.4.6.7 (Bio-Rad, USA), using ultraviolet light.

2.4.3 **Normalization and library pooling**

After all samples were indexed, the DNA was measured using Qubit and Cambrex-FLX 800 CSE. A regression line was made based on the Cambrex-Qubit measurements and used to estimate the Qubit-values for the remaining samples. The sample with the highest estimated concentration was set as a cut-off value. From this sample it was decided that 1-2µL was to be added to the pooled library. The volume to be added to the pooled library from all other samples was calculated so the concentration matched the cut-off value. Although, no larger volume than 10µL was used. Pooling and normalization were done using Biomek 3000.

2.5 Gas chromatography

2.5.1 The instrument

The instrument used was the TRACE ™ 1310 Gas Chromatograph Autosampler (Thermo Fisher Scientific), with a split injector and using helium as a carrier gas. More details about the instrument are given in Table C.1 in Appendix C.

2.5.2 Standards

Calibration standards

For calibration of the TRACE ™ 1310 Gas Chromatograph, three calibration standards were made, A, B, and C, with different acid concentrations. The acids used for calibration were acetic acid, propionic acid, butyric acid, isovaleric acid, valeric acid, and 2-methylvaleric acid. To all standards, formic acid was added in a fixed concentration of 0.2%. The acid concentrations are to be found in Table D.1 in Appendix D, and all acids are from Sigma Aldrich (USA). The calibration standards were made using a 10% stock of formic acid, and 100 mM stocks of the other acids.

Internal standard

A solution of 0.4% formic acid and 2000µM 2-methylvaleric acid added to the samples and acted as an internal standard. The internal standard was included to obtain absolute quantitative values of the SCFAs in the samples. As the samples are injected some differences in measurement might arise, however, the variations in short-chain fatty acid detection are assumed to change at the same rate as the detection of 2-methylvaleric acid. The differences in measured 2-methylvaleric acid and the true concentration in each sample were calculated to correction factors, which were used to correct the measured values of the other acids in the same sample.

The internal standard was, for the fecal samples from meconium, 3-, 6 months, and mothers, changed to contain 500µM 2-methylvaleric acid. This change was applied to gain a better correspondence between the 2-methylvaleric acid concentration in the internal standard and the calibration standard.

2.5.3 **Sample preparation**

The samples were first thawed on ice and vortexed to make a homogenized solution. A part of the sample was mixed with equal amounts of MilliQ water (Merck, USA), and to this, the internal standard was added to make a 1:1 ratio. This resulted in a total sample dilution of ¼.

This mixture was centrifuged at 13 000 rpm for 10 minutes. The supernatant was then transferred to spin columns (VWR, Germany) with a 0.2µm filter and centrifuged down at 10 000 rpm for 5 minutes. 300µL of the eluate was transferred into GC vials (Thermo Fisher Scientific, Norway), and Crimp caps (VWR, Germany) were put on before the vials were placed in the racks of the autosampler on the GC.

2.5.4 **GC runs**

All samples were run within 24 hours of preparation. To track any deviations in measurements made by the machine a calibration standard was run in between every 10th sample. The data were processed in the Chromeleon Chromatography Data System (CDS) Software (Thermo Scientific).

2.6 **Data processing**

2.6.1 **QIIME – Quantitative Insights Into Microbial Ecology**

To analyze the *16S rRNA* raw sequencing data, Quantitative Insights Into Microbial Ecology (QIIME), an open-source bioinformatics pipeline, was used. It is available at http://qiime.org/. The sequences were first preprocessed, which comprised of primers removal, demultiplexing according to barcodes, and served a function of quality filtering. Further, the data were clustered at 97% homology level, and the SILVA database (Quast et al., 2013) was used to create an operational taxonomic units (OTUs) database, and taxon-based analysis of diversity both within and between samples (Caporaso et al., 2010). QIIME generated graphical representations of the data, allowing for user interaction.

2.6.2 **Statistical analysis**

Potential outliers after SCFA determination were examined by the rule of interquartile in Microsoft Excel. Samples determined as outliers were excluded in further SCFAs analysis.

Paired T-tests were performed in RStudio to analyze statistically significant differences of SCFAs or bacterial composition between the different age groups. Additionally, a paired t-test was performed to find significant differences between thawed and frozen samples and how the gas chromatograph held over time. For all t-tests performed, a test level of 5% were used.

Spearman rank correlations (Spearman, 1904), being nonparametric, were used to determine correlations between SCFAs and bacterial composition, immune cell and bacterial composition, and SCFAs and immune cell composition. Positive correlation values were given where there was co-occurrence of high or low levels of two elements, e.g., between a given bacterial order and a SCFA. Just the same, negative correlation values were given where a high value of one element was coinciding with low levels of another element in the same samples. The probability values from the Spearman correlations were then adjusted for false discovery using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). The procedure was conducted so the adjustment of p-values was done within each SCFA, for the correlations between bacterial and SCFAs composition, as well as for the correlation between SCFAs and immune cell composition. In the case of the correlation between immune cells and bacterial composition, the p-values were adjusted within each immune cell category.

The statistical analysis and graphical visualization were done in Microsoft Excel for Mac version 16.35, and RStudio version 1.1.423, and the R-packages used were ggplot2 version 3.2.1, tidyverse version 1.3.0, dplyr version 0.8.01, tibble version 2.1.1.

3. Results

3.1 *16S rRNA* gene sequencing

Two sequencing runs were performed; the first run sequenced the 12-month samples, and the second sequenced the meconium, 3 months, 6 months, and mother samples. The data from the two runs were separately processed in the QIIME pipeline. The quality filtering in QIIME resulted in a total of 167 samples from 12 months and 225 samples from the longitudinal samples, distributed as 55 meconium samples, 56 samples at 3 months, 56 samples at 6 months and 58 of the samples from the mothers. The taxa plots retrieved from the 12 months OTU binning and longitudinal OTU binning in QIIME were merged together in RStudio. For details concerning processing in QIIME and other technical aspects see Appendix E.

3.1.1 Taxonomic distribution with age

The average taxonomic OTUs from QIIME, are presented as a bar graph of bacterial order abundance at different ages in Figure 3.1. The average distribution of bacterial phyla and genera are respectively given in Figure F.1, Figure F.2, and Table F.1 in Appendix F. Enterobacteriales appeared as the most dominant order in the meconium gut making up an average of 22.2% of the bacteria present in the gut. However, the order decreased significantly towards the 3-month samples ($p = 0.043$, paired t-test, unadjusted). Within the Eneterobacteriales order was the *Escherichia/Shigella* genus most prominent and constituted 19.9% of the gut composition (Figure F.2 in Appendix F). The relative abundance of *Escherichia/Shigella* decreased towards the 12-month samples and the mothers, with the constituting 1.3% and 0.6%, respectively. Contrarily to the Enterobaceriales order, Bifidobacteriales experienced a significant increase from 11.8% in meconium and peaked at 40.4% at 3 months ($p = 2.20*10^{-10}$, paired t-test, unadjusted), followed by a subsequent decrease in relative abundance towards gut of the mothers. The order of Clostridiales had a consequent significant increase as of 3 months and was dominating in the mothers' gut with a relative abundance of 66.4% (3 to 6 months, $p = 0.009$; 6 to 12 months, $p = 6.79*10^{-06}$; 12 months to mothers; $p = 2.11*10^{-11}$; paired t-test, unadjusted). In the mothers, *Faecalibacterium* (9.3%) and *[Eubacterium] rectale* group (6.6%) were the most dominating genera within the Clostridiales order (Figure F.2 in Appendix F).

Figure F.1 (Appendix F) showed that the Proteobacteria and Firmicutes were the most abundant phyla in the meconium samples, with a respective prevalence percentage of 44.3% and 28.0%. Proteobacteria showed a decrease throughout the sampling points, while the relative abundance of Firmicutes increased. Firmicutes constituted 70.0% of the gut microbiota of the mothers on average. The phyla Bacteroidetes followed as the second largest with a prevalence of 21.7% in the mother samples.

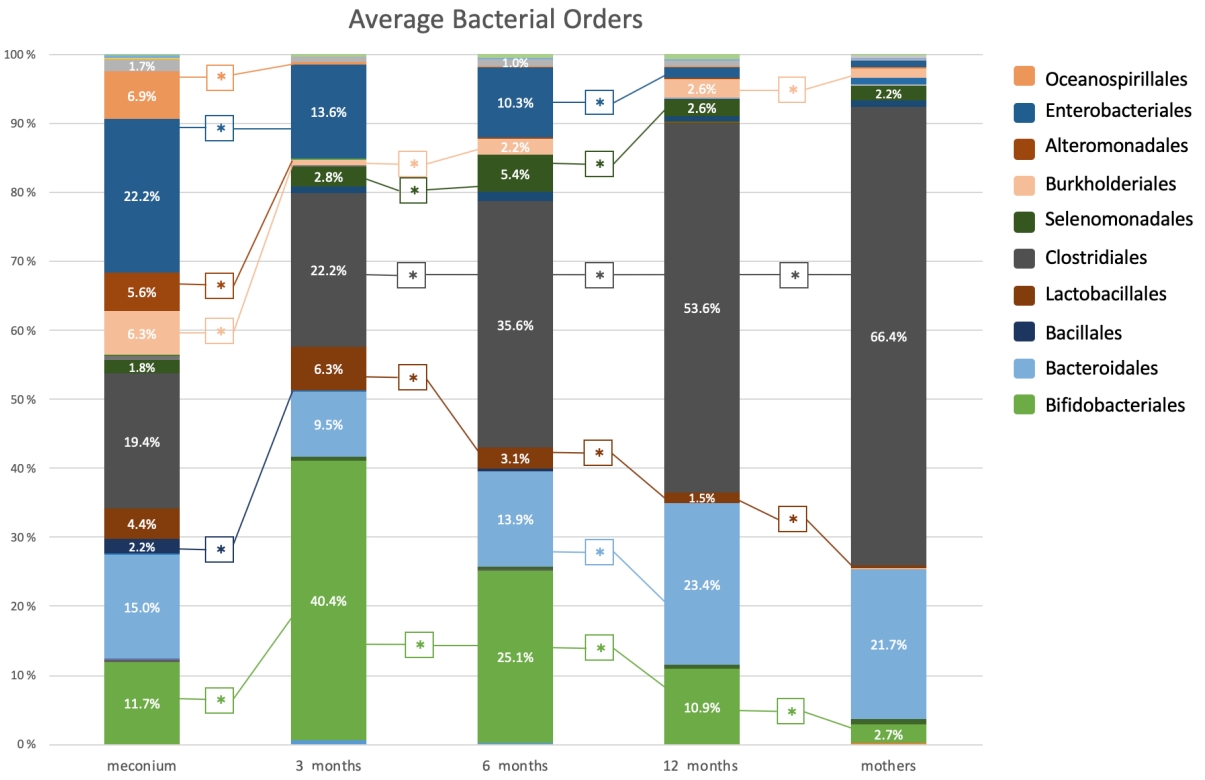


Figure 3.1: Distribution of bacterial orders within each age group. The bacterial orders are acquired from sequencing and processing using the QIIME pipeline, and the average of the dominant bacterial orders in each age group are displayed. The stars represent significant differences in average bacterial order abundance between the age groups computed by a paired t-test in RStudio using a test level of 0.05.

3.1.2 **Alpha diversity**

Observed species, Shannon-Weaver and inverted Simpson indices are presented in Figure 3.2. The lowest amount of detected species was at 3 months, with an average of 53.8 OTUs in each sample. However, the group difference was non-significant from the 6 months age group. From there on, there was an increase of OTUs towards the samples from the mothers, with the increase being significant between 6 and 12 months ($p = 9.25*10^{-32}$, paired t-test, unadjusted), and 12 months and mothers ($p = 3.89*10^{-24}$, paired t-test, unadjusted). The meconium samples had an average of 85.6 OTUs per sample, higher than both 3 and 6 months.

The samples from the mothers scored the highest values when using both the Shannon-Weaver and the inverted Simpson indices. When considering the median score for each age group, the meconium samples followed as a close second for the most diverse and even age group. Taking the average score into account, the 12-month samples (Shannon-Weaver = 4.11, Inverted Simpson = 0.89) expressed a higher diversity than the meconium samples (Shannon-Weaver index = 3.94, Inverted Simpson index = 0.79). All indices identified the 3-month samples as the age group with the least sample diversity and evenness.

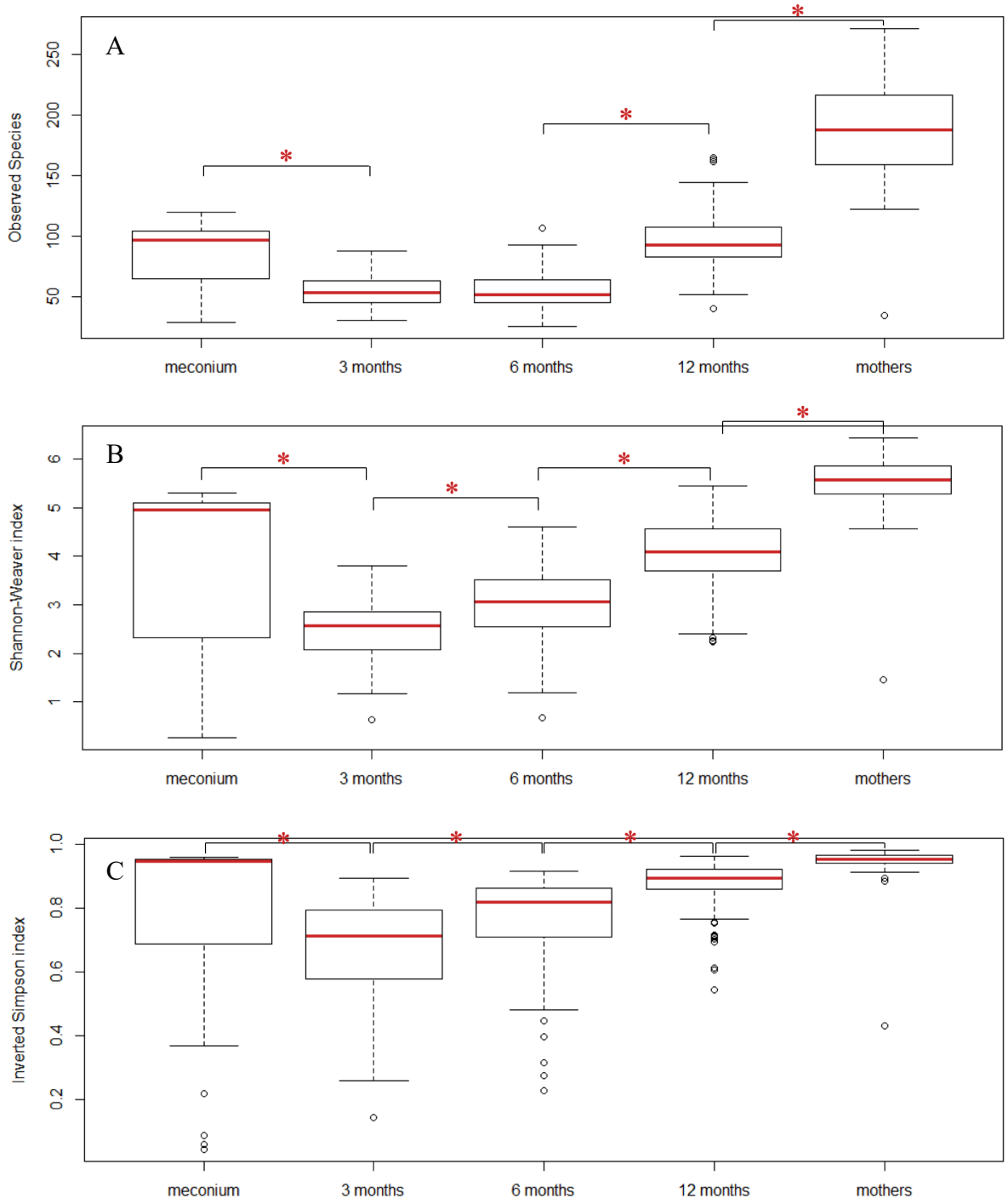


Figure 3.2: Alpha diversity index plots illustrating the diversity within each age group (Meconium (n= 55), 3 months (n = 56), 6 months (n = 56), 12 months (n = 167), mothers (n = 58)). A) illustrates the observed species richness, B) shows the distribution of the Shannon-Weaver index, while C) represents the inverted Simpson index. The index results were acquired using the QIIME pipeline and plotted in RStudio. The red lines represent the median value of the index within the groups, and the stars mark significant differences, with a p-value below 0.05, computed by running a paired t-test between the age groups. The average scores, standard deviation, and median values for all alpha indices are listed in Table G.1 in Appendix G.

3.1.3 **Beta diversity**

Binary Jaccard and Bray-Curtis beta diversity indices presented as principal coordinate analysis (PCoA) plots in Figure 3.3. PCoA plot using the Euclidian distance matrix is given in Figure H.1 in Appendix H. In the binary Jaccard plot, the mother samples clustered alone, indicating that they share a common set of OTUs that also differ from the rest of the age groups. The 12-month samples are separated into two clusters along both PCoA-axis. One of the 12 months clusters clustered with 6-month samples, implying that these 12-month samples have more OTUs common with 6 months than the other 12 months. The meconium and 3-month samples are dispersed in the plot, showing variations between samples in which OTUs that were present.

The Bray-Curtis dissimilarity also showed the meconium samples being dispersed, which illustrates dissimilarity between OTU abundance in the meconium samples. The 3-month samples were also dispersed in the plot. The two clusters of the 12-month samples observed in the binary Jaccard plot are not observed in the Bray-Curtis plot, and the 12-month samples cluster together with samples from the mothers, indicating a similar abundance of species. However, the 12-month samples are more dispersed than the mother samples. The samples from 6 months cluster together in the middle of the plot.

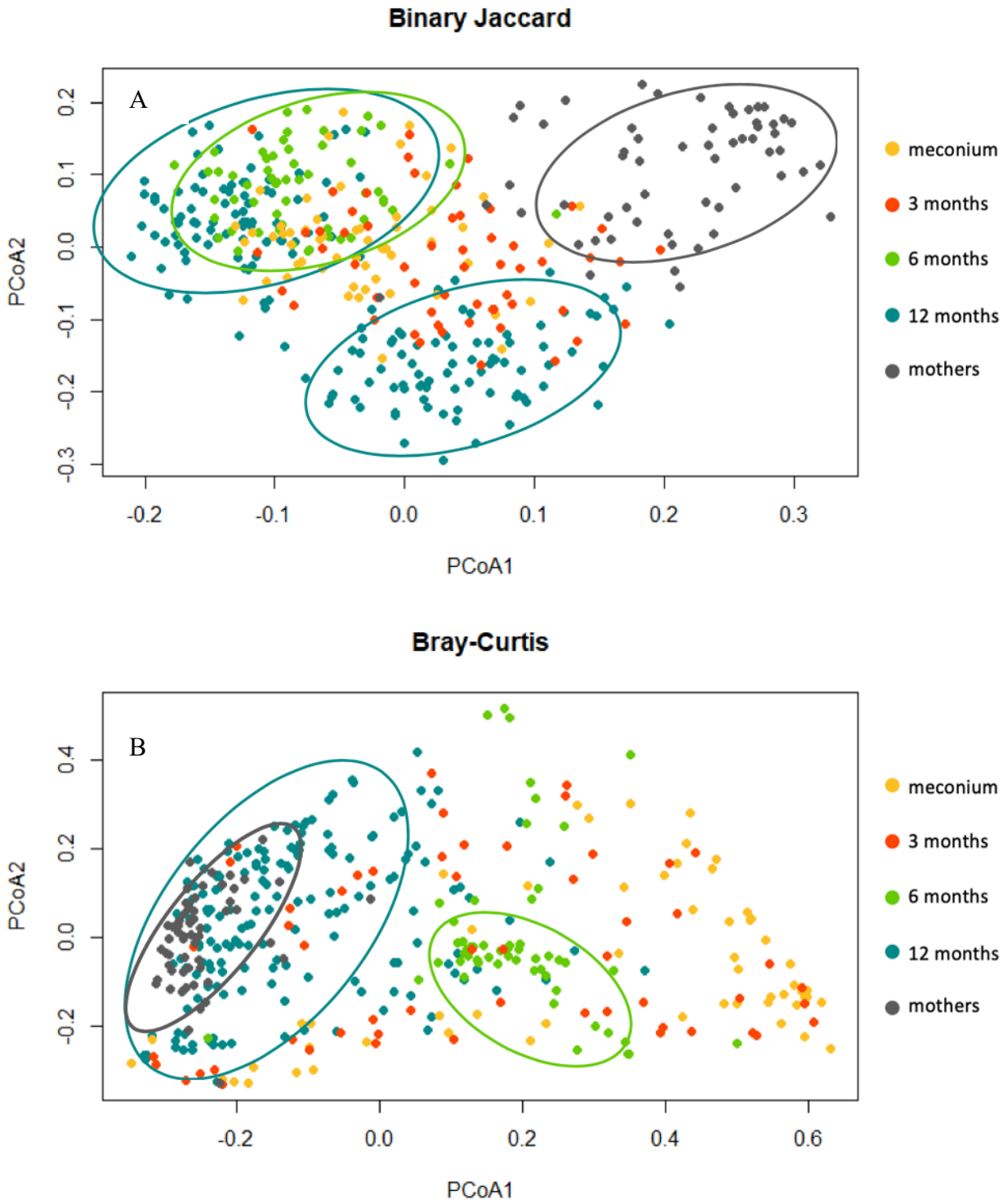


Figure 3.3: Beta diversity plots made using A) the binary Jaccard distance matrix, and B) Bray-Curtis dissimilatory matrices. The data was acquired from the sequencing results on an order level using the QIIME pipeline and further processed using RStudio. The meconium, 3-, 6-, 12 months, and mother samples were respectively colored in the yellow, red, green, blue, and gray. Circles colored with the corresponding age color, are used to mark clusters of samples. The number of samples from each age group used in these plots were 55 meconium samples, 56 samples from both 3 and 6 months, 167 samples from 12 months, and 58 mother samples.

3.2 Short-chain fatty acids

3.2.1 Short-chain fatty acid profiles

The average SCFA composition in each age group was calculated and presented in Figure 3.4. Acetic acid dominated at all ages, however, decreasing significantly as of 3 months (3 to 6 months, $p = 3.39*10^{-09}$; 6 to 12 months, $p = 2.05*10^{-13}$; 12 months to mothers; $p = 2.99*10^{-10}$; paired t-test, unadjusted). Propionic acid was most abundant in the meconium samples with 13.9%, followed by a significant decrease to 4.9% in the 3 months samples ($p = 1.21*10^{-09}$, paired t-test, unadjusted), before increasing in abundance towards the mothers, being significant from 3 to 6 months ($p = 1.19*10^{-05}$, paired t-test, unadjusted) and 6 to 12 months ($p = 0.022$, paired t-test, unadjusted). The prevalence of butyrate was constantly increasing throughout the age groups from 0.8% in the meconium samples to 22.2% in the samples of the mothers, being significant between every age group (meconium to 3 months, $p = 7.79*10^{-06}$; 3 to 6 months, $p = 2.85*10^{-05}$; 6 to 12 months, $p = 2.03*10^{-13}$; 12 months to mothers, $p = 0.003$; paired t-test, unadjusted). For the less abundant isobutyric, isovaleric and valeric acid, an increase of acid abundance was observed as the infant grew older and were at large in the mother group. Technical issues of the SCFA analysis, such as the description of the calibration process, are given in Appendix I.

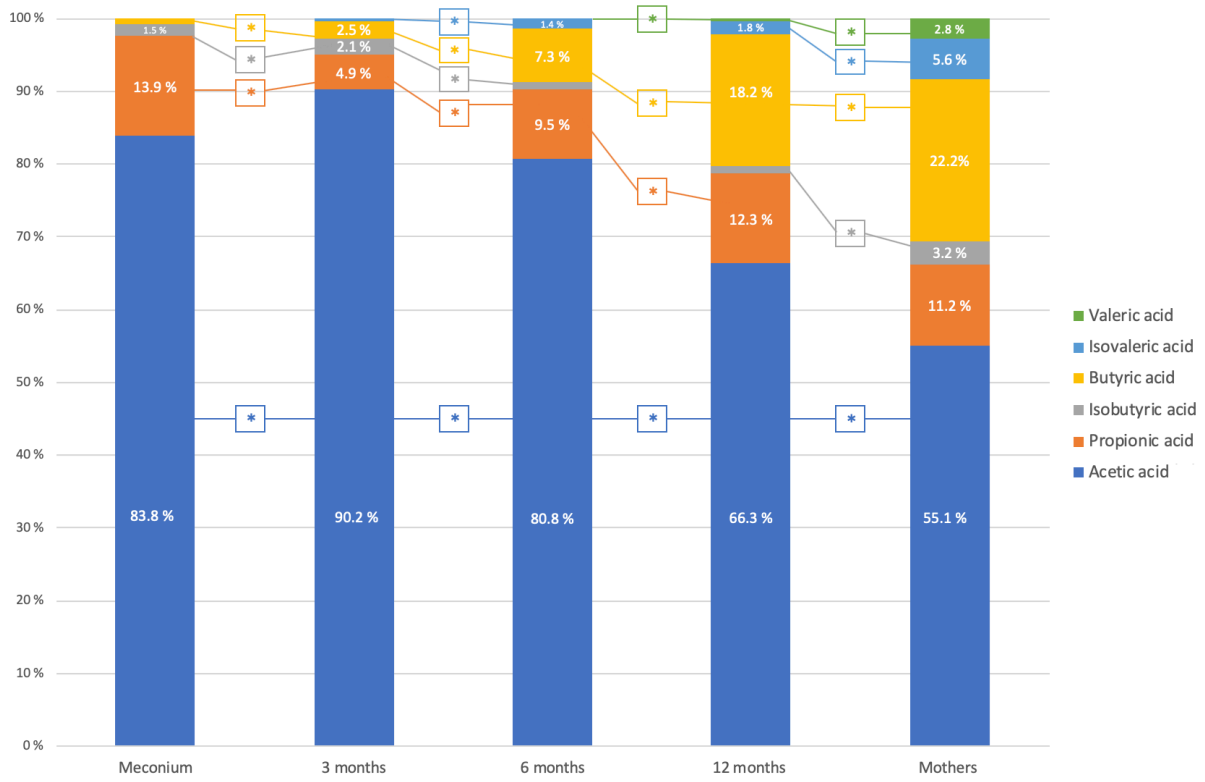


Figure 3.4: Average percentage of SCFA at different ages. In total were 60 samples form meconium, 58 samples from months, 59 samples from 6 months, 179 samples from 12 months, and 60 samples from the mothers included. Two samples, one from 3- and one from the 12-month group, were considered outliers and excluded from the statically SCFAs analysis. Significant differences between age groups, marked with a star, was determined using paired t-test in RStudio.

3.2.2 The ratio between butyrate and propionate

Logarithmic values of the butyrate/propionate-ratios were calculated and presented as boxplots for each age group in Figure 3.5. Significant differences were found between all ages, except between 3 to 6 months (meconium to 3 months, $p = 6.22*10^{-07}$; 3 to 6 months, $p = 0.59$; 6 to 12 months, $p = 2.03*10^{-04}$; 12 months to mothers; $p = 2.48*10^{-03}$; paired t-test, unadjusted). In the meconium samples, there was a predominance of samples with higher levels of propionate than butyrate, only 6.3% (3/48) of the samples showed a higher ratio of butyrate. The percentage of samples with higher levels of butyrate increased throughout the sampling times, and out of the 3 months samples, 34.2% (13/38) had a higher level of butyrate than propionate. The higher levels of butyrate increased further, however insignificantly, to comprise 37.2% (19/51) of the samples from 6 months. By 12 months, 63.7% (114/179) of the samples had higher levels of

butyrate compared to propionate, and the same applied to 91.5% (54/59) of samples from the mothers.

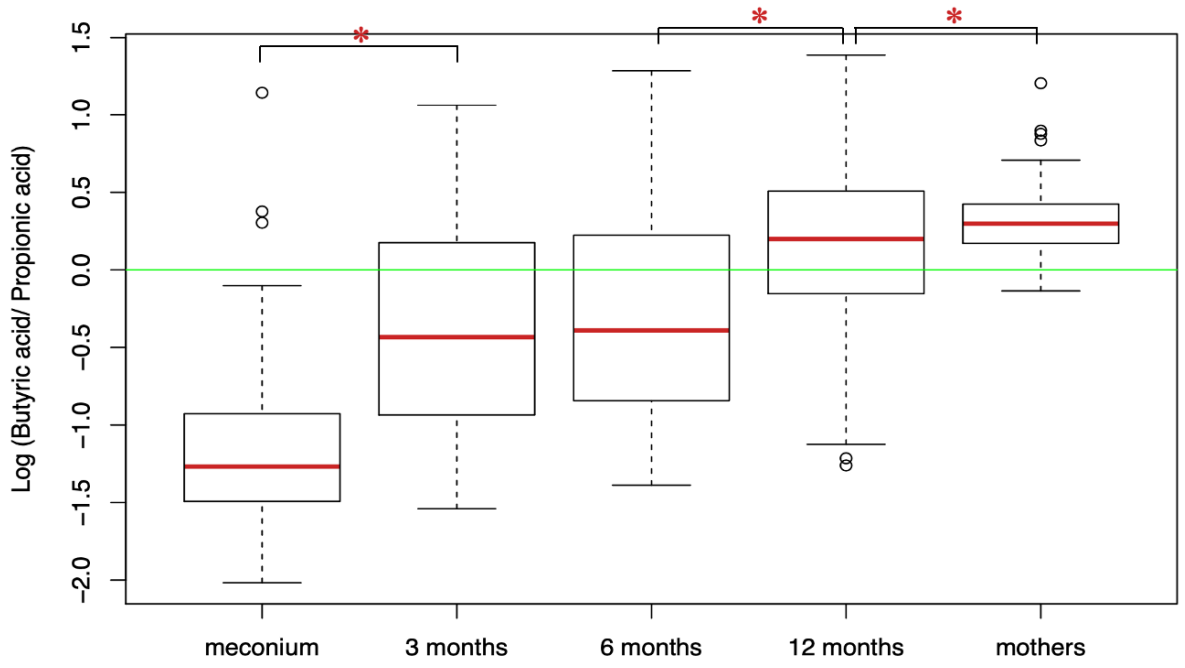


Figure 3.5: Log values of the butyrate/propionate-ratios for samples in the different age groups. Samples, where either propionic-, butyric acid, or both were undetectable by the GC, were excluded. The plot was based on 48 meconium samples, 38 samples from 3 months, 51 samples from 6 months, 179 samples from 12 months, and 59 samples from the mother group. The red line represents the median value of the age group. The green line, going through zero, illustrates the critical value of equal amounts butyric and propionic acid in the samples. Above the critical line, with positive values, are the samples with higher proportions of butyric than propionic acid, whereas underneath the green line, with negative values, are the samples with higher propionic acid concentrations compared to butyric. The stars mark significant differences in average group values, where the is p-value below 0.05, computed by running a paired t-test in RStudio.

3.3 Immune cells

3.3.1 Immune cell profile at 12-months

The immune data were categorized into 28 groups of immune cells. In total, 67 infants of the 12 months group were analyzed, and the average percentage distribution of immune cells is presented in Figure 3.6. The neutrophils dominated the samples, and on average did 30.8% of the analyzed immune cells belong to this group. Following the neutrophils, were the naïve cells

dominating, where the naïve CD4+ T-cells, naïve B-cells, and naïve CD8+ T-cell spoke for 20.9%, 12.4% and 7.7% of the total immune cell composition, respectively.

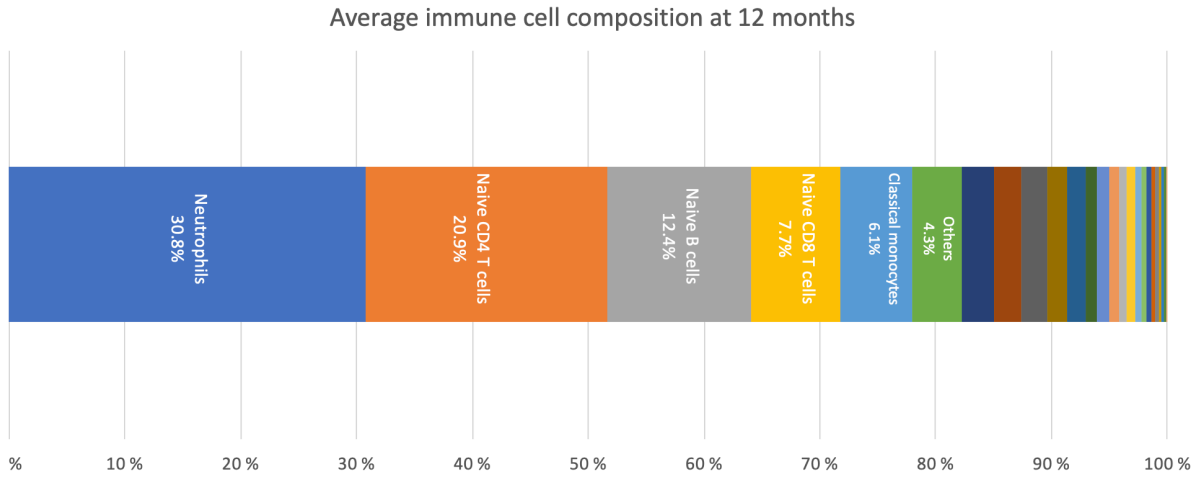


Figure 3.6: Overview of the average immune cell composition at 12 months. The cells analyzed by CyTOF2 were categorized into 28 groups of immune cells and summed to 100%. The groups with the highest average percentages are presented, and a table that lists the average values and standard deviation of all groups are given in Table J.1 in Appendix J.

3.4 Correlation analysis

3.4.1 Correlation between bacterial and short-chain fatty acid composition

Correlation between bacterial orders and short-chain fatty acid composition at different ages is presented as a heatmap in Figure 3.7. At 3 months there was a significant positive correlation between butyric acid and Lactobacillales ($p = 0.02$, Spearman, BH-adjusted). The same positive correlation was observed at 6 months ($p = 0.001$, Spearman, BH-adjusted) and 12 months ($p = 0.00$, Spearman, BH-adjusted). At 12 months there was also a significant negative correlation between Lactobacillales and propionic acid ($p = 2.63*10^{-10,}$ Spearman, BH-adjusted). For Erysipelotrichales, another order of the Firmicutes phylum, significant positive correlation with propionic acid ($p = 0.001$, Spearman, BH-adjusted), and negative correlation with butyric acid ($p = 0.004$, Spearman, BH-adjusted) were observed in samples from infants 12 months of age.

Actinomycetales were shown to have significant negative correlations with butyric acid at 3 months ($p = 2.3*10^{-04,}$ Spearman, BH-adjusted) and 6 months ($p = 0.009$, Spearman, BH-adjusted), as well as being positively correlated with acetic acid at 12 months (0.008, Spearman, BH-adjusted). Solirubrobacteriales was positively associated with propionic acid in both 6

months ($p = 0.0002$, Spearman, BH-adjusted) and 12 months ($p = 0.00$, Spearman, BH-adjusted). In the 12 months samples, Solirubrobacteriales are also found to correlate negatively with butyric acid ($p = 0.0006$, Spearman, BH-adjusted).

Mycoplasmatales, order within the phylum of Tenericutes, were found to have significant correlations with acetic acid ($p = 0.035$, Spearman, BH-adjusted), butyric acid ($p = 0.02$, Spearman, BH-adjusted) and isovaleric acid ($p = 0.036$, Spearman, BH-adjusted). Both acetic acid and isovaleric being positive associations, while butyric was negative.

Figure 3.7: Spearman correlation between bacterial order composition and tested SCFAs. Positive correlations are marked in green and negative in red. White represents no correlation. Bacterial orders retrieved from the QIIME pipeline are listed along the y-axis, while the x-axis is divided into age groups each with a set of SCFAs. The plot is based on 55 samples from the meconium group and the 3 months group, 56 samples from 6 months, 166 samples from 12 months, and 58 samples from the mother-group.

3.4.2 **Correlation between bacterial composition and immune cells at 12 months of age**

The Spearman correlation between bacterial composition and immune cells at 12 months of age are visualized in Figure 3.8. Xanthomonadales are significantly positively correlated with proinflammatory monocytes ($p = 0.045$, Spearman, BH-adjusted). Methylophiliales and classical monocytes are significantly positively correlated, with a BH-adjusted p-value of 0.049. In addition, Methylophiliales significant negatively correlated with naïve CD4+ T-cells ($p = 0.004$, Spearman, BH-adjusted) and naïve regulatory T-cells ($p = 0.018$, Spearman, BH-adjusted). Although non-significant, the order Methylococcales showed a strong trend towards being negatively correlated with proinflammatory monocytes ($p = 0.061$, Spearman, BH-adjusted) at 12 months.

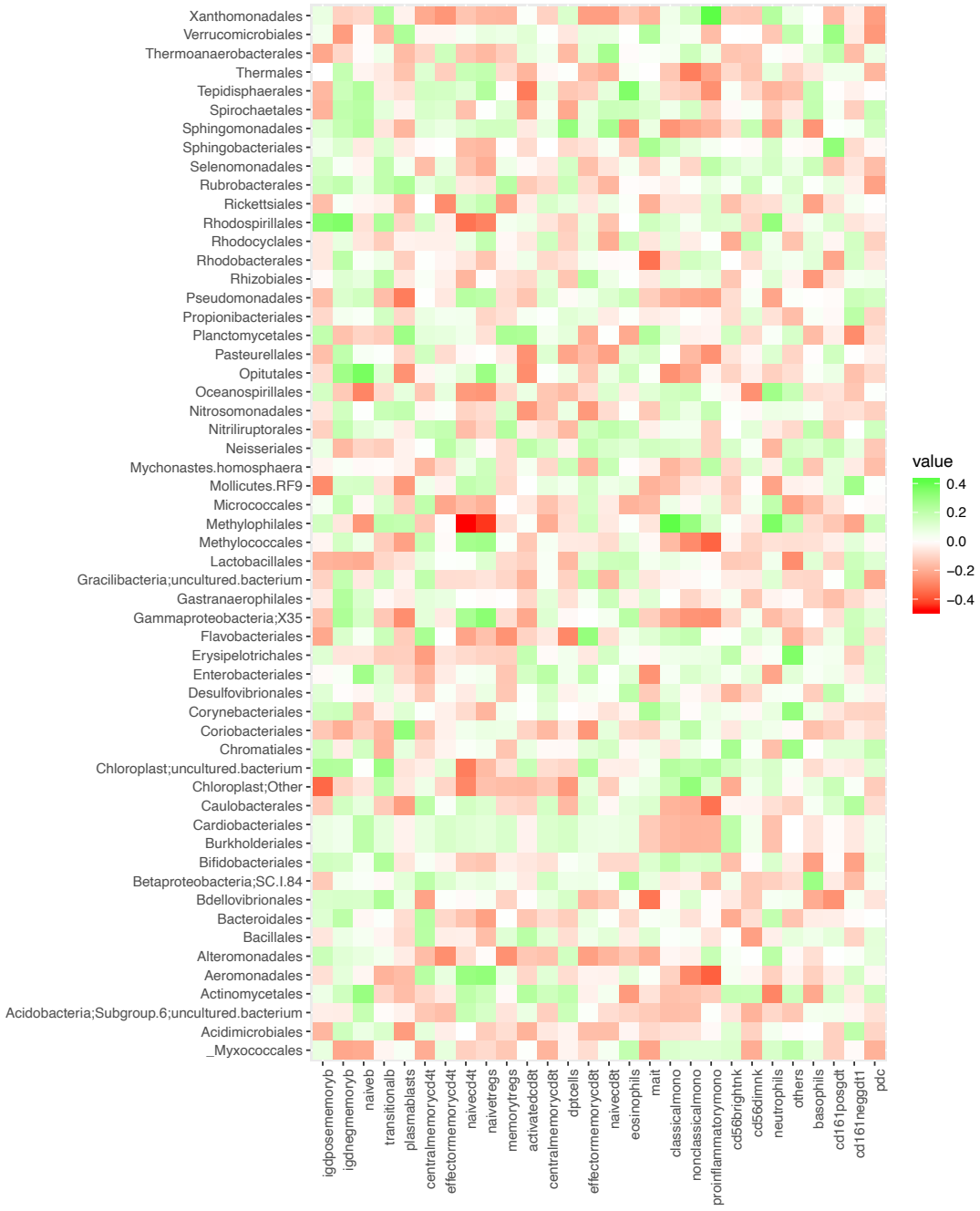


Figure 3.8: Spearman correlation between bacterial composition and immune cells at 12 months of age. Positive correlations are marked in green and negative correlations in red. White represents no correlation. Bacterial orders retrieved from the QIIME pipeline are listed along the y-axis, while the 28 groupings of immune cells are present along the x-axis. The plot is based on 63 samples.

It was finally assessed whether or not the 67 infants with immune data were representative of the taxonomic microbiota distribution for all 12 months samples. In the Euclidean distance plot of 12-month samples, Figure K.1 in Appendix K, no patterns were detected between the samples with or without immune cell data, and the 67 infants distributed evenly across the two clusters that were observed in the plot.

Spearman rank correlation between SCFA and immune cells was also conducted, however as there were no significant correlations the plot was not included. A description of trends is found in Appendix L.

4. Discussion

4.1. Bacteria and SCFA in relation to immune cells at 12-months of age

In the present study, correlations between types of immune cells and bacterial taxa and SCFAs were explored. No statistically significant correlations were found between the immune cell profiles and SCFAs, nor between any immune cells and the most dominant bacterial taxa. However, interesting significant correlations were reviled between smaller bacterial taxa, such as Methylophiliales and Xanthomonadales.

4.1.1 Environmental bacteria and the immune system

Despite the lack of significant immune correlations to the most dominant commensal bacterial orders, one interesting finding was that Methylophiliales was positively correlated to classical monocytes and had negative correlations to both naïve CD4+ T-cells and naïve regulatory T-cells (Figure 3.8). Hence, Methylophilales might have an inflammatory effect on the immune system of the host. The information on this bacterial taxon in the human gastrointestinal tract is limited, as the bacterium is primarily associated with aquatic environments, sludge and plants (Doronina, Kaparullina, & Trotsenko, 2014), however its presence in the gut has been reported in several studies (Michels et al., 2019; Stearns et al., 2011). The presence of *Methylophilaceae* on human skin was, in a study by Song et al. (Song et al., 2013), linked to dog-owning households, and a transfer route of oral-skin transmission was proposed, as the *Methylphiliaceae* was derived from canine oral bacterial communities. A study on effects of dog ownership on immune genotype development has linked exposure to a dog with increased IL-10 and IL-13 cytokine levels at one year of age, compared to infants not exposed to dogs (Gern et al., 2004). Both IL-10 and IL-13 are derived from T-cells, with anti-inflammatory properties. However, the anti-inflammatory effects of IL-13 are circumstantial as IL-13 has been linked to allergic inflammation (Wynn, 2003). The present finding of the correlation between Methylophilales and proinflammatory monocytes might be associated with allergic reactions to environmental factors, such as dog-keeping. More studies are needed to confirm this correlation, in addition to more studies on Methylophilales in general to develop the full picture of its role in the gut microbiota.

Further, the Methylococcales order showed a strong trend towards being negatively correlated with proinflammatory monocytes at 12-months. This was an interesting correlation as *Methylococcus capsulatus* Bath is a non-commensal soil bacterium found to adhere to human monocyte-derived dendritic cells, affecting their maturation and cytokine profile and the subsequent T-cell polarization (Indrelid, Kleiveland, Holst, Jacobsen, & Lea, 2017). In murine colitis models, *M. capsulatus* have exhibited anti-inflammatory effects (Kleiveland et al., 2013), which might explain lower levels of proinflammatory monocytes in the infants of the present study that constituted higher gut levels of Methylococcales.

The bacterial orders described in this section, Methylophiliales and Methylococcales, are environmental-associated bacteria and highlight the importance of assessing non-commensal environmental bacteria when considering the gut microbiota's impact on the host's immune responses, immunoregulation, and immune development.

4.1.2 **Xanthomonadales and the immune system**

The order Xanthomonadales was found to have a statistically significant positive correlation with proinflammatory monocytes (Figure 3.8). The genus *Stenotrophomonas* of Xanthomonadales was present in all age groups (Table F.1 in Appendix F) and is a known multidrug-resistant opportunistic pathogen associated with respiratory infections in immunocompromised humans (Brooke, 2012). A study of *Stenotrophomonas maltophia* infected mice showed indications of a proinflammatory cytokine response, and recruitment of monocytes (Brooke, 2012). Another study displayed an immature immune response to the same bacterial specie and reported of proinflammatory cytokine responses in neonatal monocytes (Hartel et al., 2013). The findings of these two studies both align with the correlation that was revealed in the present study. However, both underline that bacterial infection of *S. maltophia* primarily occurs in immunocompromised patients, such as neonates. Nevertheless, by 12-months the infants are no longer regarded as immunocompromised as the immune system is generally fully developed (Gleeson & Cripps, 2004). The taxon's prevalence in the gut does not necessarily implicate an infection of a *Xanthomonadales sp.* and opportunistic pathogens are known to be important contributors to the gut microbiota. Therefore, it can be questioned whether the immune response to *Xanthomonadales sp*. under non-infectious circumstances corresponds to the immune response during an infection. However, from the correlation discovered in the present study, it seems reasonable to hypothesize that the prevalence of *Xanthomonadales sp*. in the gut might be linked to a higher relative abundance of

proinflammatory monocytes, thus introducing a proinflammatory immunological response in the host.

4.1.3 **Lack of correlations to immune cell profiles**

Noteworthy was the lack of significant correlations between the most dominant bacterial taxa at 12 months and the immune cell profiles. Members of the commensal microbiota, such as Clostridia*,* Bacteroidetes, and *Bifidobacterium,* were found to affect the function and numbers of regulatory T-cells (Hart et al., 2004; Kelly, Conway, & Aminov, 2005). One example is the effect of Clostridia on activation of naïve T-cells where Clostridial species through direct or indirect measures, affect naïve CD4+ T-cells to differentiate to either effector- or regulatory T-cells depending on whether or not the gut is in a state of inflammation (Nagano, Itoh, & Honda, 2012). Another example is the *Bacteroides fragilis*, where its capsular Polysaccharide A has been shown to facilitate functional maturation of regulatory T-cells in mice, and to possess a suppressive effect of the proinflammatory cellular responses, allowing *B. fragilis'* intimate colonization on the intestinal epithelium (Jiang et al., 2017).

This particular study did not reveal any significant correlations between the groups of immune cells and SCFAs, which was surprising as SCFAs have been found to regulate the functions of almost every type of immune cell by altering gene expression, differentiation, proliferation, chemotaxis, and apoptosis (M. Sun, Wu, Liu, & Cong, 2017). SCFAs were found to facilitate T-cell differentiation through the regulation of dendritic cell functions. Exposed to butyrate, can the dendritic cells facilitate naïve CD4+ T-cell differentiation to produce regulatory T-cells and suppress differentiation to effector CD4+ T-cells (Gurav et al., 2015). Propionate (Vinolo et al., 2011) and butyrate (Chang, Hao, Offermanns, & Medzhitov, 2014) have both been connected to impairment of lipopolysaccharide-induced proinflammatory mediators (Li et al., 2018; M. Sun et al., 2017), and acetate may inhibit the secretion of inflammatory cytokines from cells such as B-cells, monocytes, natural killer cells and T-cells (Masui et al., 2013; M. Sun et al., 2017). All the above-mentioned connections of bacterial fermentation products and immune cells give an indication of the importance of SCFAs for maintaining gut homeostasis, and it would, therefore, have been expected to observe correlations between regulatory T-cells and SCFAs in the present study as well.

Nevertheless, all the studies mentioned were conducted using cell cultures, murine-, or rat models. These systems are more enclosed, where it is possible to control more of the factors interfering with immune expression (Nguyen, Vieira-Silva, Liston, & Raes, 2015). The present study was, on the other hand, conducted by analysis of material from free-living humans, with uncontrollable numbers of confounding elements and inter-individual variations, all acting on the host's immune system profiles. The system's complexity might hide correlations between SCFAs and specific immune cells and be the reason for the high p-values observed in the present study.

4.2. Bacteria and SCFAs

4.2.1 Enterobacteriales, the most dominant bacterial order in meconium

Enterobacteriales was observed to dominate the meconium samples (Figure 3.1). Contradictory to the findings of the present study, Moles and colleagues found Firmicutes to be the dominating phyla in preterm neonates, with most representatives belonging to the class of Bacilli (Moles et al., 2013). However, they observed that *Enterobacteriaceae* become the dominating family the following week after birth. Consequently, the high prevalence of Enterobacteriales in the present study might be connected to the meconium samples collection time, as samples collected later may have a higher prevalence of Enterobacteriales. Another possible explanation is that the high prevalence of Enterobacteriales could be the result of its route of transmission. *Escherichia/Shigella* was in the present study the most prominent genera in the meconium samples (19.9%, Figure F.2 in Appendix F), and previous studies have found *Escherichia* to enter the mother's genital tract by the perineum (Asnicar et al., 2017; Tameliene et al., 2012), thereby allowing for its early colonization and relatively high prevalence in neonatal gut.

As a consequence of the high prevalence of *Escherichia* and other members of Enterobacteriales, the gut lumen is depleted of oxygen, thus, constructing an anaerobic and reduced environment that favors the growth of the strictly anaerobic bacteria such as *Bifidobacterium, Clostridium*, and, *Bacteroides*. (Backhed et al., 2015; Matamoros, Gras-Leguen, Le Vacon, Potel, & de La Cochetiere, 2013; S. Wang et al., 2020).

Moles et al. (Moles et al., 2013) reported a larger prevalence of Bacilli than what was found in the present study, as Bacilliales and Lactobacilliales only constituted up 2.2% and 4.5% of the microbiome in the meconium samples, respectively (Figure 3.1). Instead, Clostridiales was in

the present study, reported as the second to largest bacterial order (19.4%) and the largest order within Firmicutes. The high abundance of Clostridiales might be a result of the mode of delivery and the hospital environment, as the *Clostridium sp.* such as *Clostridium difficile* is found to be higher in caesarian delivered infants than infants born vaginally (Penders et al., 2006). *C. difficile* is considered a bacterium of the hospital environment, and have been found on the hands and in the stool of healthy hospital personnel (Penders et al., 2006).

4.2.2 **Increase of Bifidobacteriales to 3 months**

Bifidobacteriales was demonstrated to dominate the gut at 3 months, constituting 40.4% of the microbiota on average (Figure 3.1). *Bifidobacterium'*s high prevalence is likely connected to breastfeeding as an overrepresentation of *Bifidobacterium* has been observed in breastfed infants compared to adults (Sela & Mills, 2010). This is connected to the Bifidobacterial ability to utilize non-digestible oligosaccharides from the breast milk, where the HMOs of the breast milk works as a growth factor and becomes a core element of the microbial niche (Sela & Mills, 2010). The prevalence pattern of Bifidobacteriales is accompanied by the pattern of acetic acid ratio at all ages (Figure 3.1, Figure 3.4), although no statistically significant correlations were detected in any of the age groups. One reason for the coinciding developmental pattern might be Bifidobacterial acetogenesis, as *Bifidobacteria* produce acetic acid in addition to lactic acid as a fermentative end product. Statistically significant correlations between *Bifidobacterium* and acetic acid have previously been described by Nagpal et al. (Nagpal et al., 2017). Bifidobacteirales aids in the development of the microbial niches of the gut, as the fermentative products are further metabolized by other bacteria in the gut (Moens, Verce, & De Vuyst, 2017).

4.2.3 **Correlation of Lactobacilliales and butyric acid**

Together with Bifidobacteriales, Lactobacilliales is also essential for the fermentation of non-digestible carbohydrates in the infant gut. Neither of the bacterial orders is found to be producers of butyrate. However, they are major producers of lactic acid (Moens et al., 2017). In the present study, members of Lactobacilliales order was found to be positively correlated with butyrate at 3-, 6- and 12 months (Figure 3.7). The correlation of Lactobacilliales and butyrate might stem from cross-feeding interactions between *Lactobacilli* and lactate-utilizing butyrate-producing bacteria in the gut, as described in a study by Moens et al. (Moens et al., 2017). *Lactobacillus acidophilus* first degrades oligofructose to free fructose and metabolizes this further, leaving lactate as a metabolic end-product. In the study of Moens et al., *Anaerostipes caccae* was found to utilize both the fructose and lactate as a source of energy,

thereby being an example of both substrate and metabolite cross-feeding (Moens et al., 2017; Smith, Shorten, Altermann, Roy, & McNabb, 2019). There has also been reported of cross-feeding mechanisms between lactate-producing bacteria and the butyrate-producing *Eubacterium hallii* (Duncan, Louis, & Flint, 2004). Both the genera *Anaerostipes* and *E. hallii* have been detected in all age groups in the present study with the respective prevalence range of [0.08-1.54]% and [0.003–0.74]% (Table F.1 in Appendix F). The metabolic cross-feeding could have substantial effects on the SCFA balance, as it occurs mostly to form butyric acid from acetate or lactate. More seldom is butyrate converted to propionate, and in very few cases are propionate converted to acetate (Rios-Covian, Salazar, Gueimonde, & de los Reyes-Gavilan, 2017).

4.2.4 **Clostridiales and Bacteriodetes and SCFAs**

Together with the statistically significant increase of Firmicutes from 3 months, and the order Clostridiales in particular, a simultaneous statistically significant increase of butyric acid was observed (Figure 3.1, Figure 3.4). The corresponding increase was reasonable as the main butyrate-producers belong to the *Clostridium cluster IV* and *Clostridium cluster XIVa* (Parada Venegas et al., 2019). Clostridiales was, in the present study, observed to have a steady increase throughout the first year of the infant's life and to be at large in the adult-gut of the mothers. According to literature, the shift from infant to an adult-like microbiota is represented by a shift from a gut dominated by Proteobacteria and Actinobacteria, to a gut dominated by Firmicutes and Bacteroidetes (Milani et al., 2017). The same was observed in the present study, as Proteobacteria heavily dominated the gut of the neonates, and Actinobacteria the gut of the 3 months old. Both taxa were reduced to collectively constitute only about 7% of the gut microbiota in the mothers, where the major bacterial share belonged to the phyla of Firmicutes, followed by Bacteroides (Figure F.1 in Appendix F).

In the same manner as Clostridiales, fractions of Bacteriodetes followed the fraction-pattern of propionate at all ages (Figure 3.1, Figure 3.4). The corresponding patterns are likely as Bacteroidetes is able to produce both acetic acid and propionic acid depending on the available substrate. Propionate is, however, preferred, and for that reason is Bacteroidetes considered to be the largest propionate-producing group in the gut microbiota (Rios-Covian et al., 2017). However, in the present study, no statistically significant correlations between Bacteroidales and propionate were detected at any age.

The increase of the butyrate-producing Clostridia, and Firmicutes as a whole, might be the basis for the observed increased proportions of butyrate against propionate (Figure 3.5) during the first year of life and to the samples of the mothers. The relative abundance of both SCFAs increased throughout the sampling ages. However, the increase was more profound for butyrate than propionate, reflecting the larger increase in the relative abundance of Clostridiales than Bacteroidales (Figure 3.1).

4.3 Methodological considerations

4.3.1 Strengths and limitations of study design

The present study has the strength of including a large number of sample pairs. This study is involved in the larger PreventADALL cohort, which has collected extensive biological and lifestyle information on all the infants and their families. The information gathered enables further studies of different factors in the same infants that later can be linked to generate an overall picture of infants' health in the general population. The present study is an example of this type of study collaboration, where data on 67 infants' immunological profiles were obtained to see in connection to the bacterial composition and SCFAs profiles that were analyzed.

Due to the timing and the expenses connected to immune cell profiling were not the immune profiles of all the 180 12-month infants analyzed accessible. However, the 67 infants where the immune profiles were accessible were studied longitudinally. Ideally, a complete set for all infants would have strengthened the study design with an equal number of samples of all ages. Although making the sample numbers uneven, all the 180 12-month samples were included in the statistical analysis where it was fit, as a larger number of samples results in a more representative dataset for the general age group.

4.3.2 Gas chromatographical challenges

There was no information on the exact wet weight of the fecal material before the samples were diluted in DNA/RNA shield buffer. As the initial concentration of fecal material is unknown, extraction of a given volume from each sample leads to different amounts of fecal material being analyzes. The initial concentration of the fecal sample would affect the detection rates of SCFAs, as a larger amount of inserted sample results in higher levels of detected SCFAs. It would, therefore, be inaccurate to calculate the absolute concentration of SCFAs per gram wet

weight fecal material, as all samples contained different input concentrations of fecal material. Even though the accurate absolute values of SCFAs were inaccessible, an internal standard was included during sample preparation. The internal standard of 2-methylvaleric acid was used for sample validation. However, complications occurred when an unknown compound was observed having the exact retention time as the internal standard (Appendix I). No extensive analyses of the overlapping peak were preformed, but ideally, mass spectroscopy should have been conducted for compound identification.

4.3.3 Fecal material as a proxy for bacterial and SCFAs gut composition

Fecal samples were in the present study and have frequently been used as an approximation for the luminal microbiome samples. This proxy might be inconvenient as there are significant differences in the mucosal and the fecal microbiome, and the fecal samples give an imprecise reflection of the bacterial mucosal niches across different sites in the intestine (Tang et al., 2020). Fecal material implements the need for sample preservation, as it is impossible to analyze the fresh fecal sample immediately after collection. The same challenges of using fecal material as a proxy for bacterial microbiome follows when it is used as a proxy for luminal SCFAs levels. As the SCFAs are volatile, there is a need for good sample preservation until the time of analysis to avoid sample deterioration (Primec et al., 2017). The complexity of the biological material presents a challenge in sample preparation, and different methods for pretreatment methods have been proposed, as described in 1.4.3. Lastly, only 5-10% of the SCFAs produced is estimated to be in fecal secrete, while the rest is absorbed by the colonic epithelial cells or further metabolized in bacterial cross-feeding (Garcia-Villalba et al., 2012; Henningsson et al., 2001; Primec et al., 2017; Rios-Covian et al., 2017).

The use of fecal material is, however, the best alternative for human studies, as it is cost-efficient, allows for repeated sampling, and other methods for sampling, such as biopsies, are generally considered invasive (Tang et al., 2020). Other alternatives to human fecal studies are gut microbiome studies using animal models, such as mice, where biopsies are possible. Although the murine models share anatomical similarities to humans, it cannot fully recapitulate the human systems, and differences in the gut microbiota composition have been described between murine models and humans (Nguyen et al., 2015). Besides, animal studies will never reflect the variations that occur in a human life which contributes to shaping the microbiota, since multiple factors, such as genetic background, delivery mode, diet, and social activities, are absent in these models (Nguyen et al., 2015).

The present study has presented the bacterial, SCFAs, and immune cell data as percentages representing the total sample variation. Unfortunately, were the total amount of bacterial taxa, SCFAs, or immune cell type not taken into consideration. Although it is possible to discuss the relative abundance, such an oversimplification will hamper the opportunity to investigate absolute abundance in the samples. When considering relative abundance, it is of importance to note that the decrease or increase of one SCFA or bacterial specie does not necessarily reflect the actual change in true abundance. Instead, the change in relative abundance might be caused by other SCFA or bacterial taxon becoming more or less prevalent. The different levels of biomass and SCFAs levels between age groups are also neglected. Both the biomass and the SCFA levels increase through the first period of life, which is not reflected when the data are presented by percentage. Anyhow, by doing this simplification, it facilitates a simple method for comparison of shifts in relative abundance across the age groups.

4.3.4 **Bias as result of uneven 16S library preparation**

The extraction of the 12-month samples was conducted using two plates, treated separately of each other, both during extraction and the PCR treatments that followed. The distribution of the samples between the two plates was later rediscovered as two clusters in the Euclidean distance plot (Figure H.1 in Appendix H) of the 12 months infants, as well as in the binary Jaccard plot (Figure 3.3.A). As the sample clustering coincided with the distribution of samples on the plates, no further statistical investigations were performed for pattern determination. Whether the bias stem from the extraction of DNA or PCR is uncertain, as both sample plates were processed independently of each other. Extraction of genomic DNA from complex samples is prone for bias. The wide variety of cellular membrane structures of the different taxa challenges a representative extraction of the microbial community (Teng et al., 2018). Nevertheless, a study by Teng and colleagues found the reproducibility of parallel DNA extraction on oral microbiota-samples to be of general agreement (Teng et al., 2018). The potential bias from the extraction process could be further amplified during the PCR reactions, as PCR is known to introduce artifacts such as sequencing errors and bias from unequal amplification (Acinas, Sarma-Rupavtarm, Klepac-Ceraj, & Polz, 2005). Despite the two clusters observed, is the taxonomic composition of the 12-month samples deemed trustworthy and representative for the age group in the correlation analysis with immune cell profiles. The data was considered trustworthy due to the samples with immune cell composition being evenly distrusted between the two clusters of the Euclidean distance plot (Figure K.1 in Appendix K). The differences might thereby not have had substantial effects on the output of the correlation analysis.

However, these laboratory artifacts need to be further investigated, to avoid introducing any bias from the extraction or the PCR procedures in future sample preparation.

4.3.5 Implications of contamination as a result of uneven amplification of low biomass samples

The lowest intra-individual diversity was observed at 3 months, and not in the meconium samples (Figure 3.2), as expected since meconium are generally seen to have limited diversity (Avershina et al., 2014; Milani et al., 2017). Studies have demonstrated a steady increase in alpha diversity throughout infancy (Avershina et al., 2014; Backhed et al., 2015; Hua, Goedert, Pu, Yu, & Shi, 2016). The large alpha diversity observed in the meconium samples in this study might be a result of methodical deviation, as the meconium samples were run for 5 more PCR cycles than the remainder. Thereby opening the possibility for greater amplification of bacterial contaminants in meconium samples than in the other age groups. The problem of exogenous DNA contamination in niches with low biomass is well-known (Karstens et al., 2019). Cross-contamination of simultaneous processed high biomass samples, and bacterial DNA that contaminates the reagents and components of DNA extraction kits, are both possible sources of contaminating exogenous DNA (Eisenhofer et al., 2019; Perez-Munoz et al., 2017). Because of the initially low biomass, the contaminating DNA constitutes a larger proportion of the bacterial taxa present in the meconium samples. A normal contaminant, *Bradyrhizobium,* were found to be significantly more prevalent in the meconium samples than in the other age groups (Order of Rhizobiales; meconium to 3 months; $p = 9.70*10^{-07}$; paired t-test, unadjusted), constituting of 0.50% of the meconium samples on average, while the average for the other groups is in the range of [0.002-0.08]% (Table F.1 in Appendix F). *Bradyrhizobium* is a soil-associated nitrogen fixator that is reported to contaminate laboratory PCR reagents and kits (Salter et al., 2014). It is therefore likely that the large alpha diversity observed in the meconium samples stems from a more considerable amplification of contaminants connected to the increased number of PCR cycles.

5. Conclusions and further research

The present study showed that environmental bacteria might have a considerable impact on the immune cell profile at 12 months. It reviled interesting correlations between environmental bacteria, Methylophiliales and Methylococcales, and the host's immune cells, possibly resulting in inflammatory and anti-inflammatory effects, respectively. More research should be undertaken to further investigate the effect of the non-commensal environmental bacteria on the immune system in humans. Additionally, as a continuation of this study, it would be interesting to look at the bacterial correlations at a deeper taxonomic level. Due to time limitations, this was not conducted in the present study. However, it would be advantageous as bacterial orders are heterogeneous and contain both opportunistic pathogenic, pathogenic or commensal species and strains, which may have contradicting effects on the host immune system.

References

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol, 71*(12), 8966-8969. doi:10.1128/AEM.71.12.8966-8969.2005

Angelucci, F., Cechova, K., Amlerova, J., & Hort, J. (2019). Antibiotics, gut microbiota, and Alzheimer's disease. *J Neuroinflammation, 16*(1), 108. doi:10.1186/s12974-019-1494-4

Asnicar, F., Manara, S., Zolfo, M., Truong, D. T., Scholz, M., Armanini, F., . . . Segata, N. (2017). Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems, 2*(1). doi:10.1128/mSystems.00164-16

Avershina, E., Lundgard, K., Sekelja, M., Dotterud, C., Storro, O., Oien, T., . . . Rudi, K. (2016). Transition from infant- to adult-like gut microbiota. *Environ Microbiol, 18*(7), 2226-2236. doi:10.1111/1462-2920.13248

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P., & Rudi, K. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol, 87*(1), 280-290. doi:10.1111/1574-6941.12223

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., . . . Investigators, C. S. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ, 185*(5), 385-394. doi:10.1503/cmaj.121189

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., . . . Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe, 17*(5), 690-703. doi:10.1016/j.chom.2015.04.004

Bakken, L. R., & Frostegård, Å. (2006). Nucleic Acid Extraction from Soil. In P. Nannipieri & K. Smalla (Eds.), *Nucleic Acids and Proteins in Soil* (pp. 49-73). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., . . . Tanner, S. D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem, 81*(16), 6813-6822. doi:10.1021/ac901049w

Barrett, K. E. (2014). *Gastrointestinal Physiology* (M. Weitz, Thomas, M. C. Ed. 2nd ed.). United States of America: McGraw-Hill Education.

Bengmark, S. (1998). Ecological control of the gastrointestinal tract. The role of probiotic flora. *Gut, 42*(1), 2-7. doi:10.1136/gut.42.1.2

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *J Clin Microbiol, 28*(3), 495-503. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1691208

Brandtzaeg, P. (1998). Development and basic mechanisms of human gut immunity. *Nutr Rev, 56*(1 Pt 2), S5-18. doi:10.1111/j.1753-4887.1998.tb01645.x

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs, 27*(4), 325-349. doi:10.2307/1942268

Brodin, P., & Davis, M. M. (2017). Human immune system variation. *Nat Rev Immunol, 17*(1), 21-29. doi:10.1038/nri.2016.125

Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Angel, C. J., Furman, D., . . . Davis, M. M. (2015). Variation in the human immune system is largely driven by non-heritable influences. *Cell, 160*(1-2), 37-47. doi:10.1016/j.cell.2014.12.020

Brooke, J. S. (2012). Stenotrophomonas maltophilia: an emerging global opportunistic pathogen. *Clin Microbiol Rev, 25*(1), 2-41. doi:10.1128/CMR.00019-11

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303

Chakraborty, C., Doss, C. G., Patra, B. C., & Bandyopadhyay, S. (2014). DNA barcoding to map the microbial communities: current advances and future directions. *Appl Microbiol Biotechnol, 98*(8), 3425-3436. doi:10.1007/s00253-014-5550-9

Chang, P. V., Hao, L., Offermanns, S., & Medzhitov, R. (2014). The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc Natl Acad Sci U S A, 111*(6), 2247-2252. doi:10.1073/pnas.1322269111

Clarridge, J. E., 3rd. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev, 17*(4), 840-862, table of contents. doi:10.1128/CMR.17.4.840-862.2004

Collado, M. C., Rautava, S., Aakko, J., Isolauri, E., & Salminen, S. (2016). Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci Rep, 6*, 23129. doi:10.1038/srep23129

Cooperstock, M. S. Z. A. J. (1983). Intestinal flora of infants. In e. Hentges DJ (Ed.), *Human Intestinal Microflora in Health and Disease* (pp. 79–99). New York: Academic Press.

Cummings, J. H. (1981). Short chain fatty acids in the human colon. *Gut, 22*(9), 763-779. doi:10.1136/gut.22.9.763

de Muinck, E. J., & Trosvik, P. (2018). Individuality and convergence of the infant gut microbiota during the first year of life. *Nat Commun, 9*(1), 2233. doi:10.1038/s41467-018-04641-7

Doronina, N., Kaparullina, E., & Trotsenko, Y. (2014). The Family Methylophilaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 869-880). Berlin, Heidelberg: Springer Berlin Heidelberg.

Duncan, S. H., Louis, P., & Flint, H. J. (2004). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol, 70*(10), 5810-5817. doi:10.1128/AEM.70.10.5810-5817.2004

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., & Lang, H. (1974). Proteinase K from Tritirachium album Limber. *Eur J Biochem, 47*(1), 91-97. doi:10.1111/j.1432-1033.1974.tb03671.x

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol, 27*(2), 105-117. doi:10.1016/j.tim.2018.11.003

Escobar-Zepeda, A., Vera-Ponce de Leon, A., & Sanchez-Flores, A. (2015). The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front Genet, 6*, 348. doi:10.3389/fgene.2015.00348

Garcia-Villalba, R., Gimenez-Bastida, J. A., Garcia-Conesa, M. T., Tomas-Barberan, F. A., Carlos Espin, J., & Larrosa, M. (2012). Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids in faecal samples. *J Sep Sci, 35*(15), 1906-1913. doi:10.1002/jssc.201101121

Gern, J. E., Reardon, C. L., Hoffjan, S., Nicolae, D., Li, Z., Roberg, K. A., . . . Lemanske, R. F., Jr. (2004). Effects of dog ownership and genotype on immune development and atopy in infancy. *J Allergy Clin Immunol, 113*(2), 307-314. doi:10.1016/j.jaci.2003.11.017

Gibson, M. K., Crofts, T. S., & Dantas, G. (2015). Antibiotics and the developing infant gut microbiota and resistome. *Curr Opin Microbiol, 27*, 51-56. doi:10.1016/j.mib.2015.07.007

Gleeson, M., & Cripps, A. W. (2004). Development of mucosal immunity in the first year of life and relationship to sudden infant death syndrome. *FEMS Immunol Med Microbiol, 42*(1), 21-33. doi:10.1016/j.femsim.2004.06.012

Grover, A., & Sharma, P. C. (2016). Development and use of molecular markers: past and present. *Crit Rev Biotechnol, 36*(2), 290-302. doi:10.3109/07388551.2014.959891

Guarner, F., & Malagelada, J. R. (2003). Gut flora in health and disease. *Lancet, 361*(9356), 512-519. doi:10.1016/S0140-6736(03)12489-0

Gupta, N. (2019). DNA Extraction and Polymerase Chain Reaction. *J Cytol, 36*(2), 116-117. doi:10.4103/JOC.JOC_110_18

Gurav, A., Sivaprakasam, S., Bhutia, Y. D., Boettger, T., Singh, N., & Ganapathy, V. (2015). Slc5a8, a Na+-coupled high-affinity transporter for short-chain fatty acids, is a conditional tumour suppressor in colon that protects against colitis and colon cancer under low-fibre dietary conditions. *Biochem J, 469*(2), 267-278. doi:10.1042/BJ20150242

Hart, A. L., Lammers, K., Brigidi, P., Vitali, B., Rizzello, F., Gionchetti, P., . . . Stagg, A. J. (2004). Modulation of human dendritic cell phenotype and function by probiotic bacteria. *Gut, 53*(11), 1602-1609. doi:10.1136/gut.2003.037325

Hartel, C., Scholz, T., Kuhn, M., Bendiks, M., Gopel, W., Lauten, M., & Herting, E. (2013). Innate immune responses to Stenotrophomonas maltophilia in immunocompromised pediatric patients and the effect of taurolidine. *J Microbiol Immunol Infect, 46*(2), 115-120. doi:10.1016/j.jmii.2012.04.002

Hayward, T., Hua, Y., Gras, R., & Luong, J. (2017). Industrial Applications with a New Polyethylene Glycol-Based GC Column.

Henningsson, Å., Björck, I., & Nyman, M. (2001). Short-chain fatty acid formation at fermentation of indigestible carbohydrates. *Näringsforskning, 45*(1), 165-168. doi:10.3402/fnr.v45i0.1801

Hill, J. E., Fernando, W. M., Zello, G. A., Tyler, R. T., Dahl, W. J., & Van Kessel, A. G. (2010). Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the cpn60 universal primer cocktail. *Appl Environ Microbiol, 76*(13), 4550-4552. doi:10.1128/AEM.01510-09

Hua, X., Goedert, J. J., Pu, A., Yu, G., & Shi, J. (2016). Allergy associations with the adult fecal microbiota: Analysis of the American Gut Project. *EBioMedicine, 3*, 172-179. doi:10.1016/j.ebiom.2015.11.038

Indrelid, S., Kleiveland, C., Holst, R., Jacobsen, M., & Lea, T. (2017). The Soil Bacterium Methylococcus capsulatus Bath Interacts with Human Dendritic Cells to Modulate Immune Function. *Front Microbiol, 8*, 320. doi:10.3389/fmicb.2017.00320

Jaccard, P. (1901). Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles, 37*, 241-272.

Jennewein, M. F., Abu-Raya, B., Jiang, Y., Alter, G., & Marchant, A. (2017). Transfer of maternal immunity and programming of the newborn immune system. *Semin Immunopathol, 39*(6), 605-613. doi:10.1007/s00281-017-0653-x

Jiang, F., Meng, D., Weng, M., Zhu, W., Wu, W., Kasper, D., & Walker, W. A. (2017). The symbiotic bacterial surface factor polysaccharide A on Bacteroides fragilis inhibits IL-

1beta-induced inflammation in human fetal enterocytes via toll receptors 2 and 4. *PLoS One, 12*(3), e0172738. doi:10.1371/journal.pone.0172738

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., . . . McWeeney, S. (2019). Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems, 4*(4). doi:10.1128/mSystems.00290-19

Kay, A. W., Strauss-Albee, D. M., & Blish, C. A. (2016). Application of Mass Cytometry (CyTOF) for Functional and Phenotypic Analysis of Natural Killer Cells. *Methods Mol Biol, 1441*, 13-26. doi:10.1007/978-1-4939-3684-7_2

Kelly, D., Conway, S., & Aminov, R. (2005). Commensal gut bacteria: mechanisms of immune modulation. *Trends Immunol, 26*(6), 326-333. doi:10.1016/j.it.2005.04.008

Kleiveland, C. R., Hult, L. T., Spetalen, S., Kaldhusdal, M., Christofferesen, T. E., Bengtsson, O., . . . Lea, T. (2013). The noncommensal bacterium Methylococcus capsulatus (Bath) ameliorates dextran sulfate (Sodium Salt)-Induced Ulcerative Colitis by influencing mechanisms essential for maintenance of the colonic barrier function. *Appl Environ Microbiol, 79*(1), 48-56. doi:10.1128/AEM.02464-12

Knetsch, C. W., van der Veer, E. M., Henkel, C., & Taschner, P. (2019). DNA Sequencing. In E. van Pelt-Verkuil, W. B. van Leeuwen, & R. te Witt (Eds.), *Molecular Diagnostics: Part 1: Technical Backgrounds and Quality Aspects* (pp. 339-360). Singapore: Springer Singapore.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., . . . Zoric, N. (2006). The real-time polymerase chain reaction. *Mol Aspects Med, 27*(2-3), 95-125. doi:10.1016/j.mam.2005.12.007

Lakshmikanth, T., & Brodin, P. (2019). Systems-Level Immune Monitoring by Mass Cytometry. *Methods Mol Biol, 1913*, 33-48. doi:10.1007/978-1-4939-8979-9_3

LeBlanc, J. G., Chain, F., Martin, R., Bermudez-Humaran, L. G., Courau, S., & Langella, P. (2017). Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microb Cell Fact, 16*(1), 79. doi:10.1186/s12934-017-0691-z

Li, M., van Esch, B., Wagenaar, G. T. M., Garssen, J., Folkerts, G., & Henricks, P. A. J. (2018). Pro- and anti-inflammatory effects of short chain fatty acids on immune and endothelial cells. *Eur J Pharmacol, 831*, 52-59. doi:10.1016/j.ejphar.2018.05.003

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., . . . Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol, 2012*, 251364. doi:10.1155/2012/251364

Lodrup Carlsen, K. C., Rehbinder, E. M., Skjerven, H. O., Carlsen, M. H., Fatnes, T. A., Fugelli, P., . . . study, g. (2018). Preventing Atopic Dermatitis and ALLergies in Children-the PreventADALL study. *Allergy, 73*(10), 2063-2070. doi:10.1111/all.13468

Lowe, A. J., Leung, D. Y. M., Tang, M. L. K., Su, J. C., & Allen, K. J. (2018). The skin as a target for prevention of the atopic march. *Ann Allergy Asthma Immunol, 120*(2), 145-151. doi:10.1016/j.anai.2017.11.023

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature, 489*(7415), 220-230. doi:10.1038/nature11550

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D., Hirschfield, G. M., Hold, G., . . . Hart, A. (2016). The gut microbiota and host health: a new clinical frontier. *Gut, 65*(2), 330-339. doi:10.1136/gutjnl-2015-309990

Masui, R., Sasaki, M., Funaki, Y., Ogasawara, N., Mizuno, M., Iida, A., . . . Kasugai, K. (2013). G protein-coupled receptor 43 moderates gut inflammation through cytokine

regulation from mononuclear cells. *Inflamm Bowel Dis, 19*(13), 2848-2856. doi:10.1097/01.MIB.0000435444.14860.ea

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., & de La Cochetiere, M. F. (2013). Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol, 21*(4), 167-173. doi:10.1016/j.tim.2012.12.001

McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-Generation Sequencing Technologies. *Cold Spring Harb Perspect Med, 9*(11). doi:10.1101/cshperspect.a036798

Michels, N., Van de Wiele, T., Fouhy, F., O'Mahony, S., Clarke, G., & Keane, J. (2019). Gut microbiome patterns depending on children's psychosocial stress: Reports versus biomarkers. *Brain Behav Immun, 80*, 751-762. doi:10.1016/j.bbi.2019.05.024

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., . . . 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*(4). doi:10.1128/MMBR.00036-17

Modi, S. R., Collins, J. J., & Relman, D. A. (2014). Antibiotics and the gut microbiota. *J Clin Invest, 124*(10), 4212-4218. doi:10.1172/JCI72333

Moens, F., Verce, M., & De Vuyst, L. (2017). Lactate- and acetate-based cross-feeding interactions between selected strains of lactobacilli, bifidobacteria and colon bacteria in the presence of inulin-type fructans. *Int J Food Microbiol, 241*, 225-236. doi:10.1016/j.ijfoodmicro.2016.10.019

Moles, L., Gomez, M., Heilig, H., Bustos, G., Fuentes, S., de Vos, W., . . . Jimenez, E. (2013). Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS One, 8*(6), e66986. doi:10.1371/journal.pone.0066986

Nagano, Y., Itoh, K., & Honda, K. (2012). The induction of Treg cells by gut-indigenous Clostridium. *Curr Opin Immunol, 24*(4), 392-397. doi:10.1016/j.coi.2012.05.007

Nagpal, R., Kurakawa, T., Tsuji, H., Takahashi, T., Kawashima, K., Nagata, S., . . . Yamashiro, Y. (2017). Evolution of gut Bifidobacterium population in healthy Japanese infants over the first three years of life: a quantitative assessment. *Sci Rep, 7*(1), 10097. doi:10.1038/s41598-017-10711-5

Naseem, S., & Tahir, H. M. (2018). Use of mitochondrial COI gene for the identification of family Salticidae and Lycosidae of spiders. *Mitochondrial DNA A DNA Mapp Seq Anal, 29*(1), 96-101. doi:10.1080/24701394.2016.1248428

Nguyen, T. L., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Dis Model Mech, 8*(1), 1-16. doi:10.1242/dmm.017400

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol, 5*(7), e177. doi:10.1371/journal.pbio.0050177

Parada Venegas, D., De la Fuente, M. K., Landskron, G., Gonzalez, M. J., Quera, R., Dijkstra, G., . . . Hermoso, M. A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Front Immunol, 10*, 277. doi:10.3389/fimmu.2019.00277

Parham, P. (2014). *The Immune System*: CRC Press.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., . . . Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics, 118*(2), 511-521. doi:10.1542/peds.2005-2824

Pennock, N. D., White, J. T., Cross, E. W., Cheney, E. E., Tamburini, B. A., & Kedl, R. M. (2013). T cell responses: naive to memory and everything in between. *Adv Physiol Educ, 37*(4), 273-283. doi:10.1152/advan.00066.2013

Perez-Munoz, M. E., Arrieta, M. C., Ramer-Tait, A. E., & 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*(1), 48. doi:10.1186/s40168-017-0268-4

Poole, C. F. (2012). *Gas Chromatography* (Vol. 1st ed). Oxford, UK: Elsevier.

Primec, M., Micetic-Turk, D., & Langerholc, T. (2017). Analysis of short-chain fatty acids in human feces: A scoping review. *Anal Biochem, 526*, 9-21. doi:10.1016/j.ab.2017.03.007

Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S., & Flint, H. J. (2002). The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett, 217*(2), 133-139. doi:10.1111/j.1574-6968.2002.tb11467.x

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., . . . Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-65. doi:10.1038/nature08821

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res, 41*(Database issue), D590-596. doi:10.1093/nar/gks1219

Rehbinder, E. M., Lodrup Carlsen, K. C., Staff, A. C., Angell, I. L., Landro, L., Hilde, K., . . . Rudi, K. (2018). Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? *Am J Obstet Gynecol, 219*(3), 289 e281-289 e212. doi:10.1016/j.ajog.2018.05.028

Rios-Covian, D., Salazar, N., Gueimonde, M., & de los Reyes-Gavilan, C. G. (2017). Shaping the Metabolism of Intestinal Bacteroides Population through Diet to Improve Human Health. *Frontiers in Microbiology, 8*(376). doi:10.3389/fmicb.2017.00376

Riviere, A., Selak, M., Lantin, D., Leroy, F., & De Vuyst, L. (2016). Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Front Microbiol, 7*, 979. doi:10.3389/fmicb.2016.00979

Rodriguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., . . . Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis, 26*, 26050. doi:10.3402/mehd.v26.26050

Rook, G. A., Raison, C. L., & Lowry, C. A. (2014). Microbial 'old friends', immunoregulation and socioeconomic status. *Clin Exp Immunol, 177*(1), 1-12. doi:10.1111/cei.12269

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol, 12*, 87. doi:10.1186/s12915-014-0087-z

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*(12), 5463-5467. doi:10.1073/pnas.74.12.5463

Schochetman, G., Ou, C. Y., & Jones, W. K. (1988). Polymerase chain reaction. *J Infect Dis, 158*(6), 1154-1157. doi:10.1093/infdis/158.6.1154

Segain, J. P., Raingeard de la Bletiere, D., Bourreille, A., Leray, V., Gervois, N., Rosales, C., . . . Galmiche, J. P. (2000). Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut, 47*(3), 397-403. doi:10.1136/gut.47.3.397

Sela, D. A., & Mills, D. A. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol, 18*(7), 298-307. doi:10.1016/j.tim.2010.03.008

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign, IL, US: University of Illinois Press.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol, 26*(10), 1135-1145. doi:10.1038/nbt1486

Simpson, E. H. (1949). Measurement of Diversity. *Nature, 163*(4148), 688-688. doi:10.1038/163688a0

Smith, N. W., Shorten, P. R., Altermann, E., Roy, N. C., & McNabb, W. C. (2019). The Classification and Evolution of Bacterial Cross-Feeding. *Frontiers in Ecology and Evolution, 7*(153). doi:10.3389/fevo.2019.00153

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., . . . Knight, R. (2013). Cohabiting family members share microbiota with one another and with their dogs. *Elife, 2*, e00458. doi:10.7554/eLife.00458

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72-101. doi:10.2307/1412159

Stearns, J. C., Lynch, M. D., Senadheera, D. B., Tenenbaum, H. C., Goldberg, M. B., Cvitkovitch, D. G., . . . Neufeld, J. D. (2011). Bacterial biogeography of the human digestive tract. *Sci Rep, 1*, 170. doi:10.1038/srep00170

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., . . . Petrosino, J. F. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature, 562*(7728), 583-588. doi:10.1038/s41586-018-0617-x

Sun, M., Wu, W., Liu, Z., & Cong, Y. (2017). Microbiota metabolite short chain fatty acids, GPCR, and inflammatory bowel diseases. *J Gastroenterol, 52*(1), 1-8. doi:10.1007/s00535-016-1242-9

Sun, Y., & O'Riordan, M. X. D. (2013). Regulation of Bacterial Pathogenesis by Intestinal Short-Chain Fatty Acids. *Advances in Applied Microbiology, Vol 85, 85*, 93-118. doi:10.1016/B978-0-12-407672-3.00003-4

Tameliene, R., Barcaite, E., Stoniene, D., Buinauskiene, J., Markuniene, E., Kudreviciene, A., . . . Nadisauskiene, R. (2012). Escherichia coli colonization in neonates: prevalence, perinatal transmission, antimicrobial susceptibility, and risk factors. *Medicina (Kaunas), 48*(2), 71-76. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22491384

Tang, Q., Jin, G., Wang, G., Liu, T., Liu, X., Wang, B., & Cao, H. (2020). Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices. *Front Cell Infect Microbiol, 10*, 151. doi:10.3389/fcimb.2020.00151

Teng, F., Darveekaran Nair, S. S., Zhu, P., Li, S., Huang, S., Li, X., . . . Yang, F. (2018). Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep, 8*(1), 16321. doi:10.1038/s41598-018-34294-x

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., . . . Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature, 457*(7228), 480-484. doi:10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature, 449*(7164), 804-810. doi:10.1038/nature06244

van Pelt-Verkuil, E., & te Witt, R. (2019). Principles of PCR. In E. van Pelt-Verkuil, W. B. van Leeuwen, & R. te Witt (Eds.), *Molecular Diagnostics: Part 1: Technical Backgrounds and Quality Aspects* (pp. 131-215). Singapore: Springer Singapore.

Vinolo, M. A., Rodrigues, H. G., Hatanaka, E., Sato, F. T., Sampaio, S. C., & Curi, R. (2011). Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *J Nutr Biochem, 22*(9), 849-855. doi:10.1016/j.jnutbio.2010.07.009

von Martels, J. Z. H., Sadaghian Sadabad, M., Bourgonje, A. R., Blokzijl, T., Dijkstra, G., Faber, K. N., & Harmsen, H. J. M. (2017). The role of gut microbiota in health and disease: In vitro modeling of host-microbe interactions at the aerobe-anaerobe interphase of the human gut. *Anaerobe, 44*, 3-12. doi:10.1016/j.anaerobe.2017.01.001

Wang, L., Zhu, L., & Qin, S. (2019). Gut Microbiota Modulation on Intestinal Mucosal Adaptive Immunity. *J Immunol Res, 2019*, 4735040. doi:10.1155/2019/4735040

Wang, S., Ryan, C. A., Boyaval, P., Dempsey, E. M., Ross, R. P., & Stanton, C. (2020). Maternal Vertical Transmission Affecting Early-life Microbiota Development. *Trends Microbiol, 28*(1), 28-45. doi:10.1016/j.tim.2019.07.010

Weiner, H. L., da Cunha, A. P., Quintana, F., & Wu, H. (2011). Oral tolerance. *Immunol Rev, 241*(1), 241-259. doi:10.1111/j.1600-065X.2011.01017.x

Weir, T. L., Manter, D. K., Sheflin, A. M., Barnett, B. A., Heuberger, A. L., & Ryan, E. P. (2013). Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One, 8*(8), e70803. doi:10.1371/journal.pone.0070803

Wong, J. M., de Souza, R., Kendall, C. W., Emam, A., & Jenkins, D. J. (2006). Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol, 40*(3), 235-243. doi:10.1097/00004836-200603000-00015

Wynn, T. A. (2003). IL-13 effector functions. *Annu Rev Immunol, 21*, 425-456. doi:10.1146/annurev.immunol.21.120601.141142

Yu, Y., Lee, C., Kim, J., & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng, 89*(6), 670-679. doi:10.1002/bit.20347

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Sci Transl Med, 6*(237), 237ra265. doi:10.1126/scitranslmed.3008599

Appendices

Appendix A – CyTOF methods_MECFS manuscript by Petter Brodin

"Immune cell phenotyping by Mass Cytometry

Cryopreserved and stabilized whole blood (blood mixed with 'Stabilizer' component of Whole blood processing kit; Cytodelics AB, Sweden) collected from …. patients sampled during the study period were thawed, and cells were fixed and RBCs lysed using Wash # 1 and # 2 buffers (Whole blood processing kit; Cytodelics AB, Sweden) as per the manufacturer's recommendations. This was performed a few days prior to barcoding and staining of cells. Post fix/lysis of cells, ~1x106 cells/sample were plated onto a 96 well 'U' bottom plate using standard cryoprotective solution (10% DMSO and 90% FBS) and cryopreserved at -80oC.

On the day of barcoding and staining of cells, cells were thawed at 37oC using RPMI medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin and benzonase (Sigma-Aldrich, Sweden). Briefly, cells were barcoded using automated liquid handling robotic system (Agilent technologies, Santa Clara, CA, USA (REF Mikes et al, Methods Mol Biol, 2019) using the Cell-ID 20-plex Barcoding kit (Fluidigm Inc.) as per the manufacturer's recommendations. Following cell pooling batch wise (with samples from placebo and treatment groups equally represented in each batch), cells were washed, FcR blocked using blocking buffer (in-house developed recipe) for 10min at room temperature, following which cells were incubated for another 30 min at 4°C after addition of a cocktail of metal conjugated antibodies targeting the surface antigens. Following two washes with CyFACS buffer, cells were fixed overnight using 4% formaldehyde made in PBS (VWR, Sweden). The broad extended panel of antibodies used for staining are listed in Supplementary Table X. For acquisition by CyTOF (within 2 days after staining), cells were stained with DNA intercalator (0.125 µM Iridium-191/193 or MaxPar® Intercalator-Ir, Fluidigm) in 4% formaldehyde made in PBS for 20 min at room temperature. After multiple washes with

CyFACS, PBS and milliQ water, cells were filtered through a 35µm nylon mesh and diluted to 750,000 cells/ml. Cells were acquired at a rate of 300-500 cells/s using a super sampler (Victorian Airship, USA) connected to a CyTOF2 (Fluidigm) mass cytometer, CyTOF software version 6.0.626 with noise reduction, a lower convolution threshold of 200, event length limits of 10-150 pushes and a sigma value of 3 and flow rate of 0.045 ml/min.

Antibodies and reagents

Purified antibodies for mass cytometry were obtained in carrier/protein-free buffer and then coupled to lanthanide metals using the MaxPar antibody conjugation kit (Fluidigm Inc.) as per the manufacturer's recommendations. Following the protein concentration determination by measurement of absorbance at 280 nm on a nanodrop, the metal-labeled antibodies were diluted in Candor PBS Antibody Stabilization solution (Candor Bioscience, Germany) for long-term storage at 4°C. Antibodies used are listed in Supplementary Table X."

Appendix B – PRK Illumina primer sequences for Index PCR

Table B.1: PCR index primer information. Listed is the name of the primer, sequence and target region, together with primer direction.

Primer name	Sequence, 5' -> 3'	Target region/gene	Direction
F1	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F2	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F3	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F4	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F5	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F6	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F7	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F8	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F9	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F10	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F11	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F12	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F13	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F14	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F15	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
F16	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	Forward
R1	caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R2	caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R3	caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R4	caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R5	caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R6	caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R7	caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R8	caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R9	caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R10	caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R11	caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R12	caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R13	caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R14	caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R15	caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R16	caagcagaagacggcatacgagatGGACGGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R17	caagcagaagacggcatacgagatCTCTACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R18	caagcagaagacggcatacgagatGCGGACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R19	caagcagaagacggcatacgagatTTTCACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R20	caagcagaagacggcatacgagatGGCCACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R21	caagcagaagacggcatacgagatCGAAACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R22	caagcagaagacggcatacgagatCGTACGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R23	caagcagaagacggcatacgagatCCACTCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse
R24	caagcagaagacggcatacgagatGCTACCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	Reverse

Appendix C – Information on the Gas Chromatograph

Table C.1: Technical information on the Trace 1310 with Autosampler gas chromatograph.

Instrument	Trace 1310 with Autosampler (Thermo Fisher Scientific, Norway)
Injector	
Mode:	Split
Temperature:	250*C
Carrier gas:	Helium
Column flow:	2.5 mL/min
Split flow:	200 mL/min
Purge flow:	3 mL/min
Injector volume:	0.2 uL
Liner:	4mm x 6.3mm x 78.5mm (Restek, USA)
Syringe:	10 µl syr FN 50 mm C, Ga 23, cone tip (Thermo Fisher Scientific, Norway)
Column	
	Stabilwax DA 30m, 0.25 mm ID, 0.25 µm (Restek, USA)
Temperature program:	90 °C to 150 °C (6 min), 150 °C to 245 °C (1.9 min).
Time per sample:	14.9 min
Detector	
Type	FID Temperature: 275°C
Hydrogen:	30 mL/min
Air:	300 mL/min
Makeup gas:	30mL/min

Appendix D – Calibration standards

Table D.1: Calibration standards concentrations

Standard No.	**Standard A**	**Standard B**	**Standard C**
Acetic acid	3000 µM	1500 µM	300 µM
Propionic acid	120 µM	60 µM	12 µM
Isobutyric acid	80 µM	40 µM	8 µM
Butyric acid	120 µM	60 µM	12 µM
Isovaleric acid	80 µM	40 µM	8 µM
Valeric avid	80 µM	40 µM	8 µM
2-methylvaleric acid	250 µM	125 µM	25 µM
Formic acid	0.2%	0.2%	0.2%

Appendix E – Technical aspects of *16S rRNA* gene sequencing

Sequence results generated from the QIIME pipeline

For the sequencing of the 12 months samples the total number of sequences was 6,163,507, and the sequencing depth was set to 10500 dsDNA fragments per sample. The total OTU count was 724. The longitudinal samples (meconium, 3-, 6 months, and mothers) the total sequence number was 10,672,069, and the cut-off was 5000 dsDNA fragments per sample. In total 1286 OTUs were counted.

Technical considerations concerning PCR

For the first qualitative PCR amplification all samples were run for 25 cycles, except from meconium samples that were run for 30 cycles to ensure sufficient amounts of amplicons.

Positive and negative controls were included in all PCR reactions. These controls followed the same treatment as all other samples, and showed sufficient results, meaning they did not report any methodical errors or major contaminations.

Normalization and library pooling

After all the samples had been indexed, the sample DNA concentrations were measured by Qubit and Cambrex-FLX 800 CSE. In the first sequencing run the samples were measured using different Qubit working solutions, and two separate standard curves were made for each solution. One having a regression line with the formula $y = 0.0257x - 9.3475$ and a $R^2 = 0.9108$, the other $y = 0.0171x - 2.3537$ and a $R^2 = 0.9189$. The regression lines were used to estimate the Qubit-values for all samples on the sample plate. Based on the estimated Qubit values the sample with the highest concentration was used as a cut off value. From this sample it was decided that 2μL was to be added to the pooled library. The added volume to the pooled library from all other samples was calculated so the concentration matched the cut off value. Although, no larger volume than 10μL was used. Pooling and normalization were done using the Beckman Biomek 3000 robot. In total, 188 samples, including positive and negative controls, were pooled and prepared for sequencing. 110 of the samples used the first regression line for Qubit estimation and 78 samples used the second regression line.

As the two libraries had different concentrations of DNA, they had to be normalized. The DNA concentrations of the two pooled libraries were measured using Qubit. The library with the highest concentration was then diluted to equal concentrations in the two libraries, and the libraries were pooled together, taking the number of samples in each library into account.

For the second sequencing run all samples were measured using the same Qubit working solution using Cambrex-FLX 800 CSE, and a regression line with the formula y = 0.0107x - 0.9652 and an $R^2 = 0.872$. The highest estimated Qubit concentration was set as a cut-off value. As more samples were pooled together before this sequencing run, the total volume of the pooled sample would exceed what one 1.5mL Eppendorf tube would allow. To account for this, the cut off value was set to 1µL instead of 2µL, and no volume larger than 10µL was used in the pooling. Pooling and normalization were done using the Beckman Biomek 3000 robot. In total, 245 samples, including positive and negative controls were pooled.

Appendix F – Bacterial phyla and genera

Bacterial OTUs are taxonomically defined by QIIME using the SILVA database. The average percentage abundance of phyla and genera for the different ages are illustrated in Figure F.1 and Figure F.2, respectively. The average bacterial genera abundance in percentage are listed in Table F.1.

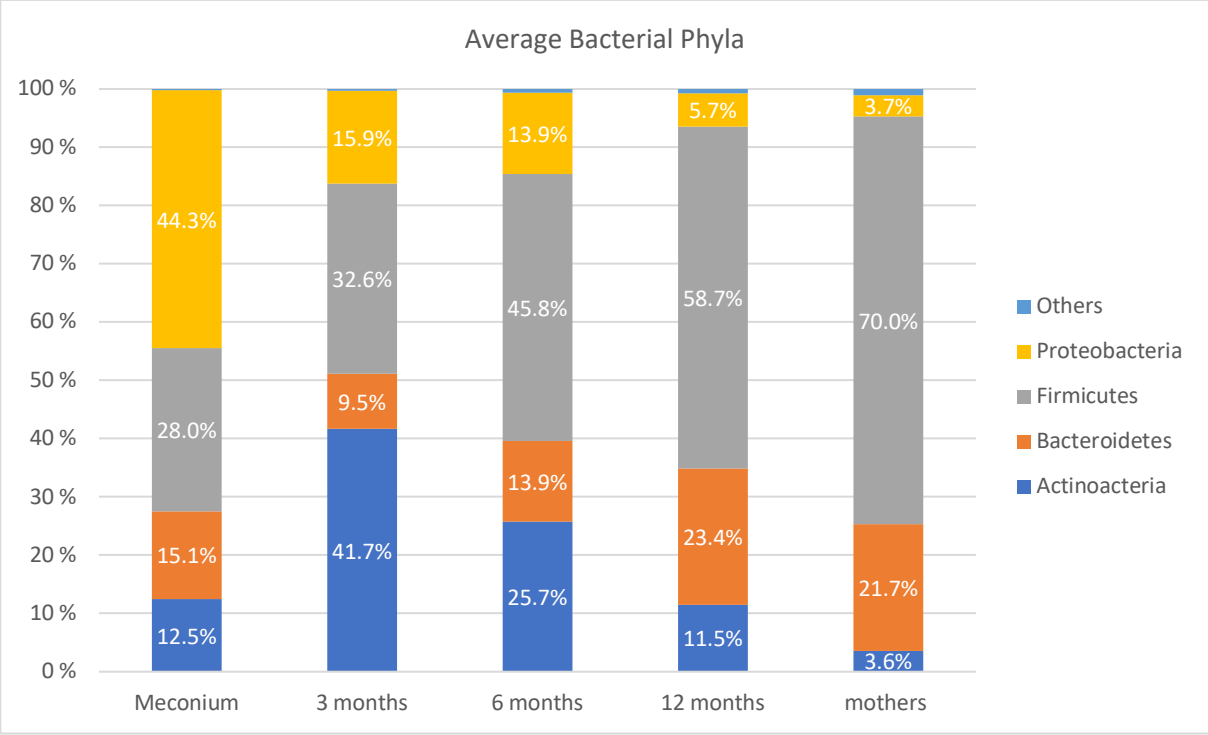


Figure F.1: Distribution of bacterial phyla within each age group. The bacterial phyla are acquired from sequencing and processing using the QIIME pipeline, and the average of the most dominant bacterial phyla in each age group are displayed.

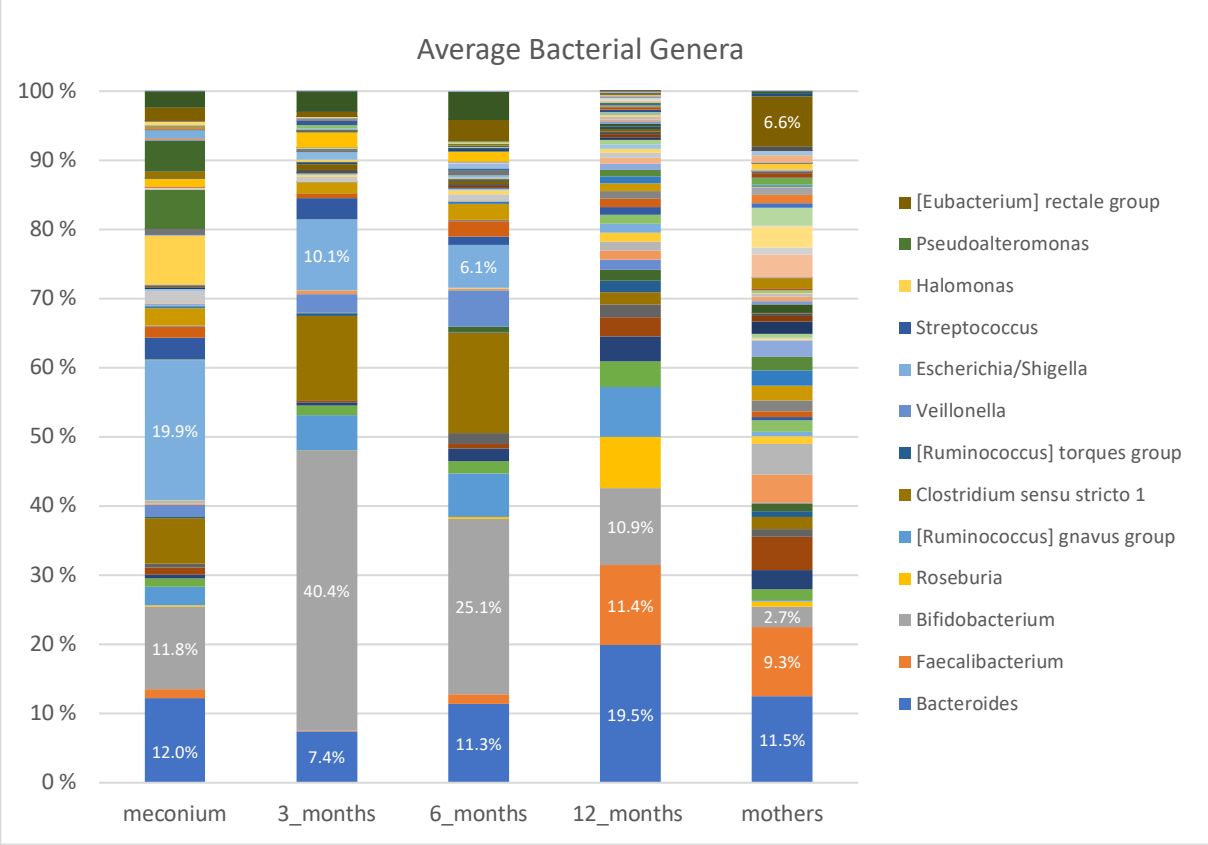


Figure F.2: Distribution of bacterial genera within each age group. The bacterial genera are acquired from sequencing and processing using the QIIME pipeline, and the average of the most dominant bacterial genera in each age group are displayed. More detailed information on the labeled bacterial genera is listed in Table F.1.

Table F.1: The average distribution of selected bacterial genera within each age group. The bacterial orders are acquired from sequencing and processing using the QIIME pipeline and given as average abundance in percentages (%) in each age group.

Order	***Genera***	***Meconium***	***3 months***	***6 months***	***12 months***	***mothers***
Alteromonadales	*Pseudoalteromonas*	5. 50	0.16	0.071	0.15	0.034
Bacteroidales	*Bacteroides*	12.00	7.39	11.35	19.53	11.54
Bifidobacteriales	*Bifidobacterium*	11.75	40.38	25.11	10.92	2.72
Clostridiales	*Anaerostipes*	0.11	0.08	0.68	1.54	1.02
	[Eubacterium] hallii group	0.06	0.003	0.28	0.40	0.74
	[Eubacterium] rectale group	1.90	0.70	3.04	0	6.63
	Faecalibacterium	1.23	0.05	1.36	11.41	9.31
	[Ruminococcus] torques group	0.15	0.33	0.10	1.61	0.78
	[Ruminococcus] gnavus group	2.70	5.09	6.32	7.11	0.15
	Clostridium sensu stricto 1	6.40	12.18	14.41	1.73	1.58
	Roseburia	0.11	0.005	0.27	7.28	0.66
Enterobacteriales	*Escherichia/Shigella*	19.90	10.16	6.15	1.25	0.59
Lactobacillales	*Streptococcus*	3.09	3.05	1.21	1.15	0.38
Oceanospirillales	*Halomonas*	6.88	0.28	0.10	0.27	0.056
Rhizobiales	*Bradyrhizobium*	0.50	0.08	0.08	0.002	0.005
Selenomonadales	*Veillonella*	1.61	2.71	5.22	1.453	0.073
Xanthomonadales	*Stenotrophomonas*	0.03	0.001	0.0002	0.11	0.0001

Appendix G – Alpha diversity index table

Table G.1: The average values, standard deviation, and median alpha diversity indices for the different age groups.

		Average	*Standard deviation*	*Median*
Observed Species	meconium	85.5836364	26.1754275	96.4
	3 months	53.7910714	13.987549	53.3
	6 months	55	15.0072245	51.8
	12 months	95.5808383	20.5601846	92.4
	mothers	186.958621	42.1144951	187.55
Shannon-Weaver	meconium	3.93949052	1.64953372	4.9535917
	3 months	2.40592105	0.73460843	2.4856818
	6 months	2.93694019	0.8629056	3.06735965
	12 months	4.11437733	0.60884415	4.130469
	mothers	5.62528243	0.41636544	5.700417
Inverted Simpson	meconium	0.78749771	0.26795548	0.94716043
	3 months	0.66399969	0.17071985	0.7082703
	6 months	0.7476758	0.17918895	0.8194322
	12 months	0.88512025	0.0584986	0.8983061
	mothers	0.9538194	0.019396	0.9580333

Appendix H – Euclidian distance plot of all age groups

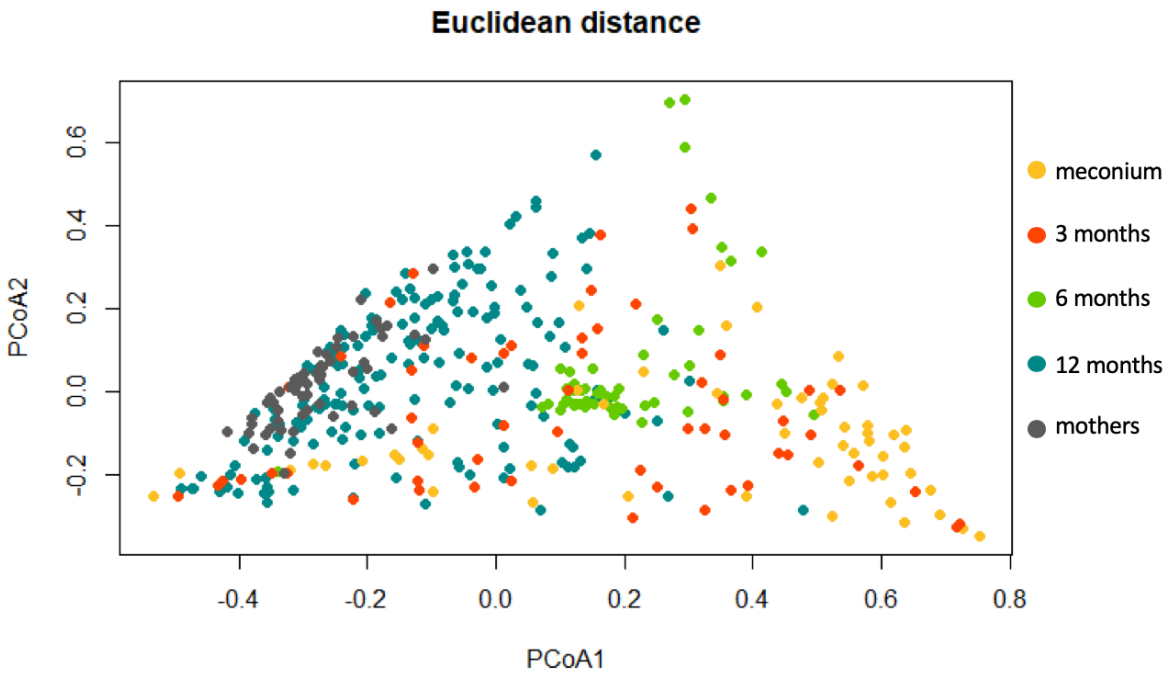


Figure H.1: Principal coordinate analysis plot using the Euclidian distance matrix. The mother samples cluster together with the 12-month samples, indicating similarity in the samples. The 6-month samples cluster in the middle of the plot, while the 3-month and meconium samples are dispersed.

Appendix I – Technical issues of SCFA analysis

95% confidence interval between thawed and frozen 12 months samples

When retrieving the samples from Oslo University Hospital (OUS) some of the 12 months samples were found to be stored in a wrecked freezer. As the SCFAs are volatile, a paired t-test was performed to determine if there were any significant differences in the SCFA composition of the samples that were thawed and the samples that had stayed frozen the entire time. No significant effects were detected, and the samples that had been thawed were included in the study. The 95% confidence intervals for all SCFA analyzed are visualized in Figure I.1.

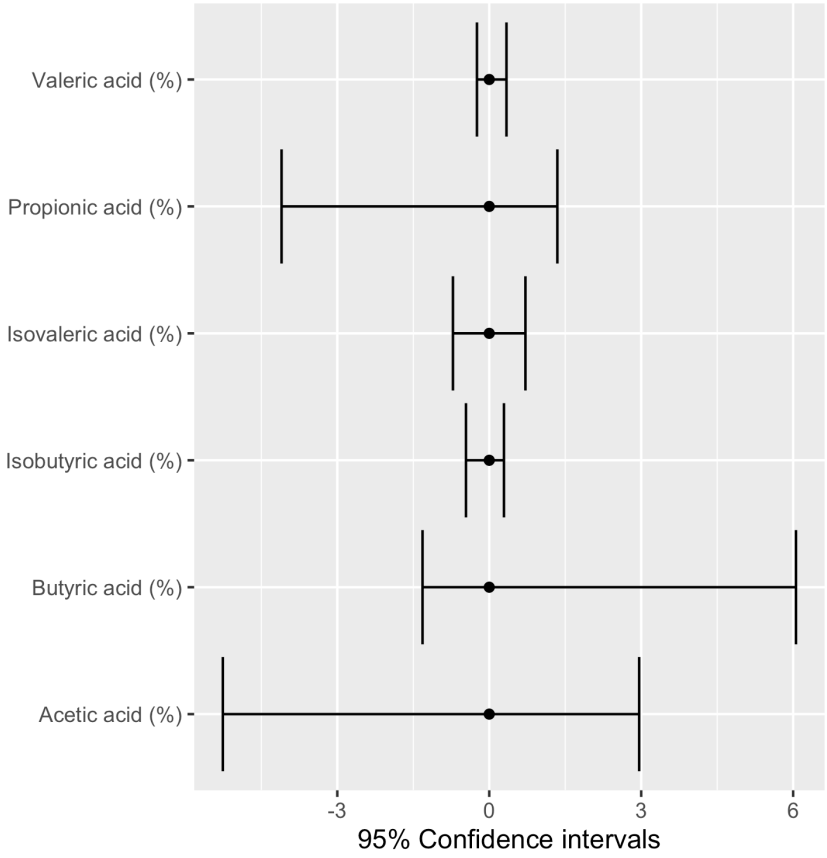


Figure I.1: 95% confidence interval illustrating the differences in SCFA composition between the 12 months samples that had thawed and the ones that had stayed frozen the whole time. The paired t-tests were computed in RStudio. All confidence intervals include zero, therefore the difference between the thawed and frozen samples is seen as statistically non-significant.

Calibration and optimization

The first set calibration standards consisted of five standards, 1 to 5, with the acids in equal concentrations of 250µM, 500µM, 1000µM, 2000µM and 4000µM, to the respective standard. The acids used were acetic acid, propionic acid, butyric acid, isovaleric acid, valeric acid, and 2-methylvaleric acid. To all these standards formic acid was added in a fixed concentration of 0.2%.

Ten samples from mothers that had dropped out of the study were prepared and analyzed. The SCFAs distribution of the ten samples are listed in Figure I.2 below, together with their calculated average.

The ratios in the samples did not correspond to the ratios listed in the literature, with acetic acid, propionic acid, and butyric acid in a 60:20:20 ratio (Cummings, 1981). Seemingly, there was an overestimation of isobutyric, isovaleric, and valeric acid concentration and underestimation of acetic acid. As the ratios did not comply with the literature, a new set of standards was made. This time the calibration standards were made with different concentrations of the acids, listed in Table D.1 in Appendix D. After the calibration was computed, it was applied to the ten mother samples run previously, and the distribution of SCFAs is visualized in Figure I.2. With this calibration, the ratios of the sample corresponded to the ratios given in the literature, 60:20:20 (Cummings, 1981). As the ratio was correct for the mothers, we assume that the gas chromatograph will detect the correct ratios in the infant samples as well.

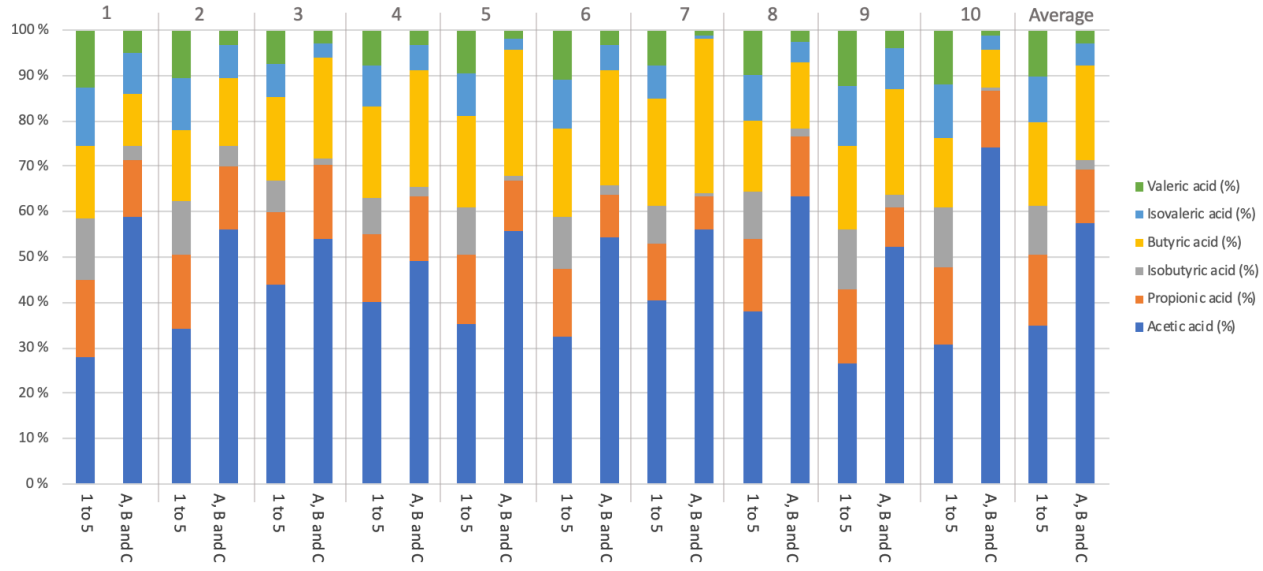


Figure I.2: Comparison of the two sets of calibration standards based on 10 mother samples, and their average.

To see how the concentration measurements varied over time, three runs of the 10 mother samples were conducted, with a standard in-between every set. From these values the average ratios of the short-chain fatty acids for each run were calculated, and presented in Figure I.3, bellow. In addition, a paired t-test was performed and showed no significant differences in SCFAs composition in between runs, using a test level of 0.05.

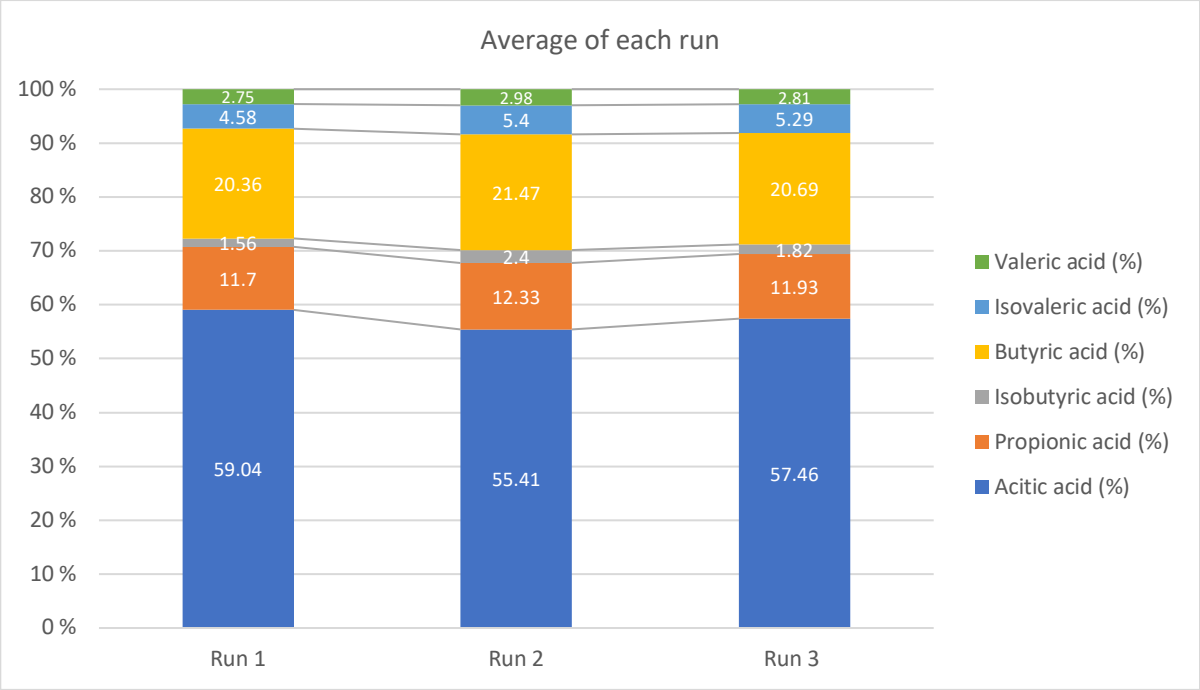


Figure I.3: A set of ten samples from mothers has been analyzed for SCFA composition through gas chromatography three times. The average concentration of all ten samples from one run was calculated, with no significant changes between runs as determined by a paired t-test with a test level of 0.05.

After a few samples had been run, it was noticed that the internal 2-methylvaleric acid standard was measured to be higher than the concentration injected in the sample, however 2-methylvaleric acid was measured precisely in the standards that were run in-between. As the concentration of 2-methylvaleric acid injected in the sample is higher than in the highest of the standards, two new standards were prepared to investigate any extrapolation of the standard curve. Without including the two new standards into the calibration, the GC detected almost the exact concentration of the acids in the newly made standards, thereby removing any suspicion of extrapolation.

This calibration was used throughout the analysis of the 12-month samples, anyhow the column was replaced before analyzing the longitudinal samples of meconium, 3- and 6 months, and the mothers. As the column was replaced it needed to be calibrated, and standards with equivalent concentration to A, B, and C were used, as was an internal standard with the same concentration (1000µM) as previously. After a few of the longitudinal samples had been analyzed an overestimation of 2-methylvaleric acid in the samples was observed. As the highest value of 2-methylvaleric acid in samples was 250µM, extrapolation was again suspected, and a new internal standard with 500µM 2-methylvaleric acid was prepared. The 2-methylvaleric acid in the new internal standard would in the samples coincide with the concentration of 2-methylvaleric acid in standard A. Despite the new standard, high levels of 2-methylvaleric acid were still observed in the samples. A suspicion that there may be another compound with the same retention time as the 2-methylvaleric acid arose. A sample that had been run previously was prepared again. However, this time, the sample was diluted ¼ and had a 0.2% formic acid concentration, and the 2-methylvaleric acid was not included in the preparation. The analysis of this sample showed a top at the same retention time as where the 2-methylvaleric acid should have been and thereby implying that the high 2-methylvaleric acid concentrations are a result of overlapping peaks with an unknown compound.

Appendix J – Average immune cell

Table J.1: Average immune cell compositions, and standard deviations of 67 infants at 12 months of age.

Immune cells	*Average (%)*	*Standard deviation (%)*
neutrophils	30.79	8.33
naivecd4t	20.88	5.69
naiveb	12.40	3.47
naivecd8t	7.73	2.43
classicalmono	6.15	1.98
others	4.34	1.77
effectormemorycd4t	2.75	0.85
eosinophils	2.37	1.63
cd56dimnk	2.20	1.04
effectormemorycd8t	1.73	1.35
naivetregs	1.60	0.50
centralmemorycd4t	1.00	0.35
igdnegmemoryb	1.00	0.38
cd161posgdt	0.86	0.49
cd161neggdt1	0.72	0.35
nonclassicalmono	0.65	0.63
memorytregs	0.55	0.16
cd56brightnk	0.47	0.20
basophils	0.44	0.19
centralmemorycd8t	0.32	0.20
mait	0.28	0.14
activatedcd8t	0.27	0.45
igdposememoryb	0.19	0.16
pdc	0.17	0.07
dptcells	0.05	0.02
proinflammatorymono	0.05	0.03
transitionalb	0.02	0.02
plasmablasts	$8.96*10^{-4}$	$4.2*10^{-3}$

Appendix K – Euclidean distance PCoA plot illustrating the diversity of bacterial orders of 12-month samples

In Figure K.1 two clusters of 12-month samples were observed. By analyzes of the two clusters it became apparent that the sample distribution between the clusters was coinciding with the distribution of samples on the extraction- and PCR-plates, and therefore no further investigations of the two clusters were performed. Figure K.1 showed that the infants with immune cell data connected were representative of the taxonomic microbiota distribution for all 12 months samples. The samples with immune data were represented in both clusters and no pattern was detected between them and the rest of the 12-month samples.

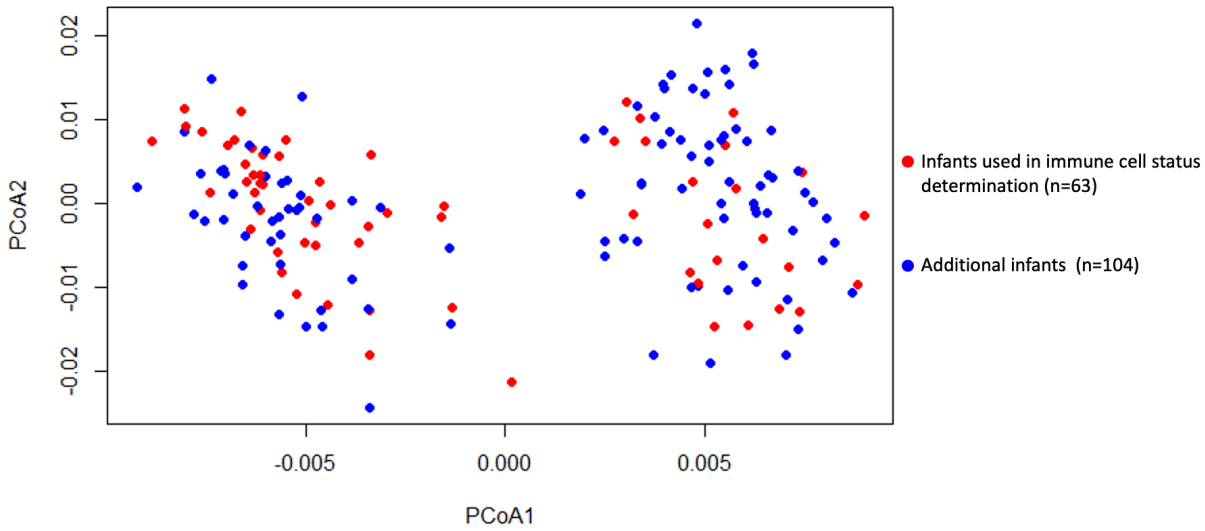
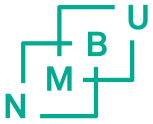


Figure K.1: Euclidean distance PCoA plot illustrating the diversity of bacterial orders of 12 months samples. Samples marked in red, represent the 12 months samples from infants used in the determination of immune cells, while samples marked in blue are the additional infants, making up a total of 167 infants. In total there are 63 infants with taxonomic information on bacterial order at 12 months and immune cell profiles, and 104 infants without immune cell profiles.

Appendix L - Correlation between short-chain fatty acid and immune cell composition at 12 months of age

The Spearman correlation between bacterial composition and immune cells at 12 months was conducted, however after BH adjustments were there no statistically significant correlations. By looking at the unadjusted p-values several trends were revealed. There was an indication of acetic acid being positively correlated with naïve B-cells ($p = 0.035$, Spearman, unadjusted). Butyric acid showed a trend of being negatively correlated with naïve B-cells ($p = 0.044$, Spearman, unadjusted). Isobutyric acid tended to correlate negatively with memory regulatory T-cells ($p = 0.028$, Spearman, unadjusted), and mucosal-associated invariant T-cells (Mait) ($p = 0.048$, Spearman, unadjusted). Valeric acid showed a trend towards being positively correlated with eosinophils ($p = 0.044$, Spearman, unadjusted) and IgD positive memory B-cells ($p = 0.033$, Spearman, unadjusted). However, after the FDR correction the smallest p-value was 0.62, implying that all the above-mentioned trends were weak, and it is difficult to state that these weak trends are consistent for all 12 months old or just coincidental in this set of samples.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway