Norwegian University of Life Sciences

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

The Origin of the First Colonizers of the Human Infant Gut

Regina Sørensen
MSc Biotechnology

Master thesis in Biotechnology

60 credits

The Origin of the First Colonizers of the Human Infant Gut

The Norwegian University of Life Sciences (NMBU),

Faculty of Chemistry, Biotechnology and Food Science

Supervisor Knut Rudi

Acknowledgements

This master thesis was performed at the Norwegian University of Life Sciences in Ås at the Faculty of Chemistry, Biotechnology and Food Sciences under the supervision of Professor Knut Rudi. Also, PhD Morten Nilsen and PhD Eva Maria Rehbinder as my secondary supervisors.

First, I would provide a big thanks to my supervisor Professor Knut Rudi. Thank you for always helping me with challenges in the lab or during the writing process. I am impressed by all your clever ideas and your scientific enthusiasm. I am deeply grateful for having such a dedicated and talented supervisor.

I would also like to thank PhD Morten Nilsen for helping me with the bioinformatics and statistics, and for always answering all my questions about everything. Thank you for your supportive and cheering attitude! Also, thanks to PhD Eva Maria Rehbinder for helping me with the protocol for the skin samples.

Also, I would like to thank the whole MiDiv group for making me feel so included and for all the knowledge you have shared throughout the year. You made me look forward to going to the lab. Especially thanks to Laboratory Engineers Inga Leena Angell and Ida Ormaasen for being so helpful in the lab and for all the good conversations.

Furthermore, a big thanks to the master students Cecilie Fredheim, Mari Raudstein, and Unni Lise Albertsdóttir Jonsmoen. Together we made a great team both in and outside the lab, I am grateful for every day we got to spend together. Thank you for making this year special.

I would also like to thank Karin C. Lødrup Carlsen and the PreventADALL group giving me the opportunity to analyze your samples and the information connected to these.

Finally, a big thanks to my boyfriend Eirik André Fredheim for your patience and for always supporting me, and to my friends and family for cheering me up and believing in me throughout my education.

Hamarøy, May 2020

Regina Sørensen

Abbreviations

ASV	Amplicon Sequence Variant
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
DADA	Divisive Amplicon Denoising Algorithm
DNA	Deoxyribonucleic Acid
ESwab	Elution Swab
dPCR	Digital PCR
ddPCR	Droplet Digital PCR
HMO	Human Milk Oligosaccharide
HRM	High Resolution Melting
MLST	Multilocus Sequence Typing
NGS	Next Generation Sequencing
OTU	Operational Taxonomic Unit
PCoA	Principal Coordinate Analysis
PC	Principal Component
PCR	Polymerase Chain Reaction
rRNA	Ribosomal Ribonucleic Acid
rpoB	RNA polymerase β-subunit
SNP	Single Nucleotide Polymorphism
qRT-PCR	Quantitative Real Time PCR
Qiime	Quantitative Insights Into Microbial Ecology

Abstract

Several discoveries have unraveled the correlation of the human gut microbiota with both diseases and health. The first colonizers of the infant's gut are crucial for gut microbiota development, immune system, and health later in life. Microbes are shared between mother and child. Hence, it is unclear whether the colonization starts inside the uterus or after the rapture of the amniotic membrane. There still lacks detailed information about the time and mechanisms for bacterial transfer and whether the transmitted bacteria stable colonizes the gut. Previous studies have shown association of *Bacteroides* in mother-child pairs. Thereby, this study aimed to identify the direct mother-child transmission of *Bacteroides* species.

Meconium samples from 464 newborns were screened for *Bacteroides.* For the *Bacteroides*-positive infants (n=51), the bacterial composition was analyzed in meconium and from skin samples. In addition, their respective mother's stool was screened (n=48). The overlap of *Bacteroides* was investigated in 48 mother-child pairs. The samples included in this study were received from the PreventADALL cohort. The bacterial composition in stool and on the skin was studied by multilocus sequence typing (MLST), 16S rRNA sequencing, and quantitative PCR. Further amplicon sequence variant (ASV) profiles were used to study the overlap in the *Bacteroides* species in mothers and children.

Bacteroides were present in meconium, on newborn's skin, and in mother's stool. *Bacteroides vulgatus* was the predominant species. Several species of *Bacteroides* were observed in mother-children pairs and *B. vulgatus* has a significant ($p<0.05$) association between mother and infant. The ASV profiles indicated the presence of one abundant sequence variant of *B. vulgatus*. All *Bacteroides*-positive infants were delivered vaginally. *B. vulgatus* discovered in infants was directly linked to the mother's stool, which may indicate her gut is an important plausible reservoir of the initial colonizers of the infant's gut.

In conclusion, *B. vulgatus* is transferred from mother to child, and one sequence variant of this species was predominant in the pairs. The increase in content of *B. vulgatus* in three days indicates the ability of *B. vulgatus* to proliferate in the gut. In addition, the amount of *B. vulgatus* was significantly ($p<0.05$) higher compared to most other *Bacteroides* species. These findings point towards that *B. vulgatus* possibly form stable colonies in infants' gut. Further research is essential to investigate if the sequence variant of *B. vulgatus* is stable in the gut over a longer time period and if the transferred bacteria perform important functions for the development of the gut microbiota and the immune system.

Sammendrag

Flere studier har avdekket korrelasjon mellom menneskets tarmmikrobiota med sykdom og helse. De første koloniserende mikrobene i nyfødtes tarm er essensielt for utviklingen av tarmmikrobiotaen, immunsystemet og helse senere i livet. Mikrober overføres fra mor til barn. Det er likevel uklart om koloniseringen starter når barnet er inni livmoren, eller om det skjer når fostervannsmembranen er brutt. Det mangler fortsatt detaljert informasjon om tidspunkt og mekanismene for bakteriell overføring, og om de overførte bakteriene stabilt koloniserer tarmen. Tidligere studier har vist en sammenheng av *Bacteroides* i mor-barn par. Derfor er målet med denne studien å identifisere direkte mor-barn overføring av *Bacteroides*-arter.

Mekonium (barnebek) fra 464 nyfødte ble screenet for *Bacteroides.* For de *Bacteroides*-positive barna (n=51), ble den bakterielle komposisjonen analysert i mekonium og fra hudprøver. I tillegg ble avføringsprøven fra deres respektive mødre screenet (n=48). Overlapp av *Bacteroides* ble studert i 48 mor-barn par. Prøvene inkludert i denne studien er en del av «PreventADALL» kohorten. Den bakterielle komposisjonen i avføring og på hud ble studert ved multilokus sekvensering (MLST), 16S rRNA sekvensering og kvantitativ PCR. Videre ble amplikon sekvensvariant (ASV) profiler brukt for å studere overlapp av *Bacteroides*-arter hos mødre og hennes barn.

Bacteroides-arter ble påvist i de nyfødtes mekonium, deres hud, og i avføringsprøvene hos deres mødre. *Bacteroides vulgatus* var den dominerende arten. Flere *Bacteroides*-arter ble observert i mor-barn parene hvor *B. vulgatus* viste en signifikant ($p<0.05$) sammenheng mellom mødrene og hennes nyfødte barn. ASV profilene resulterte i en dominant sekvensvariant av *B. vulgatus*. Alle de *Bacteroides*-positive barna var født vaginalt. *B. vulgatus* hos de nyfødte var direkte linket til mødrenes avføring, noe som indikerer at morens tarm sannsynligvis er en viktig kilde for de første koloniserende bakteriene i nyfødtes tarm.

Konklusjonen er at *B. vulgatus* overføres fra mor til barn, og at en sekvensvariant av denne arten ser ut til å dominere i mor-barn parene. I mekonium ble det observert en økende mengde av *B. vulgatus* i en tidsperiode på tre dager, noe som indikerer evnen bakterien har til å formere seg i tarmen. I tillegg var mengden *B. vulgatus* signifikant høyere ($p<0.05$) sammenlignet med de fleste andre *Bacteroides*-arter. Dette peker mot at *B. vulgatus* selekteres for og kan danne stabile kolonier i nyfødtes tarm. Videre studier er nødvendig for å undersøke om sekvensvarianten av *B. vulgatus* er konsistent i tarmen over tid og om overførte bakterier utfører viktige funksjoner for videre utviklingen av tarmmikrobiotaen og immunsystemet.

Table of Contents

1.0 Introduction

1.1 Human microbiota

1.1.1 Human gut microbiota

The human microbiota comprises all microorganisms that live in different niches in and on the human body (Thursby & Juge, 2017). These microorganisms belong to the domains Bacteria, Archaea, and Eukarya. In humans, the most important microbial colonization site is in the colon, and this intestinal segment has the highest density of microorganisms (Alonso & Guarner, 2013). It is estimated that there are around 10^{14} microorganisms in the human colon, and bacterial cells are 10 times more compared to human cells (Thursby & Juge, 2017). Another article has reported that the human to cell ratio could be 1:1 (Sender et al., 2016). As for the amount of bacterial species in an adult human gut, there are around 400 species (Vael & Desager, 2009). However, lower numbers such as 160 have also been suggested (Faith et al., 2013). The number of bacterial cells and species in the human gut is still not clearly defined.

The human adult microbiota is proposed to be relatively stable unless an imbalance is evoked by antibiotic usage or drastic changes in diet. The hypothesis of a core gut microbiota has been presented which indicates that most of the adults have many common bacterial species (Lozupone et al., 2012). Further, these microorganisms are suggested to be persistent in the gut for years (Faith et al., 2013). The dominating bacterial phyla in adults' gut are primarily Bacteroidetes and Firmicutes, but also taxonomic groups such as Actinobacteria, Proteobacteria, and Verrucomicrobia are present (Eckburg et al., 2005). The human gut microbiota has been stratified into enterotypes which are distinct compositional bacterial types; three enterotypes are observed in gut microbiome data collected from around the world. The variation in enterotypes is based on the abundance of the genus *Bacteroides* and *Prevotella,* from the phylum Bacteroidetes, and that of *Ruminococcus* belonging to phylum Firmicutes (Wu et al., 2011; Costea et al., 2018). Although classification based on enterotypes can provide information about biological differences, key microbial variation cannot be uncovered by analyzing only the enterotypes (Costea et al., 2018).

The gut microbiome is the collection of all genomes from microbes inhabiting the gut (Walker et al., 2017). The gut genome has unique sets of genes encoding important metabolic

capabilities. The “core microbiome” or the “island” theory considers each individual to be inhabited by a unique collection of gut microbiota, which overall share core functions (Turnbaugh et al., 2009; Vael & Desager, 2009). Core metabolic functions in the human gut are consistent and include important metabolic functions, such as the amino-acid and carbohydrate metabolism or the biosynthesis of secondary metabolites (Turnbaugh et al., 2009). The metabolic functions of the gut microbiota help them to thrive in the human gut (Tremaroli & Bäckhed, 2012).

Many of the microorganisms in the gut have a mutualistic relationship with the human body and such a symbiotic bond is critical for the host health. Mutualistic gut microbiota offers several benefits to the human body, for example, produce vitamins and short-chain fatty acids (SCFAs). Examples of mutualistic bacteria are species of *Bacteroides* that produce vitamins such as biotin, riboflavin pantothenate and ascorbate, and *Prevotella* that generate thiamine and folate (Arumugam et al., 2011). The gut microbiota that works as a metabolic organ provides energy to the host; by producing SCFAs through the fermentation of complex dietary carbohydrates, including cellulose, xylans, inulin, and resistant starch (Tremaroli & Bäckhed, 2012). The most abundant SCFAs produced by bacteria are acetate, propionate, and butyrate (den Besten et al., 2013). According to Alonso & Guarner (2013), the intestinal epithelial cells derive most of their energy from SCFAs fermented by bacteria. Likewise, SCFAs are important for human health as the largest and most complex part of the human immune system is located in the gut. The colonization of commensal microbes in the gut also stimulates the immune system and bestow them the ability to respond and ward off invading pathogens (Alonso & Guarner, 2013). The metabolic functions provided by commensal bacteria are key for a healthy human body.

The first microbial colonizers are involved in the bacterial host interactions which are established during infancy. These interactions are critical for human health as the first colonizers are important for gut maturation, development of the immune system, the intestinal tract, and the associated metabolism (Dominguez-Bello et al., 2010; Milani et al., 2017). Inadequate development of the infant gut microbiota is linked to allergic diseases such as asthma, obesity, cardiovascular disorders, chronic disorders like inflammatory bowel disease (IBD), irritable bowel syndrome (IBS), and diabetes. (Milani et al., 2017). The establishment of the infant gut microbiota is crucial for health later in life.

1.1.2 Infant gut microbiota establishment

During and after birth, the nearly sterile infant gut develops and changes rapidly (Hansen et al., 2015). Some phyla that dominate the infant's gut are Actinobacteria, Bacteroidetes, and Proteobacteria (Rodríguez et al., 2015; Yassour et al., 2018). During infancy and early childhood, the bacterial composition changes from facultative anaerobes to obligate anaerobes. Facultative anaerobes, like enterobacteria, deplete oxygen in the gut and make the environment more favorable for obligate anaerobes such as *Bacteroides, Bifidobacterium,* and *Clostridium* (Del Chierico et al., 2015). The compositional changes in the infant gut microbiota are dependent on the anaerobic environment and the bacteria that are part of the ecosystem.

Changes in the infant gut bacterial proportion and composition are mainly driven by diet and can be divided into three phases: developmental, transitional, and stable (Stewart et al., 2018; Derrien et al., 2019). The compositional changes are based on the content of Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, and Verrucomicrobia (Derrien et al., 2019). The developmental phase of the gut microbiota happens during 3-14 months of age. At this stage, the infant's diet will be either breastmilk or formula. The gut of breastfed infants is dominated by *Bifidobacterium* (Stewart et al., 2018). The prevalence of *Bifidobacterium* increases during breastfeeding since human milk oligosaccharides (HMOs) from the mother's milk propel the selective pressure to favor the establishment of *Bifidobacterium* species (Milani et al., 2017; Stewart et al., 2018). In addition, several *Bacteroides* species can utilize milk oligosaccharides (Thursby & Juge, 2017). The development phase is mainly driven by the introduction of breast milk to the diet.

Subsequent to the development phase, the transitional and stable phases occur. They represent the compositional changes in the infant gut microbiota. Alpha diversity is in relation to human gut microbiota the bacterial diversity within one person's gut. The transitional phase starts from around 15 months of age and extends to 20 months, during which period the diet consists of both breast milk and solid food. The introduction of solid food increases the gut microbiota diversity where Bacteroidetes and Proteobacteria continue to develop (Stewart et al., 2018; Derrien et al., 2019). The next phase, namely stable phase, occurs around 31- 46 months of age. At this time, the diet consists of solid food (Milani et al., 2017; Stewart et al., 2018). The stable phase is associated with an increased alpha-diversity. Likewise, the abundance of *Clostridia* increases as the conditions are more favorable to the bacteria belonging to Firmicutes, at the same time as the prevalence of *Bifidobacterium* decreases

(Avershina et al., 2016; Stewart et al., 2018; Derrien et al., 2019). Thus, three bacterial compositions and development in the infant's gut is strongly modified by diet.

1.1.3 Human skin microbiota

The human skin is another ecological niche for microbial colonization. The skin comprises different conditional sites such as sebaceous, moist, or dry areas. Dominating genera in moist areas are *Corynebacterium* and *Staphylococcus*, and *Propionibacterium* is the dominant type in sebaceous areas. The dry areas have the most diverse bacterial diversity among the skin sites; the predominant phyla are Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes (Grice et al., 2009). The microbiota of newborn babies is homogeneous across the body compared to adults (Dominguez-Bello et al., 2010). Infant's skin microbiota is dominated by phyla (in the order of their abundance): Firmicutes, Actinobacteria, Proteobacteria, and Bacteroidetes. The dominating genera on infants' skin are *Streptococcus* and *Staphylococci* (Capone et al., 2011). The skin microbiota at different body sites is homogenous in infants.

1.2 Mother-child associated bacteria

1.2.1 Initial bacterial transmission from mother to infant

The timing and mechanism of bacterial transfer are key to understanding the establishment of the gut microbiome (Walker et al., 2017). The timing of bacterial transfer from mother to child is not fully understood. Sterility of the intrauterine environment is a contentious subject (Figure 1.1) (Kuperman et al., 2020). To date, there are two hypotheses about the timing of bacterial colonization in the infant's gut. First is the non-sterile *in utero* hypothesis that states that the bacterial colonization starts inside the womb during pregnancy (Walker et al., 2017). However, some studies suggest a non-sterile environment with bacteria present in the placenta, amniotic fluid, and the umbilical cord (Jiménez et al., 2005; Aagaard et al., 2014; Collado et al., 2016). As for the non-sterile hypothesis, there are suggested mechanisms for bacterial transfer.

Pregnant women undergo physical changes and the gestation period may favor translocation of bacteria to the intrauterine environment. According to Walker et. al. (2017) proposed mechanisms behind the establishment of a non-sterile *in utero* environment (Figure 1.1) are the transfer of bacteria either through the placental barrier or via amniotic fluid or by vaginal

ascension or hematogenous translocation through the cord blood. However, intrauterine sites, such as the placenta and amniotic fluid, contain low bacterial content which necessitates the DNA analysis to be sensitive (avoid false negative detection) to contamination. In addition, sequence-based technologies do not distinguish between living and dead bacteria (Walker et al., 2017; Kuperman et al., 2020). The non-sterile hypothesis is doubtful even though there are proposed mechanisms for bacterial transfer to the *in utero* environment (Walker et al., 2017).

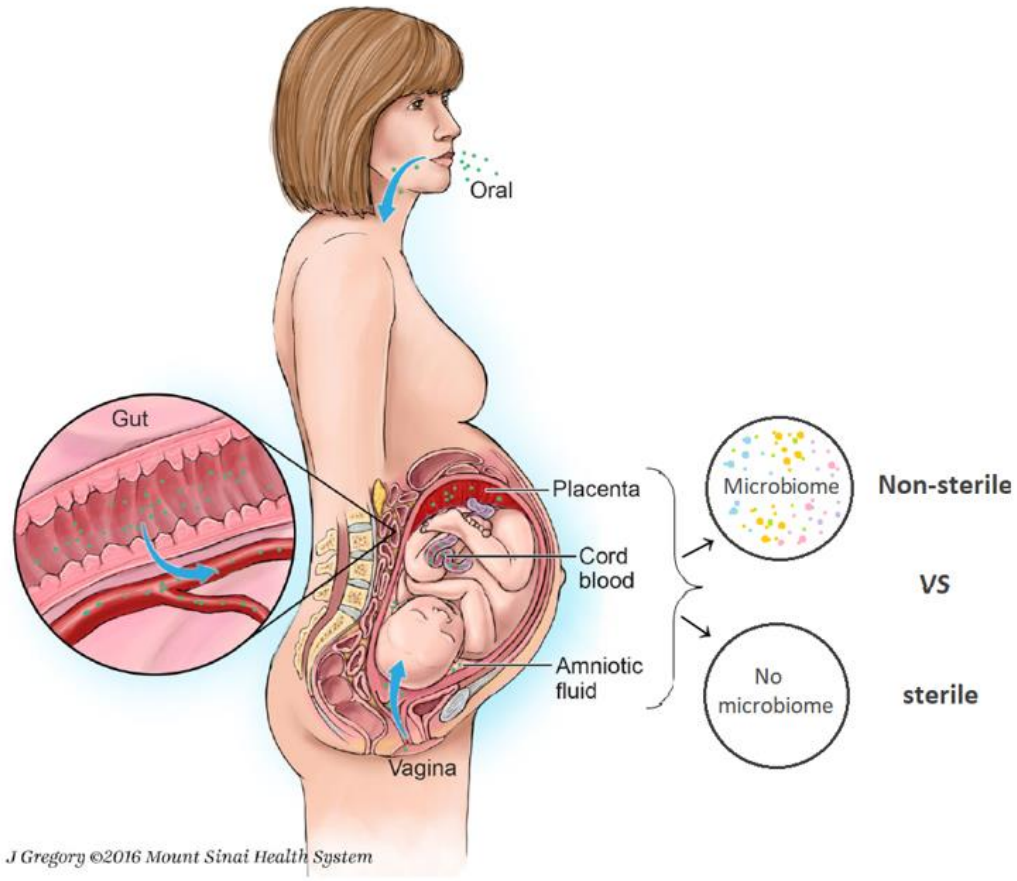


Figure 1.1. Is the environment *in utero* sterile or not? According to the proposed mechanisms of Walker et. al. (2017) there occurs vertical transfer of bacteria from mother to offspring; bacteria cross the epithelial barrier and access the lymphatic and circulatory system and further translocate to intrauterine sites. Another suggestion is the translocation of bacteria from inflammatory gums to the placenta via the bloodstream, although, mothers and their infants do not share oral bacterial species (Ferretti et al., 2018). Likewise, bacterial ascension from the vagina has been proposed (Walker et al., 2017). In the context of healthy pregnancies, the origin of bacteria present *in utero* is still unknown. A report suggests the presence of bacteria in the placenta during a healthy pregnancy (Zhu et al., 2018). However, there is no evidence of bacterial colonization in amniotic fluid, fetal membrane, etc. during a healthy pregnancy. Bacteria established at these sites are associated with infections and pre-term delivery (Walker et al., 2017). Source: retrieved and edited from Walker. et.al. (2017).

By contrast, the sterile *in utero* hypothesis is that the environment intrauterine is sterile (Figure 1.1). Intrauterine, the fetus is physically separated from the lumen by the amniotic membrane (Hornef & Penders, 2017). Previous studies indicate that the placenta and amniotic membrane are sterile. Microbial colonization starts after the rapture of the amniotic membrane and the uterine contractions have started (Lim et al., 2018; Rehbinder et al., 2018; de Goffau et al., 2019; Theis et al., 2019; Kuperman et al., 2020). There is an ongoing debate concerning the non-sterile and sterile hypothesis. However, the time and mechanism of transmission could be explained by understanding where the overlapping bacteria are originated from.

1.2.2 Bacterial overlap from mother to infant

The mode of delivery, either vaginal or caesarean section (C-section), determines the type of bacteria that are transmitted from mother to infant. The vagina and gut microbiota represent important sources of bacteria that potentially could colonize the infant's gut. Therefore, vaginal birth is key to the establishment of infant gut microbiota (Walker et al., 2017). Species of *Bacteroides* are thought to be directly transmitted from mother to child as the prevalence of *Bacteroides* (*B. vulgatus, B. ovatus, B. fragilis, B.xylaisolvens, B. thetaiotaomicron)* was shown to be higher in early gut colonizers of vaginally infants compared to infants delivered by C-section (Bäckhed et al., 2015; Stewart et al., 2018; Eck et al., 2020). Colonization of vertically transmitted bacteria like *Bacteroides* and *Bifidobacterium* are delayed in C-section delivered infants. In addition, vaginally-delivered infants show a higher prevalence of *Lactobacillus* acquired from the maternal vaginal microbiota (Vael & Desager, 2009; Nagpal et al., 2016). Infants delivered by C-section are more often colonized by skin and hospital linked microbes compared to vaginally-delivered. When it comes to the gut microbiota, C-section infants show higher colonization of *Clostridium* and *Enterococcus* compared to vaginally-delivered newborns (Dominguez-Bello et al., 2010; Milani et al., 2017). The mode of delivery governs the establishment of transmitted bacteria from mother to child, where the mother's vaginal and gut microbiota are sources for vertically transmitted bacteria in the case of delivery through the birth canal.

Vertically transmitted bacteria, especially *Bacteroides*, from mother to infant are more likely to form stable colonization in the infant's gut over time compared to transmitted bacteria not associated with the mother (Korpela et al., 2018). Mother-child associated bacteria have traits making them able to colonize the infant gut (Ferretti et al., 2018). The oral cavity presents a

likely transmission route, by swallowing, for bacteria that colonize the infant's gut. Bacterial sharing between the infant's oral cavity and gut microbiota has been illustrated as enhanced close after birth and decreased with time (Ferretti et al., 2018). The environment infants are first exposed to modifies which bacteria that can form stable colonization in the gut, which is illustrated by the oral-gut microbiota sharing. The mother's gut has been suggested as the main source of vertically transferred bacteria from mother to infant (Ferretti et al., 2018).

1.2.3 Genus *Bacteroides*

Bacteria from genus *Bacteroides*, belonging to the phyla Bacteroidetes, are dominant in the human gut. *Bacteroides* are non-spore forming, gram-negative bacilli (Wexler, 2007). The bacteria are considered as obligate anaerobes, however, *Bacteroides* are also aerotolerant and able to divide in the presence of oxygen in nanomolar concentrations (Bacic & Smith, 2008). By representing 25-50% of the gut microbiota, *Bacteroides* are the most prevalent gram-negative bacteria in the human gut (Townsend et al., 2020). In infants' gut, species of *Bacteroides* are important in the immune and microbial development. Interestingly, bacteria from the phylum Bacteroidetes have been suggested as the main producer of the SCFA propionate (Salonen et al., 2014; Aguirre et al., 2016). Species of *Bacteroides* such as *B. fragilis*, are saccharolytic organisms that harbor carbohydrate utilizing genes. These genes make *Bacteroides* able to degrade dietary fibers and thereby acquire energy and to colonize the human gut; in addition to providing other microbiota and immune epithelial cells with energy (Wexler, 2007; Townsend et al., 2020). Under commensal conditions, *Bacteroides* have capabilities that make them beneficial bacteria both in the infant and adult gut.

Interestingly, some bacteria possess genes which may benefit them during colonization in infants' gut. Species of *Bacteroides*, especially *B. fragilis* and *B. vulgatus*, have been described to have commensal colonization factor genes (ccf). These genes, ccfA-E, helps in polysaccharide utilization; gene clusters called polysaccharide utilization loci. These commensal colonizing factors are upregulated during gut colonization at the colonic surface (Lee et al., 2013). Some bacteria may benefit from certain capabilities that are required to form a stable microbiota in infants' gut.

Some mice studies present indications of how *Bacteroides* species could influence the immune system. Among the *Bacteroides* species in the human gut, the commensal *B. fragilis* has been shown to influence the development of thymic lymphocytes. The thymic

lymphocytes are derived from the bone marrow and further developed in the thymus. A study showed that by colonizing germ-free (GF) neonate mice with only *B. fragilis* increased the number of thymocytes compared to mice lacking monocolonization (Ennamorati et al., 2020). Another notable species is *B. ovatus* which has been found to induce IgA production in the mice gut (Yang et al., 2020). Mice studies show how *Bacteroides* could be beneficial for the immune system.

Bacteroides serve other purposes in the human gut; produce anti-inflammatory components and antimicrobial peptides (Mazmanian et al., 2008; Roelofs et al., 2016). Polysaccharide A (PSA) is an anti-inflammatory component produced by *Bacteroides* which protects the host from immune diseases (Mazmanian et al., 2008). *Bacteroidales* species secrete a total of five antimicrobial peptides which are beneficial for the competitive fitness in the gut (Roelofs et al., 2016; Shumaker et al., 2019). According to Shumaker et al. (2019), *B. fragilis* has been suggested as the major producer of antimicrobial peptides among the *Bacteroides* species. The antimicrobial peptides antagonizes closely related bacteria and potential pathogens. Antagonism between closely-related species could give a better understanding of the importance of maintaining gut microbial composition and their stability (Shumaker et al., 2019). In contrast, the production of toxins could be a potential virulence factor that threatens the host health. The peptides and components produced by *Bacteroides* are beneficial for the bacterial competition, but could also potentially harm the host.

Bacteroides are normally commensals in the human gut. According to Wexler (2007), the bacterial species also have virulence factors such as protecting itself from the immune system, destroying of or adherence to gut tissue. The virulence factors could cause serious infections like bacteremia. Especially *B. fragilis* is known to cause harm, for instance, its capsule could initiate an immune response in the host that could lead to the formation of abscesses. The bacterial capsule can also evade the host immune response and help the bacteria to become resistant to phagocytic killing. In addition, *Bacteroides* could modulate the surface polysaccharides which makes it able to avoid the host immune system. Other virulence factors are enzymes produced by *Bacteroides.* For example, proteinase could destroy brush border enzymes in the gut with important functions in the final digestion and absorption of nutrients. Similarly, *Bacteroides* produce enterotoxins which could cause loss of tight junctions between the epithelial cells and thereby generate a leak between the gut lumen and lamina propria (Wexler, 2007). Treatment of infections caused by *Bacteroides* could be a challenge as the

species is resistant to a wide range of antibiotics (Bacic & Smith, 2008). *Bacteroides* have several virulence factors with the potential to cause serious harm to the host.

1.2.4 Potential bacterial colonization of meconium

The first colonizers of the infant's gut can be analyzed by studying meconium, which is the very first stool delivered by a neonate. Meconium contains substances such as amniotic fluid that the fetus has ingested from the environment *in utero* and during birth (Gosalbes et al., 2013). As neonate stool is already colonized by microorganisms, analysis of meconium will give an indication of the microorganisms that can first reach and possibly colonize the gut. However, the microbial complexity is low in meconium compared to the community composition in adults' gut (Yassour et al., 2018). Some bacterial genera discovered in meconium are *Bacteroides, Bifidobacterium, Enterobacteria, Klebsiella, Enterococcus, Streptococcus,* and *Staphylococcus* (Jiménez et al., 2008; Hansen et al., 2015; Chu et al., 2017). Cultivation-based studies have shown that there are living cells of bacteria in meconium. Bacteria isolated from meconium were dominated by species of *Enterococcus, Escherichia, Staphylococcus,* and *Streptococcus* (Jiménez et al., 2008; Nagpal et al., 2016; Moles et al., 2020). Similarly, *Bifidobacterium, Enterobacter, Klebsiella, Bacteroides, Parabacteroides, Rothia,* and the lactic acid bacteria *Leuconostoc* have been isolated from meconium (Jiménez et al., 2008). Bacteria in meconium represents species that can potentially colonize infants' gut and can provide information about the origin of the bacteria.

1.3 Analysis of taxonomic composition

Culture-dependent and culture-independent techniques are employed in microbiology. On the one hand, bacterial traits like morphology could be observed by cultivation. However, one limitation with cultivation studies is the challenge to fulfill all the elements required for bacterial growth. The conditions which must be considered are nutrients, temperature, atmosphere, and the incubation time necessary for growth (Lagier et al., 2015). Cultivating microorganisms from diverse environments like the human gut could be difficult because several fundamental demands must be fulfilled. The culturable bacteria will thereby present a limited overview of the community composition. In contrast, culture-independent techniques do not have cultivation challenges. Some culture-independent approaches to examine

taxonomic composition in the gut is by 16S ribosomal RNA (16S rRNA) sequencing and the analysis of genetic markers.

1.3.1 16S ribosomal RNA sequencing

The 16S ribosomal RNA (rRNA) gene is commonly used for taxonomic classification of the prokaryotic community composition. Bacterial species can easily be distinguished by 16S rRNA sequencing. The gene is highly conserved, present in all prokaryotes, and protected from horizontal gene transfer events. The 16S rRNA gene contains 9 hypervariable regions (V1-V9) which allows bacteria to be distinguished at a higher taxonomic level. While the entire 16S rRNA gene can be used to identify prokaryotes at lower taxonomic levels (Thursby & Juge, 2017). However, 16S rRNA gene sequencing-based analyzes will not allow capturing sequence variation to separate phylogenetically close bacteria because the sequence divergence is rather low (Ko et al., 2007). 16S rRNA sequencing is a widely used tool for bacterial compositional analysis, however, the approach fails to separate closely-related bacterial species.

1.3.2 Genetic markers for phylogenetic studies

Genetic markers are used for assigning taxonomy by detecting variation within genomes of prokaryotic environments. Among the genetic typing approaches are multilocus sequence typing (MLST) which is based on single nucleotide polymorphism (SNP) (Maiden et al., 1998). SNP is a rare variation within single nucleotides in the genome caused by point mutations. SNPs are the most abundant polymorphisms in the human genome and is often used as a genetic marker because of a low mutation rate. Genetic markers can, for instance, be used to identify severe diseases or to analyze complex microbial environments (Sobrino et al., 2005). Both SNP-based strain typing and MLST is based on comparing nucleotide variation within the genome (Van Belkum et al., 2007). Sequence-based genetic markers are based on analyzing sequences at the nucleotide level to detect taxonomic composition.

1.3.3 Multilocus sequencing typing

MLST is a sequence-based typing approach used in bacterial community studies. MLST was first proposed for the identification of the pathogen *Neisseria meningitis.* However, MLST is

also applicable for taxonomic characterization of other known pathogenic and non-pathogenic bacterial species (Maiden et al., 1998). The concept of MLST is to type and identify DNA sequence variation between alleles at multiple loci of housekeeping genes (Maiden et al., 1998; Van Belkum et al., 2007). Housekeeping genes are genes essential for a cell to survive, such as genes encoding enzymes involved in protein synthesis. The biological variation between microorganisms can be detected by analyzing alleles at unambiguous housekeeping genes. Further, strains with the same alleles are defined as the same sequence type (Sakamoto & Ohkuma, 2011). The sequence variant profiles derived from multilocus sequencing can be used to study sequence variation and thereby decide the bacterial taxonomic composition, both at the species and strain level (Maiden et al., 1998). MLST is a favorable approach when closely related bacterial species are studied as the technique helps to detect the variation at the nucleotide level.

1.3.4 Genetic markers to examine *Bacteroides*

Genetic markers are used in diversity studies to separate bacteria up to the highest taxonomic level. Some protein-encoding genes used for classification of *Bacteroides* include *recA, dnaJ, gyrB, hsp60,* and *rpoB* (Sakamoto & Ohkuma, 2011). The housekeeping gene RNA polymerase β-subunit (*rpoB*) is used in phylogenetic studies to separate similar strains within the genus *Bacteroides* (Ko et al., 2007; Sakamoto & Ohkuma, 2011) and is shown as a promising marker in other diversity studies (Ogier et al., 2019). Closely-related species are easier to separate by analyzing the *rpoB* gene since the gene provides a greater resolution and higher genetic variation compared to the V3-V4 region of 16S rRNA. In addition, the rate of false positives is lower in *rpoB* gene-based typing (Ogier et al., 2019). Sequences of protein-encoding genes are ideal to distinguish bacteria such as gut-dwelling *Bacteroides* species that are closely-related (Sakamoto & Ohkuma, 2011).

1.4 Assigning taxonomy

1.4.1 Operational taxonomic units

The sequences of housekeeping genes can be assigned taxonomy by clustering DNA barcodes into operational taxonomic units (OTUs). Sequences are assigned a taxonomy (e.g. species) if the similarity between the sequences is above a given similarity threshold, commonly the

value is set to 97%. However, this phylogenetic assignment may not always be accurate. A disadvantage of accepting a 3% sequence dissimilarity is that variation on a nucleotide level may be lost which could represent an important phylogenetic separation (Callahan et al., 2017). Bioinformatics tools are used to assign and visualize phylogenetic composition. The Qiime (Quantitative Insight Into Microbial Ecology) pipeline is a bioinformatic tool that is widely used to assign taxonomy to 16S rRNA and other sequenced data. The pipeline processes raw sequencing data into an interpretable microbial community (Caporaso et al., 2010). Bacterial taxonomy is studied by assigning OTUs, however, it does not necessarily distinguish all phylogenetic variation.

1.4.2 Amplicon sequence variants

Another approach that can separate sequences down to the nucleotide level is amplicon sequence variants (ASVs). ASVs implies the original DNA sequence before amplification and can thereby give a sequence variation profile. ASVs operates with a higher resolution and is more precise than similarity-based OTUs. In addition, ASVs-based methods discover all sequence variations in the data better compared to closed-referenced OTUs (Callahan et al., 2017). ASV profiles could be obtained using bioinformatical tools such as the DADA2 pipeline. Amplicon sequence variant profiles give a broader understanding of community diversity within and between samples. DADA2 is an extension of Divisive Amplicon Denoising Algorithm (DADA) which was originally designed for correcting errors and identifying variation from pyrosequencing amplicons. DADA2 is developed for Illumina sequencing amplicons. Compared to OTU-constructing algorithms, DADA2 is more accurate in detecting variations in microbial communities and obtaining fewer incorrect sequences (Callahan et al., 2015). ASV profiles are handy in the phylogenetic analysis where closely related bacterial species are studied.

1.4.3 Databases and classification tools for sequencing data

There are several databases for assigning taxonomy to sequences retrieved from high-throughput sequencing approaches like Illumina. Examples of classification tools are BLAST (Basic Local Alignment Search Tool), SILVA (from Latin silva, forest) ribosomal RNA gene, and Kraken databases. A commonly used tool for 16S rRNA sequences is the quality-controlled SILVA database, which contains databases for the small and large subunit rRNA

for domains Archaea, Eukarya, and Bacteria (Quast et al., 2012). Kraken, another system that assigns taxonomic labeling, provides genus-level precision and sensitivity and is often used in metagenomic studies to assign taxonomy to shorter DNA sequences. The algorithm used for Kraken is based on assigning sequences into weighted *k*-mers used for sequence classification by the last common ancestor (LCA) (Wood & Salzberg, 2014). The available databases are made for different purposes and contents, and users should consider these aspects when choosing the right database for assigning sequence taxonomy.

1.5 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a highly sensitive method that is often used for studying microbial communities. The three main steps in each PCR cycle include denaturation, annealing, and extension. During denaturation, double-stranded DNA (dsDNA) is melted by high temperature into single-stranded DNA (ssDNA). The annealing step renders the primer bind to DNA at an optimum temperature. The final step involves the DNA polymerase which extends the primer sequence. The three steps are repeated for a given amount of cycles, during which the amount of DNA theoretical grows exponentially (Mullis et al., 1986). PCR can either be quantitative or qualitative.

1.5.1 Qualitative PCR

Both quantitative and qualitative PCR approaches should be considered while deciding the method that is appropriate for the study goal. Qualitative PCR includes an end point analysis, such as gel electrophoresis, which will confirm or disprove the presence of the sequence of interest and give an indication about the amplicon size. Contrarily, quantitative real-time PCR (qRT-PCR) measures the amount of DNA in real-time at each PCR stage (Kubista et al., 2006).

1.5.2 Quantitative PCR

Employing qRT-PCR, the concentration of amplified PCR products can be monitored during the thermal cycles by using fluorescent probes. The fluorescent signals increase after every thermal cycle according to the number of amplified amplicons. The number of cycles

necessary to reach the fluorescent threshold value is defined as the Cq-value. A lower yield of DNA results in a higher Cq-value, and for high DNA yield the Cq-value will be low. A standard curve is included to calculate the DNA concentration. Further, a melt curve analysis could be included for fragment comparison (Kubista et al., 2006). Analysis by qRT-PCR gives information about DNA yield and gives the opportunity to investigate different fragments.

Digital PCR (dPCR) is another PCR approach used to quantify the number of DNA fragments in samples (Hindson et al., 2011; Pinheiro et al., 2011). dPCR is based on dividing the sample into several portions where each act as individual reactions (Rački et al., 2014). The chance of potential inhibitors disrupting the polymerase or primer annealing decreases by generating the sample into several droplets (Rački et al., 2014). As for qRT-PCR, fluorescent signals from amplified PCR products gives information about the DNA concentration in the samples (Hindson et al., 2011). However, while qRT-PCR measures fluorescent real-time, fragments amplified by dPCR are directly quantified in an end point analysis based on whether the droplets contain PCR products or not. By an end point analysis, the need for a standard curve or gel electrophoresis is excluded (Pinheiro et al., 2011). dPCR is unique by dividing samples into several droplets which are analyzed in an end point analysis.

1.6 Nucleic acid sequencing techniques

1.6.1 First-generation sequencing

Fredrick Sanger and his colleagues reported a breakthrough in the first-generation sequencing in 1977. The sequencing technique is also referred to as the Sanger's chain termination or the dideoxy technique. Sanger sequencing is based on using deoxynucleotide triphosphates (dNTPs) to a sequence where the dNTPs are specific for each DNA base. In addition, fluorescent dideoxynucleotide triphosphates (ddNTPs), are added. The ddNTPs are analog of the monomer dNTPs but unlike dNTP, they lack hydroxyl (OH) group in the 3'-position which causes the elongation to stop. One reaction is performed per ddNTP which results in DNA fragments with different lengths. The fragments can then be separated employing gel electrophoresis which identifies the DNA sequence (Sanger et al., 1977). Sanger sequencing technology was the first breakthrough in the history of nucleic acid sequencing techniques.

1.6.2 Second-generation sequencing

The second-generation, also referred to as next-generation sequencing (NGS) makes DNA analysis efficient by sequencing high amounts of DNA at once. A widely used NGS technique is Illumina sequencing which involves three steps: preparation, sequencing, and data analysis. Firstly, sequencing libraries are prepared and mixed with unique combinations of forward and reverse Illumina index primers attached to the DNA fragments. Due to unique sets of primers, different samples are sequenced together on the same Illumina flow cell. The index primers contain an adapter and a barcode. The adapter is attached to each side of the DNA fragment and necessary for hybridization to oligonucleotides on the flow cell. Barcodes are unique sequences making it possible to map fragments back to the original sample. During sequencing, DNA fragments are amplified into millions of single-stranded DNA copies in the process called cluster generation. Moreover, during sequencing by synthesis, nucleotides bind complementarily to template DNA. The nucleotides contain a fluorescent tag signaling the type of attached nucleotide and a terminator inhibiting the next nucleotide from adding and allows for the fluorescent signal to be read. The terminator is reversible and is cleaved before the next nucleotide binds to the template. The forward reads are first read and washed away when the process is finished. Finally, the reverse DNA strand is read in the same way as the forward reads. Nucleotides are identified after sequencing is finished, and the sequenced data can further be analyzed by bioinformatic tools (Illumina, 2020). The steps included in Illumina sequencing allows a high amount of DNA to be sequenced and read fast.

1.6.3 Third-generation sequencing

Third-generation sequencing is a technique which is also referred to as long-read sequencing. This sequencing is different from second-generation sequencing because the DNA amplification step is skipped. Instead, the DNA fragment is sequenced as one read. Examples of third-generation sequencing techniques are Oxford Nanopore and Pacific Biosciences. In addition to sequencing single DNA molecules, third-generation sequencing technologies present the opportunity to bring the sequencing directly out in the field making it more accessible for any scientist, anywhere at any time (Heather & Chain, 2016). Third-generation sequencing gives several opportunities in the field and sequence information about long reads.

1.7 PreventADALL cohort

The samples analyzed in this study are retrieved from the ongoing multinational population-based birth cohort study “Preventing Atopic Dermatitis and Allergies in children” (PreventADALL). The PreventADALL cohort is a general population-based study highlighting the increase of allergic and immune-related diseases in the western world. The study aims to determine if early life factors impact the development of allergies and other immune-related diseases like asthma and obesity; likewise, chronic diseases such as diabetes. Recently, a study showed that skin emollients and introduced complementary feeding did not have any effect on the development of atopic dermatitis by age 12 months (Skjerven et al., 2020). To examine early exposures and life factors data was collected data from 2397 mother-child pairs. The women were recruited between December 2014 to October 2016, and infants were enrolled from April 2015-2017. Data included detailed questioners and biological samplings such as skin and stool samples. The first biological samples were taken from 18 weeks pregnant mothers, and from the infants aged 0 (meconium), 4, 6, 12, 24, and 36 months. Follow-up studies are planned from children at year 4 (Lødrup Carlsen et al., 2018). Samples at different age categories make it possible to investigate the establishment and development of the infant gut microbiota.

1.8 Aim of thesis

The origin of the first colonizers of the human infant gut is still not completely understood. This is also true for the species of *Bacteroides,* which are identified as important early colonizers in infants' gut (Bäckhed et al., 2015; Ravi et al., 2018; Yassour et al., 2018). Similarly, mothers and their infants often share the same species/strains of *Bacteroides.* *Bacteroides* species in the gut of vaginally-delivered infants can be directly linked to their mother's microbiota (Stewart et al., 2018; Eck et al., 2020). Thereby, it is hypothesized that *Bacteroides* species are directly transmitted from mother to child during vaginally delivery.

Therefore, the main aim of this thesis was to determine the direct mother-child association based on the presence of *Bacteroides* species.

To achieve the main goal, we had the following subgoals:

- Examine the association of *Bacteroides* sequence variants in mothers and their children
- Evaluate the origin of bacteria that potentially can colonize the infant's gut

2.0 Materials and Methods

The overview of this study experiments is illustrated in Figure 2.1.

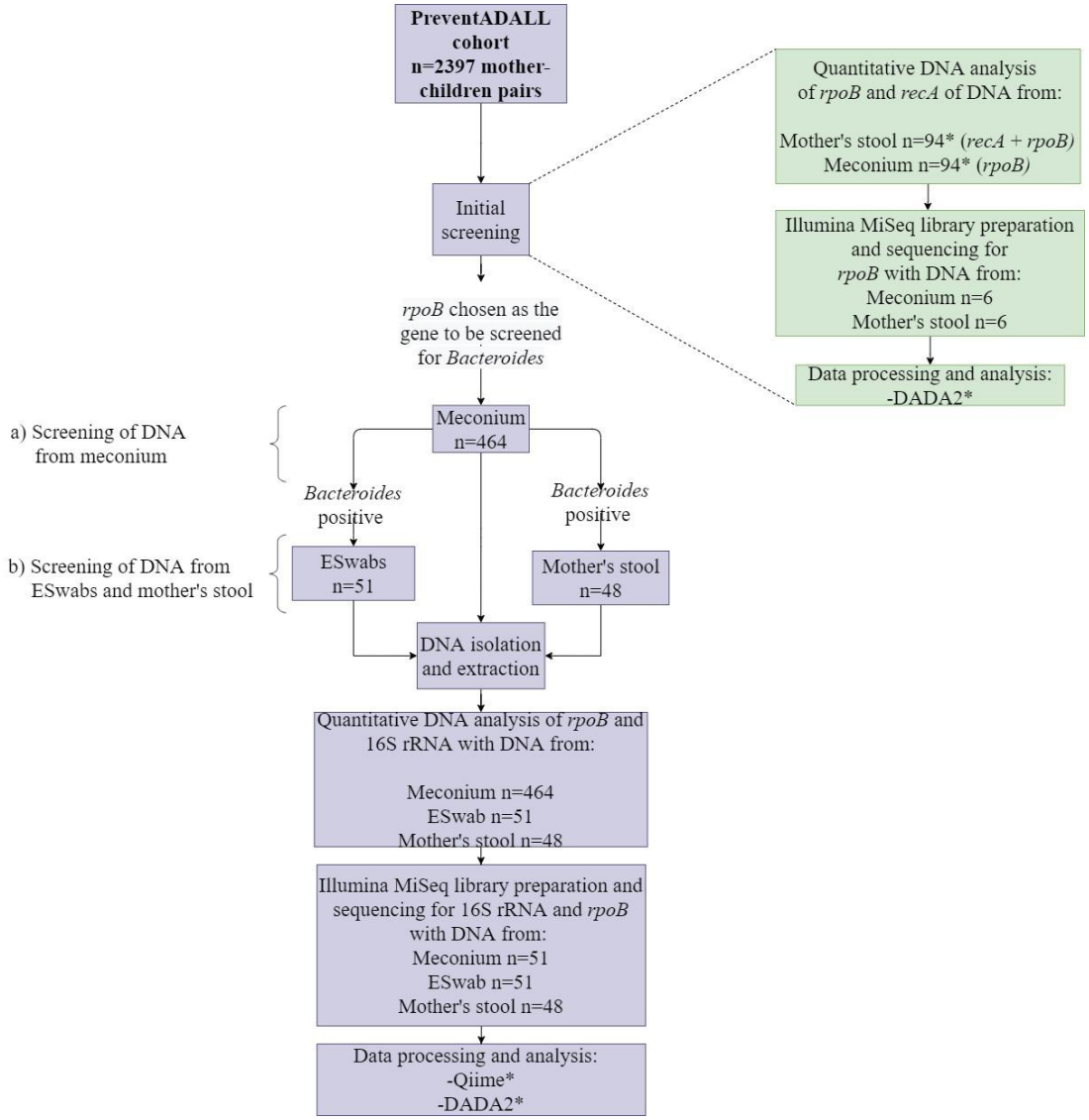


Figure 2.1. Workflow overview of this master thesis. Samples were collected by the PreventADALL cohort. An initial screening was performed to establish *Bacteroides* screening strategies which resulted in 6 mother-children pairs. However, the pairs lacked information about the mode of delivery, thereby sequencing results for these pairs are not presented. In addition, 10 meconium samples and 10 ESwabs were DNA extracted and analyzed qualitatively during the initial screening. However, these samples were excluded from the study due to a lack of information about stool samples from the infants and their mothers. Further in the study, 2 screenings were undertaken. After understanding the presence of *Bacteroides* in the meconium samples (2a), individual ESwabs from *Bacteroides*-positive infants and the respective mother's stool samples were selected and screened (2b).

*DNA extraction was performed by Morten Nilsen, he also helped to conduct the bioinformatic analysis and performed Chi-square and Spearman correlation tests in RStudio (RStudio, Inc., USA).

2.1 Sample description

The samples used in this project were retrieved from the PreventADALL cohort (see the introduction for details); they included meconium, elution swabs (ESwabs) from newborn's skin, and mother's stool (Figure 2.1). Mother's fecal samples were collected from 18-week pregnant women. Meconium from the newborns was collected immediately after passed by the infant. To avoid DNA degradation, the fecal samples were stored in DNA stabilizing buffer (1:10) (Nordic Biolabs, Sweden); these samples were used for DNA analysis. Additionally, fecal samples were stored without any buffer for cultivation studies. Skin samples (ESwabs) from the elbow hook of newborns within 10 minutes after birth were stored in 1 mL Amies transport media. Stool and skin samples were collected and delivered to The Norwegian University of Life Sciences and stored at -80°C until the samples were used for DNA analysis.

2.2 DNA isolation and purification

2.2.1 Initial handling and lysis of bacterial cells

The frozen stool and skin samples were thawed on ice and homogenized by using a vortex machine. Next, 1.2 ml of the stool was transferred from the stool sample tube to a 1.5 ml Eppendorf tube. The fecal samples were pulse centrifuged up to 12 000 rpm to remove the biggest particles. For the skin samples, the initial handling and further DNA extraction were carried out employing the protocols described in Rehbinder et al. (2018). From the ESWABs, 1 mL of the liquid solution was transferred to a 1.5 ml Eppendorf tube and further centrifuged at 13 000 rpm for 15 minutes to harvest the bacterial cells. The cell pellet was resuspended and homogenized in 400 µl S.T.A.R. (Stool Transport and Recovery) buffer (Roche Molecular Systems, USA) (Rehbinder et al., 2018). S.T.A.R. buffer has important properties such as the ability to inactivate infectious organisms, minimize degradation of nucleic acids, and increase the binding of nucleic acids to magnetic particles (Espy et al., 2006). The initial handling was necessary to prepare the samples for bacterial cell lysis.

The bacterial cells were further lysed mechanically. In this study, the bacterial cells were lysed in FastPrep tubes containing acid-washed beads in the size range (0.2 g) <106 µm, (0.2 g) 425-600 µm and (2 beads) 2.5-3.5 mm (Sigma-Aldrich, USA). Different sizes of glass beads were used to ensure the lysis of most of the bacterial cells. FastPrep 96 (MP

Biomedicals, USA) was used for mechanical cell lysis; tubes containing bacterial cells and beads were processed twice at 1800 rpm for 40 sec by the FastPrep robot. Lysed bacterial cells were further centrifuged at 13000 rpm for 5 min at 4˚C. Contrarily, the skin samples were centrifuged at 25˚C as the S.T.A.R. buffer solidifies at low temperatures. The centrifugation step prepared the DNA for further treatment, by clearing beads and excess cell materials from the supernatant.

After centrifugation, 50 µl supernatant from each sample was transferred to a KingFisher 96 well plate. For chemical lysis, 50 µl lysis buffer was added to wells containing supernatant. In addition, 5 µl proteinase was added to the samples to protect DNA from being degraded by cellular proteins. The lysis and proteinase treatment were incubated for 10 minutes at 55ºC in the KingFisher Flex robot (Thermo Fischer Scientific, USA) by the procedure "ProteinaseLGCmini". Further, genomic DNA was extracted.

2.2.2 Genomic DNA extraction

DNA extraction was performed both manually and automatically. The KingFisher Flex robot (Thermo Fischer Scientific, USA) was used for automatic extraction. DNA was extracted employing the MagMidi LGC Kit (LGC Biosearch Technologies, UK). Magnetic particles were used in combination with salt to extract DNA. The positively charged salt made a cation bridge with the magnetic particles which absorbed negatively charged DNA to the particle surface (Tian et al., 2000). While DNA was bonded to the particles, impurities were washed away by alcohol and buffers with salts. Further, nuclease-free water (VWR, USA) was used to elute DNA. DNA was stored at -20˚C until further use.

2.2.3 DNA purification

All PCR products were purified manually following the AMPure- clean up protocol. Sera-Mag Speed beads (0.1%) (Sigma-Aldrich, USA) were used to bind and isolate PCR products. The amplicon size was considered to decide the Sera-Mag Bead concentration. A higher concentration of Sera-Mag Beads bonded smaller PCR products. The ratio between Sera-Mag Speed Beads and PCR products was 1.2:1. The exception was for the mother's 16S rRNA PCR products where the ratio was set to 1:1. The ratio was adjusted due to a higher content of bacterial DNA from the mother's stool. The amplicons were cleaned twice by freshly made

80% ethanol to ensure the removal of larger DNA fragments and inhibitors. The cleaned PCR products were eluted in nuclease-free water (VWR, USA).

2.3 Polymerase chain reaction

Both qualitative and quantitative PCR were employed to reveal the *Bacteroides* species that are common in mothers and infants. All primers used in PCR are listed in table 2.1.

Table 2.1. Overview of the primers used in this study. For all the primers used, F stands for forward primer and R is the reverse primer.

Target area	Primer name F/R	Amplicon length (bp)	Primer sequence F/R (5'-3')	Annealing temperature	Reference
V3-V4 region of the 16S rRNA gene	PRK 341F PRK 806R	466	CCTACGGGRBGCASCAG GGACTACYVGGGTATCTAAT	55˚C	(Yu et al., 2005)
***rpoB* (RNA polymerase β-subunit)**	BF BR	358	CACTTGAGCAAYCGTCGTRT CCTTCAGGAGTYTCAATNGG	55˚C	(Ko et al., 2007)
***recA* (recombinase A)**	recAF recAR	669	GAATCITCCGGTAARACIACI CCAIGAICCGCTYTTYTTGAT	50˚C	(Sakamoto & Ohkuma, 2011)
V4 region of the eukaryotic 18S rDNA	3NDF V4_Euk_R2	450	GGCAAGTCTGGTGCCAG ACGGTATCTRATCRTCTTCG	59˚C	(Cavalier-Smith et al., 2009) (Bråte et al., 2010)

2.3.1 Qualitative PCR

The reaction mix contained 1x HotFirePool Blend Master Mix Ready to Load (Solis BioDyne, Estonia) and primer sets of 0.2 µM concentration (Table 2.1). Template DNA was added in a volume of 2 µl. Amplification was achieved by the thermal cycler (Applied Biosystems, USA). The PCR program used for amplification for one of the housekeeping genes, *rpoB*, 16S rRNA and 18S rDNA were 95˚C for 15 min followed by 30-40 cycles each 30 seconds denaturation at 95˚C and annealing from 55-59˚C (Table 2.1) and elongation at 72˚C for 45 seconds. The final steps were 72˚C for seven minutes and 10˚C for storage. The number of PCR cycles for the amplification of *rpoB* was set to 40 for all samples. For 16S

rRNA, the number of cycles was 30 for DNA from meconium and ESwabs, and 25 for the mother's fecal DNA samples. The 18S rDNA gene was amplified by 30 cycles. Another *Bacteroides* housekeeping gene, *recA*, was amplified for 3 minutes at 94˚C, followed by 40 cycles of 94˚C for 30 seconds, 50˚C for 30 seconds and 72˚C for one minute (Sakamoto & Ohkuma, 2011).

2.3.2 Quantitative PCR

The reaction mix contained 1x HOT FIREPol EvaGreen qPCR supermix (Solis BioDyne, Estonia), primer sets of 0.2 µM, and 2 µl template DNA with a final volume of 20 µl. PCR programs for amplification of 16S rRNA and *rpoB* were 95˚C for 15 min, followed by 40 cycles of 95˚C for 30 seconds, 55˚C for 30 seconds and 72˚C. The fluorescence was measured by LightCycler480 II (BioRad, USA) which also included a High Resolution Melting (HRM) analysis.

2.3.3 Droplet digital PCR

The master mix for ddPCR contained EvaGreen QX200^{TM} ddPCR Supermix (BioRad, USA) and primers with a final concentration of 1 and 0.2 µM, respectively. Droplets were generated before PCR amplification. By mixing the sample with EvaGreen generation oil QX200^{TM} (BioRad, USA), a water-oil emulsion was made and further distributed into thousands of droplets by the droplet generator QX200^{TM} from BioRad (USA). The ddPCR plate containing the water-oil emulsion was sealed in the plate sealing instrument (BioRad, USA) by Foil Heat Seals 180˚C for 5 seconds. Droplets that contained the desired DNA fragments were amplified by the thermal cycler (Applied Biosystems, USA). Target DNA was amplified by 95˚C for 5 minutes, followed by 40 cycles of 95˚C and 55-59˚C for both 30 seconds, further at 72˚C for 45 seconds, and final steps at 4˚C and 90˚C for both 5 minutes and 4˚C for storage. Droplets were assigned as positive or negative by the droplet reader QX200^{TM} from BioRad (USA). The droplet reader measured the fluorescent signal from droplets that contained amplified products. Further, DNA concentration was calculated based on the number of positive and negative droplets (Hindson et al., 2011).

2.3.4 Index PCR for Illumina sequencing

Index primers made it possible to map all sequences back to the original sample. Each PCR product was attached by a unique primer combination. To attach *rpoB*-fragments, 16 forward and 12 reverse index primers (Appendix A, Table A1) were designed. In addition, 16 forward and 15 reverse index primers (Appendix A, Table A2) were used to attach to the 16S rRNA fragments. The reaction mix for index PCR contained 1x FIREPol® Master Mix Ready to Load (Solis BioDyne, Estonia) and unique primer sets (0.2 µM) for each sample. To each well, there were added 5 µl extracted DNA from the skin and meconium samples, whereas the volume was adjusted to 2 µl for the mother's samples. Indexed PCR samples were amplified in the thermal cycler by the following program: 5 minutes at 95°C, 10-12 cycles of 95°C in 30 seconds, 55°C for 1 minute, 45 seconds at 72°C, and a final step for 7 minutes at 72°C before storage at 10°C. The PCR cycles were set to 12 for the DNA from meconium and ESwabs, whereas the mother's fecal DNA index PCR was set to 10 cycles. Further, indexed PCR products were normalized and pooled together into one amplicon library according to their DNA concentration.

2.3.5 Illumina sequencing set up

The pooled amplicon library was prepared for sequencing by the lab engineer. Preparations included quantification by q-RT PCR, dilution, and loading to the Illumina Miseq.

2.5 DNA quantity and quality measurements of PCR products

2.5.1 Quantification by Qubit dsDNA High Sensitivity Assay

Qubit measurements were used to quantify DNA concentration. DNA was measured according to the Qubit dsDNA High Sensitivity Assay. Quant-iT dyes were selective for dsDNA and employed in the concentration of 1:200 per µl DNA sample. The Qubit fluorometer estimated DNA concentration.

For qualitative DNA measurements, the Cambrex FLX 800 CSE robot was used. Qubit solution (same concentration as described above) was prepared for all the samples. A standard curve based on fluorescence given from Cambrex was used to estimate the samples' DNA concentration. Fluorescent values included in the standard curve were the highest and lowest, in addition to regular values in between. Further, the samples were normalized according to

the highest DNA concentration. The Biomek 3000 robot (Beckman Coulter, USA) was used for normalization and pooling of DNA samples into one library.

2.5.2 Agarose gel electrophoresis

PCR products were visualized by a 1.5% agarose gel (Invitrogen, Thermo Fischer Scientific) in a 1x tris-acetate EDTA (TAE) buffer. The gel was run at 80V, 400mA for approximately 35 minutes (Bio-Rad, USA). The electric current separated the negatively charged DNA fragments based on their size. A 100 bp ladder (New England BioLabs, USA) was used as a reference. Further, the gel was visualized by Gel Doc XR (Bio-Rad, USA).

2.5.3 Melt point analysis

HRM analysis was added to the qRT-PCR program. HRM made it possible to separate PCR products based on their melting point. Because fluorescence was measured as a function of time, the HRM analysis could successfully separate false positives like primer dimers from the PCR products. The melting point of the different fragments could be observed by taking the 1st derivate of the melting curve. The target sequence typically showed a higher melting temperature (Tm) than the shorter primer-dimers (Kubista et al., 2006). The HRM analysis was performed with the LightCycler480 III by changes in temperature which ranged from 60-95°C. The real-time data and the melt curve analysis obtained from the LightCycler were analyzed with software BioRad CFX Maestro 1.0, version 4.0.2325.0418 (BioRad 2017, USA).

2.6 Data processing of Illumina sequences

2.6.1 Qiime for 16S rRNA sequences

The Qiime pipeline was used to assign taxonomy to 16S rRNA sequences (see the introduction for details about Qiime) which included several steps. The 16S rRNA sequences retrieved from Illumina MiSeq were pre-processed; removed primers, demultiplexed, and filtered by quality. The purpose of demultiplexing was to separate sequences with unique indexes to make it possible to map samples back to the origin. Further, the OTU table was made with a 97% or higher sequence similarity (Rapin et al., 2017). The SILVA database (see

the introduction for details) was used to assign bacterial taxonomy, as the database covers up-to-date data information of the prokaryotic rRNA gene (Quast et al., 2012). Bacterial diversity within and between samples was determined using Qiime. Bar charts were prepared to visualize the bacterial taxonomy composition and multivariate analysis tool Principal Coordinates Analysis (PCoA), was employed to understand the β-diversity.

2.6.2 DADA2 for *rpoB* sequences

The DADA2 pipeline was applied for phylogenetic analysis of the *rpoB* sequences retrieved from Illumina sequencing. The DADA2 pipeline was run in R Studio version 3.6.1. Background for DADA2 is described in the introduction. The input to DADA2 was demultiplexed (described above) FASTQ files of *rpoB* sequences without barcodes. The pipeline itself served several functions: filtered sequence by quality, dereplication which sorted unique sequences and their abundance, denoising which kept sequence variants that were not just random noise, removed chimeras, and merged forward and reverse reads. The output from DADA2 was amplicon sequence variant profiles of *rpoB* sequences. The ASV profile provided information about how many times the exact ASV was observed in the different samples (Callahan et al., 2015). ASVs retrieved from the DADA2 pipeline were run through the Kraken database (see the introduction for details) to assign bacterial taxonomy to the different sequence variants.

2.7 *Bacteroides* cultivation strategies

Several cultivation strategies were adopted to investigate the presence of living cells of *Bacteroides* in meconium. The media and solutions used for the cultivation strategies are described in Appendix B and Table B1. Cultures and plates were incubated anaerobically in chambers with 2.5 L AnaeroGen (Oxoid, Thermo Scientific) at 37°C. Lab strains *B. vulgatus* (no. 1447) and *B. thetaiotaomicron* (no. 279) retrieved from DSMZ (Germany) were used as positive controls. Both positive controls were isolated from human stool (DSMZ, Germany). Colonies growing on media and DNA from the enrichment culture were prepared for Sanger sequencing of both the *rpoB* and 16S rRNA gene. Sequences retrieved from Sanger were identified by Nucleotide BLAST from the NCBI database.

2.7.1 Enrichment culture for growing *Bacteroides*

The positive meconium sample was added to a chopped medium (DSMZ, Germany) with hemin and vitamin K_1 solution (Appendix B, Table B1) in the volumes of 10 µl per ml medium. After 14 days, 200 µl from the chopped meat medium with meconium was prepared for DNA extraction according to the MagMidi LGC kit (Biosearch Technologies, USA). qRT-PCR and preparation for Sanger sequencing were done for both 16S rRNA and *rpoB*.

2.7.2 Plate incubation for isolating *Bacteroides*

Selective media were made to isolate *Bacteroides* species. Media used for cultivation and isolation were brain heart infusion medium (BHIS), anaerobe basal broth (ABB) (Oxoid, Thermo Scientific), and *Bacteroides fragilis* selective medium (BFS) (Appendix B).

2.8 Statistical approaches and data analysis

2.8.1 Handling of 16S rRNA and *rpoB* sequences data

The abundance of processed sequences data of the 16S rRNA and *rpoB* gene were studied in each sample. In addition, the mean and standard error of the mean (SEM) were calculated for the ratio between *rpoB* and other bacteria (16S rRNA), and for bacteria connected to the different registered sample time points of meconium.

2.8.2 Spearman correlation

Spearman correlation is a nonparametric approach that measures a monotone correlation instead of a linear association between two variables (Hauke & Kossowski, 2011). Spearman correlation test was performed to investigate the association of *Bacteroides* in mother and child. To conduct the statistical test, ASVs were merged into their belonging *Bacteroides* species. Further, the content in percentage was calculated for the different *Bacteroides* species within the sample categories. Spearman correlation test was performed with a basis in the percentage within the sample categories. P-values were corrected by the Benjamini Hochberg method with a 95% significance level. Benjamini Hochberg procedure controls the false discovery rate and thereby reduce the false positives (Benjamini & Hochberg, 1995). Also, a rho (ρ)- matrix was made to investigate positive or negative correlation.

2.8.3 Chi-square (X^2) test

The Chi-square test was performed to analyze the *Bacteroides* association between mother and child. Before the Chi-square test, the *rpoB* data with merged *Bacteroides* amplicon variants, and were assigned 1 if they contained any sequences and 0 if no sequences were present. A Chi-square test with Yates correction was conducted for the binarized data with a 95% confidence interval. The Chi-square test assumed continuously distributed data, where Yates correction allowed discrete datasets to be approximated by Chi-square (Yates, 1934).

2.8.4 T-test

Two-tailed t-tests were performed in Microsoft Excel to detect a significant difference in the bacterial content, for the three sample time points registered for meconium. The tests were done with the assumption about unequal variance within the tested groups. T-tests were conducted with a 95% significance level.

Paired t-tests were employed to test the significant difference between *B. vulgatus* and other *Bacteroides* species within the different sample time points. The tests were performed using a 95% significance level. The paired t-test were performed in Microsoft Excel.

3.0 Results

3.1 Establishment of *Bacteroides* screening strategies

An initial screening was essential to optimize MLST primers and approaches and to decide the sample size for the main study (Figure 2.1). Samples were assigned as positive if they contained the *rpoB* gene (the are terms described later). For the initial *rpoB* screening, we employed DNA from 10 meconium samples in DNA stabilizing buffer, which resulted in 10% (n=1) positives. In addition, DNA from 10 ESwabs was screened for *rpoB*. Likewise, DNA from 94 meconium and 94 stool samples from the mother's stool were included in the study. Screening results showed that all mother's fecal DNA was positive for *rpoB* (n=94) but only 6.40% (n=6) of the DNA from meconium was positive. Moreover, a positive meconium sample without any DNA stabilizer was used for cultivation. In addition, the 18S rDNA gene was attempted amplified to estimate the ratio of eukaryotic DNA compared to bacterial DNA. However, because the 18S rDNA was not amplified it was not further studied. DNA from salmon was used as a positive control for 18S rDNA amplification, whereas DNA from the mother's stool was used as the positive control. Results from the initial screening were the basis for the main study and it was expected that around 10% of the meconium samples were positive for *rpoB.*

3.2 Sample size and approaches used in main study

Two screenings were performed for the *rpoB* gene in the main study. In the first screening of *rpoB,* DNA from 464 meconium samples were included (Figure 2.1) to get around 50 positives. The second screening was conducted employing extracted DNA from the infants' ESwabs (n=51) and their corresponding mother's stool (n=48). Likewise, the 16S rRNA gene was quantified for all the positive samples of extracted DNA from meconium, ESwabs, and mother's stool (Figure 2.1). Positive samples were prepared for Illumina sequencing for both the *rpoB* and the 16S rRNA gene. The positive controls in the main study was a positive meconium sample identified in the initial screening. The exception was for the mother's 16S rRNA samples where *E. coli* was used instead because of missing amplified DNA from meconium from first-stage PCR.

3.3 Identification of *rpoB* positive samples

3.3.1 Screening by quantitative PCR

Regarding the quantitative PCR approaches used in this study, ddPCR was used for the initial screening, whereas qRT-PCR was used throughout the whole study. ddPCR was applied to quantify the 18S rDNA, 16S rRNA, and the *rpoB* gene from 10 DNA extracted meconium samples. The dilution series used for the DNA analysis of meconium ranged from 10^{-1} to 10^{-3}. Due to technical issues with the droplet generator, qRT-PCR was used for DNA quantification. qRT-PCR was firstly used in the initial *rpoB* screening of DNA from meconium (n=94) and mother's' samples (n=94). In this study, we attempted to amplify two *Bacteroides'* housekeeping genes, namely *rpoB* (358 bp) and *recA* (669 bp) genes. Screening results from the light cycler of qRT-PCR produced spurious amplification products and melt point graphs for *recA* which is presented in Appendix C, Figure C1a-b. Likewise, the amplification of *recA* resulted in neither Cq values or melting points. The amplification of *rpoB* resulted in Cq values ranging from around 28.00-38.00 for amplified DNA from meconium and 20.00-32.00 for DNA from mothers' stool. DNA from meconium resulted in melting points from 84.00-84.50 for amplified *rpoB,* and 83.00-85.50 for DNA from mothers' stool (Appendix C, Figure C1c-f). The presented Cq values and melt point temperatures are marked in Figure C1c-f (Appendix C). Based on results from the initial screening, *rpoB* was chosen as the housekeeping gene to be screened for *Bacteroides.*

3.4 Quantifying the *rpoB* and 16S rRNA gene

Based on qRT-PCR results from the initial screening, the Cq threshold value was set to 38.00. The threshold value was decided from the positive meconium samples from the initial screening with the Cq values described above. The Cq value of 38.00 was the highest value visible on an agarose gel. Cq threshold values equal to or lower than 38.00 were expected to contain *rpoB* and thereby assigned as positive samples. The threshold value was set above the background signal and below the negative control, the value was set to the same for all samples.

Screening for *rpoB* in the main study gave the following results for the different sample categories. DNA from meconium yielded 11% (n=51, total=464) positive results, where all positive individuals were delivered vaginally. Out of the total 86.90% (n=403) newborns that

were delivered vaginally, 12.70% (n=51) were *rpoB*-positive (based on the analysis of meconium samples). The remaining 13.10% (n=61) infants that were born were delivered by C-section. Results from the second screening showed that all mother's fecal samples (n=48) positive for the *rpoB* gene. Similarly, 23.50% (n=12, total=51) of the skin samples were positive for the *rpoB*-gene. With the given threshold value (Cq=38.00), the average ratio and SEM was calculated for *rpoB*/16S rRNA in DNA from positive mother's stool samples (1.29 $\pm$ 0.01), meconium (1.21 $\pm$ 0.018) and newborn's skin (1.25 $\pm$ 0.04).

3.5 Analysis of 16S rRNA sequencing data

3.5.1 Bacterial composition in stool and skin samples

The Qiime pipeline was employed for the analysis of taxonomic composition. The threshold sequence depth was set before any taxonomic analysis. The threshold value was set to 5000 which defined the number of randomly picked sequences from all sequenced samples for further analysis. Samples that contained more than 5000 sequences were further analyzed; they included 93.75% (n=45, total n=48) of the mother's stool samples, 96.00% (n=49, total n=51) of the meconium samples and 82.35% (n=42, total n=51) of the newborn's skin samples. Further, they were assigned to 1100 OTU's in the SILVA database, employing Qiime. The OTU table contained information about bacterial abundance from kingdom to species level. The bacterial order composition within each sample category is presented in Figure 3.1.

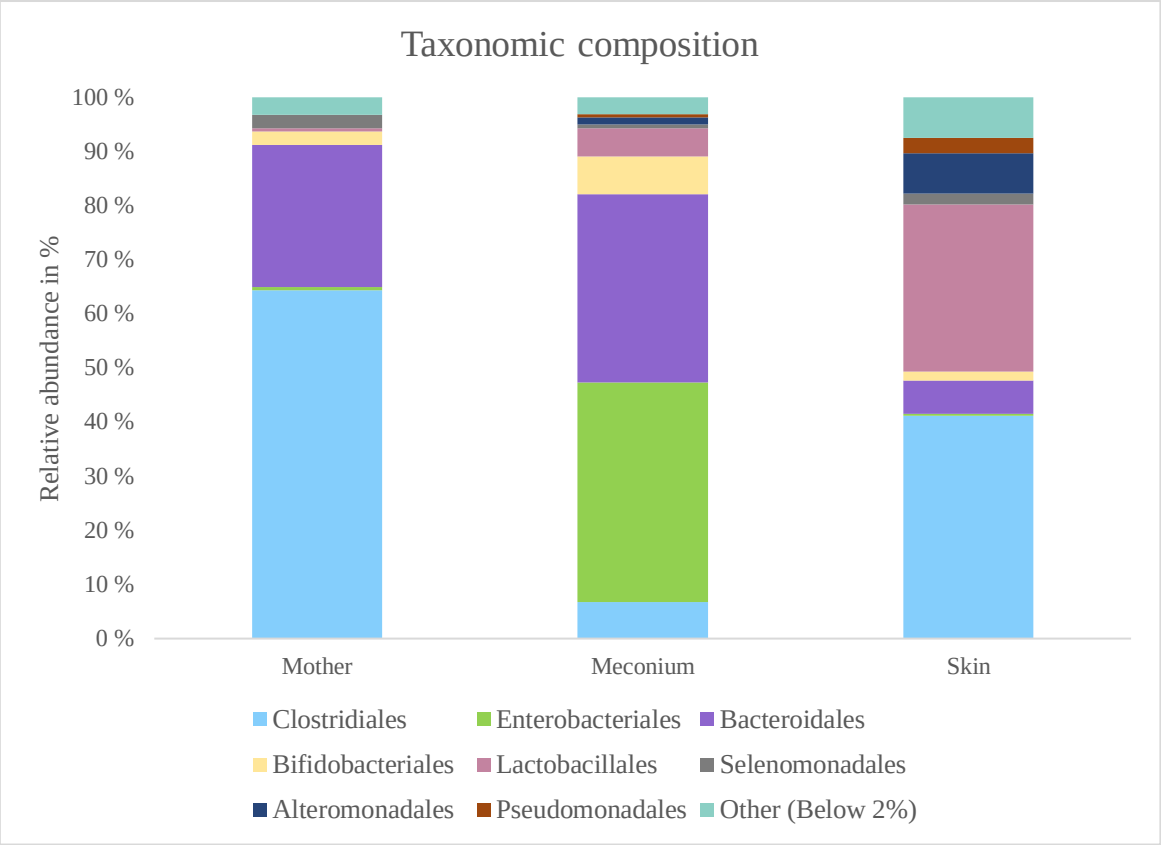


Figure 3.1. Composition of bacterial orders. Based on 16S rRNA analysis bacterial taxonomic composition (order-level) within the sample categories (X-axis) – mother's stool sample, meconium, and newborn's skin – were obtained. The relative abundance of bacteria within each sample category was used as the basis for the content of bacterial orders. The relative abundance is presented on the Y-axis. The colors represent bacterial orders described at the bottom of the figure.

The bacterial composition varied among the sample categories (Figure 3.1). The predominant bacterial orders in mother's stool were Clostridiales (62.76%) and Bacteroidales (25.60%, phylum Bacteroidetes). However, the highest abundance of Bacteroidales (30.78%) was observed in meconium. Another predominant order in meconium was Enterobacteriales (35.91%). The highest amount of Bifidobacteriales (6.10%, phylum Actinobacteria) was detected in meconium. The lowest amount of Bacteroidales (5.17%) was present in the newborn's skin, where Clostridiales (34.81%) followed by Lactobacillales (26.07%) were the predominant bacteria. The species *Bacteroides* accounted for 13.47% of the Bacteroidales in the mother's stool, 26.39% in meconium, and 2.00% on newborn's skin. *Bacteroides* were not further separated into the belonging species by 16S rRNA sequence analysis. In general, the dominance of the orders Clostridiales, Enterobacteriales, Bacteroidales, and Lactobacillales varied in stool and skin.

3.5.2 β-diversity of bacteria in the stool and skin samples

The differences in bacterial diversity in the mother's stool, meconium, and newborn's skin were deciphered by analyzing the β-diversity. PCoA plots were generated with different metrics, namely Jaccard, Bray Curtis, unweighted, and weighted UniFrac (Figure 3.2a-d). The qualitative Binary Jaccard similarity (Figure 3.2a) considers the presence of OTU to display the similarity between samples (Lozupone et al., 2007). As observed in Figure 3.2a, the total variation in the data explained by the PC1-3 of the Binary Jaccard similarity plot was lesser compared to the other plots. Shown in Figure 3.2b is the Bray Curtis dissimilarity plot that takes quantitative differences in OTU composition between the samples into account (Bray & Curtis, 1957). Diversity metrics Binary Jaccard and Bray Curtis measured diversity taking similarity and dissimilarity, respectively into consideration.

Weighted and unweighted UniFrac distances were also employed to understand bacterial diversity. The unweighted UniFrac (Figure 3.2c) is a qualitative measurement that evaluates the presence or abundance of sequences in samples (Lozupone et al., 2007). The weighted UniFrac is a quantitative diversity metric that considers the relative abundance of sequences shared between the samples (Lozupone et al., 2007). The first principal component in the weighted Unifrac plot (Figure 3.2d) explains most of the variation in the samples; 35.99%. The PCoA plot (Figure 3.2d) also captured the highest variation in the data; PC2 and PC3 explains 28.29% and 8.63% of the variation, respectively. The quantitative weighted UniFrac diversity metric was able to explain the highest total variation in the samples, through its principal components.

All the metrics were able to describe the β-diversity of the bacterial communities in the samples (Figure 3.2a-d); to cluster different sample categories. The mother's samples were closely clustered in all the plots. The PC1 (35.99%) in the weighted UniFrac plot (Figure 3.2d) separates the mothers from the newborn's samples. Skin samples show the greatest dispersion compared to the other sample categories (Figure 3.2a-d), an example is marked in Figure 3.2a. Some skin samples have clustered close to the mothers' samples in all the plots (marked in Figure 3.2c). Most of the variation between the infants' samples is explained by PC2 (28.29%) in the weighted UniFrac (Figure 3.2d). The cluster of meconium samples distinct in Figure 3.2a. In the other plots (Figure 3.2b-d) meconium samples are not closely clustered but disperse and lie adjacent to the skin samples. Examples are marked in Figure 3.2b and c, respectively. The variation within and between groups is described differently by the different distance metrics.

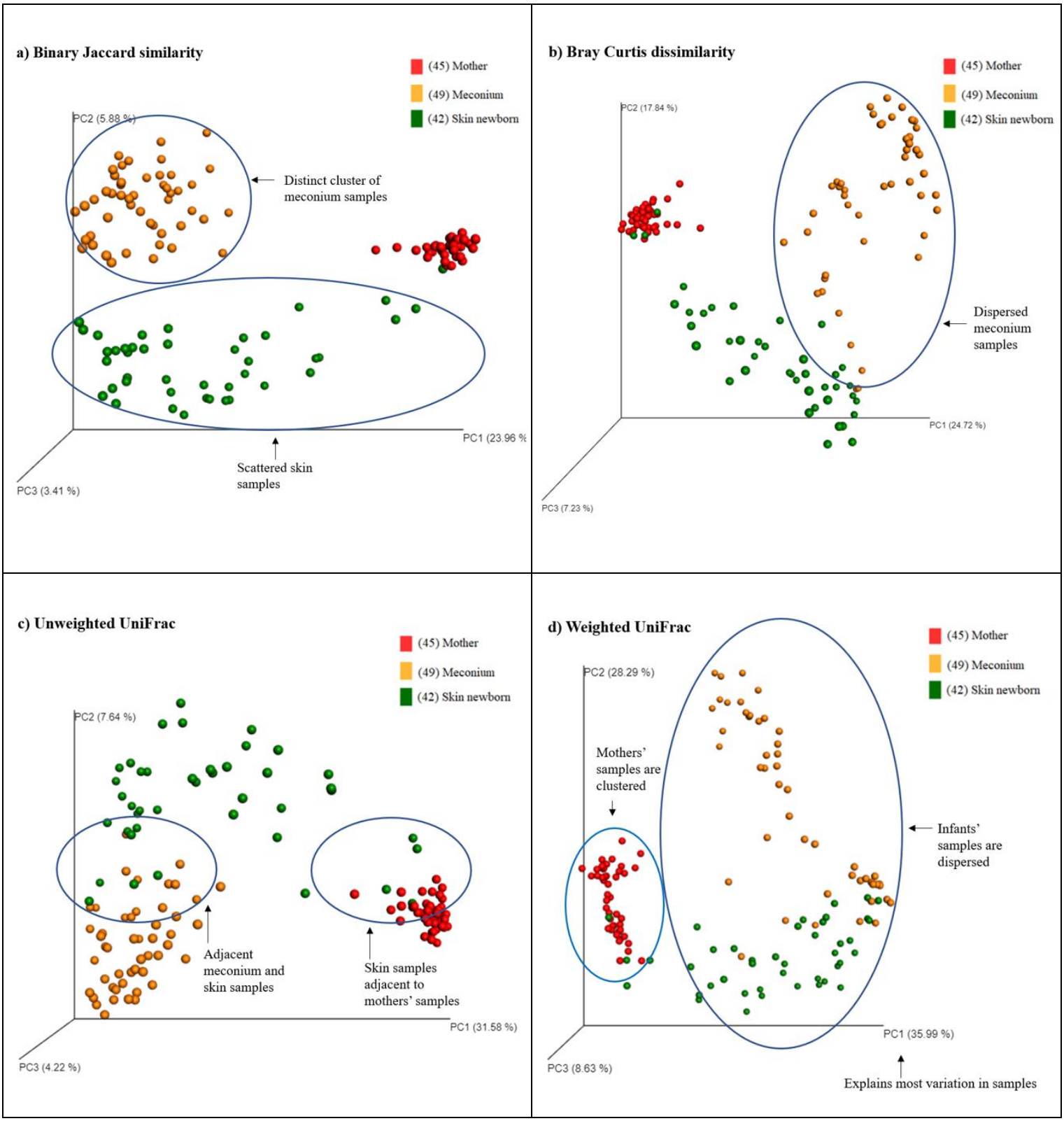


Figure 3.2 a-d. Principal coordinate analysis (PCoA) showing the dissimilarities/similarities of the samples. β-diversity was assessed using the a) Binary Jaccard similarity b) Bray Curtis dissimilarity c) unweighted UniFrac and d) weighted UniFrac indexes. The color codes for the sample categories are: red, mother; yellow, meconium; green, skin newborn. The marked clusters are described above. The figures are retrieved and edited from Qiime.

3.6 *rpoB* sequencing data

3.6.1 Composition of *Bacteroides* species in stool and skin samples

Sequences of the *rpoB* gene retrieved from Illumina sequencing were analyzed using the DADA2 pipeline. See the introduction for details about DADA2. Before forward and reverse sequences were merged and taxonomy assigned, the *rpoB* sequences were quality checked and filtered to retain the sequences with high-quality scores. A quality score profile from the DADA2 pipeline contained information for each base position for both the forward and reverse reads (Callahan et al., 2015). With quality score information, the forward reads were cut at position 200, and position 220 for the reverse read. The final output from the DADA2 pipeline was an amplicon sequence variant profile containing 1390 unique ASVs.

The ASV profile was run through the Kraken database to assign taxonomy. In total 27.27% (n=379) ASVs were identified to different species of *Bacteroides*. The *Bacteroides* species are presented with their average abundance. In total, through this study, 13 *Bacteroides* species were identified (Figure 3.3): *B. vulgatus* (33.60%), *B. thetaiotaomicron* (12.17%), *B. dorei* (7.94%), *B. fragilis* (7.14%), *B. salanitronis* (6.88%), *B. caccae* (6.35%), *B. intestinalis* (6.35%), *B. ovatus* (5.30%), *B. uniformis* (4.76%), *B. caecimuris* (4.76%), *B. cellulosilyticus* (2.91%), *B. helcogenes* (1.60%), and *B. heparinolyticus* (0.26%). The composition of *Bacteroides* species in mother's stool, meconium, and newborn's skin is presented in Figure 3.3.

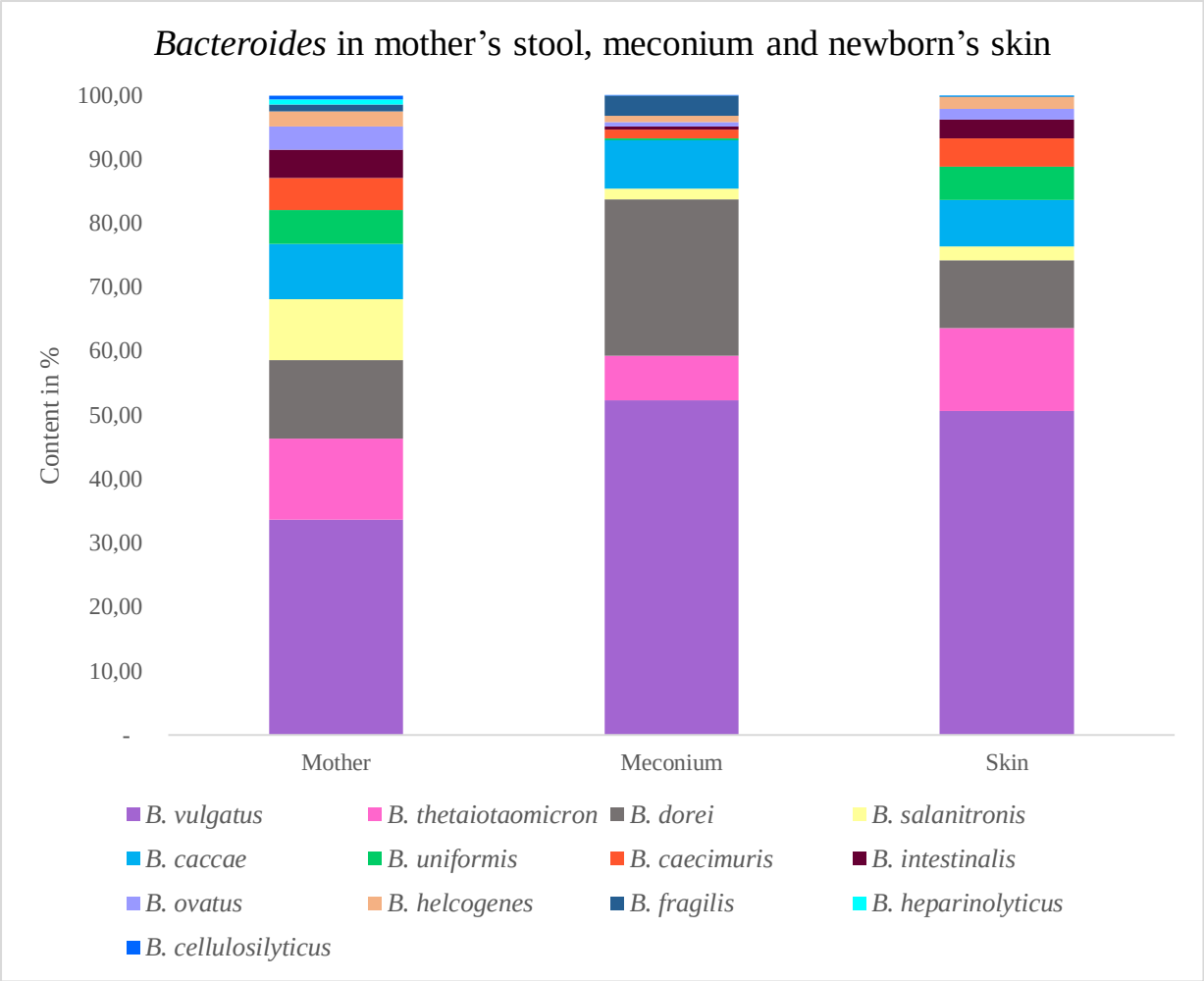


Figure 3.3. Composition of *Bacteroides* species. The bacterial amount is based on average numbers for each *Bacteroides* species within each sample category. The sample categories are presented on the X-axis and the bacterial species content in % on the Y-axis. The colors codes for the different *Bacteroides* species are provided in the figure legend.

B. vulgatus was the most abundant *Bacteroides* species in all sample categories (Figure 3.3). Meconium presented the sample category with the highest abundance of *B. vulgatus* (52.31%) of all *Bacteroides* ASVs discovered in the sample categories. Similarly, *B. vulgatus* was identified in 96% (n=49, of total n=51) of the meconium samples (Appendix D, Table D1). Likewise, *B. vulgatus* was discovered in all mother's stool samples (n=48) (Appendix D, Table D1) with an average content of 33.63% (Figure 3.3). Further, *B. vulgatus* was observed in 88% (n=45, of total n=51) of the skin samples (Appendix D, Table D1) and represented 50.69% of the bacterial content on newborn's skin (Figure 3.3). The second dominating species in mother's stool and newborn's skin was *B. thetaiotaomicron*, whereas *B. dorei* occupied the place in meconium. All *Bacteroides* species were present in the three sample

categories, except for *B. heparinolyticus* which was observed only in the mother's stool and on newborn's skin. The presence of *Bacteroides* and the ASV quantity within each sample category is presented in Table D1 (Appendix D).

3.6.2 Presence of *Bacteroides* species in mother-child pair

In total, 48 complete mother and children pairs were analyzed in this study. Complete pairs mean that information about the *rpoB*-positive samples for meconium, infant's ESwab, and the corresponding mother's stool is available. The quantity of *Bacteroides* species in the complete 48 mother-children pairs were studied (Figure 3.4). The sequences within each sample category were binarized, either presenting the presence or absence of the given *Bacteroides* species. Thereafter the presence of each bacterial species was calculated within each category which is indicated in the figure legend (Figure 3.4). The complete mother-child pair is shown as the category named "Present in all". As shown in Figure 3.4, *B. vulgatus* was the dominating *Bacteroides* within the 48 complete mother-child pairs. Furthermore, *B. dorei* and *B. caccae* were second and third dominating species in the complete pairs. *B. thetaiotaomicron* was the most abundant species in the category of mother and skin. In the category meconium and mother, *B. caccae* was the predominant species. Most *Bacteroides* species found in only one category were in mother's stool. The presence of different *Bacteroides* was varying in the complete mother-children pairs, however, *B. vulgatus* was dominating.

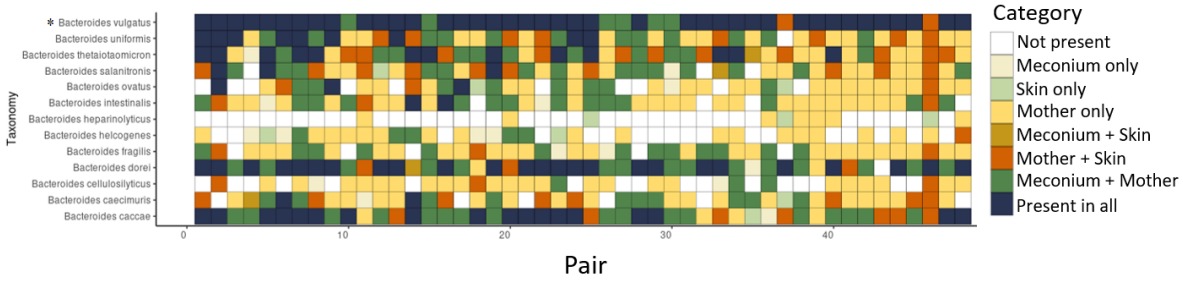


Figure 3.4 Presence of *Bacteroides* species in the 48 complete mother-children pairs. The figure illustrates the presence of the different *Bacteroides* species found in the mother's stool, meconium, and on newborn's skin. The X-axis represents the 48 mother-children complete sets, Y-axis shows the *Bacteroides* taxonomy, and the colors codes for each sample category are given in the legend on the right side. The figure was made by PhD Morten Nilsen.

*Indicates significance ($p<0.05$) overlap in bacteria between mother and child.

Both the Chi-square test and Spearman correlation were performed to test the association between mother and child, based on the presence of *Bacteroides*. As shown in Figure 3.4, *B. vulgatus* caused a significant overlap in the presence of *Bacteroides* in mothers and their children. The association of *Bacteroides* was tested within three groups: mother's stool-meconium, mother's stool-newborn's skin, and meconium-skin. The Chi-square tested if the *Bacteroides* species were significant present in mothers and their infant, and on infant's skin and meconium. Results from Chi-square tests indicated a significant association of *B. vulgatus* in two of the tested groups: mother's stool and meconium (p=2.1e-10), and mother and newborn's skin (p=2.03e-7). Based on the Spearman correlation test, there exists a positively correlation (ρ=0.38) trend (p= 0.053) between *B. vulgatus* in meconium and skin. Other ρ-values are not presented. All p-value results, for both Chi-square tests and Spearman correlation, are presented in Appendix E, Table E1. Results from statistical tests indicated that the species *B. vulgatus* is associated between mothers and their children.

3.6.3 Association between *Bacteroides* ASV in mother and child

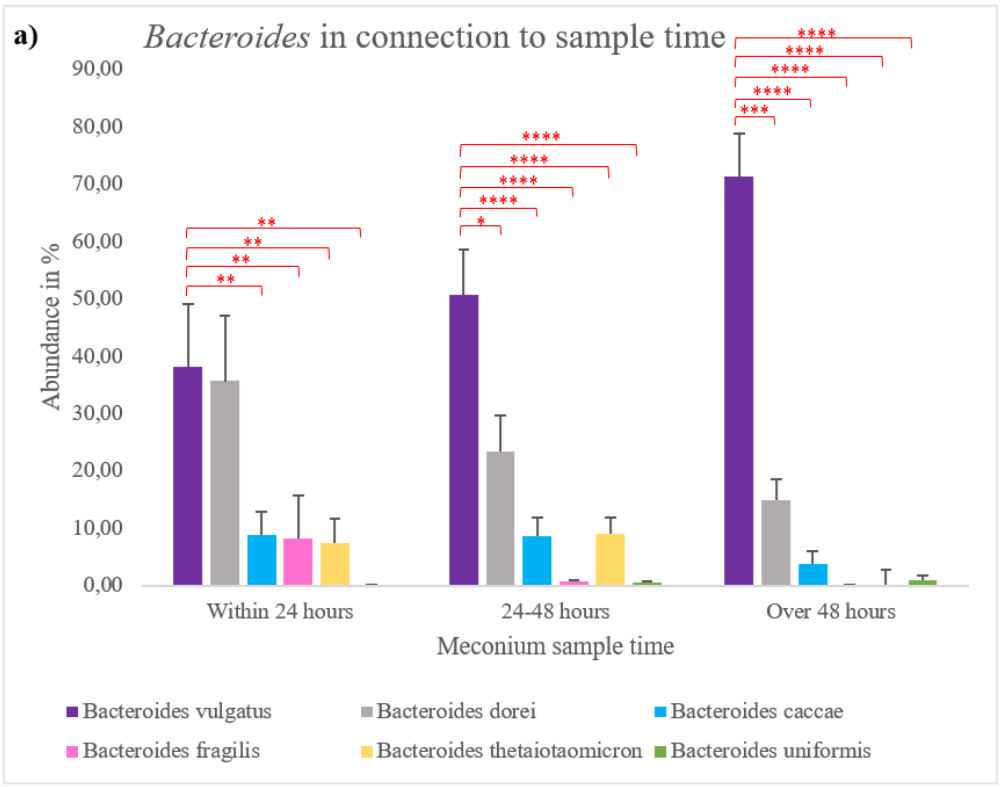
The sequencing of the *rpoB* gene helped in identifying a total of 109 sequence variants of *B. vulgatus*. The sequence variant with the highest richness (named sequence variant 1) was, in general, the most abundant ASV discovered for all *Bacteroides* species. One hundred and eight of the other 109 sequence variants showed low quantity with no clear mother-child overlapping patterns. The quantity for all *B. vulgatus* sequence variants with log-transformed values is presented in Appendix F, Figure F1. The number of observed sequence variants of *B. vulgatus* in mother's stool was lowest (n=16) and increased with meconium (n=38) and newborn's skin (n=70). From Figure F1 (Appendix F) it is observed that the amount of *B. vulgatus* sequence variants is most abundant in the mother's stool (mean 3.23, median 3.83, interquartile range (IQR) 0.70-5.49). The lowest amount is in meconium (mean 2.43, median 2.36, IQR 0.48-5.56), and on the skin (mean 2.48, median 2.37, IQR 0.60-5.12). SV1 presents the maximum value for the mother's samples, and the highest abundant outliers for the infant's samples (Figure F1, Appendix F). In addition, SV1 was present in 98.00% (n=47) of mother's stool samples, and 94.10% (n=48) and 74.50% (n=38) in meconium and skin, respectively. Similarly, SV1 was observed in 70.83% (n=34) of the complete (n=48) mother and children pairs (mother's stool, meconium, and newborn's skin). One *B. vulgatus* sequence variant was observed as the most abundant ASV among all sample categories, likewise, SV1 showed a clear mother-child association pattern.

3.7 Presence of bacteria in connection to meconium collection time

Meconium collection time was registered for most samples. Of all meconium samples (n=464) included in the main study, collection time was registered for 65.73% (n=305) of the samples. The results that are presented hereafter are based on meconium samples with registered sampling time. Most of the meconium samples were registered within 24 hours after birth (71.80%, n=219). Further, 20.66% (n=63) were collected within 48 hours after birth and 7.54% (n=23) had collection time greater than 48 hours. As for the *rpoB*-positive meconium samples, the collection time was registered for 76.47% (n=39) of all samples (n=51). Of the positive registered meconium samples, 30.77 % (n=12) were collected within 24 hours, 51.28% (n=20) between 24-48 hours and 17.95% (n=7) after 48 hours. The content of bacterial orders and *Bacteroides* species were also studied in connection with the meconium collection time (Figure 3.5a-b).

Presented in Figure 3.5a are *Bacteroides* species that were present in over 5% of the meconium samples. *B. vulgatus* was observed to increase with sampling time and was the predominant *Bacteroides* species at all the sampling time points (Figure 3.5a). The percentage of different *Bacteroides* species at different registered sampling time points is presented in Appendix G, Table G1. Two-tailed t-tests were performed to test if there were any differences in bacterial content in meconium according to the different sampling time points. Of the *Bacteroides* species, the content of *B. fragilis* resulted in a significant decrease (p=0.048) from the 2nd to the 3rd sampling time point. Whereas, the content of *B. uniformis* was significantly higher (p=0.015) in meconium collected at the 2nd sample time point compared to those collected within 24 hours. All p-values for *Bacteroides* and the tested groups are presented in Appendix G, Table G2. Because *B. vulgatus* was predominant in the meconium samples, paired t-tests were performed to test if the content of *B. vulgatus* differed from other *Bacteroides* species (Figure 3.5a). *B. vulgatus* resulted in a significant difference compared to the other *Bacteroides* species at all sampling time points, except for *B. dorei* at the first sample time point. All p-values for the paired t-tests are presented in Appendix G, Table G3.

Figure 3.5b shows the presence of Enterobacteriales and Bacteroidales. The bacterial orders are presented that were present in over 10% of the meconium samples. Enterobacteriales was higher when the sampling time of meconium was over 48 hours, Bacteroidales was not stable. Results from a two-tailed t-test did not detect any significant increase or decrease of Bacteroidales or Enterobacteriales.



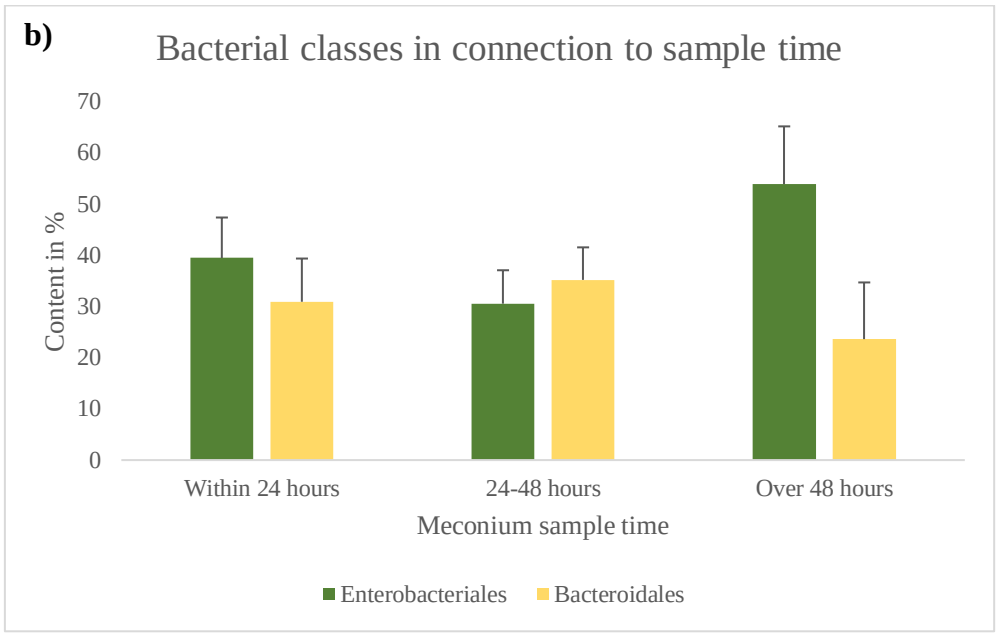


Figure 3.5a-b. The content of bacteria at different meconium collection time points. The numbers are based on average values within each category of time. The standard error of the data is shown in each bar. The X-axis shows the different registered sample times for meconium, the y-axis represents the bacterial content. The figures present the quantity of a) some *Bacteroides* species and b) two bacterial orders (Enterobacteriales and Bacteroidales) in connection to meconium collection time. The colors code for the bacterial specie/order provided in the figure legend.

The asterisks codes for the significant difference of *B. vulgatus* from other *Bacteroides* species; *, $p<0.05$; **, $p<0.01$; ***, $p < 0.001$; ****, $p<0.0001$.

3.8 Cultivating and isolating *Bacteroides*

3.8.1 Enrichment culture for growing *Bacteroides* species

The extracted DNA from the enrichment culture with a *rpoB*-positive meconium sample was prepared for Sanger sequencing. Sequences retrieved from Sanger sequencing were identified, by Nucleotide BLAST search, as *B. vulgatus* (e-value=$2*10^{-114}$) with a query coverage of 96%. The e-value given from a BLAST search is the number of expected matches with a score equal to or greater than the calculated score (Kerfeld & Scott, 2011). DNA from the *rpoB*-positive meconium sample, with *B. vulgatus,* was used in the main study as a positive control.

3.8.2 Selective media for isolating *Bacteroides* species

Selective media were used to isolate *Bacteroides* from the *rpoB*-positive meconium sample. The media for cultivation were prepared considering the traits of the bacterial species. For instance, *Bacteroides* are resistant to several antibiotics (Bacic & Smith, 2008). However, no colonies were identified as *Bacteroides* species. DNA sequences from the growing colonies in BHIS and ABB were identified by BLAST as *E. coli* (e-value=$6*10^{-154}$) and *E. faecalis* (e-value =$2*10^{-143}$). The query coverage was 97% for both bacterial species. In this study, the presented species were the nucleotide matches with the lowest e-value.

BFS was made with to test the antibiotic resistance of *E. faecalis*, *E. coli*, and lab strains of *Bacteroides.* The results are presented in Table H1 (Appendix H). *E. faecalis* was the only bacteria that was able to grow in BHIS with tetracycline. *B. vulgatus*. *E. faecalis* and *E. coli* grew on BFS media with gentamicin. *B. vulgatus* and *E. coli* grew on BHIS with gentamicin and kanamycin. *E. faecalis* and *E. coli* were the only bacteria that grew on ABB media without any antibiotics. *Bacteroides* species from the *rpoB*-positive meconium sample were not isolated in this study, however, species such as *E. faecalis* and *E. coli* were identified from the samples.

4.0 Discussion

4.1 The origin of potential transmission of bacteria from mother to children

4.1.1 The birth canal as the first bacterial exposure site for infants

In this study, all sequencing data retrieved from both meconium and newborns' skin samples originated from vaginally-delivered infants. The order Lactobacillales was abundant on newborn's skin, but also some Lactobacillales were observed in meconium. Infants delivered vaginally are more often colonized (both skin and meconium) by bacteria from the mother's vaginal microbiota, compared to infants delivered by C-section. Similarly, studies show that infants delivered through the birth canal show a higher abundance of especially Lactobacillales (Dominguez-Bello et al., 2010; Nagpal et al., 2016; Ravi et al., 2018). The diversity and richness of the vaginal microbiota decrease during pregnancy. Interestingly, Lactobacillales predominate in the vaginal microbiota this period; mainly with *Lactobacillus* species, but also species from the orders like Clostridiales and Bacteroidales have been observed (Aagaard et al., 2012). The birth canal serves as a natural site where the infants are exposed to microbes for the first time. (Vael & Desager, 2009; Dominguez-Bello et al., 2010; Chu et al., 2017). In this study, the origin of Lactobacillales observed in the samples collected from newborn babies is most likely from the mother's vaginal microbiota.

4.1.2 Mother's gut is the most important bacterial reservoir of bacteria from where transmission from mother to infant occurs

Bacteria in the infant's gut can be directly linked to the mother's vagina and gut. The bacteria belonging to orders Bifidobacteriales, Clostridiales, Bacteroidales, and Enterobacteriales that were present in the mothers' stool were also observed in infants' samples. Of the bacterial orders, Clostridiales that were dominating on newborn's skin had less abundance in meconium. Clostridiales can possibly be vertically transmitted from mother to their child because the bacteria are able to form spores, and the abundance of Clostridiales has been observed to increase by age (Avershina et al., 2016). Likewise, Bacteroidales were abundant in meconium, and these microbes together with Clostridiales were the predominant bacteria in the mother's gut (Eckburg et al., 2005). Previous studies showed that some vertically transmitted bacterial species from mother to infants, were variants of and *E. coli* (Enterobacteriales), *Bacteroides* (Bacteroidales) and *Bifidobacterium* (Bifidobacteriales)

(Bäckhed et al., 2015; Ferretti et al., 2018). The mother's gut serves as an important habitat for bacteria that potentially can colonize the infant's gut. Evidence about the bacterial overlap between mother and child indicates that the bacteria are transmitted from the mother's vagina and gut to the infant during vaginally-delivery.

4.1.3 Indication of living bacterial cells in meconium

Cultivation results point towards the presence of living bacterial cells in meconium. In this study, DNA analysis indicated that species as *E. coli, E. faecalis,* and *B. vulgatus* are present in meconium. Further, there were indications of bacterial growth, because *E. coli* and *E. faecalis* were identified from growing colonies. Likewise, *B. vulgatus* was identified by amplifying *rpoB* which points towards an increased abundance of the bacterial species in meconium. However, *E. coli* and *E. faecalis* were competitive over *Bacteroides* when attempted isolated on selective media because the latter was not able to colonize in the presence of the other two. The dominance of Enterobacteriales in meconium is described in other studies (Nagpal et al., 2016; Chu et al., 2017). Likewise, the indication of transmission of *E. coli* and *Enterococcus* has previously been reported (Jiménez et al., 2008). The *rpoB*-positive meconium which was used for cultivation lacked information about the mother's stool, thereby no conclusion could be drawn about vertically transmission from the cultivation results. Due to a lack of time, cultivation was not performed for *rpoB*-positive meconium samples with available information about their respective mother's stool. There are indications of living bacterial cells in meconium, however, this was not confirmed and should be further investigated.

4.2 *Bacteroides* that are common in mothers and their children

4.2.1 *Bacteroides* species associated between mothers and their children

Some *Bacteroides* species were abundant in the complete mother-children pairs. Especially, *B. vulgatus, B. uniformis, B. dorei, B. thetaiotaomicron,* and *B. caccae* were most often observed in both mother's stool and newborn's samples. *Bacteroides* species that were observed in meconium and on newborn's skin could be directly linked to their mother's stool, especially *B. vulgatus* were present in most of the complete mother-child pairs. The above mentioned *Bacteroides* species, except *B. caccae* and *B. heparinolyticus,* were shown to be

linked to the mother-child association (Bäckhed et al., 2015; Ravi et al., 2018; Yassour et al., 2018). Likewise, *Bacteroides* species are more prevalent in vaginally-delivered infants (Bäckhed et al., 2015; Stewart et al., 2018). Infants delivered vaginally have several *Bacteroides* species in the gut that are associated with their mother.

The abundance of some species of *Bacteroides* in meconium was shown to increase with increased collection time, while those of others decreased as the sampling time increased. The abundance of *B. vulgatus* increased with the collection time, whereas those of other species as *B. dorei* and *B. fragilis* decreased. Likewise, *B. uniformis* showed a significant increase for samples registered within 24 to the 2nd sampling time point. Interestingly, there could be a selective advantage exercised by some bacteria from the mother's non-dominating species. Yassour et al. (2018) observed that bacteria from the mother's non-dominating strains could colonize the infant's gut. The non-dominating strains can utilize glycans from HMOs due to a starch utilizing system. However, the transferred mother's dominating strains miss the starch utilizing systems. Transmission of *B. uniformis* which is a non-dominating strain in the mothers indicates the advantage of this strain to utilize starch and colonize the infant gut (Yassour et al., 2018). However, bacterial traits were not investigated in this study. Similarly, the presence of non-dominating strains was not studied over a longer time period, and hence, the inheritance of the mother's non-dominating species cannot be described. Some *Bacteroides* species show indications of selection over time in the infant's gut, however, the mechanisms for selection needs further investigation.

4.2.2 Transmission of *B. vulgatus* from mothers to their children

B. vulgatus was observed as the dominating species in all sample categories of mothers and infants. Likewise, *B. vulgatus* increased over meconium sample time points and significantly differed from the other *Bacteroides* species, which indicates a selection for the bacterial species in the infant's gut. Similarly, one sequence variant (SV1) of *B. vulgatus* was observed as the predominating variant in all sample categories and in the complete mother-children pairs. The fraction of SV1 was highest in the mother's stool compared to the infant's samples. The results indicate that one species variant of *B. vulgatus* is present and predominant in the human gut. However, the presence of SV1 should be further investigated over a longer timeline to confirm the mother-child overlap of the sequence variant. The non-dominant species variants of *B. vulgatus* could be considered as sequencing errors. For Illumina

sequencing, the most common error is base substitution which in theory will cause pollution in the fluorescent signal and decrease the read quality (Kircher et al., 2009). In this study, substitution could have potentially caused false sequence variants. However, the dominating sequence variant (SV1) of *B. vulgatus* is considered as a true variant that is predominant in the human gut.

Subsequently, the presence of *B. vulgatus* contributed to the significant association between mothers and their children. There are several studies which support the vertical transmission of *B. vulgatus* (Bäckhed et al., 2015; Ravi et al., 2018; Yassour et al., 2018; Eck et al., 2020; Koo et al., 2020). Similarly, *B. vulgatus* is the main species of *Bacteroides* that are transmitted from mother to child (Yassour et al., 2018). Likewise, *B. vulgatus* is highly abundant in newborns' gut and they are known as stable colonizers of the infant's gut at least up to 4 months of age (Ferretti et al., 2018; Ravi et al., 2018). Thereby, *B. vulgatus* is suggested as an important early colonizer in infants' gut microbiota development (Ravi et al., 2018). Similarly, the newly created database, HumGut, indicates that the most prevalent genomes in the human gut are that of *Bacteroides* and among them, *B. vulgatus,* lies at the top (Hiseni et al., 2020). Among the *Bacteroides* species, *B. vulgatus* is dominating in the human gut and vertically transmitted from mothers to their infants.

4.3 Methodological considerations

4.3.1 DNA extraction from samples with low DNA yield

Both meconium and newborn's skin samples contain less bacterial DNA compared to the mother's stool. Some factors may influence the quality of DNA extracted from meconium and skin. Meconium exists in various textures of which some variants could be difficult to dissolve. We fail to extract a representative amount of DNA from poorly dissolved samples. Likewise, meconium could contain PCR inhibitors such as urea, bile acids, and glycolipids (Stinson et al., 2018). Besides, the analysis of samples with low DNA yield could detect more false positives (Kuperman et al., 2020). Therefore, controls are important to reduce such risk for false positives. In this study, the negative controls were evaluated after each step during the sequencing library preparations.

4.3.2 Assigning samples positive for the *rpoB* gene

Most meconium samples were not positive for the *rpoB* gene. However, based on the Cq values we had to exclude some samples from the *rpoB*-gene study. The excluded samples with Cq > 38.00 might also be *rpoB*-positive. Information regarding the number of meconium samples that were originally positive is unknown. Likewise, *rpoB*-positive meconium samples could have been excluded if they were not successfully DNA extracted and further identified as *rpoB*-positive by qPCR. Factors as Cq values, DNA yield, and extraction may influence the number of meconium samples assigned as positive for the *rpoB* gene.

4.3.3 Sequence analysis of the 16S rRNA gene

The sequencing depth was set before sequences were run through the Qiime pipeline to assign taxonomy. The number of analyzed sequences from the samples is modified by the chosen sequence depth. On the one hand, samples expected to contain low bacterial DNA yield, like meconium, require fewer sequences to represent bacterial diversity. On the other hand, samples with higher amounts of bacterial DNA, such as the mother's stool, are dependent on more sequences to represent the complete bacterial diversity. By setting a sequence depth threshold, samples containing less than the threshold are excluded. In this study, by taking the threshold value into consideration, the skin sample size was reduced. The chosen sequence depth influences the diversity profiles retrieved from Qiime.

The 16S rRNA gene is commonly employed in taxonomic studies. 16S rRNA sequencing data provides information about the bacterial diversity within and between samples. However, species such as *Bacteroides* were not separated at the lower taxonomic level. Thereby, 16S rRNA sequencing was not suited to detect bacterial overlap on species or strain level between mother and children due to lower taxonomic resolution which fails to separate closely related species (Ravi et al., 2018). The 16S rRNA sequencing approach is suitable for diversity studies, however, its poor resolution should be considered when investigating bacteria at a lower taxonomic level.

4.3.4 Evaluation of MLST for screening of *Bacteroides*

The non-specific *rpoB* gene was chosen over *recA* for the screening of *Bacteroides.* The *recA* gene was not used further because its amplicon length creates problems for the light cycler, and it will not give credible Cq values and melt point graphs. However, *rpoB* is also present in other species than *Bacteroides.* Although Kraken assigned *rpoB* to other species, the information regarding the *Bacteroides* was obtained to make inferences for this study. Other housekeeping genes of *Bacteroides*, like *gyrB* (approx. 1200 bp) and *dnaJ* (927 bp), have shown to be more informative and have a higher discrimination power compared to both the *rpoB* and 16S rRNA gene (Sakamoto & Ohkuma, 2011). However, longer read lengths require other sequencing platforms such as third-generation sequencing technologies that can handle longer reads (Heather & Chain, 2016). Even though the *rpoB* gene is not specific for only *Bacteroides*, amplification of the *rpoB* gene is suitable for the approaches used in this study.

4.3.5 Strengths and weaknesses of this study

There are several strengths in this study. One strength is the sample size which made it possible to examine the *Bacteroides* association between 48 mother and children pairs. In addition, the positive meconium samples were all delivered vaginally which strengthens the theory about the vertical transfer of bacteria through the birth canal. Furthermore, the *Bacteroides* species were studied at the sequence level which confirms the mother-child association based on the presence of *Bacteroides.*

When it comes to the weaknesses, no conclusions could be drawn about the differences in the bacterial composition in the gut between vaginally and C-section delivered infants. In addition, the mothers' vaginal microbiota was not investigated, which could have provided information about vertical transferred bacteria.

5.0 Conclusions and Future Perspectives

There were several mother-child associated *Bacteroides* species; these findings corroborate with those of previous studies. Our results indicated the significant association of *B. vulgatus* in mothers and their infants. Interestingly, the investigation of the ASV profiles suggested the presence of one predominant sequence variant of *B. vulgatus.* Likewise, the sequence variant of *B. vulgatus* discovered in the infant's gut and on their skin was clearly linked to the respective mother's gut microbiota which indicates transmission during birth. In addition, the infant's gut seems favorable for the colonization of *B. vulgatus*. Future studies should investigate if the predominant sequence variant of *B. vulgatus* stable colonizes the infant gut over a longer time period. Similarly, further cultivation studies could examine the transmission of living cells of the sequence variant of *B. vulgatus.*

More research could also be done by including samples from C-section delivered infants. With sequence information of bacteria from both vaginally and C-section delivered infants, the differences in how the mode of delivery modifies the bacterial composition in the infant could be investigated. In addition, the functional genes of transferred bacteria could be analyzed. *Bacteroides* species could be beneficial for the development of the gut microbiota and immune system. The knowledge about functional genes could provide information about the importance of *Bacteroides* in the infant's gut.

References

Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.-A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I. & Milosavljevic, A. (2012). A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PloS One*, 7 (6).

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J. & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science Translational Medicine*, 6 (237): 237ra65-237ra65.

Aguirre, M., Eck, A., Koenen, M. E., Savelkoul, P. H., Budding, A. E. & Venema, K. (2016). Diet drives quick changes in the metabolic activity and composition of human gut microbiota in a validated in vitro gut model. *Research in Microbiology*, 167 (2): 114-125.

Alonso, V. R. & Guarner, F. (2013). Linking the gut microbiota to human health. *British Journal of Nutrition*, 109 (S2): S21-S26.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T. & Batto, J.-M. (2011). Enterotypes of the human gut microbiome. *Nature*, 473 (7346): 174.

Avershina, E., Lundgård, K., Sekelja, M., Dotterud, C., Storrø, O., Øien, T., Johnsen, R. & Rudi, K. (2016). Transition from infant-to adult-like gut microbiota. *Environmental Microbiology*, 18 (7): 2226-2236.

Bacic, M. K. & Smith, C. J. (2008). Laboratory maintenance and cultivation of *Bacteroides* species. *Current Protocols in Microbiology*, 9 (1): 13C. 1.1-13C. 1.21.

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*, 57 (1): 289-300.

Bray, J. R. & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27 (4): 325-349.

Bråte, J., Logares, R., Berney, C., Ree, D. K., Klaveness, D., Jakobsen, K. S. & Shalchian-Tabrizi, K. (2010). Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The ISME Journal*, 4 (9): 1144.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H. & Zhong, H. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe*, 17 (5): 690-703.

Callahan, B., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. & Holmes, S. P. (2015). DADA2: High resolution sample inference from amplicon data. . *bioRxiv*.

Callahan, B. J., McMurdie, P. J. & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11 (12): 2639.

Capone, K. A., Dowd, S. E., Stamatas, G. N. & Nikolovski, J. (2011). Diversity of the human skin microbiome early in life. *Journal of Investigative Dermatology*, 131 (10): 2026-2032.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K. & Gordon, J. I. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7 (5): 335.

Cavalier-Smith, T., Lewis, R., Chao, E. E., Oates, B. & Bass, D. (2009). Helkesimastix marina n. sp.(Cercozoa: Sainouroidea superfam. n.) a gliding zooflagellate of novel ultrastructure and unusual ciliary behaviour. *Protist*, 160 (3): 452-479.

Chu, D. M., Ma, J., Prince, A. L., Antony, K. M., Seferovic, M. D. & Aagaard, K. M. (2017). Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine*, 23 (3): 314.

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. *Scientific Reports*, 6: 23129.

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., De Vos, W. M., Ehrlich, S. D., Fraser, C. M. & Hattori, M. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, 3 (1): 8-16.

de Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J., Parkhill, J., Charnock-Jones, D. S. & Smith, G. C. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature*: 1.

Del Chierico, F., Vernocchi, P., Petrucca, A., Paci, P., Fuentes, S., Praticò, G., Capuani, G., Masotti, A., Reddel, S. & Russo, A. (2015). Phylogenetic and metabolic tracking of gut microbiota during perinatal development. *PloS One*, 10 (9): e0137347.

den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J. & Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of Lipid Research*, 54 (9): 2325-2340.

Derrien, M., Alvarez, A.-S. & de Vos, W. M. (2019). The gut microbiota in the first decade of life. *Trends in microbiology*.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107 (26): 11971-11975.

Eck, A., Rutten, N. B., Singendonk, M. M., Rijkers, G. T., Savelkoul, P. H., Meijssen, C. B., Crijns, C. E., Oudshoorn, J. H., Budding, A. E. & Vlieger, A. M. (2020). Neonatal microbiota development and the effect of early life antibiotics are determined by two distinct settler types. *PLoS One*, 15 (2): e0228133.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308 (5728): 1635-1638.

Ennamorati, M., Vasudevan, C., Clerkin, K., Halvorsen, S., Verma, S., Ibrahim, S., Prosper, S., Porter, C., Yeliseyev, V. & Kim, M. (2020). Intestinal microbes influence development of thymic lymphocytes in early life. *Proceedings of the National Academy of Sciences*.

Espy, M., Uhl, J., Sloan, L., Buckwalter, S., Jones, M., Vetter, E., Yao, J., Wengenack, N., Rosenblatt, J. & Cockerill, F. (2006). Real-time PCR in clinical microbiology:

applications for routine laboratory testing. *Clinical Microbiology Reviews*, 19 (1): 165-256.

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., Clemente, J. C., Knight, R., Heath, A. C. & Leibel, R. L. (2013). The long-term stability of the human gut microbiota. *Science*, 341 (6141): 1237439.

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D. T., Manara, S. & Zolfo, M. (2018). Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe*, 24 (1): 133-145. e5.

Gosalbes, M., Llop, S., Valles, Y., Moya, A., Ballester, F. & Francino, M. (2013). Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical Experimental Allergy*, 43 (2): 198-211.

Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., Bouffard, G. G., Blakesley, R. W., Murray, P. R. & Green, E. D. (2009). Topographical and temporal diversity of the human skin microbiome. *Science*, 324 (5931): 1190-1192.

Hansen, R., Scott, K. P., Khan, S., Martin, J. C., Berry, S. H., Stevenson, M., Okpapi, A., Munro, M. J. & Hold, G. L. (2015). First-pass meconium samples from healthy term vaginally-delivered neonates: an analysis of the microbiota. *PloS One*, 10 (7): e0133320.

Hauke, J. & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30 (2): 87-93.

Heather, J. M. & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107 (1): 1-8.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L. & Legler, T. C. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83 (22): 8604-8610.

Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T. & Snipen, L. (2020). HumGut: A comprehensive Human Gut prokaryotic genomes collection filtered by metagenome data. *bioRxiv*.

Hornef, M. & Penders, J. (2017). Does a prenatal bacterial microbiota exist? *Mucosal Immunology*, 10 (3): 598.

Illumina. (2020). *Understanding the NGS workflow* Available at: https://emea.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html (accessed: 11. February 2020).

Jiménez, E., Fernández, L., Marín, M. L., Martín, R., Odriozola, J. M., Nueno-Palop, C., Narbad, A., Olivares, M., Xaus, J. & Rodríguez, J. M. (2005). Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Current Microbiology*, 51 (4): 270-274.

Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernández, L. & Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology*, 159 (3): 187-193.

Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to teach "E-value-tionary" concepts. *PLoS biology*, *9*(2).

Kircher, M., Stenzel, U. & Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology*, 10 (8): R83.

Ko, K., Kuwahara, T., Haehwa, L., Yoon, Y. J., Kim, B. J., Lee, K. H., Ohnishi, Y. & Kook, Y. H. (2007). RNA polymerase β-subunit gene (rpoB) sequence analysis for the identification of *Bacteroides* spp. *Clinical Microbiology Infection*, 13 (1): 48-54.

Koo, H., McFarland, B. C., Hakim, J. A., Crossman, D. K., Crowley, M. R., Rodriguez, J. M., Benveniste, E. N. & Morrow, C. D. (2020). An individualized mosaic of maternal microbial strains is transmitted to the infant gut microbial community. *Royal Society Open Science*, 7 (4): 192200.

Korpela, K., Costea, P., Coelho, L. P., Kandels-Lewis, S., Willemsen, G., Boomsma, D. I., Segata, N. & Bork, P. (2018). Selective maternal seeding and environment shape the human gut microbiome. *Genome Research*, 28 (4): 561-568.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B. & Strömbom, L. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27 (2-3): 95-125.

Kuperman, A. A., Zimmerman, A., Hamadia, S., Ziv, O., Gurevich, V., Fichtman, B., Gavert, N., Straussman, R., Rechnitzer, H. & Barzilay, M. (2020). Deep microbial analysis of multiple placentas shows no evidence for a placental microbiome. *BJOG: An International Journal of Obstetrics Gynaecology*, 127 (2): 159-169.

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M. & Raoult, D. (2015). Current and past strategies for bacterial culture in clinical microbiology. *Clinical microbiology reviews*, 28 (1): 208-236.

Lee, S. M., Donaldson, G. P., Mikulski, Z., Boyajian, S., Ley, K. & Mazmanian, S. K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*, 501 (7467): 426-429.

Lim, E. S., Rodriguez, C. & Holtz, L. R. (2018). Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome*, 6 (1): 87.

Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, 73 (5): 1576-1585.

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.

Lødrup Carlsen, K. C., Rehbinder, E. M., Skjerven, H. O., Carlsen, M. H., Fatnes, T. A., Fugelli, P., Granum, B., Haugen, G., Hedlin, G. & Jonassen, C. M. (2018). Preventing Atopic Dermatitis and ALLergies in Children—the Prevent ADALL study. *Allergy*, 73 (10): 2063-2070.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & Caugant, D. A. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, 95 (6): 3140-3145.

Mazmanian, S. K., Round, J. L. & Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*, 453 (7195): 620.

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Palacio, S. D., Montes, S. A. & Mancabelli, L. (2017). The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiology and Molecular Biology Reviews*, 81 (4): e00036-17.

Moles, L., Gómez, M., Moroder, E., Bustos, G., Melgar, A., del Campo, R. & Rodríguez, J. M. (2020). Staphylococcus epidermidis in feedings and feces of preterm neonates. *PloS One*, 15 (2): e0227823.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 51: 263-273.

Nagpal, R., Tsuji, H., Takahashi, T., Kawashima, K., Nagata, S., Nomoto, K. & Yamashiro, Y. (2016). Sensitive quantitative analysis of the meconium bacterial microbiota in healthy term infants born vaginally or by cesarean section. *Frontiers in Microbiology*, 7: 1997.

Ogier, J.-C., Pages, S., Galan, M., Barret, M. & Gaudriault, S. (2019). rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiology*, 19 (1): 171.

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S. & Emslie, K. R. (2011). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical Chemistry*, 84 (2): 1003-1011.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41 (D1): D590-D596.

Rački, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A. & Ravnikar, M. (2014). Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. *Plant Methods*, 10 (1): 42.

Rapin, A., Pattaroni, C., Marsland, B. J. & Harris, N. L. (2017). Microbiota analysis using an illumina MiSeq platform to sequence 16S rRNA genes. *Current Protocols in Mouse Biology*, 7 (2): 100-129.

Ravi, A., Avershina, E., Angell, I. L., Ludvigsen, J., Manohar, P., Padmanaban, S., Nachimuthu, R., Snipen, L. & Rudi, K. (2018). Comparison of reduced metagenome and 16S rRNA gene sequencing for determination of genetic diversity and mother-child overlap of the gut associated microbiota. *Journal of Microbiological Methods*, 149: 44-52.

Rehbinder, E. M., Carlsen, K. C. L., Staff, A. C., Angell, I. L., Landrø, L., Hilde, K., Gaustad, P. & Rudi, K. (2018). Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? *American Journal of Obstetrics Gynecology*, 219 (3): 289. e1-289. e12.

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Avershina, E., Rudi, K., Narbad, A. & Jenmalm, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health Disease*, 26 (1): 26050.

Roelofs, K. G., Coyne, M. J., Gentyala, R. R., Chatzidaki-Livanis, M. & Comstock, L. E. (2016). Bacteroidales secreted antimicrobial proteins target surface molecules necessary for gut colonization and mediate competition in vivo. *MBio*, 7 (4): e01055-16.

Sakamoto, M. & Ohkuma, M. (2011). Identification and classification of the genus *Bacteroides* by multilocus sequence analysis. *Microbiology*, 157 (12): 3388-3397.

Salonen, A., Lahti, L., Salojärvi, J., Holtrop, G., Korpela, K., Duncan, S. H., Date, P., Farquharson, F., Johnstone, A. M. & Lobley, G. E. (2014). Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *The ISME Journal*, 8 (11): 2218-2230.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74 (12): 5463-5467.

Sender, R., Fuchs, S. & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology*, 14 (8): e1002533.

Shumaker, A. M., McEneany, V. L., Coyne, M. J., Silver, P. A. & Comstock, L. E. (2019). Identification of a fifth antibacterial toxin produced by a single *Bacteroides fragilis* strain. *Journal of Bacteriology*, 201 (8): e00577-18.

Skjerven, H. O., Rehbinder, E. M., Vettukattil, R., LeBlanc, M., Granum, B., Haugen, G., Hedlin, G., Landrø, L., Marsland, B. J. & Rudi, K. (2020). Skin emollient and early complementary feeding to prevent infant atopic dermatitis (PreventADALL): a factorial, multicentre, cluster-randomised trial. *The Lancet*.

Sobrino, B., Brión, M. & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 154 (2-3): 181-194.

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., Ross, M. C., Lloyd, R. E., Doddapaneni, H. & Metcalf, G. A. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*, 562 (7728): 583.

Stinson, L. F., Keelan, J. A. & Payne, M. S. (2018). Comparison of meconium DNA extraction methods for use in microbiome studies. *Frontiers in Microbiology*, 9: 270.

Theis, K. R., Romero, R., Winters, A. D., Greenberg, J. M., Gomez-Lopez, N., Alhousseini, A., Bieda, J., Maymon, E., Pacora, P. & Fettweis, J. M. (2019). Does the human placenta delivered at term have a microbiota? Results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics. *American Journal of Obstetrics Gynecology*, 220 (3): 267. e1-267. e39.

Thursby, E. & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, 474 (11): 1823-1836.

Tian, H., Hühmer, A. F. & Landers, J. P. (2000). Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturized format. *Analytical Biochemistry*, 283 (2): 175-191.

Townsend, G. E., Han, W., Schwalm, N. D., Hong, X., Bencivenga-Barry, N. A., Goodman, A. L. & Groisman, E. A. (2020). A Master Regulator of *Bacteroides thetaiotaomicron* Gut Colonization Controls Carbohydrate Utilization and an Alternative Protein Synthesis Factor. *MBio*, 11 (1).

Tremaroli, V. & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, 489 (7415): 242-249.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A. & Affourtit, J. P. (2009). A core gut microbiome in obese and lean twins. *Nature*, 457 (7228): 480-484.

Vael, C. & Desager, K. (2009). The importance of the development of the intestinal microbiota in infancy. *Current Opinion in Pediatrics*, 21 (6): 794-800.

Van Belkum, A., Tassios, P., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N., Fussing, V., Green, J., Feil, E. & Gerner-Smidt, P. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology Infection*, 13: 1-46.

Walker, R. W., Clemente, J. C., Peter, I. & Loos, R. J. (2017). The prenatal gut microbiome: are we colonized with bacteria in utero? *Pediatric Obesity*, 12: 3-17.

Wexler, H. M. J. (2007). *Bacteroides*: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews*, 20 (4): 593-621.

Wood, D. E. & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15 (3): R46.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A. & Knight, R. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334 (6052): 105-108.

Yang, C., Mogno, I., Contijoch, E. J., Borgerding, J. N., Aggarwala, V., Li, Z., Siu, S., Grasset, E. K., Helmus, D. S. & Dubinsky, M. C. (2020). Fecal IgA Levels Are Determined by Strain-Level Differences in *Bacteroides ovatus* and Are Modifiable by Gut Microbiota Manipulation. *Cell Host Microbe*.

Yassour, M., Jason, E., Hogstrom, L. J., Arthur, T. D., Tripathi, S., Siljander, H., Selvenius, J., Oikarinen, S., Hyöty, H. & Virtanen, S. M. (2018). Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe*, 24 (1): 146-154. e4.

Yates, F. (1934). Contingency tables involving small numbers and the χ 2 test. *Supplement to the Journal of the Royal Statistical Society*, 1 (2): 217-235.

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. *Biotechnology Bioengineering*, 89 (6): 670-679.

Zhu, L., Luo, F., Hu, W., Han, Y., Wang, Y., Zheng, H., Guo, X. & Qin, J. (2018). Bacterial communities in the womb during healthy pregnancy. *Frontiers in microbiology*, 9: 2163.

Appendix

Appendix A: Illumina index-primers

Table A1. The index-primers used for library preparation for sequencing *rpoB.*

Primer name	Primer sequence	Target gene	Reference
F1 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F2 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F3 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F4 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F5 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F6 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F7 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F8 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F9 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F10 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F11 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F12 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F13 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F14 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F15 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
F16 rpoB	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagCACTTGAGCAAYCGTCGTRT	rpoB	(Ko et al., 2007)
R1 rpoB	caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R2 rpoB	caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R3 rpoB	caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R4 rpoB	caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)

R5 rpoB	caagcagaagacggcatacgagatCACTCTgtgactggagttcaga cgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R6 rpoB	caagcagaagacggcatacgagatATTGGCgtgactggagttcaga cgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R7 rpoB	caagcagaagacggcatacgagatGATCTGgtgactggagttcaga cgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R8 rpoB	caagcagaagacggcatacgagatTCAAGTgtgactggagttcaga cgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R9 rpoB	caagcagaagacggcatacgagatCTGATCgtgactggagttcaga cgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R10 rpoB	caagcagaagacggcatacgagatAAGCTAgtgactggagttcag acgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R11 rpoB	caagcagaagacggcatacgagatGTAGCCgtgactggagttcag acgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)
R12 rpoB	caagcagaagacggcatacgagatTACAAGgtgactggagttcag acgtgtgctcttccgatctCCTTCAGGAGTYTCAATNGG	rpoB	(Ko et al., 2007)

Table A2. Overview of the index-primers used for library preparation for sequencing 16S rRNA.

Primer name	**Primer sequence**	**Target gene**	**Reference**
F1	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctagtcaaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F2	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctagttccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F3	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctatgtcaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F4	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctccgtccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F5	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgtagagCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F6	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgtccgcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F7	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgtgaaaCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F8	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgtggccCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F9	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgtttcgCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F10	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctcgtacgCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F11	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctgagtggCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F12	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctggtagcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)

F13	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctactgatCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F14	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctatgagcCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F15	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctattcctCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
F16	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc cgatctcaaaagCCTACGGGRBGCASCAG	16S rRNA (V3-V4)	(Yu et al., 2005)
R1	caagcagaagacggcatacgagatCGTGATgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R2	caagcagaagacggcatacgagatACATCGgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R3	caagcagaagacggcatacgagatGCCTAAgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R4	caagcagaagacggcatacgagatTGGTCAgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R5	caagcagaagacggcatacgagatCACTCTgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R6	caagcagaagacggcatacgagatATTGGCgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R7	caagcagaagacggcatacgagatGATCTGgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R8	caagcagaagacggcatacgagatTCAAGTgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R9	caagcagaagacggcatacgagatCTGATCgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R10	caagcagaagacggcatacgagatAAGCTAgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R11	caagcagaagacggcatacgagatGTAGCCgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R12	caagcagaagacggcatacgagatTACAAGgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R13	caagcagaagacggcatacgagatTTGACTgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R14	caagcagaagacggcatacgagatGGAACTgtgactggagttcag acgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)
R15	caagcagaagacggcatacgagatTGACATgtgactggagttcaga cgtgtgctcttccgatctGGACTACYVGGGTATCTAAT	16S rRNA (V3-V4)	(Yu et al., 2005)

Appendix B: Media and solutions for cultivation

The media used in this study are described below. After 4 days of incubation, 100 µl enrichment culture was added with concentration 10^{-5}-10^{-7} to plates with BHIS (Bacic & Smith, 2008) and ABB (Oxoid, Thermo Scientific). BHIS and ABB plates with enrichment culture were incubated anaerobically at 37°C for 3 days. Plates with BFS were incubated anaerobically at 37°C for 2 days with the lab strains *B. vulgatus* (no. 1447) and *B. thetaiotaomicron* (no. 279) retrieved from DSMZ (Germany). In addition, *E. faecalis* and *E. coli* that were isolated from the enrichment culture of meconium were also incubated to culture them. *Bacteroides* are reported to form yellow colonies and blacken the medium (Ho et al., 2017).

Supplemented Brain Heart Infusion Medium (BHIS)

For 1 L BHIS medium:
37.0 g Brain Heart Infusion broth powder (Sigma-Aldrich, USA)
1.0 g cysteine (Sigma-Aldrich)
10 mL hemin solution
1 mL resazurin solution
15 g agar (VWR, USA)
970 mL distilled water.
20 mL $NaHCO_3$ *
10 µg/ml tetracycline * (Sigma-Aldrich)

*Added after autoclaving

B. fragilis selective medium (BFS)

BFS was prepared for 0.5 L:

22.25 g bile aesculin (Sigma-Aldrich)
2.5 g yeast extract (Merck, USA)
0.25 g cysteine hydrochloride (Sigma-Aldrich)
0.5 glucose
5 mL vitamin K solution
5 mL hemin solution
2 mg gentamicin * (Sigma-Aldrich)
7.5 mg kanamycin * (Sigma-Aldrich)

* Added after autoclaving.

Plates were incubated with gentamicin or gentamicin and kanamycin.

Table B1. Solutions used in the cultivation studies. Hemin and vitamin K_1 solution were sterilized with a 0.2 µM filter and stored in the dark at 4°C. Resazurin was stored under the same conditions.

Solution	Content
Hemin	50 mg hemin dissolved in 1 ml 1 M NaOH and diluted with distilled water in a final volume of 10 mL.
Vitamin K_1	0.1 ml soluble K_1 20 ml 95% ethanol
Resazurin (0.1%)	0.1 g resazurin 10 mL distilled water.
$NaHCO_3$ (10%)	10 g solid $NaHCO_3$ 100 mL distilled water Sterilized by autoclaving.

Appendix C: Quantification data and melt curve graphs for *recA* and *rpoB*

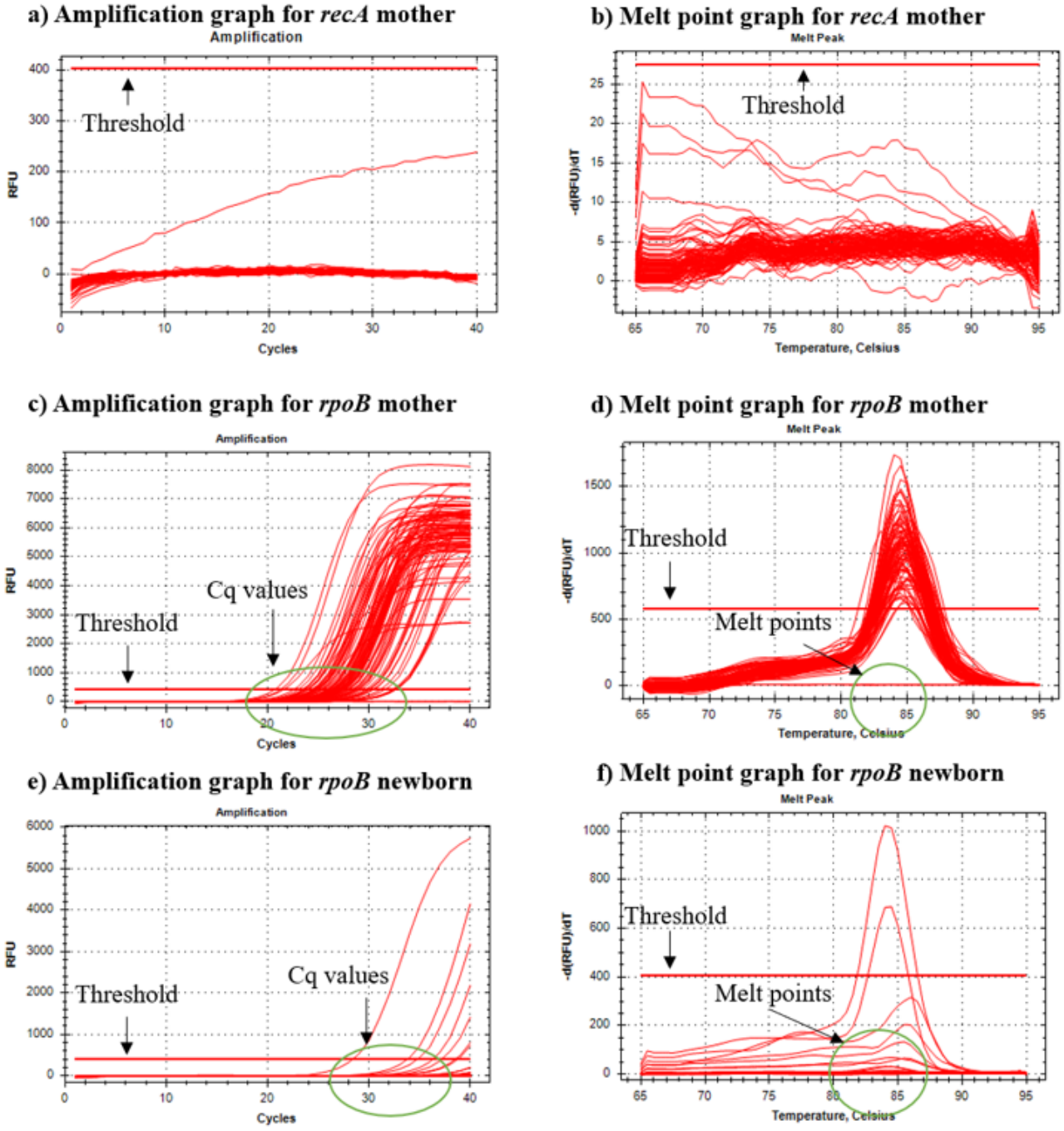


Figure C1a-f. The amplification plots and melt curve graphs for the *rpoB* and *recA* genes. The genes were amplified from DNA from either mother's stool or meconium. The amplification plots (a, c and e) are presented with RFU (relative fluorescence units) on the y-axis and the number of cycles on the x-axis. The Cq values presented are the number of cycles needed to reach the threshold line. The threshold line is set above the background fluorescence signals. The melt point graphs (b, d, and f) are explained by the first derivate of the RFU values on the y-axis and temperatures on the x-axis. Melt points over the threshold line are presented.

Appendix D: Presence of *Bacteroides* in the different sample categories

Table D1. Observed *Bacteroides* species in mother's stool, meconium, and newborn's skin. For each species, the percentage is based on the presence of ASV in the sample categories. The percentage presence of ASV within each sample category is shown on the right side of the observed species amount, the numbers are based on the average amount of ASV within the sample category. The % ASV quantity for each *Bacteroides* species is visualized in Figure 3.3.

	Mother's stool (n=48)		**Meconium** (n=51)		**Newborn's skin** (n=51)	
Specie	% presence in samples	% ASV quantity of total	% presence in samples	% quantity ASV of total	% presence in samples	% quantity ASV of total
B. vulgatus	100	33.63	96.08	52.31	88.24	50.69
B. uniformis	97.92	5.30	45.10	0.35	31.37	5.30
B. caccae	95.83	8.65	74.51	7.58	50.98	7.24
B. thetaiotaomicron	95.83	12.67	56.86	7.02	52.94	12.90
B. dorei	93.75	12.36	88.24	24.48	58.82	10.66
B. intestinalis	87.50	4.38	35.29	0.55	11.76	2.95
B. fragilis	85.42	1.12	29.41	3.20	5.88	0.14
B. salanitronis	85.42	9.50	39.22	1.60	29.41	2.11
B. caecimuris	81.25	4.99	33.33	1.29	23.53	4.36
B. ovatus	70.83	3.70	27.45	0.62	13.73	1.69
B. cellulosilyticus	60.42	0.62	13.73	0.02	7.84	0.10
B. helcogenes	50.00	2.33	19.61	0.99	3.92	1.84
B. heparinolyticus	14.58	0.75	0.00	0.00	5.88	0.00

Appendix E: P-values that indicate the association of *Bacteroides* species in mothers and children

Table E1. Results of Chi-square tests with Yates correction or Spearman's rank correlation corrected by Benjamini-Hochberg.

	Chi-square			Spearman Correlation		
Species	**Mother-meconium**	**Mother-skin**	**Meconium-skin**	**Mother-meconium**	**Mother-skin**	**Meconium-skin**
B. caccae	1.00	0.99	0.50	0.93	0.87	0.97
B. caecimuris	0.92	1.00	1.00	0.96	0.72	0.91
B. cellulosilyticus	1.00	0.25	1.00	0.91	0.43	0.78
B. dorei	0.82	0.70	0.91	0.33	0.40	0.23
B. fragilis	1.00	1.00	0.57	0.86	0.76	0.54
B. helcogenes	1.00	1.00	0.99	0.97	0.97	0.77
B. heparinolyticus [b]	9.23E-07*	1.00	1.35E-09*	NA	0.56	NA
B. intestinalis	1.00	0.86	1.00	0.95	0.68	0.98
B. ovatus	0.46	0.72	0.58	0.55	0.36	0.68
B. salanitronis	1.00	1.00	0.36	0.96	0.96	0.94
B. thetaiotaomicron	0.55	1.00	1.00	0.21	0.98	0.82
B. uniformis	1.00	0.99	0.47	0.43	0.98	0.75
B. vulgatus	2.14E-10*	2.04E-07*	1.00	0.91	0.92	0.053[a]

*Significant p-value (<0.05) by Chi-square test or Spearman's correlation

[a]Indicates correlation between tested groups.

[b]Specie was not observed in meconium (only in mother's stool- and newborn's skin samples)

Appendix F: *B. vulgatus* amplicon sequence variants

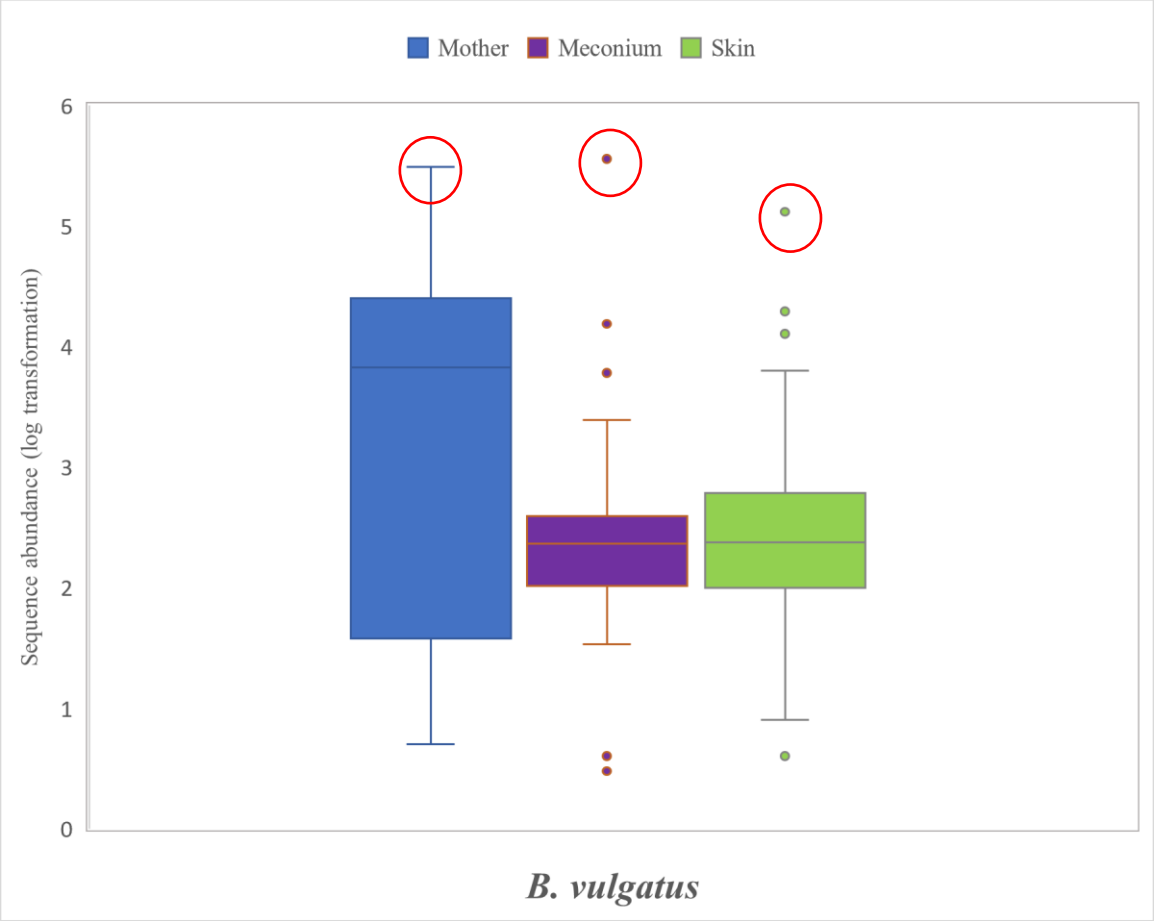


Figure F1. Sequence variants quantity of *B. vulgatus* in the different sample categories. Values presented on the Y-axis are log-transformed (log10) fragment content for all sequence variants of *B. vulgatus* in mother's stool, meconium, and skin samples. The colors indicate the sample category. The boxplot shows the median, the interquartile range, minimum and maximum values of the data. The marked maximum value from the mother's data and the two outliers from the infant's data present the predominant sequence variant of *B. vulgatus* (SV1).

Appendix G: Collection time of meconium and abundance of *Bacteroides*

Table G1. Observed percentage of *Bacteroides* at different meconium collection time points. Values are calculated based on the total average of *Bacteroides* species within one time point.

Specie	%Within 24 hours	%24-48 hours	%Over 48 hours
B. vulgatus	38.07	50.75	71.30
B. dorei	35.61	23.47	14.89
B. caccae	8.90	8.63	3.71
B. fragilis *	8.29	0.64	0.11
B. thetaiotaomicron	7.43	9.05	0.06
B. ovatus	1.41	0.38	0.02
B. intestinalis	0.16	0.18	4.26
B. cellulosilyticus	0.04	0.01	0.07
B. uniformis *	0.04	0.44	0.86
B. salanitronis	0.02	3.01	0.01
B. caecimuris	0.02	2.01	2.37
B. helcogenes	0.01	1.44	2.34

* indicates a significant increase or decrease in abundance with a significant p-value<0.05 which is further described in table G2.

Table G2: P-values that indicate the differences in the *Bacteroides* composition at two collection time points. The p-values are results from two-tailed t-tests.

	Groups (in hours)		
***Bacteroides* species**	**Within 24 and 24-48**	**24-48 and over 48**	**Within 24 and over 48**
B. caccae	0.9478	0.6603	0.7631
B. caecimuris	0.2954	0.4471	0.4588
B. cellulosilyticus	0.3594	0.4532	0.8640
B. dorei	0.5643	0.4949	0.3147
B. fragilis	0.3217	0.0475*****	0.2844
B. helcogenes	0.3387	0.9544	0.2545
B. heparinolyticus [a]	-	-	-
B. intestinalis	0.2825	0.0826	0.0541
B. ovatus	0.8652	0.6727	0.7472
B. salanitronis	0.0588	0.0577	0.2600
B. thetaiotaomicron	0.6896	0.2931	0.6758
B. uniformis	0.0153*****	0.8924	0.2354
B. vulgatus	0.8706	0.1952	0.2225

*****Significant employing a two sample t-test (p<0.05)

[a]Species was not observed in meconium (only in mother's stool and newborn's skin samples)

Table G3. P-values that indicate the differences between *B. vulgatus* and other *Bacteroides* species at the different sample time points. The results are from a paired t-test.

B. vulgatus			
***Bacteroides* species**	**Within 24 hours**	**24-48 hours**	**Over 48 hours**
B. caccae	0.0030**	0.00004****	0.00009****
B. caecimuris	0.0010**	0.00006****	0.00009****
B. cellulosilyticus	0.0009**	0.000004****	0.00009****
B. dorei	0.1964	0.0113*	0.0003***
B. fragilis	0.0061**	0.000005****	0.00009****
B. helcogenes	0.00091**	0.000006****	0.00009****
B. heparinolyticus [a]	-	-	-
B. intestinalis	0.00089**	0.000005****	0.00009****
B. ovatus	0.0009**	0.000005****	0.00009****
B. salanitronis	0.0009**	0.000010****	0.00009****
B. thetaiotaomicron	0.00279**	0.00006****	0.00007****
B. uniformis	0.0009**	0.000006****	0.00009****

*****Significant employing a paired t-test ($p<0.05$)
**$p<0.01$

***$p<0.001$

****$p<0.0001$
[a]Species was not observed in meconium (only in mother's stool and newborn's skin samples)

Appendix H: Selective media for isolating *Bacteroides*

Table H1. Bacteria growth on selective media. The lab strains *B. vulgatus* and *B. thetaiotaomicron* were used as controls in the plate incubation studies. The media were supplemented with different antibiotics. The plus (+) sign shows that the bacterial were able to grow and the minus (-) points to the absence of bacterial growth.

Incubated bacteria	**BHIS with tetracycline**	**BFS with gentamicin**	**BFS with gentamicin and kanamycin**
B. vulgatus	-	+	+
B. thetaiotaomicron	-	-	-
E. faecalis	+	+	-
E. coli	-	+	+

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

Postboks 5003
NO-1432 Ås
Norway