

**MASTEROPPGAVE**  
**Masterstudium i Biomedisin**  
**Mai/Juni 2019**

Undersøkelse av mor-barn assosiasjon av tarmbakterier ved bruk av  
Redusert Metagenomisk Sekvensering (RMS)

Asima Lokmic



**OsloMet – storbyuniversitetet**

**Fakultet for helsevitenskap**  
**Institutt for naturvitenskapelige helsefag**





OsloMet – storbyuniversitet

Fakultet for helsevitenskap, Institutt for naturvitenskapelige helsefag

Masterstudium i Biomedisin

Masteroppgave, 60sp

Mai/Juni 2019

**Undersøkelse av mor-barn assosiasjon av tarmbakterier  
ved bruk av Redusert Metagenomisk Sekvensering (RMS)**

Asima Lokmic

Veiledere: Knut Rudi og Karin C. Lødrup Carlsen

Utført ved Norges Miljø- og Biovitenskapelige Universitet (NMBU)

Institutt for kjemi, bioteknologi og matvitenskap

I samarbeid med Universitetet i Oslo og Oslo universitetssykehus



## **Forord**

Denne oppgaven ble utført ved Institutt for kjemi, bioteknologi og matvitenskap, ved Norges Miljø- og Biovitenskapelige Universitet (NMBU), med professor Knut Rudi som hovedveileder og professor Karin C. Lødrup Carlsen som medveileder.

Jeg vil først og fremst rette en stor takk til min hovedveileder professor Knut Rudi som først og fremst ga meg muligheten til å skrive en ekstern oppgave ved NMBU. Jeg vil takke for alle de gode og konstruktive tilbakemeldingene, og for den hurtige responsen på alle utkast som ble sendt. Din positive innstilling og ditt store faglige engasjement har lett smittet over. Tusen takk for at jeg fikk muligheten til å ha en så dyktig veileder som deg!

Jeg vil rette en stor takk til medveileder Karin C. Lødrup Carlsen som hjalp meg i prosessen med valg av masteroppgave og som prosjektleder for PreventADALL-studien, har gitt meg muligheten til å benytte deres prøver til analyse. Jeg vil takke deg for god oppfølging underveis i oppgaven og for at du var behjelpelig med det som måtte trenge.

En spesiell takk til bioinformatiker Lars Snipen for all behandling av sekvensdata og for alle de gode besvarelsene på alle mine spørsmål med svar fra A til Å, uavhengig hvilken dag og når på dagen spørsmålene ble stilt. Denne oppgaven hadde så å si ikke vært mulig uten deg.

En stor takk til min kjære labingeniør Inga Leena Angell for all den gode opplæringen jeg fikk på laboratoriet og for at du alltid stilte opp, uansett hva. En like stor takk til stipendiat Morten Nilsen som hjalp med oppstarten av analysene og med all oversikt av prøver.

Jeg vil også takke masterstudentene Ida Ormaasen, Zuzanna Gulczynska og Didrik Villard for utrolig godt selskap på laboratoriet og for et godt samarbeid.

Avslutningsvis vil jeg takke min kjære familie og venner for all støtte i en ellers hektisk periode av livet.

Ås, mai 2019

Asima Lokmic

Asima Lokmic

## Sammendrag

En stor utfordring med undersøkelse av mor-barn assosierte tarmbakterier er at det kreves analysemetoder av høy oppløselighet som er i stand til å spore bakterier på arts- og stammenivå, og som samtidig er billige og effektive. Ved 16S rRNA gensekvensering studeres konservative områder av bakteriene som er tilstede i prøven, noe som vanskeliggjør sporing av bakterier på arts- og stammenivå. Helgenom-shotgun sekvensering, derimot, benytter en mye større andel av genomet og er derfor altfor kompleks og dyr. En nyutviklet høyoppløselig analysemetode, Redusert Metagenomisk Sekvensering (RMS), som har kapasitet til sporing av bakterier på arts- og stammenivå, ble derfor benyttet i denne studien. Hovedmålet med denne studien var dermed å undersøke mor-barn assosiasjon av tarmbakterier, samt tarmbakterier assosiert med alder, ved bruk av RMS for sporing av bakteriearter.

Materialet som har blitt undersøkt er genomisk DNA fra avføringsprøver fra 100 mor-barn par knyttet til PreventADALL-studien. Avføringsprøvene ble samlet fra mor i 18 ukers svangerskap og deres barn i ulike alderskategorier (nyfødt/mekonium, tre måneder, seks måneder og tolv måneder). For å evaluere kvaliteten til RMS-metoden i detektering av kjente bakteriearter, ble det gjennomført ytterligere analyser av genomisk DNA fra Mock-samfunn. RMS-biblioteket ble fremstilt til sekvensering på Illumina MiSeq for å sikre tilstrekkelig kvalitet av RMS-biblioteket før ytterligere dypere sekvensering på Illumina HiSeq 3000. Totalt ble 543 prøver analysert.

Resultatene viste at *Proteobacteria* og *Actinobacteria* dominerer tarmmikrobiota signifikant i ung alder ( $p < 0,05$ ), til forskjell fra *Firmicutes* som signifikant dominerer i høy alder ( $p < 0,05$ ). Det ble observert en jevn relativ mengde av *Bacteroides* blant spedbarn i ulike alderskategorier, inkludert nyfødt/mekonium, noe som indikerer at *Bacteroides* stabilt koloniserer spedbarnets tarm. *Bacteroides* var også rikelig tilstede hos mødre, noe som tyder på at bakterien er mor-barn assosiert. Tilstedeværelsen av artene i mekonium indikerer at mødrene er en potensiell kilde til den tidlige koloniseringen av spedbarnets tarm. Imidlertid trengs det ytterligere analyser på stammenivå for å bekrefte at mor-barn assosierte tarmbakterier også er tilknyttet direkte mor-barn overføringer av tarmbakterier.

## Abstract

A major challenge with studies of mother-infant associated gut bacteria is that high resolution analysis methods, that are capable of tracing bacteria at species and strain level, and which are at the same time cheap and effective are required. In 16S rRNA gene sequencing, conserved areas of the bacteria present in the sample are studied, which complicates the detection of bacteria at species and strain level. Whole-genome shotgun sequencing, on the other hand, utilizes a much larger proportion of the genome and is therefore far too complex and expensive. A newly developed high-resolution analytical method, Reduced Metagenome Sequencing (RMS), which has the capacity to detect bacteria at species and strain level, was therefore used in this study. The main aim of this study was thus to investigate mother-infant association of gut bacteria, as well as gut bacteria associated with age, using RMS for tracking bacterial species.

Genomic DNA from stool samples from 100 mother-infant pairs associated with the PreventADALL study has been investigated. The stool samples were collected from mothers in 18th week of pregnancy and their children in different age categories (newborn/meconium, three months, six months and twelve months). In order to evaluate the quality of the RMS method in detection of known bacterial species, further analyses of genomic DNA from Mock-community were carried out. The RMS library was prepared for sequencing on Illumina MiSeq to ensure sufficient quality of the RMS method before further deeper sequencing on Illumina HiSeq 3000. A total of 543 samples were analyzed.

The results showed that *Proteobacteria* and *Actinobacteria* dominate intestinal microbiota significantly at a young age ( $p < 0,05$ ), unlike *Firmicutes* which dominate significantly at high age ( $p < 0,05$ ). A steady relative amount of *Bacteroides* was observed among infants in various age categories, including newborns/meconium, indicating that *Bacteroides* stably colonize the infant's intestine. *Bacteroides* were also abundantly present in mothers, suggesting that the bacterium is mother-infant associated. The presence of the species in meconium indicate that the mothers are a potential source of the early colonization of the infant's intestine. However, additional strain-level analyses are needed to confirm that mother-infant associated gut bacteria are also associated with direct mother-infant transmission of gut bacteria.

## Forkortelser

HMO	Humane melkeoligosakkarider (Human milk oligosaccharides)
rRNA	Ribosomal ribonukleinsyre (Ribosomal ribonucleic acid)
PCR	Polymerase Kjedereaksjon (Polymerase Chain Reaction)
OTU	Operasjonell Taksonomisk Enhet (Operational Taxonomic Unit)
MetaHIT	Metagenomics of the Human Intestinal Tract
HMP	The Human Microbiome Project
DNA	Deoksyribonukleinsyre (Deoxyribonucleic acid)
RMS	Redusert Metagenomisk Sekvensering
AFLP	Amplification Fragment Length Polymorphism
ddRAD	double digest Restriction Site Associated DNA
RefSeq	Reference Sequence
PreventADALL	Preventing Atopic Dermatitis and ALLergies
ddNTP	Dideoksynukleotid trifosfat (Dideoxynucleotide triphosphate)
NGS	Neste Generasjons Sekvensering (Next Generation Sequencing)
PTP	Piko Titer Plate (Picotiter plate)
dNTP	Deoksynukleotid trifosfat (Deoxynucleotide triphosphate)
BLAST	Basic Local Alignment Search Tool
LCA	Lowest Common Ancestor
NCBI	National Center for Biotechnology Information
REK	Regional Etisk Komité
NMBU	Norges Miljø- og Biovitenskapelige Universitet
rpm	revolutions per minute
qPCR	Kvantitativ Polymerase Kjedereaksjon (Quantitative Polymerase Chain Reaction)
ddPCR	Droplet Digital Polymerase Chain Reaction
bp	basepar
PacBio	Pacific Biosciences
SPSS	Statistical Package for the Social Sciences

# Innholdsfortegnelse

<b>Forord</b> .....	<b>I</b>
<b>Sammendrag</b> .....	<b>II</b>
<b>Abstract</b> .....	<b>III</b>
<b>Forkortelser</b> .....	<b>IV</b>
<b>Innholdsfortegnelse</b> .....	<b>V</b>
<b>1. Innledning</b> .....	<b>1</b>
1.1. Den humane mikrobiota .....	1
1.1.1 Tarmmikrobiotas sammensetning og funksjon .....	1
1.1.2 Utviklingen av tarmmikrobiota fra spedbarn til voksen alder .....	2
1.2 Metoder for å studere tarmmikrobiotas sammensetning og funksjon .....	4
1.2.1 Markørgen metagenomikk (16S rRNA gensekvensering) .....	5
1.2.2 Helgenom-shotgun metagenomikk .....	6
1.2.3 Redusert Metagenomisk Sekvensering .....	6
1.3 DNA-sekvensering .....	9
1.3.1 Førstegenerasjons sekvensering .....	9
1.3.2 Andregenerasjons sekvensering .....	9
1.3.3 Tredjegerasjons sekvensering .....	10
1.4 Taksonomisk klassifisering av metagenomiske data .....	10
1.4.1 Kraken2 .....	11
1.5 PreventADALL-studien .....	12
1.6 Mål med studien .....	13
<b>2. Materiale og metode</b> .....	<b>14</b>
2.1 Prøvemateriale .....	14
2.1.1 Mor-barn analyser .....	14
2.1.2 Mock-analyser .....	14
2.2 Forbehandling av prøver .....	17
2.2.1 Mor-barn prøver .....	17
2.2.2 Mock-prøver .....	18
2.3 Klargjøring av RMS-bibliotek til sekvensering .....	19
2.3.1 Restriksjonskutting .....	19
2.3.2 Adapterligering .....	19
2.3.3 Kvalitativ PCR .....	20



2.3.4 Rensing av PCR-produkt .....	22
2.3.5 Index-PCR.....	22
2.3.6 Normalisering av RMS-bibliotek .....	23
2.4 DNA kvantitet- og kvalitetssikring .....	24
2.4.1 Qubit.....	24
2.4.2 Droplet Digital PCR (ddPCR).....	24
2.4.3 Agarose gelelektroforese .....	25
2.5 Dataanalyse.....	25
2.5.1 Prosessering av sekvensdata .....	25
2.5.2 Kraken2 HumGut database .....	26
2.5.3 Statistiske analyser .....	26
<b>3. Resultat.....</b>	<b>28</b>
3.1 Tillaging av RMS-bibliotek.....	28
3.2 Illumina sekvensering av RMS-bibliotek .....	31
3.2.1 Illumina MiSeq sekvensering.....	31
3.2.2 Illumina HiSeq 3000 sekvensering.....	31
3.3 Vurdering av Kraken2-databaser .....	32
3.4 RMS-analyse av Mock-samfunn.....	34
3.5 RMS-analyse av mor-barn prøver .....	36
3.5.1 Seleksjon av bakteriearter for videre analyse.....	36
3.5.2 Den relative mengden av bakteriearter hos spedbarn og mødre.....	38
<b>4. Diskusjon.....</b>	<b>43</b>
4.1 RMS-analyse av mor-barn prøver .....	43
4.1.1 <i>Proteobacteria</i> og <i>Actinobacteria</i> dominerer tarmmikrobiota signifikant i ung alder .....	43
4.1.2 Direkte mor-barn overføring av <i>E. coli</i> ?.....	44
4.1.3 <i>Bifidobacterium longum</i> høyest representert ved seks måneders alder .....	44
4.1.4 Mulig kolonisering av spedbarnstarm før fødsel .....	45
4.1.5 <i>Bacteroides</i> - mor-barn assosiert.....	46
4.1.6 Utfører <i>Bifidobacterium</i> en nisjeokkupasjon?.....	46
4.2 Metodiske betraktninger .....	47
4.2.1 Tillaging av RMS-bibliotek .....	47
4.2.2 Validering av RMS-bibliotek og sekvensering .....	47
4.2.3 Vurdering av Kraken2-databaser.....	48
4.2.4 RMS-analyse av Mock-samfunn .....	49
4.3 Studiedesign.....	50
4.3.1 Styrker ved studien .....	50
4.3.2 Begrensninger ved studien .....	50

<b>5. Konklusjon og videre forskning .....</b>	<b>51</b>
<b>6. Kilder.....</b>	<b>52</b>
<b>Vedlegg.....</b>	
Vedlegg 1. REK-godkjenning .....	
Vedlegg 2. Oversikt over Illumina index-primere .....	
Vedlegg 3. «QScore Heatmap» fra Illumina MiSeq sekvenseringen.....	
Vedlegg 4. «QScore Distribution» fra Illumina MiSeq sekvenseringen .....	
Vedlegg 5. «Per Sequence Quality Scores» fra Illumina HiSeq sekvenseringen .....	
Vedlegg 6. «Mean Quality Scores» fra Illumina HiSeq sekvenseringen .....	
Vedlegg 7. Taksonomisk rangering av utvalgte bakteriearter .....	
Vedlegg 8. Kruskal-Wallis test med <i>Dunn's</i> PostHoc og Bonferroni korrigerte p-verdier .....	



# 1. Innledning

## 1.1. Den humane mikrobiota

Den humane mikrobiota er en samlebetegnelse på alle mikroorganismer, inkludert bakterier, sopp, virus, eukaryoter og arker som koloniserer ulike kroppshabitat, typisk vagina, hud, munn og tarm (1). Tarmen representerer det største habitatet, primært for bakterier, mens virus, eukaryoter, sopp og arker er også tilstede i et mindretall (2). Mikroorganismenes samlede genom utgjør videre sammen det humane mikrobiom.

I løpet av de første årene av livet danner mikroorganismene stabile samfunn på ulike kroppshabitat. Det er bare mikroorganismer som er tilpasset forholdene i og på de ulike kroppshabitatene som kan overleve, noe som er en indikasjon på at miljøet selekterer for mikroorganismer (3). Noen mikroorganismer, typisk kommensale bakterier, regnes som vennlige som ko-eksisterer i mennesket. Ofte er det også vanlig med såkalte mutualistiske forhold mellom kommensale bakterier og mennesket, hvor de er til gjensidig nytte for hverandre. I tillegg kan interaksjonen mellom kommensale bakterier og mennesket bli så tett at forholdet blir symbiotisk. Omvendt kan noen mikroorganismer og deres metabolske produkter være skadelige for mennesket (4).

### 1.1.1 Tarmmikrobiotas sammensetning og funksjon

Et bredt mangfold av mikroorganismer, hovedsakelig bakterier, som omfavner mage-tarm kanalen, vil sammen danne et mikrobielt samfunn kjent som tarmmikrobiota (3). Det er antatt at tarmen er kolonisert av omtrent 160 ulike bakteriearter per individ per avføringsprøve (5), men som tilhører bare noen få bestemte phyla, hovedsakelig *Firmicutes* og *Bacteroidetes*, men også *Proteobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Fusobacteria* og *Cyanobacteria* i mindretall (6). *Bifidobacteria* og *Lactobacilli* ansees som helsefremmende bakterier, mens bakterier som *Staphylococcus* og *Clostridium* regnes som potensielt patogene (7). I tillegg til mikroorganismer omfatter tarmmikrobiota viktige metabolske produkter som essensielle vitaminer, kortkjedede fettsyrer og andre næringsstoffer produsert av mikroorganismer (8, 9). Kortkjedede fettsyrer er blant annet tarmcellenes viktigste energikilde (9).

Tarmmikrobiota spiller en viktig rolle i mange humane fysiologiske prosesser som blant annet fordøyelse (10), immunutvikling- og funksjon (11). I tillegg beskytter tarmmikrobiota mot kolonisering av patogener (11). På den måten antas tarmmikrobiota å spille en viktig rolle i menneskets helse og sykdom (3). For å forstå dens viktige roller er det derfor nødvendig å studere innholdet, mangfoldet og funksjonen av tarmmikrobiota. Ulike mikrobiologiske prosjekter (12, 13) over hele verden har dermed blitt lansert med et felles formål å undersøke

rollen til den humane mikrobiota, hovedsakelig tarmmikrobiota, og dens innvirkning på menneskets helse. En av prosjektene, som er en del av Metagenomics of the Human Intestinal Tract (MetaHIT) prosjektet (12), karakteriserte tarmbakterienes funksjonelle egenskaper ved hjelp av metagenomisk sekvensering (dypsekvensering av komplekse prøver), og avslørte også at mange bakteriearter deles mellom ulike individer. Det andre prosjektet, The Human Microbiome Project (HMP) (13), karakteriserte det humane mikrobiomet mer fullstendig i prøver generert fra flere ulike kroppshabitat, inkludert tarm, munn, vagina og hud. Sammen har prosjektene fungert som et grunnlag for all den etterkomne humane mikrobiologiske forskningen.

Et bredt spekter av faktorer, typisk diett- og miljøforandringer, kan forstyrre tarmmikrobiotas homeostase og på den måten forårsake dysbiose. Imidlertid er den nøyaktige betydningen av dysbiose kontrovers på grunn av mangelen på en nøyaktig definisjon på en ellers «normal» eller «sunn» mikrobiota (3). Dysbiose er vanligvis forbundet med ulike patologiske tilstander hos voksne som inflammatorisk tarmsykdom (14), diabetes type II (15), tykktarm- og endetarmskreft (16). På lignende måte kan dysbiose tidlig i livet disponere for sykdommer som inflammatorisk tarmsykdom (17), Crohns sykdom (18), diabetes type I (19), nekrotiserende enterokolitt (20), allergi (21) og astma (22).

### **1.1.2 Utviklingen av tarmmikrobiota fra spedbarn til voksen alder**

Ved fødsel er tarmene nesten sterile og umiddelbart etter fødsel blir spedbarn raskt kolonisert av et stort mangfold av mikroorganismer, hovedsakelig bakterier, fra mor og miljøet. Passasjen gjennom fødselskanalen vil sammen med amming og konsekvent eksponering av maternell tarm-, vaginal- og hudmikrobiota, representere en mulighet for mikrobiell kolonisering av den nyfødtes mage-tarm kanal (23-27). Imidlertid er det fortsatt uklart når denne første koloniseringen inntreffer. Det har blitt observert at den inntreffer så snart fostermembranen nedbrytes (28). Det som er tydeligere er at en mer utbredt mikrobiell kolonisering begynner postpartum, betegnet etter fødsel (25).

Den første koloniseringen synes å være avgjørende for den videre utviklingen og modningen av spedbarnets tarmmikrobiota og tilhørende immunsystem (29). Den kan sannsynligvis påvirke utviklingen av ulike sykdommer som blant annet fedme, inflammatorisk tarmsykdom og allergi i barndoms- og voksen alder (29).

I starten av livet er tarmmikrobiotas sammensetning svært ustabil, dynamisk og påvirkes av ulike faktorer som fødselsmåte (vaginalt versus keisersnitt) (24), diett (morsmelk versus

morsmelkerstatning) (30), antibiotika (31) og graviditetsalder (32). Dette er også viktige årsaker til at det er vanskelig å definere en «sunn» eller «normal» tarmmikrobiota. Basert på tidligere studier kan det likevel utledes noen trender som karakteriserer hva som er typisk for spedbarn og for voksne (33). Nyfødtes tarmmikrobiota er i starten stort sett preget av en redusert diversitet som hovedsakelig domineres av phyla *Actinobacteria* og *Proteobacteria*. Deretter blir tarmmikrobiota med tiden mer variert og domineres av phyla *Firmicutes* og *Bacteroidetes* (5, 12), som karakteriserer den voksne tarmmikrobiota. Det er fortsatt uklart når en voksenassosiert tarmmikrobiota vil utvikles, men generelt vil dette kunne skje i en alder av tre år (34). På dette stadiet vil tarmmikrobiotas sammensetning være relativt stabil (35), men kan også endres som følge av antibiotika, livsstil og langsiktig endring i kosthold (36).

Ved vaginal fødsel er det svært vanlig at fakultative anaerobe bakterier som *Enterobacteriaceae*, *Enterococcus* og *Streptococcus* koloniserer spedbarnets tarm og på den måten produserer anaerobe miljøer de første dagene av livet slik at strikt anaerobe bakterier som *Bifidobacterium*, *Bacteroides* og *Clostridium* kan trives (37). Imidlertid vil fødsel ved keisersnitt kunne føre til at spedbarnets tarm først blir kolonisert av bakterier avledet fra mors hudmikrobiota, hovedsakelig *Staphylococcus*, *Corynebacterium*, *Propionibacterium* og andre miljøbakterier (24).

*Bifidobacterium* synes å dominere hos ammende spedbarn inntil det første året av livet (38, 39), og har viktige roller, blant annet i utviklingen av spedbarnets immunsystem. Veksten av *Bifidobacterium* stimuleres av komplekse prebiotiske sukkerpolymerer, Humane Melkeoligosakkarider (HMO'er), som er tilstede i morsmelk (40). Som et resultat vil bakteriene kunne trives i tarmen under spedbarnets amming, men mengden av bakterier endres proporsjonalt med mengden av morsmelk tilstede. I overgangen fra morsmelk til fast føde vil dermed mengden av *Bifidobacterium* reduseres (30), mens mengden av typisk *Bacteroides* og *Clostridium* vil kunne øke (41).

Bakterier som overføres fra mor til barn synes å ha en beskyttende rolle som forhindrer tilstrømming av miljøbakterier som kan ha en høyere risiko for å bære patogener og dermed uønskede egenskaper til spedbarnet (27). Noen bakteriearter, typisk medlemmer av *Clostridia* og *Akkermansia muciniphila*, vil sannsynligvis ikke kolonisere spedbarns tarm før etter det første året av livet. Bakterier som rekrutteres på et seinere stadium ansees til å være spesialiserte til forholdene funnet i voksen tarm, noe som gjør at de ikke er i stand til å kolonisere spedbarnets tarm umiddelbart etter fødsel (27). Det kan også være en indikasjon på at ulike bakterier har

utviklet ulike overføringsmekanismer, der noen er avhengig å bli overført fra mor til barn, mens andre er avhengig å bli overført fra miljøet seinere i livet (27).

## **1.2 Metoder for å studere tarmmikrobiotas sammensetning og funksjon**

Å studere tarmmikrobiotas sammensetning og funksjon har vært krevende på grunn av blant annet mangel på optimale analysemetoder. Det siste tiåret har dyrkningsuavhengige metoder som 16S rRNA gensekvensering (42) og helgenom-shotgun sekvensering (26, 43) ført til økt kunnskap om tarmmikrobiotas sammensetning og funksjon, og fortsetter ytterligere å dominere feltet. Imidlertid har ikke 16S rRNA gensekvensering tilstrekkelig oppløsning for sporing av bakterier på arts- og stammenivå (42), mens helgenom-shotgun sekvensering krever omfattende og komplekse analyser som begrenser storskala studier (26, 44). En effektiv og høyoppløselig metode, Redusert Metagenomisk Sekvensering (RMS), har dermed nylig blitt utviklet (45, 46).

Metagenomisk sekvensering omfatter dypsekvensering av komplekse prøver (for eksempel miljøprøver, avføringsprøver) som består av nukleinsyrer fra flere organismer (47). En slik metagenomisk sekvensering har mange fordeler som relativ analysehastighet, evnen til å oppdage organismer som er vanskelige eller umulige å dyrke og det faktum at den krever ingen eller svært lite forkunnskap om organismene som studeres. Data som frembringes fra en metagenomisk sekvensering kan være av ulik kvalitet og kompleksitet avhengig av hva og hvordan det studeres (47).

Ved metagenomisk sekvensering er man som regel interessert i hva som er tilstede i en prøve og i hvilken grad. En grundig profilering av mikroorganismer er blant annet avhengig av at analysemetoden som anvendes har tilstrekkelig dybde/oppløsning. Dette er viktig for å oppdage organismer på potensielt lave nivåer i en prøve og for å oppdage eventuelle forskjeller i og mellom prøver, for eksempel skille to lignende bakterier fra hverandre. Dybden til en analysemetode vil dermed bestemme antall sekvenser som hentes fra en bestemt prøve og er unik for enhver sekvenseringsplattform. En eventuell mangel på tilstrekkelig dybde kan i verste fall føre til en mangel på passende referansegenomer for sammenligning av prøver (47).

### **1.2.1 Markørgen metagenomikk (16S rRNA gensekvensering)**

Markørgen-metagenomikk er en rask og kostnadseffektiv analysemetode for taksonomisk profilering og karakterisering av diversiteten i mikrobiota. Prinsippet til markørgen-metagenomikk baserer seg på en selektiv PCR-amplifisering og etterfølgende sekvensering av konserverte markørgener, typisk det prokaryotiske 16S rRNA markørgenet, som er tilstede i bakterier og arker (48). Etter sekvensering blir 16S rRNA gensekvenser kategorisert i Operasjonelle Taksonomiske Enheter (OTU-er). Kategoriseringen i OTU-er utføres vanligvis ved 97% sekvenslikhet (49). På den måten omfatter hver OTU flere beslektede sekvenser som representerer en taksonomisk enhet, for eksempel en bakterieslekt. Hver OTU blir taksonomisk identifisert via et klassifiseringsverktøy som for eksempel Basic Local Alignment Search Tool (BLAST). En OTU-tabell blir så fremstilt på bakgrunn av de genererte OTU-ene, og kan benyttes til visualisering av fylogenetiske forhold mellom de identifiserte OTU-ene. Det er vanlig at OTU-tabeller blir kvalitetsfiltrert for eventuelle kimære sekvenser som kan oppstå i forbindelse med PCR og sekvensering (49).

Andre markørgener som inkluderer rRNA-gener eller proteinkodende gener, som har en tendens til å være felles for mikrobielle genomer, kan benyttes og avhenger av hva som studeres. I og med at det ved markørgen-metagenomikk fokuseres på ett eller noen få universelle gener, kan det ikke direkte identifisere metabolske eller andre funksjonelle evner hos mikroorganismer som studeres (48).

Til tross for 16S rRNA gensekvenseringens fordeler, har den også noen begrensninger. Sekvensene som dannes er ofte korte og stammer fra konserverte 16S-regioner, noe som gjør det vanskelig å skille nært relaterte bakterier fra hverandre. Til dette kreves det analyser av høyere oppløselighet. Pålitelig taksonomisk klassifisering av nåværende 16S rRNA gensekvenser er imidlertid begrenset til slektsnivå (42). Generelt kan det også oppstå forstyrrelser tilknyttet PCR, typisk en produksjon av kimære sekvenser (et resultat av ufullstendig polymerisering). Kimære sekvenser kan være vanskelige å identifisere og i verste fall kan de føre til at mengden bakterier blir overestimert (48). Dette kan igjen forårsake sekvenseringsfeil (50), som kan ha en negativ innvirkning på vurderingen av OTU-er. Videre er bruken av et enkelt markørgen for vurdering av den mikrobielle diversiteten svært utfordrende, gitt utbredelsen av horisontal genoverføring mellom mikroorganismer og vanskeligheten ved definering av bakteriearter (51, 52).



### **1.2.2 Helgenom-shotgun metagenomikk**

Helgenom-shotgun metagenomikk er en analysemetode som benyttes til en kompleks genomisk proliferering av en miljøprøve. Ved helgenom-shotgun sekvensering blir genomisk DNA tilfeldig fragmentert i mange små biter som sekvenseres. De sekvenserte fragmentene blir så samlet i lengre «contigs», sammenhengende sekvenser, ved bruk av dataprogrammer som finner regioner der fragmenter overlapper (53). På denne måten er helgenom-shotgun sekvensering i stand til å gjennomføre en fullstendig sekvensering av de fleste tilgjengelige genomer i en miljøprøve. Den vil dermed ikke målrette seg kun til et bestemt genetisk locus, posisjon, for amplifikasjon, slik tilfellet for 16S rRNA gensekvensering er. På den måten kan det dannes en biologisk diversitetsprofil, som i tillegg til taksonomi på stammenivå (44), kan knyttes til mikroorganismenes funksjonelle egenskaper. Helgenom-shotgun metagenomikk vil dermed kunne besvare hvilke mikroorganismer som er tilstede i en miljøprøve, deres funksjonelle egenskaper og hvordan samspillet mellom mikroorganismene er i opprettholdelsen av en balansert økologisk nisje (48).

Til tross for fordelene med helgenom-shotgun metagenomikk, er det også rapportert flere begrensninger og utfordringer. Metagenomiske data kan være relativt komplekse gitt den høye oppløsningen til metoden og det store antallet sekvenser generert, noe som kan komplisere de videre bioinformatiske analysene som igjen vil kunne kreve spesialiserte programmer (54). I tilfeller hvor sekvenser overlapper vil det ikke alltid være tydelig at de stammer fra samme eller ulike genom, noe som videre kan utfordre sekvenssammenstillingen (55). I tillegg er analysene svært dyre og ikke like effektive sammenlignet med 16S rRNA gensekvensering (56).

### **1.2.3 Redusert Metagenomisk Sekvensering**

Redusert Metagenomisk Sekvensering (RMS) er en nyutviklet høyoppløselig metode for analyse av mikrobielle genomer (45). Ved RMS blir genomisk DNA først kuttet i fragmenter ved bruk av restriksjonsenzymer som kutter hyppig (MseI, 4 bp) og sjeldent (EcoRI, 6 bp). Deretter følger adapterligering til endene av fragmentene, selektiv PCR-amplifisering og til slutt sekvensering. Det er bare fragmenter som flankeres av begge enzymer som til slutt sekvenseres, noe som utgjør ca. 10% av det totale arvestoffet. På denne måten har RMS, i likhet med helgenom-shotgun sekvensering, kapasitet til taksonomisk profilering av mikrobielle genomer på arts- og stammenivå, i tillegg til karakterisering av mikroorganismenes funksjonelle egenskaper (45).

RMS har blitt videreutviklet fra Amplification Fragment Length Polymorphism (AFLP) metoden, men ved bruk av sistnevnte metode blir de selektivt PCR-amplifiserte fragmentene

fra mikrobielle genom kun analysert på gel, med etterfølgende visualisering ved autoradiografi (57, 58). Foruten om er RMS en alternativ metode til double digest Restriction Site Associated DNA (ddRAD) metoden (59). Det som skiller de mest fra hverandre er enzymkombinasjonen hvor det i sistnevnte metode benyttes en enzymkombinasjon av N1aIII og HpyCH4IV for kutting av genomisk DNA i fragmenter.

**Et mer detaljert prinsipp for RMS-analysemetoden er beskrevet nedenfor og illustrert i figur 1.**

Ved RMS blir genomisk DNA først kuttet i fragmenter ved hjelp av spesifikke restriksjonsenzymmer. Deretter følger en adapterligering til endene av fragmentene for å danne templat for videre PCR-amplifisering. Adapter- og restriksjonskuttesequensen vil dermed sammen fungere som bindesete for primeren benyttet til PCR-amplifiseringen. På den måten blir det en selektiv PCR-amplifisering av et sett av restriksjonsfragmenter på en definert lengde og antall avhengig av den valgte enzymkombinasjonen (45, 58).

Enzymkombinasjonen som benyttes betegnes gjerne som «sjeldne» og «hyppige» kuttere av genomisk DNA. En hyppig kutter vil kunne danne små DNA-fragmenter der amplifiseringen sannsynligvis vil være vellykket og størrelsen optimal for sekvensering. En sjelden kutter, derimot, vil redusere antallet fragmenter da det er kun de fragmentene flankert av begge enzymmer som amplifiseres og sekvenseres.

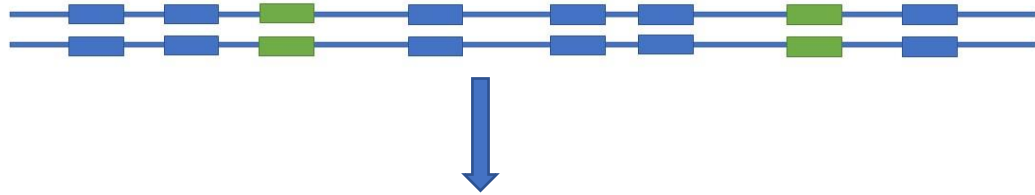
Ved valg av restriksjonsenzymmer kan det være lurt å vurdere hvor GC-fattig eller GC-rikt genomet er for at kuttingen skal bli mest mulig optimal. I og med at EcoRI og MseI har A- og T-rike gjenkjennelsessekvenser, henholdsvis GAATTC og TTAA, kan de dermed være foretrukket for GC-fattige genomer (57). Riktig valg av type restriksjonsenzymmer er avgjørende for analysen da det vil på forhånd bestemme størrelsen av og antallet fragmenter.

Adapterne som liggeres til endene av restriksjonsfragmentene består av en kjernesekvens og en enzymspesifikk sekvens. Kjernesekvensen karakteriserer det som er til felles for forward og revers adapter. Den enzymspesifikke sekvensen muliggjør ligering av adapterne til DNA-fragmentsekvensen som har oppstått etter restriksjonskutting. På denne måten vil de ligerte adapterne opprette et bindingssted for primerne i etterfølgende PCR-amplifisering. Primeren vil derfor være komplementær til kjernesekvensen, enzymspesifikke sekvensen og nukleotidrestene oppstått etter restriksjonskutting. De amplifiserte restriksjonsfragmentene blir deretter dypt sekvensert ved bruk av en egnet sekvenseringsplattform. Overlapping av fragmenter kan knyttes opp mot grad av overlapping i mikrobiota (arts- og stammedeling mellom mor og barn) (46).

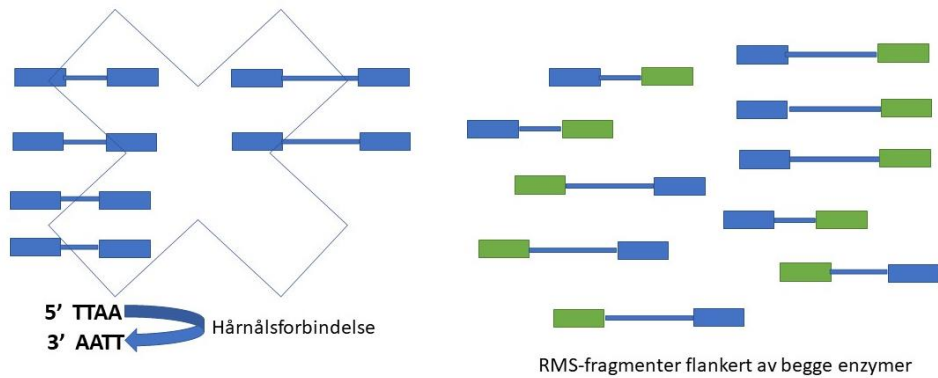
## A) Prøvepreparering

### Restriksjonsenzymmer

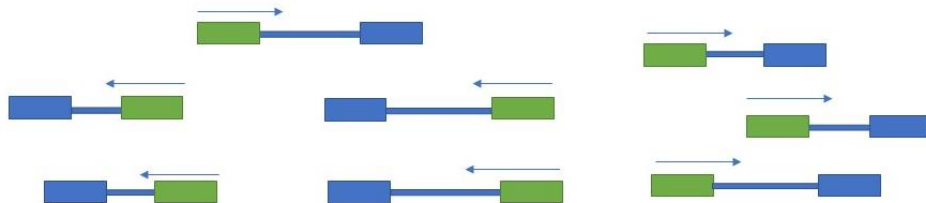
- Hyppig (MseI)
- Sjelden (EcoRI)



## B) PCR-amplifisering



## C) Sekvensering



Figur 1. **Illustrasjon av prinsippet til Redusert Metagenomisk Sekvensering (RMS).** A) Prøvepreparering; Kutting av genomisk DNA i fragmenter ved hjelp av en enzymkombinasjon av hyppig (MseI, her illustrert med blå farge) og sjelden (EcoRI, her illustrert med grønn farge) kutter. B) Dominant PCR-amplifisering av fragmenter flankert av begge enzymer. MseI/MseI fragmenter har en tendens til å danne hårnålsstrukturer ved at endene på fragmentene basepares, og blir dermed utkonkurrert (illustrert med kryss). C) De amplifiserte restriksjonsfragmentene flankert av begge restriksjonsenzymmer blir deretter dypt sekvensert (piler angir retning til avlesningen av forward DNA- tråden). Figur laget av student.

### **1.3 DNA-sekvensering**

#### **1.3.1 Førstegenerasjons sekvensering**

Førstegenerasjons sekvensering ble først utviklet i 1977, herunder Sanger sekvensering. Sanger sekvensering er kjent som kjedeterminerings-, dideoksynukleotid trifosfat (ddNTP)- eller sekvensering-ved-syntese metoden. Prinsippet til metoden baserer seg på bruk av radioaktivt merket ddNTP som med sin manglende 3' OH-gruppe fører til en kjedeterminering av DNA-tråden i den gitte ddNTP-posisjon. Dette vil gi en blanding av fragmenter i ulike størrelser som kan detekteres ved elektroforese på en denaturerende polyakrylamidgel og videre visualiseres ved autoradiografi (60). Metoden er i stand til å produsere sekvenser av relativt høy kvalitet og lengde (1000-1200 bp), men har også begrensninger med hensyn til å være tidkrevende, dyr og med redusert gjennomstrømming (61). Begrensningene med første generasjons sekvensering har dermed vært et grunnlag for utvikling av neste generasjons sekvensering.

#### **1.3.2 Andre generasjons sekvensering**

Andre generasjons sekvensering, også kalt for Neste Generasjons Sekvensering (NGS), er en teknologi som har revolusjonert genomisk forskning det siste tiåret. Den beskriver en rekke ulike moderne sekvenseringsplattformer som Roche 454, Ion Torrent, SOLiD/ABI og Illumina (62). Plattformene tilbyr en massiv parallell og dyp sekvensering av nukleinsyrer av høy kvalitet, som samtidig er effektiv og billig sammenlignet med Sanger-sekvensering. Imidlertid er Roche 454 og Illumina de plattformene som har blitt mest benyttet i metagenomiske studier (48).

Roche 454 benytter en såkalt pyrosekvenseringsteknikk, som baserer seg på deteksjon av pyrofosfat i form av lys som blir frigjort etter enhver nukleotid-inkorporering til den voksende DNA-tråden. DNA blir først fragmentert og hvert fragment blir festet til en kule som har oligonukleotider festet på overflaten som er komplementære til fragmentet. Hver kule blir så isolert og amplifisert ved bruk av emulsjons-PCR som produserer millioner av DNA-fragment kopier. Kulene blir så overført til en plate, PikoTiterPlate (PTP), som inneholder millioner av brønner, før pyrosekvenseringen settes i gang. Når den tilsatte deoksynukleotid trifosfat (dNTP) er komplementær til templat-DNA i en gitt posisjon, forlenges den voksende DNA-tråden ved hjelp av DNA-polymerase. Lys blir avgitt og registrert av et kamera. Lysemisjonen som oppstår etter enhver nukleotid-inkorporering er med å utlede sekvensen til DNA-fragmentet (63).

Illumina prinsippet, derimot, starter med at DNA blir fragmentert og adaptere som er nødvendige for sekvenseringen blir så festet på endene av fragmentene. DNA-fragmentene vil videre binde seg til komplementære oligonukleotider som er festet til kanaler på «flowcellen».

Dette muliggjør amplifisering av fragmentene og såkalte «clustere» dannes, der hvert cluster er en kopi av et gitt DNA-fragment. Clusterne dannes ved samtidig tilsetning av alle fire fluorescensmerkede nukleotider (ddNTP). Ved sekvensering vil nukleotidene sammen konkurrere om binding til templat-DNA. Hvilket nukleotid som blir festet avhenger av hvilket nukleotid som befinner seg i akkurat den posisjonen på templatet, alt som er ubundet vaskes vekk. Et nukleotid blir festet på den voksende tråden om gangen, noe som skjer samtidig på alle clustere som har identiske tråder (=massiv parallell sekvensering). Etter hver syklus vil de bundede fluorescensmerkede nukleotidene bli eksitert av laser og lys vil bli avgitt og fanget opp av et kamera. Et blokkeringstrinn av 3' ende er også inkludert for å forhindre uønsket priming. Prosessen repeteres helt til den hele DNA-sekvensen til ethvert fragment utledes (48).

Illumina tilbyr også såkalt «paired-end» sekvensering hvor begge ender av fragmentet blir sekvensert. Ved sekvensering på MiSeq vil dette kunne gi en sekvens på 2 x 300 bp, mens ved sekvensering på for eksempel HiSeq 3000 vil det kunne gi en sekvens på 2 x 150 bp. Illumina er også billig, men analysetiden er noe lenger sammenlignet med Roche 454. Foruten om tilbyr Illumina en høy dekning. Disse fordelene har blant annet gjort Illumina til den mest foretrukne sekvenseringsplattformen i metagenomiske studier (48).

### **1.3.3 Tredjegerasjons sekvensering**

Tredjegerasjons sekvensering er det nyeste på markedet og inkluderer metoder som Nanopore og Pacific Biosciences (PacBio). Prinsippet til førstnevnte baserer seg på DNA-fragment som passerer gjennom en proteinpore, og variasjon av ionisk strøm måles og er avhengig av hvilket nukleotid som passerer gjennom porene. Informasjon om variasjonen i ionestrømmen registreres og sekvensen til fragmentet identifiseres (64). PacBio, derimot, benytter en DNA-polymerase som er festet til strukturer på en matrise og fluorescensmerkede nukleotider for sekvensering av DNA. Disse strukturene finnes i et høyt antall i samme matrise og muliggjør dermed sekvensering av flere unike DNA-fragmenter samtidig (65).

### **1.4 Taksonomisk klassifisering av metagenomiske data**

En metagenomisk prøve vil sannsynligvis inneholde arter som er identiske eller til dels like når de sammenlignes med arter som er tilgjengelige i databaser. En slik tildeling av taksonomi fra databaser kan utføres ved hjelp av flere ulike klassifiseringsverktøy som for eksempel BLAST (66) og Kraken2 (67). Førstnevnte er en av de mest kjente og beste metodene for tildeling av taksonomi til ukjente sekvenser. Den er i stand til å klassifisere sekvenser mot data fra store databaser. Selv om BLAST i utgangspunktet ikke er etablert for klassifikasjon av metagenomiske data, er den fortsatt en av de beste tilgjengelige metodene (67). Kraken2,

derimot, har blitt omtalt til å være en av de mest foretrukne klassifiseringsverktøyene for metagenomiske data, men er ikke like sensitiv som BLAST. Imidlertid er Kraken2 mer effektiv i sekvensklassifiseringen enn BLAST (67), og er detaljert beskrevet under.

#### **1.4.1 Kraken2**

Kraken2 er et effektivt verktøy for taksonomisk klassifisering av metagenomiske korte sekvenser kalt K-merer. En K-mer er et ord av lengde K i antall baser. Kraken2 baserer seg på at alle sekvenser, fra helgenomer og ned til korte sekvenser, kan splittes opp i en mengde 31-merer. Kraken2 opererer dermed med  $K=31$ . Når hele genomet eller en enkelt sekvens splittes opp i K-merer, kan sekvensene klassifiseres raskt ved å «matche» alle K-merer fra søkesekvensen mot K-merer som er tilgjengelige i databasen. Når det skal undersøkes om en sekvens kommer fra et spesielt genom, så har man på forhånd splittet opp genomet i 31-merer og lagret alle de unike 31-merene i en hash-tabell (en måte å strukturere data på). Når man så splitter søkesekvensen opp i 31-merer, så kan man raskt slå opp i hash-tabellen for å se om lignende 31-merer har blitt funnet i genomet (67).

Dersom Kraken2 finner en 31-mer i to eller flere genomer, så tilordnes ikke denne 31-meren til noen av disse genomene, men til deres siste felles forfader (Lowest Common Ancestor (LCA)). Dette vil si at dersom 31-meren finnes i både art A og art B i databasen, og disse to artene er fra samme slekt – så tilordnes 31-meren til deres felles slekt. For en gitt sekvens er det mange 31-merer, og den taksonomiske klassifiseringen følger da en «flertall-avstemning», det vil si dit som flest 31-merer matcher. Dersom søkesekvensens 31-merer ikke matcher noen av de tilgjengelige genomenes 31-merer i databasen, forblir 31-merene merket som uklassifiserbare. Dette gjør Kraken2 svært presis, men samtidig ikke fullt så sensitiv (67).

En av Krakens2 store fordeler er at den er utstyrt med kommandoer for å opprette sin egen database. For å opprette en Kraken2 database må det først selekteres ut genomer fra National Center for Biotechnology Information (NCBI); Reference Sequence (RefSeq)-databasen. Deretter må genomene splittes opp i 31-merer ved hjelp av et egnet verktøy som JellyFish. Sekvensene må så behandles i NCBI taksonomi-databasen for å generere taxon-informasjon. Alle 31-merer blir tilordnet en taksonomi fra sekvensen de kommer fra (67).

Det finnes ulike varianter av Krakens2 databaser som Mini Kraken og Standard Kraken. Forskjellen mellom dem er at førstnevnte inneholder et mindre antall genomer sammenlignet med Standard Kraken som inneholder omtrent 2256 RefSeq-genomer. Dette gjør Mini Kraken mer presis, men ikke fullt så effektiv. En database som inneholder færre genomer/k-merer,

krever flere klassifiseringer av en sekvens for å finne en match. Dermed er Standard Kraken mer effektiv (67).

### **1.5 PreventADALL-studien**

Preventing Atopic Dermatitis and ALLergies (PreventADALL) er en stor internasjonal randomisert faktorielt designet prospektiv fødselskohortstudie, der barn ble rekruttert i Norge og Sverige. Studien har som formål å undersøke om tiltak i tidlig liv (eksponering av mat og hudpleie) kan redusere utviklingen av astma og andre allergiske sykdommer seinere i livet. I tillegg søker studien å identifisere faktorer i tidlig liv, gjennom svangerskap, fødsel og barndom som kan påvirke utviklingen av allergiske og andre folkesykdommer som diabetes, fedme og hjerte-karsykdommer. Samlet er dette basert på innsamlet data fra 18 ukers svangerskap, biologiske tester og oppfølging av barn på lang sikt.

Gravide kvinner ble bedt om å utfylle spørreskjema relatert til generell helse, livsstil (alkohol, tobakk, medisiner, kosthold, fysisk aktivitet), stress og livskvalitet. Oppfølgingen av barna inkluderte spørreskjema om generell helse, medisiner, sykdommer, boforhold, kosthold, stress og livskvalitet. Det ble sendt ut brev til alle kvinner som skulle ut til kontroll med invitasjon til å være med i prosjektet. De inkluderte deltakerne ble randomisert til fire grupper av tilsvarende størrelse, basert på hud-intervensjon, tidlig introduksjon av tilleggs kost ved siden av amming, begge intervensjoner eller ingen intervensjoner. Hudintervensjonen besto av olje-bad/bruk av hudkrem i ansiktet minst fire dager per uke fra 2-26 ukers alder, mens matvareintroduksjonen omfattet introduksjon etterfulgt av regelmessig bruk av peanøtt, kumelk, egg og hvete i form av smaksprøver minst fire dager/uke mellom 3 og 4 måneders alder. Ut over disse tiltakene ble deltagerne oppfordret til å følge generelle råd og anbefalinger fra Helsedirektoratet. Disse innebærer helst fullamming til 6 måneders alder, og ingen generelle råd om hudpleie.

Fra desember 2014 – november 2017 rekrutterte studien kvinner i 18 ukers svangerskap og deres biologiske materiale (urin, blod, morsmelk, spytt og avføring blant annet) ble innsamlet, mens deres nyfødte barn ble inkludert ved fødselen. Totalt ble 2400 mor-barn rekruttert og studien har innsamlet data via spørreskjema, dagbøker og kliniske undersøkelser, samt biologisk materiale fra barn i ulike alderskategorier (nyfødt, tre måneder, seks måneder, tolv måneder, 24 måneder og 36 måneder). Deretter skal barna forhåpentligvis undersøkes med noen års mellomrom frem til voksen alder.

## 1.6 Mål med studien

En stor utfordring med undersøkelse av mor-barn assosiasjon av tarmbakterier er at det kreves analysemetoder av høy oppløselighet som er i stand til å spore bakterier på arts- og stammenivå, og som samtidig er billige og effektive (26).

Ved 16S rRNA gensekvensering studeres konserverte områder av bakteriene som er tilstede i prøven, noe som vanskeliggjør sporing av bakterier på arts- og stammenivå. Helgenom-shotgun sekvensering, derimot, benytter en mye større andel av genomet og er derfor altfor kompleks og dyr (68). En høyoppløselig og effektiv analysemetode, Redusert Metagenomisk Sekvensering (RMS), som har kapasitet til sporing av bakterier på arts- og stammenivå, ble derfor benyttet i denne studien.

Hovedmålet med denne studien var dermed *å undersøke mor-barn assosiasjon av tarmbakterier ved bruk av RMS for sporing av bakteriearter* med utgangspunkt i genomisk DNA fra avføringsprøver fra 100 mor-barn par knyttet til PreventADALL-studien. Avføringsprøvene ble samlet fra mor i 18 ukers svangerskap og deres barn i ulike alderskategorier (nyfødt/mekonium (spedbarnets første avføring), tre måneder, seks måneder og tolv måneder).

Følgende delmål ble også inkludert i studien:

- Validering av kvaliteten på RMS-biblioteket ved MiSeq sekvensering før ytterligere dypere sekvensering på HiSeq 3000
- Analyse av Mock-samfunn for evaluering av kvaliteten til RMS-metoden i sporing av kjente bakteriearter
- Sporing av bakteriearter assosiert med alder



## 2. Materiale og metode

### 2.1 Prøvemateriale

Alle aspekter av rekruttering, prøvemateriale, analyse og databehandling ble godkjent av Regional Etisk Komité (REK) for medisinsk og helsefaglig forskningsetikk (2014/518), og er gitt i vedlegg 1. Informert skriftlig samtykke ble innhentet fra mødrene. Alle analysemetoder ble utført i henhold til produsentens spesifikasjoner og anbefalinger.

#### 2.1.1 Mor-barn analyser

Materialet som har blitt undersøkt er genomisk DNA fra avføringsprøver fra 100 mor-barn par knyttet til PreventADALL-studien. Avføringsprøvene ble samlet fra mor i 18 ukers svangerskap og deres barn i ulike alderskategorier (nyfødt/mekonium, tre måneder, seks måneder og tolv måneder).

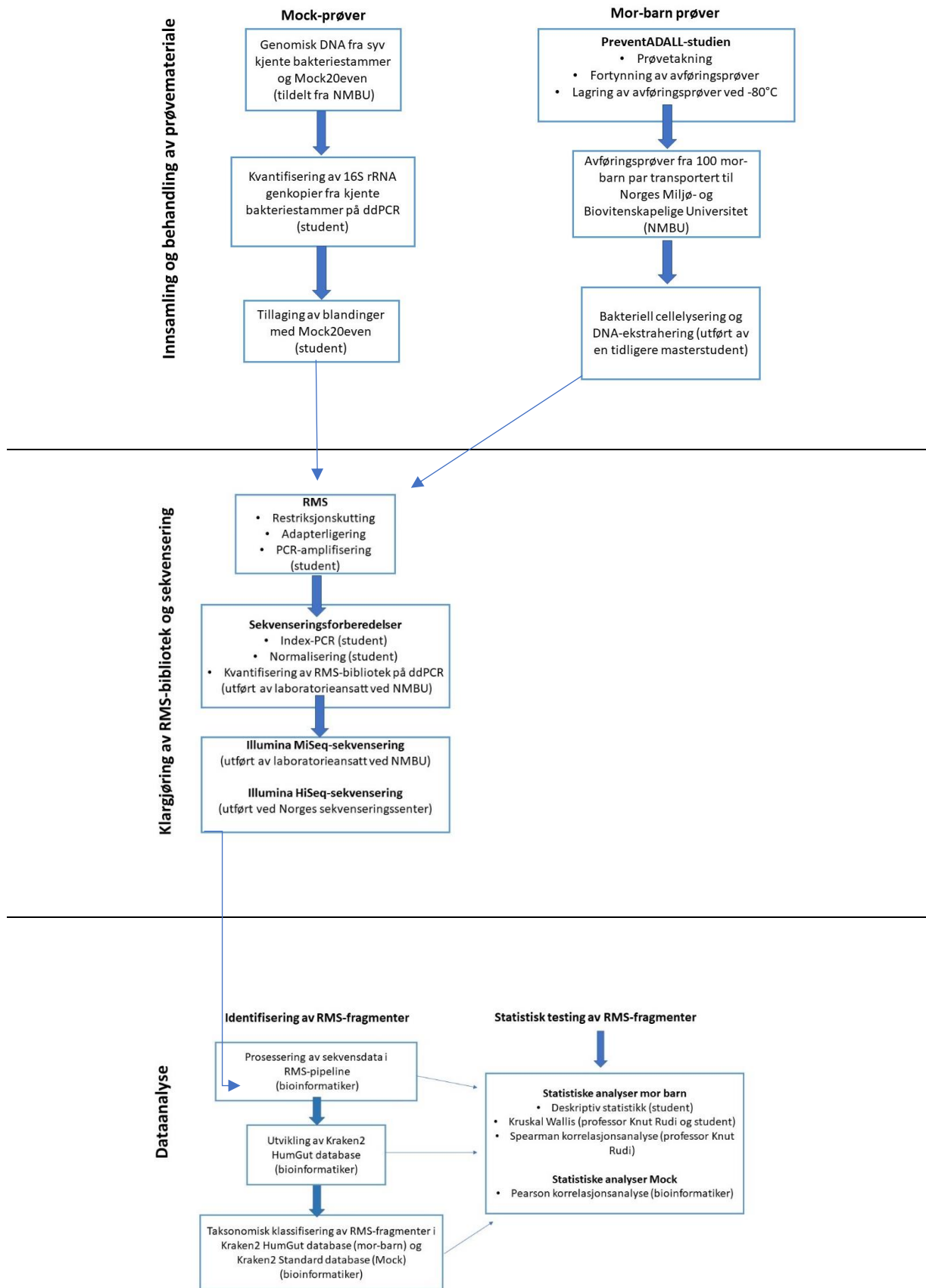
Umiddelbart etter Prevent-ADALLs prøvetakning ble avføringsprøvene samlet i sterile avføringstuber og fortynnet 1:10 i DNA-buffer for å unngå degradering av DNA. Avføringsprøvene ble lagret ved -80 °C. I 2017 ble et antall på 500 avføringsprøver transportert til Norges Miljø- og Biovitenskapelige Universitet (NMBU) for ytterligere lagring ved -60°C. I forkant av denne studien var DNA fra avføringsprøvene allerede blitt ekstrahert, rensset, kvantifisert og lagret i fryser.

#### 2.1.2 Mock-analyser

For å evaluere kvaliteten til RMS-metoden i detektering av kjente bakterier, ble det gjennomført ytterligere analyser av genomisk DNA fra følgende syv kjente bakteriestammer; *Streptococcus agalactiae* DSM 2134; *Streptococcus mutans* DSM 20523; *Streptococcus pneumoniae* DSM 14378; *Enterococcus faecalis* DSM 20478; *Helicobacter pylori* DSM 21031; *Enterococcus faecalis* DSM 2570; *Lactobacillus gasseri* DSM 20243 og Mock-samfunn (Mock20even). Mock20even består av genomisk DNA fra ytterligere tjue kjente bakteriestammer; *Acinetobacter baumannii*, stamme 5377, *Actinomyces odontolyticus*, stamme 1A.21, *Bacillus cereus*, stamme NRS 248, *Bacteroides vulgatus*, stamme ATCC® 8482™, *Clostridium beijerinckii*, stamme NCIMB 8052, *Deinococcus radiodurans*, stamme R1, *Enterococcus faecalis*, stamme OG1RF, *Escherichia coli*, stamme K12, substamme MG1655, *Helicobacter pylori*, stamme 26695, *Lactobacillus gasseri*, stamme 63 AM, *Listeria monocytogenes*, stamme EGDe, *Neisseria meningitidis*, stamme MC58, *Propionibacterium acnes*, stamme KPA171202, *Pseudomonas aeruginosa*, stamme PAO1-LAC, *Rhodobacter sphaeroides*, stamme ATH 2.4.1, *Staphylococcus aureus*, stamme TCH1516, *Staphylococcus epidermidis*, FDA stamme PCI 1200, *Streptococcus agalactiae*, stamme 2603 V/R, *Streptococcus mutans*, stamme UA159,

*Streptococcus pneumoniae*, stamme TIGR4, som inneholder samme antall (Even) 16S rRNA genkopier (100 000 kopier per organisme per  $\mu\text{L}$ ).

En oversikt av studiens arbeidsflyt er illustrert i figur 2.



Figur 2. Illustrasjon av studiens arbeidsflyt med hensyn til mor-barn- og Mock-analyser.

## **2.2 Forbehandling av prøver**

### **2.2.1 Mor-barn prøver**

#### **Lysering av bakterieceller**

I forkant av lyseringen ble avføringsprøvene tint på is og vortexet for å homogenisere prøvene. Prøvene ble så sentrifugert ved 1200 rpm i åtte sekunder for å simplifisere ekstraksjonen ved å aggregere større fekale partikler.

Det ble utført en kombinasjon av kjemisk og mekanisk lysering av bakterieceller. Prøvene ble blandet med glasskuler av tre ulike størrelser. Det ble tilsatt 0,2g av syrevaskede glasskuler (<106  $\mu\text{M}$ , Sigma-Aldrich, Germany) 0,2g syrevaskede glasskuler (425-600  $\mu\text{M}$ , Sigma-Aldrich, Germany) og 2x 2,5-3,5mm syrevaskede glasskuler (Sigma-Aldrich, Germany) til 300  $\mu\text{L}$  prøve. Denne kombinasjonen har som hensikt å sikre et høyere mangfold og redusere bias som ellers kan oppstå ved en mekanisk lysering av bakterieceller. Prøvene ble prosessert to ganger på en FastPrep 96 (MP Biomedicals, USA) ved 1800 rpm i 40 sekunder, med etterfølgende sentrifugering ved 13 000 rpm i 5 minutter. Sentrifugeringen tilrettelegger for videre DNA-behandling ved å aggregere cellerester og glasskuler.

Den kjemiske lyseringen ble gjennomført ved bruk av lyseringsbuffer og proteinase K. De mekanisk lyserte prøvene ble behandlet med lyseringsbuffer og proteinase K på KingFisher Flex Robot. En lyseringsbuffer vil føre til en lysering av cellene grunnet den høye saltkonsentrasjonen, mens proteinase K bryter ned kontaminerende proteiner i prøven, som for eksempel nuklease som bryter ned DNA (69).

#### **DNA-ekstrahering**

DNA-ekstrahering av lyserte prøver ble utført på KingFisher Flex robot (Thermo Scientific, USA) ved bruk av paramagnetiske kuler. DNA bindes til paramagnetiske kuler i nærvær av en høy saltkonsentrasjon, som danner en bro mellom DNA og de paramagnetiske kulene (70). Etanol ble tilsatt for presipitering av DNA. Prøvene ble så vasket med vaskebuffer som inneholder salt som binder DNA til paramagnetiske kuler. DNA ble eluert ved bruk av nukleasefritt vann som forstyrrer saltbroen mellom DNA og de paramagnetiske kulene. Eluatet med DNA ble så kvantifisert ved bruk av kvantitativ PCR (qPCR).

## 2.2.2 Mock-prøver

### Tillaging av blandinger med Mock20even

For å se om RMS-metoden klarer å skille lignende bakterier fra hverandre ble det laget åtte ulike blandinger i triplikater (for å øke påliteligheten og redegjøre for eventuelle avvik) av kjente bakteriestammer og Mock20even som vist i tabell 1. Blanding 1 inneholdt alle syv bakteriestammer og Mock20even, mens blanding 10 inneholdt kun Mock20even. Nukleasefritt vann (VWR, USA) ble benyttet som negativ kontroll for å utelukke forurensning fra blant annet reagenser. Det var et totalt antall på 30 prøver.

I forkant av tillagingen av blandinger med Mock20even ble antall 16S rRNA genkopier fra alle de syv kjente bakteriestammene kvantifisert på Droplet Digital PCR (ddPCR) (se punkt 2.4.2). Det var ønskelig at hver bakteriestamme besto av tilsvarende 100 000 16S rRNA genkopier/ $\mu\text{L}$  lik Mock20even. For å oppnå det ble de syv bakteriestammene videre fortynnet 1:10 med nukleasefritt vann (VWR, USA) etter kvantifiseringen.

I og med at det var unike bakterielle sammensetninger i 9/10 blandinger besto blandinger av ulikt volum genomisk DNA og dermed også av ytterligere ulikt volum nukleasefritt vann (VWR, USA). Til blanding 1 ble det derfor tilsatt 8  $\mu\text{L}$  genomisk DNA (1  $\mu\text{L}$  per bakteriestamme og 1  $\mu\text{L}$  Mock20even), mens til blanding 2 ble det tilsatt 2  $\mu\text{L}$  genomisk DNA. Blandingene fungerte dermed som templat for videre RMS-analyse. I og med at det allerede ble tilsatt nukleasefritt vann (VWR, USA) til blandinger, besto mastermixen i punkt 2.3.1 av kun resterende reagenser nødvendig for analyse.

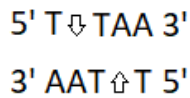
Tabell 1. Oversikt av ulike blandinger til Mock-analyse, samt bakteriestammenes antall 16S rRNA genkopier/ $\mu\text{L}$  som et resultat av kvantifiseringen på ddPCR.

	Antall 16S rRNA genkopier/ $\mu\text{L}$	Blanding 1	Blanding 2	Blanding 3	Blanding 4	Blanding 5	Blanding 6	Blanding 7	Blanding 8	Blanding 9	Blanding 10
Mock20even	100 000X	X	X	X	X	X	X	X	X		X
<i>Streptococcus agalactiae</i> (DSM 2134)	91 900 000X	X									
<i>Streptococcus mutans</i> (DSM 20523)	5 050 000X			X							
<i>Streptococcus pneumoniae</i> (DSM 14378)	139 250 000X				X						
<i>Enterococcus faecalis</i> (DSM 20478)	62 550 000X					X					
<i>Enterococcus faecalis</i> (DSM 2570)	142 250 000X						X				
<i>Helicobacter pylori</i> (DSM 21031)	33 550 000X							X			
<i>Lactobacillus gasseri</i> (DSM 20243)	70 650 000X								X		
Nukleasefritt vann	-	X	X	X	X	X	X	X	X	X	

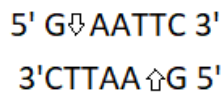
## 2.3 Klargjøring av RMS-bibliotek til sekvensering

### 2.3.1 Restriksjonskutting

Genomisk DNA (~10 ng) ble kuttet i fragmenter til en størrelse mellom 500 og 550 bp ved bruk av en enzymkombinasjon av MseI og EcoRI. Enzymene kutter DNA på ulike steder hyppig og sjeldent, henholdsvis på 5' T (TTAA) og 5' G (GAATTC) som vist i figur 3 for MseI og figur 4 for EcoRI. Til restriksjonskutting ble det laget en reaksjonsmix (20 µL) bestående av 8U EcoRI (New England Biolabs, USA), 4U MseI (New England Biolabs, USA), 1x Cut Smart Buffer (New England Biolabs, USA) og nukleasefritt vann (VWR, USA). Prøvene ble inkubert i én time ved 37 °C for å sikre tilstrekkelig kutting av genomisk DNA i fragmenter.



Figur 3. MseI gjenkjenner en palindromisk sekvens 5' TTAA 3' og kutter fosfodiesterbindingen mellom to T på begge tråder for å danne to base 5' forlengelse T/TAA.



Figur 4. EcoRI gjenkjenner en palindromisk sekvens 5' GAATTC 3' og kutter fosfodiesterbindingen mellom G og A på begge tråder for å danne fire base 5' forlengelse G/AATTC.

Ved bruk av en enzymkombinasjon av MseI og EcoRI vil genomisk DNA bli kuttet i fragmenter av ulik størrelse og ulikt antall. Fragmentene kategoriseres i «MseI/MseI, EcoRI/EcoRI og MseI/EcoRI», og avhenger av om de flankeres av begge enzymer eller ikke. Det store flertallet forventes å være MseI/MseI fragmenter, og dobbelt så mange MseI/EcoRI fragmenter som EcoRI/EcoRI fragmenter (57, 58).

### 2.3.2 Adapterligerings

For mulig gjøring av etterfølgende PCR-amplifisering ble det utført adapterligerings til endene av restriksjonsfragmentene. Adapterligerings innebærer å introdusere nukleotider som er komplementære til 5' overheng som oppstår etter kutting med restriksjonsenzymer. Adapterne består av en kjernesekvens og en enzymspesifikk sekvens som vist i figur 5 for EcoRI adapter og figur 6 for MseI adapter. Kjernesekvensen karakteriserer det som er til felles for forward og revers adapter. Den enzymspesifikke sekvensen muliggjør ligerings av adapterne til DNA-

fragmentsekvensen som har oppstått etter restriksjonskutting. På denne måten vil de ligerte adapterne opprette et bindingssted for primerne i etterfølgende PCR-amplifisering.

Til adapterligering ble det laget en reaksjonsmix (5µL) bestående av 0,5 µM EcoRI adaptermix (Invitrogen, USA), 5 µM MseI adaptermix (Invitrogen, USA), 1 µL T4 DNA ligase (New England Biolabs, USA) og 1x T4 reaksjonsbuffer (New England Biolabs, USA). Adaptermixen ble laget av like store mengder av forward adaptere (EcoRI; 5'-CTCGTAGACTGCGTACC-3', MseI; 5'-GACGATGAGTCCTGAG-3') og reverse adaptere (EcoRI; 5'-AATTGGTACGCAGTCTAC-3', MseI; 5'-TACTCAGGACTCAT-3'). Prøvene ble inkubert i tre timer ved 37 °C.

#### EcoRI adapter

```
5'CTCGTAGACTGCGTACCAATTC3'  
3'CATCTGACGCATGGTTAAG5'
```

Figur 5. EcoRI adapter (forward og revers) består av en kjemesevens (svart) og en enzymspesifikk sevens (grønn) som er komplementær til nukleotidrestene/DNA-kuttesevensen oppstått etter restriksjonskutting (blå).

#### MseI adapter

```
5'GACGATGAGTCCTGAGTAA3'  
3'TACTCAGGACTCATT5'
```

Figur 6. MseI adapter (forward og revers) består av en kjernesevens (svart) og en enzymspesifikk sevens (grønn) som er komplementær til nukleotidrestene/DNA-kuttesevensen oppstått etter restriksjonskutting (blå).

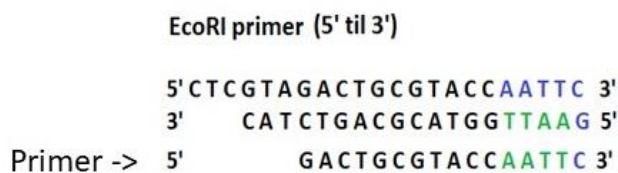
### 2.3.3 Kvalitativ PCR

Kvalitativ PCR ble utført for amplifisering av adapterligert genomisk DNA, med spesifikke EcoRI forward (5'-GACTGCGTACCAATTC-3') og MseI reverse (5'-GATGAGTCCTGAGTAA-3') primerpar komplementære til adapter- og kuttesevensen (58), som vist i figur 7 for EcoRI og figur 8 for MseI. Av ulike grunner vil det foregå en dominant amplifisering av fragmenter flankert av begge enzymer. MseI-primer har lavere hybridiseringstemperatur enn EcoRI-primer, noe som gjør at MseI/MseI fragmentamplifisering blir mindre effektiv sammenlignet med EcoRI/MseI fragmentamplifisering. MseI/MseI fragmenter er korte og de har i tillegg en «invertert repeat» i endene. De kan dermed ha en

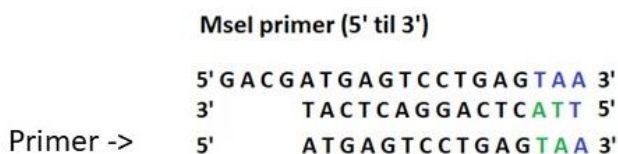
tendens til å danne hårnålsstrukturer ved at endene på fragmentene basepares, noe som sannsynligvis vil utkonkurrere primerbindingen (57, 58). EcoRI/EcoRI fragmenter, derimot, er lenger og det vil dermed være mindre sannsynlig at hårnålsstrukturer dannes. Det er svært få av EcoRI/EcoRI fragmenter ved PCR-amplifisering på grunn av lengden deres.

Det ble laget en reaksjonsmix (20 µL) bestående av 1x HotFirePol Blend Master Mix Ready To Load (Solis BioDyne, Estonia), 0,2 µM EcoRI forward primer (Invitrogen, USA), 0,2 µM MseI reverse primer (Invitrogen, USA) og nukleasefritt vann (VWR, USA). Reaksjonsmixen ble først tilsatt til hver brønn i 96-brønnersplate med etterfølgende tilsetning av adapterligert DNA-templat (5 µL) til et totalt volum på 25 µL. Prøvene ble amplifisert opp ved følgende oppsett; 95 °C oppvarming i 15 minutter etterfulgt av 25 sykluser bestående av denaturering ved 95 °C i 30 sekunder, hybridisering ved 55 °C i 1 minutt og elongering ved 72 °C i 1 minutt, før et avslutningssteg ved 72 °C i 7 minutter og nedkjøling ved 10°C ∞. Antall sykluser ble økt til 30 for mekoniumsprøvene. Alle PCR-reaksjoner ble gjennomført på en 2720 Thermal Cycler (Applied Biosystems, USA).

For å utelukke kontaminasjon fra PCR-reagenser ble det til enhver PCR-reaksjon inkludert en negativ kontroll bestående av nukleasefritt vann (VWR, USA).



Figur 7. EcoRI forward primer (5' til 3') benyttet i første PCR er komplementær til adapter-kjernesekvensen (svart), adapter-enzym spesifikk sekvensen (grønn) og nukleotidrestene/DNA-kuttsekvensen oppstått etter restriksjonskuttingen (blå).



Figur 8. MseI revers primer (5' til 3') benyttet i første PCR er komplementær til adapter-kjernesekvensen (svart), adapter-enzym spesifikk sekvensen (grønn) og nukleotidrestene/DNA-kuttsekvensen oppstått etter restriksjonskuttingen (blå).



#### **2.3.4 Rensing av PCR-produkt**

For fjerning av overflødig primer, primer-dimer, nukleotider, salt, polymerase og andre kontaminanter etter PCR ble PCR-produktet rensed ved bruk av AMPure® XP kuler (Beckman Coulter, USA). Et rensed PCR-produkt danner et større utbytte for sekvenseringen. Rensingen selekterte også for fragmenter av ønsket lengde ved å rense bort de minste fragmentene. Rensingen ble gjennomført på Biomek 3000 (Beckman Coulter, USA) i forhold 1:1 av volum kuler og DNA i henhold til produsentens anbefalinger.

AMPure® XP kuler er paramagnetiske og har den betydning at kulene er magnetiske i nærvær av en magnet. DNA vil først bindes til kulene og separeres fra mulige kontaminanter i løsningen ved hjelp av magnet. Deretter vil kulene bundet til DNA bli vasket to ganger med 70-80% etanol for fjerning av ytterligere kontaminanter. Til slutt vil det rensede DNA elueres fra kulene ved hjelp av for eksempel nukleasefritt vann.

Rensingen av RMS-biblioteket ble utført manuelt ved bruk av AMPure® XP kuler (Beckman Coulter, USA) i forhold 2:1 av volum kuler og DNA på grunn av problemer med instrument.

#### **2.3.5 Index-PCR**

Index-PCR ble utført for en ytterligere adapterligering (Illumina-spesifikk) til de amplifiserte RMS-fragmentene. Adapterligeringen er nødvendig for massiv parallell sekvensering på Illumina MiSeq og Illumina HiSeq plattformene. Index-PCR primernes 3' ende er komplementær til fragmentenes adapter- og enzymkuttesevens, mens 5' ende er komplementær til oligonukleotidene festet til flowcellen i Illumina sekvenseringsplattformen sammen med Illumina sekvenseringsregion (bindesete for sekvenseringsprimer) og en unik barcode/index. Indekseringen er nødvendig for å skille de ulike prøvene fra hverandre etter at de er blandet sammen til en prøve, slik at de kan sekvenseres parallelt. Til dette ble en kombinasjon av unike 16 forward og 36 revers primere benyttet, noe som gav muligheten å indeksere 576 (16x36) prøver til sammen. En oversikt av Illumina index-primere er gitt i vedlegg 2.

Til index-PCR ble det laget en reaksjonsmix (13 µL) bestående av 1x FirePol Master Mix Ready to Load (Solis BioDyne, Estonia) og nukleasefritt vann (VWR, USA). Reaksjonsmixen ble tilsatt til hver brønn i 96-brønnersplate etterfulgt av tilsetning av 0,2 µM unik EcoRI forward index-primer (Invitrogen, USA) og 0,2 µM unik MseI revers index-primer (Invitrogen, USA) på epMotion 5070 (VWR, USA). Til slutt ble det tilsatt rensed PCR-produkt (2 µL) til et totalt volum på 25 µL. Fragmentene ble amplifisert og indeksert ved følgende oppsett; 95°C oppvarming i 5 minutter, etterfulgt av 10 sykluser bestående av denaturering ved 95°C i 30

sekunder, hybridisering ved 56°C i ett minutt og elongering ved 72°C i ett minutt før et avslutningssteg på 72 °C i 7 minutter og nedkjøling ved 10°C ∞.

### **2.3.6 Normalisering av RMS-bibliotek**

Normalisering av RMS-bibliotek ble utført for å sikre at like mengder av indexerte PCR-produkter blir fremstilt for Illumina sekvensering, og for å unngå over- eller underrepresentasjon av prøver. Normaliseringen ble gjennomført på Biomek 3000 (Beckman Coulter, USA) i henhold til produsentens anbefalinger. I forkant av normaliseringen ble fluorescens fra alle prøvene målt på Cambrex-FLX 800 CSE. Til målingen ble det laget en Qubit reaksjonsmix (Invitrogen, Burlington, Ontario) av Qubit dsDNA HS Reagent fortynnet 1:200 i Qubit dsDNA HS Buffer. 70 µL reaksjonsmix og 2 µL templat ble tilsatt til hver brønn i en Nunc plate til et totalt volum på 72 µL.

Deretter fulgte en kvantifisering på Qubit™ fluorometer (Life Technologies, USA) av tilfeldig utvalgte prøver (fem prøver per alderskategori og Mock) som representerte både lav, middels og høy fluorescens. Disse utvalgte prøvene ble videre benyttet til fremstilling av en standardkurve. På den måten kunne konsentrasjonen av resterende prøver beregnes. Et volum fra hver prøve i én 96-brønnersplate som tilsvarte 50 ng ble overført til et samlet eppendorfrør på Biomek 3000 (Beckman Coulter, USA). Instrumentet normaliserer én og én plate, henholdsvis 12 mnd, 6 mnd etc. hver for seg.

Bibliotekene (fra hvert eppendorfrør) ble så blandet til én prøve manuelt tilsvarende 700 ng DNA fra hvert bibliotek fra alle alderskategorier og Mock. Dette var basert på en ytterligere måling av bibliotekenes konsentrasjon på Qubit™ fluorometer (Life Technologies, USA) i etterkant av normaliseringen på Biomek 3000 (Beckman Coulter, USA).

Det endelige biblioteket ble så rensert ved bruk av AMPure® XP kuler (Beckman Coulter, USA) som forklart i 2.3.4, kontrollert på agarose gelelektroforese som forklart i 2.4.3 og kvantifisert på ddPCR som forklart i 2.4.2. DNA-konsentrasjonen ble målt til å være 23 nM. Prøven ble så fortynnet til 7pM i kombinasjon med PhiX (15%) og fremstilt til 2x300bp sekvensering på MiSeq (Illumina, USA) plattformen ved NMBU (Ås, Norway). Etter en kvalitetsvurdering av RMS-data fra MiSeq sekvenseringen, ble RMS-biblioteket videre fremstilt til 2x150bp sekvensering på HiSeq3000 (Illumina, USA) plattformen ved Norges Sekvenseringscenter (Oslo, Norway).

## **2.4 DNA kvantitet- og kvalitetssikring**

### **2.4.1 Qubit**

Genomisk DNA og PCR-produkter ble kvantifisert på Qubit™ fluorometer (Life Technologies, USA) ved bruk av Qubit dsDNA HS assay kit (Invitrogen, Burlington, Ontario). Det ble laget en Quant-iT™ reaksjonsmix bestående av Quant-iT™ reagens fortynnet 1:200 i Quant-iT™ buffer. Det ble tilsatt 2 µL prøve til 198 µL reaksjonsmix i henhold til produsentens anbefalinger.

### **2.4.2 Droplet Digital PCR (ddPCR)**

Kvantifiseringen av de unike bakteriestammene nevnt i punkt 2.1.2 ble gjennomført tilsvarende det for RMS-biblioteket (som forklart under), med unntak av at det til førstnevnte ble benyttet andre primere (0,2 µM PRK 341 forward primer og 0,2 µM PRK 806 reverse primer).

Til kvantifiseringen av RMS-biblioteket ble det laget reaksjonsmix bestående av 1x Super mix for EvaGreen (Bio-Rad), 0,2 µM Illumina Colony forward primer (5'-AATGATACGGCGACCACCGAGATCT-3'), 0,2 µM Illumina Colony reverse primer (5'-CCAGCAGAAGACGGCATACGAGAT-3') og nukleasefritt vann (VWR, USA). Reaksjonsmixen (20,7 µL) ble først tilsatt i små eppendorfrør, med etterfølgende tilsetning av DNA-templat (2,3 µL) til et totalt volum på 23 µL. Deretter ble 20 µL av prøven og 70 µL EvaGreen olje (Bio-Rad) overført til «cartridge» i ulike posisjoner for dråpegenerering. Prøven ble partisjonert tilfeldig i omtrent 20.000 dråper gjennom vann-olje emulsjon i QX 200 Droplet Generation instrumentet (BioRad, USA). 40 µL av dråpene ble så overført til en PCR-plate og forseglet med aluminium ved 180°C i fem sekunder på PCR Plate Sealer PX1 (Bio-Rad). PCR-amplifisering ble så utført individuelt i hver enkelt dråpe ved følgende oppsett; 95°C oppvarming i 5 minutter, etterfulgt av 40 sykluser bestående av denaturering ved 95°C i 30 sekunder, hybridisering ved 60°C i 30 sekunder og elongering ved 72°C i 45 sekunder, før to avslutningssteg ved 4°C i 5 minutter og 90°C i 5 minutter. Alle reaksjoner ble gjennomført på 2720 Thermal Cycler (Applied Biosystems, USA).

Dråpene ble kvantifisert ved hjelp av en programvare (Quantisoft; Bio-Rad) hvor de ble karakterisert som positiv (en eller flere mål-DNA-kopier tilstede) eller negativ (ingen mål-DNA-kopier tilstede), på bakgrunn av om fluorescens ble avgitt eller ikke. Antall mål-DNA-kopier presentert kan beregnes ut fra fraksjonen av antall positive dråper ved bruk av Poisson-statistikk i henhold til formel 1. Fra  $\lambda$  og sammen med volumet av hver PCR- dråpe og det totale antall dråper analysert, beregnes et estimat av den absolutte mål-DNA-konsentrasjonen. På den måten kan absolutte DNA-konsentrasjoner beregnes nøyaktig uten behov for standardkurve

(71). Jo flere dråper som representeres jo mer nøyaktig og presis blir kvantifiseringen, og da spesielt i lave konsentrasjonsområder.

$$\lambda = -\ln(1 - p) \quad (\text{Formel 1})$$

der,

$\lambda$  representerer gjennomsnittet av antall mål-DNA-molekyler/kopier per dråpe

$p$  representerer antall positive dråper med amplifisert mål-DNA-produkt tilstede

### **2.4.3 Agarose gelelektroforese**

For kontrollering av blant annet forventet størrelse, ble PCR-produktene validert på en 1,5% agarosegel ved gelelektroforese (BioRad, USA) i 30 minutter ved 85 volt. For tillaging av agarosegel ble 4,5 gram agarose (Invitrogen, USA) oppløst i 300 mL 1x tris-acetat EDTA (TAE) buffer. Etter at gelen var nedkjølt ble det tilsatt 12  $\mu$ L PeqGreen (PeqLab, Germany). PeqGreen er et fluorescerende fargestoff som binder DNA. 100 bp DNA-ladder (New England Biolabs, USA) ble analysert sammen med PCR-produktene på gelen for å kunne bestemme deres størrelse. I forkant av gelelektroforesen inneholdt PCR-produktet allerede en farge fra polymerasen benyttet ved kvalitativ PCR. Denne fargen muliggjorde overvåking av fragmentvandringen i agarosegelen.

Når agarosegelen blir omgitt av et elektrisk felt, vil negativt ladet DNA vandre mot den positivt ladede anoden (72). Fragmentene ble visualisert ved hjelp av Molecular Imager<sup>®</sup> Gel Dox<sup>™</sup> XR Imaging System med Quantity One 1-D programvare v.4.6.7 (BioRad, USA) ved bruk av ultrafiolett lys.

## **2.5 Dataanalyse**

### **2.5.1 Prosessering av sekvensdata**

Sekvensdata fra både MiSeq og HiSeq ble prosessert i en RMS-pipeline, som fortsatt er under utvikling. Første steg i prosesseringen var de-multiplexing i Cutadapt, som involverer bruk av barcode/index-informasjon, til å kunne splitte sekvenser fra ulike prøver fra hverandre til separate filer, etter at prøvene har blitt sekvensert samtidig. Etter de-multiplexing dannes det dermed FASTQ-filpar for hver prøve. For hvert FASTQ-filpar ble det så gjennomført søkt opp mot Kraken2 HumGut databasen for taksonomisk klassifisering.

Videre ble antallet sekvenser normalisert for det antallet fragmenter som var tilstede fra hver bakterie i en prøve (=gjennomsnittlig dybde) ved bruk av R-script. Grunnen til det er at DNA-et til ulike bakterier består av et ulikt antall fragmenter. På den måten kan signalet bli lett over-

eller underestimert på bakgrunn av antallet fragmenter, noe som igjen kan se ut som at en bakterie er mindre eller mer tilstede enn en annen bakterie.

### **2.5.2 Kraken2 HumGut database**

Kraken2 HumGut database er en nyutviklet database og inneholder en samling av bakterielle genomer som er vanlige å finne i den humane tarm. Per dags dato består databasen av 412 bakteriearter, i tillegg til *Homo sapiens*. Prinsippet for utviklingen av databasen følger flere steg. Det ble først laget en liste over alle arter som sannsynligvis forekommer i den humane tarm. Disse opplysningene ble hentet fra «Human Microbiome Project» og EzBioCloud (en taksonomisk database) (73). Videre ble det selektert ut genomer fra NCBI- RefSeq-databasen av artene. Alle RefSeq-genomer av disse artene ble samlet, noe som resulterte i rundt 45 000 genomer. Det ble så beregnet en «avstand» mellom genomer av samme art ved hjelp av et verktøy som kalles Mash (74). Grunnen til dette er at mange genomer innenfor en art kan være ekstremt like, og vi ikke trenger alle slike nesten-kopier i databasen. Genomer ble så clustret med avstand 0,01 – det vil si at grupper av «like» genomer ligger innenfor en radius av 0,01, og bare ett genom fra hver gruppe utgjør databasen. Dette gjorde at antallet genomer ble redusert til omtrent 2500. En Kraken2 HumGut database ble så laget og var basert på denne samlingen av bakterielle genomer, i tillegg til *Homo sapiens*.

### **2.5.3 Statistiske analyser**

For å få et overblikk av de viktigste egenskapene til data produsert i studien, og for å kontrollere at variablene oppfyller gitte forutsetninger for bruk av spesifikke statistiske tester, ble data først presentert ved hjelp av deskriptiv statistikk. Deskriptiv statistikk for ikke-parametriske data inkluderte bruk av median og interkvartilbredde. Datasett fra alle alderskategorier ble visualisert i individuelle box-plotter. Box-plottene sammenligner medianen og spredningen av data for hver alderskategori.

Ikke-parametrisk Kruskal-Wallis test ble utført for å teste eventuelle statistisk signifikante forskjeller i den relative mengden av bakteriearter (kontinuerlig avhengig variabel) mellom ulike alderskategorier (kategorisk uavhengig variabel). I tilfeller hvor testen var statistisk signifikant, ble det videre utført Dunn's Post-Hoc test for parvise sammenligninger for å undersøke hvilke bestemte alderskategorier av den kategoriske uavhengige variabelen som er statistisk signifikant forskjellig fra hverandre, med hensyn til den relative mengden av bakteriearter av den kontinuerlige avhengige variabelen. Alle p-verdier ble videre Bonferroni-korrigert for å redusere sjansen for falske positive resultater.

En Spearman-korrelasjonsanalyse ble utført av mor-barn prøver for å teste for bakterier som er korrelert med alder (to kontinuerlige variabler).

I tillegg ble det utført en Pearson-korrelasjonsanalyse av Mock-prøvene for å teste hvorvidt det er korrelasjon mellom forventet og bestemt bakteriell sammensetning (to kontinuerlige variabler) i Mock. Korrelasjonskoeffisienten,  $R$ , kan variere fra 1 til -1. En verdi på 1 indikerer svært god positiv korrelasjon mellom to variabler.  $R = 0$  dersom  $H_0$  er sann, noe som tilsier ingen korrelasjon mellom to variabler.

Alle statistiske analyser ble utført i Statistical Package for the Social Sciences (SPSS), med et signifikansnivå på 5%.

### 3. Resultat

#### 3.1 Tillaging av RMS-bibliotek

Før tillaging av RMS-biblioteket ble DNA-konsentrasjonene til 11 prøver, med henholdsvis fire prøver fra mor, to prøver fra tolv måneder, fire prøver fra seks måneder og én prøve fra tre måneder, kvantifisert på Qubit. DNA-kvantifiseringen ble gjennomført for å kunne bestemme prøvevolumet som skulle tilsettes tilsvarende 10 ng genomisk DNA. DNA-konsentrasjonene fra Qubit-målingene var i området 0,1-0,8 ng/μL.

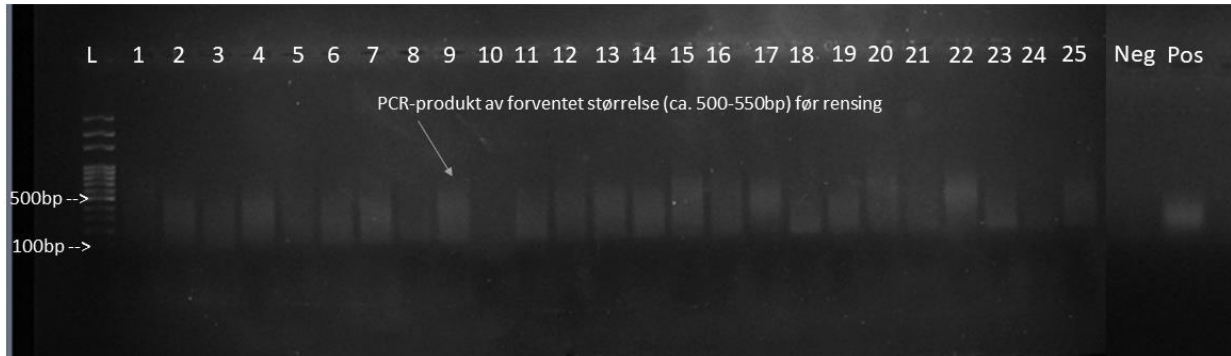
Etter første PCR og index-PCR fulgte kvalitetssjekk av alle 543 prøver (mor-barn prøver og Mock-prøver med tilhørende negative og positive kontroller) på agarose gelelektroforese. Kvalitetssjekken var nødvendig for å kontrollere for fragment av forventet størrelse, for å utelukke forurensning fra blant annet reagenser og for å kontrollere at rensingen av PCR-produktet var tilstrekkelig.

Gelresultatet fra tolv måneders prøver etter første PCR og index-PCR viste bånd av forventet størrelse på cirka 500-550 bp for henholdsvis 66% og 73% av prøvene. En høyere andel av seks måneders prøver fikk bånd av forventet størrelse etter første PCR og index-PCR, henholdsvis 95% og 84%. Når det gjelder gelresultatet fra tre måneders prøver etter første PCR og index-PCR, ble det observert bånd av forventet størrelse for henholdsvis 93% og 92% av prøvene. Andelen av mødre-prøver som fikk bånd av forventet størrelse etter første PCR og index-PCR var henholdsvis 85% og 81%. Imidlertid var andelen prøver som fikk bånd av forventet størrelse lavest for mekonium-prøvene etter første PCR og index-PCR, henholdsvis 8% og 19%. Ingen av prøvene ble utelukket fra videre analyse.

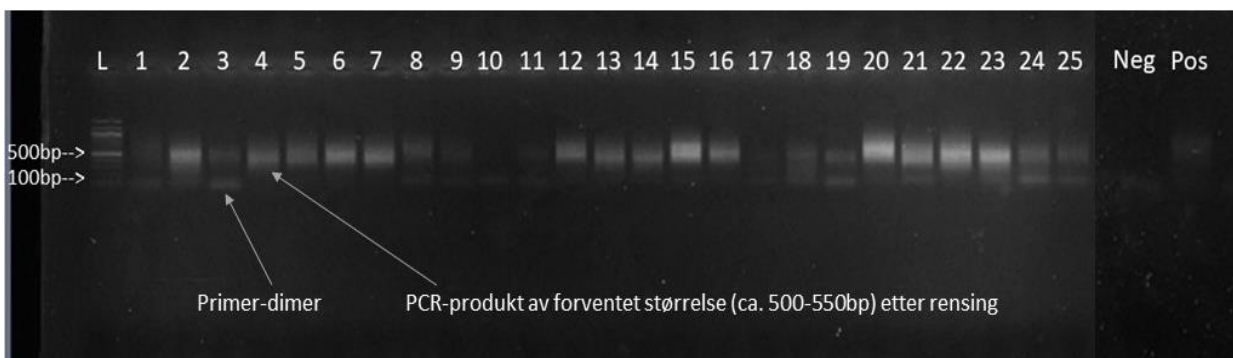
Et eksempel av gelresultatet fra tre måneders prøver (25/100 prøver) med tilhørende negativ (nukleasefritt vann) og positiv (*Escherichia coli*) kontroll fra første PCR og fra index-PCR er vist i henholdsvis figur 9 og figur 10. Fra figurene observeres det bånd av forventede størrelser på omtrent 500-550 bp for alle prøver. Det er ingen bånd i negativ kontroll til forskjell fra positiv kontroll som har et klart bånd. Begge PCR-reaksjonene viser tegn på «smearing». Index-PCR reaksjonen viser spesielt dannelsen av primer-dimer på omtrent 100 bp. Det er også tydelig at etter index-PCR er båndene sterkere. Her er også de minste fragmentene rensset vekk da det var et rensetrinn mellom første PCR og index-PCR.

Gelresultatet fra Mock-blandingene med tilhørende negativ kontroll fra første PCR og fra index-PCR, er gitt i henholdsvis figur 11 og figur 12. Her vises triplikater (a-c) av hver av de ti blandningene, inkludert negativ kontroll (nukleasefritt vann). Det observeres bånd av forventede

størrelser på omtrent 500-550 bp for alle prøver. Det er ingen bånd i negativ kontroll. Begge PCR-reaksjonene viser tegn på «smearing». Primer-dimer på omtrent 100 bp er også tydelig her etter index-PCR.

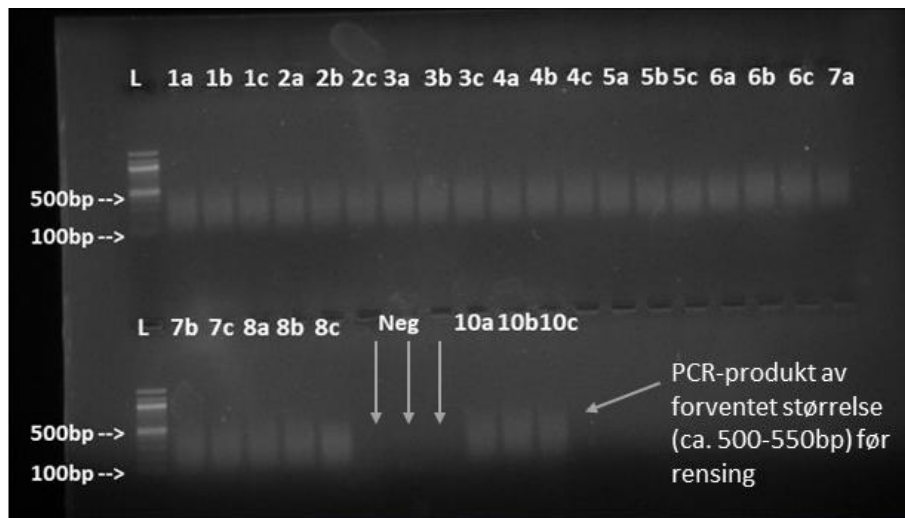


Figur 9. PCR-produkter (før rensing) av forventet størrelse på ca. 500-550 bp fra 3 mnd prøver (25/100 prøver) fra første PCR, visualisert på 1,5% agarose gelelektroforese, ved hjelp av Molecular Imager® Gel Dox™ XR Imaging System med Quantity One 1-D programvare v.4.6.7 (BioRad, USA), ved bruk av ultrafiolett lys. Helt til venstre vises 100bp DNA-ladder (L) og helt til høyre vises positiv (*Escherichia coli*) og negativ (nukleasefritt vann) kontroll.

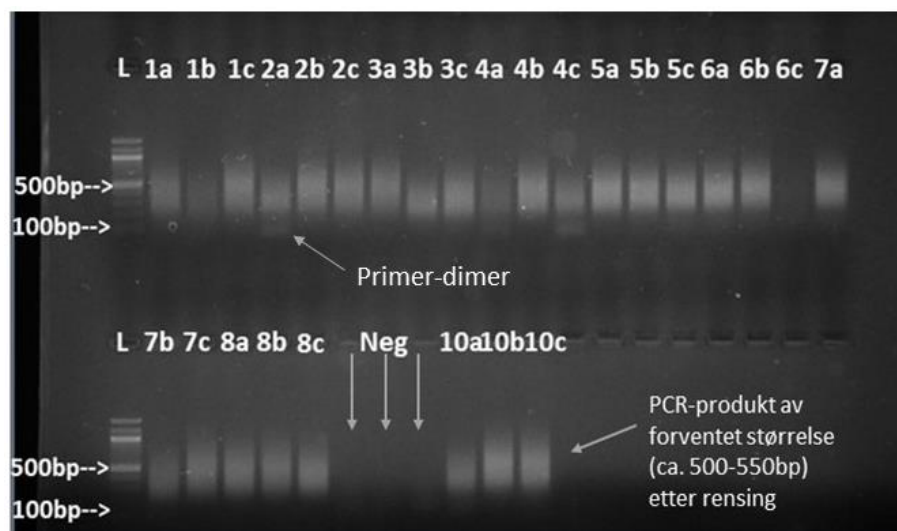


Figur 10. PCR-produkter (etter rensing) av forventet størrelse på ca. 500-550 bp fra 3 mnd prøver (25/100 prøver) fra index-PCR, visualisert på 1,5% agarose gelelektroforese ved hjelp av Molecular Imager® Gel Dox™ XR Imaging System med Quantity One 1-D programvare v.4.6.7 (BioRad, USA), ved bruk av ultrafiolett lys. Helt til venstre vises 100bp DNA-ladder (L) og helt til høyre vises positiv (*Escherichia coli*) og negativ (nukleasefritt vann) kontroll. Primer-dimer observert på omtrent 100 bp.





Figur 11. PCR-produkter (før rensing) fra Mock (blanding 1-10 i triplikater (a-c)), inkludert negativ kontroll (neg), fra første PCR, visualisert på 1,5% agarose gelelektroforese ved hjelp av Molecular Imager® Gel Dox™ XR Imaging System med Quantity One 1-D programvare v.4.6.7 (BioRad, USA), ved bruk av ultrafiolett lys. Helt til venstre vises 100bp DNA-ladder (L). Negativ kontroll (neg) besto av nukleasefritt vann.



Figur 12. PCR-produkter (etter rensing) fra Mock (blanding 1-10 i triplikater (a-c)), inkludert negativ kontroll (neg), fra index-PCR, visualisert på 1,5% agarose gelelektroforese ved hjelp av Molecular Imager® Gel Dox™ XR Imaging System med Quantity One 1-D programvare v.4.6.7 (BioRad, USA), ved bruk av ultrafiolett lys. Helt til venstre vises 100bp DNA-ladder (L). Ladder i nederste gel er litt utydelig ved 100bp. Negativ kontroll (neg) besto av nukleasefritt vann.

### **3.2 Illumina sekvensering av RMS-bibliotek**

Det ble først gjennomført en sekvensering på Illumina MiSeq (2x300 bp) som kontroll på at RMS-biblioteket var av tilstrekkelig kvalitet, før ytterligere dypere sekvensering på Illumina HiSeq 3000 (2x150 bp). Et totalt antall på 543 prøver, inkludert mor-barn prøver og Mock-prøver med tilhørende positive og negative kontroller ble sekvensert. Det ble fokusert på fragmenter i lengdeintervallet 50bp-500bp, da det er her man får tilnærmet 90% av signalet (=antall sekvenser).

#### **3.2.1 Illumina MiSeq sekvensering**

Fra sekvenseringen på Illumina MiSeq ble et totalt antall på 26 225 864 sekvenser av høy kvalitet produsert. Informasjon om kvaliteten på RMS-biblioteket fra Illumina MiSeq sekvenseringen er hentet fra BaseSpace Illumina, og er angitt i ulike diagrammer «Qscore Heatmap» og «Qscore distribution» som er illustrert i henholdsvis vedlegg 3 og vedlegg 4.

Fra «Qscore Heatmap» observeres det en høyere andel av sekvenser med lavere kvalitet (Qscore < 35) i starten av sekvenseringen, med etterfølgende økning i kvalitet (Qscore > 35). I tillegg observeres det en høyere andel av sekvenser med en høyere Qscore for «read 1» til forskjell fra «read 2», som forventet. Fra «Qscore distribution» observeres det at 65,9% av sekvensene har en Qscore > 30.

#### **3.2.2 Illumina HiSeq 3000 sekvensering**

Fra sekvenseringen på Illumina HiSeq 3000 ble et totalt antall på 266 800 000 sekvenser av høy kvalitet produsert. Sekvenskvaliteten er hentet fra «MultiQC», tilsendt fra Norges sekvenseringssenter, og er angitt i ulike diagrammer «Per Sequence Quality Scores» og «Mean Quality Scores» som er illustrert i henholdsvis vedlegg 5 og vedlegg 6.

Fra «Per Sequence Quality Scores» observeres det at majoriteten av sekvensene har en Phred Score/Qscore > 35. Fra «Mean Quality Scores» ser det ut til at de første syklusene med sekvensering er av lavere kvalitet (Phred Score < 30) til forskjell fra resterende sykluser, omtrent fra syklus 10, som er av høyere kvalitet (Phred Score > 30). Som forventet tyder det også på her at kvaliteten av sekvenseringen reduseres mot slutten.

### 3.3 Vurdering av Kraken2-databaser

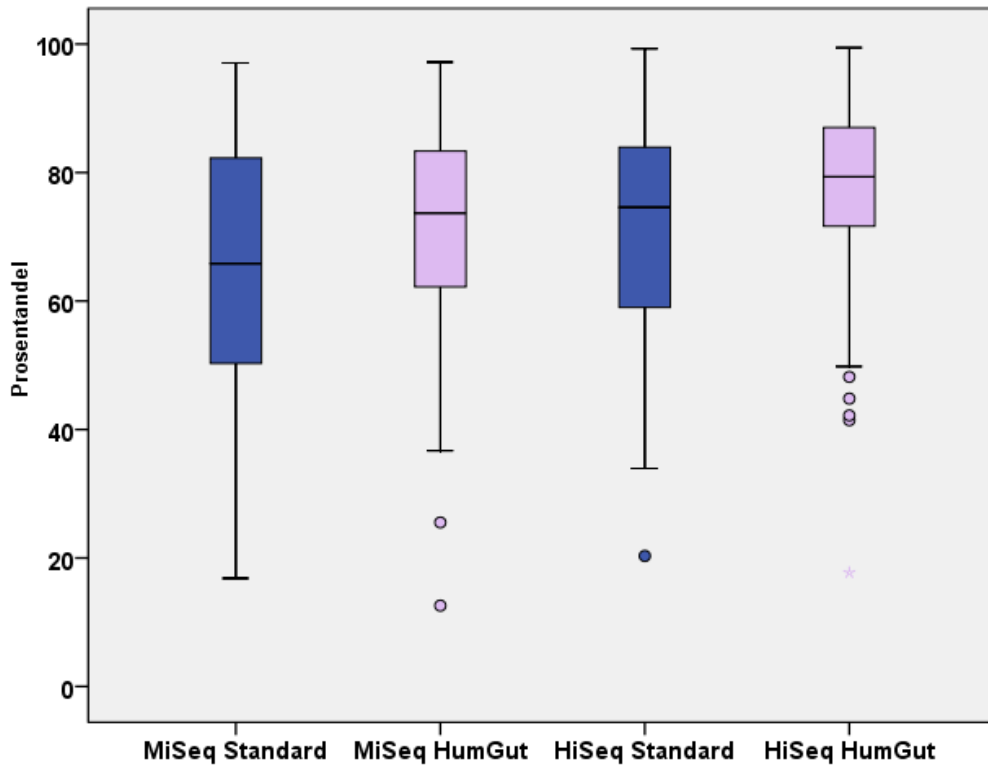
Det ble gjennomført en sammenligning av andel sekvenser som mappet til Kraken2 HumGut database og til Kraken2 Standard database, fra både MiSeq data og HiSeq data, som vist i figur 13. Dataene er presentert som median og interkvartilområde.

Fra figuren observeres det at en gjennomgående høyere andel av sekvenser fra MiSeq data mappet til Kraken2 HumGut databasen (median 73,67, interkvartilområde 62,19-83,36) til forskjell fra Kraken2 Standard database (median 65,81, interkvartilområde 50,26-82,28). Tilsvarende observasjon gjaldt også HiSeq data hvor en gjennomgående høyere andel av sekvenser mappet til Kraken2 HumGut databasen (median 79,36, interkvartilområde 71,60-87,15) til forskjell fra Kraken2 Standard databasen (median 74,60, interkvartilområde 59,01-83,95).

Kraken2 HumGut database inneholder 412 bakteriearter til forskjell fra Kraken2 Standard database som inneholder 3100 bakteriearter. Av de 412 bakterieartene som er tilgjengelige i Kraken2 HumGut database, så mappet bakteriesekvensene til et antall på 269 bakteriearter, i tillegg til *Homo sapiens*.

I denne studien er alle resultater videre basert på HiSeq-data med taksonomisk klassifisering ved bruk av Kraken2 HumGut database. Imidlertid ble den taksonomiske klassifiseringen av Mock-samfunn (Mock20even) gjennomført ved bruk av Kraken2 Standard database.

### Vurdering av Kraken2-databaser

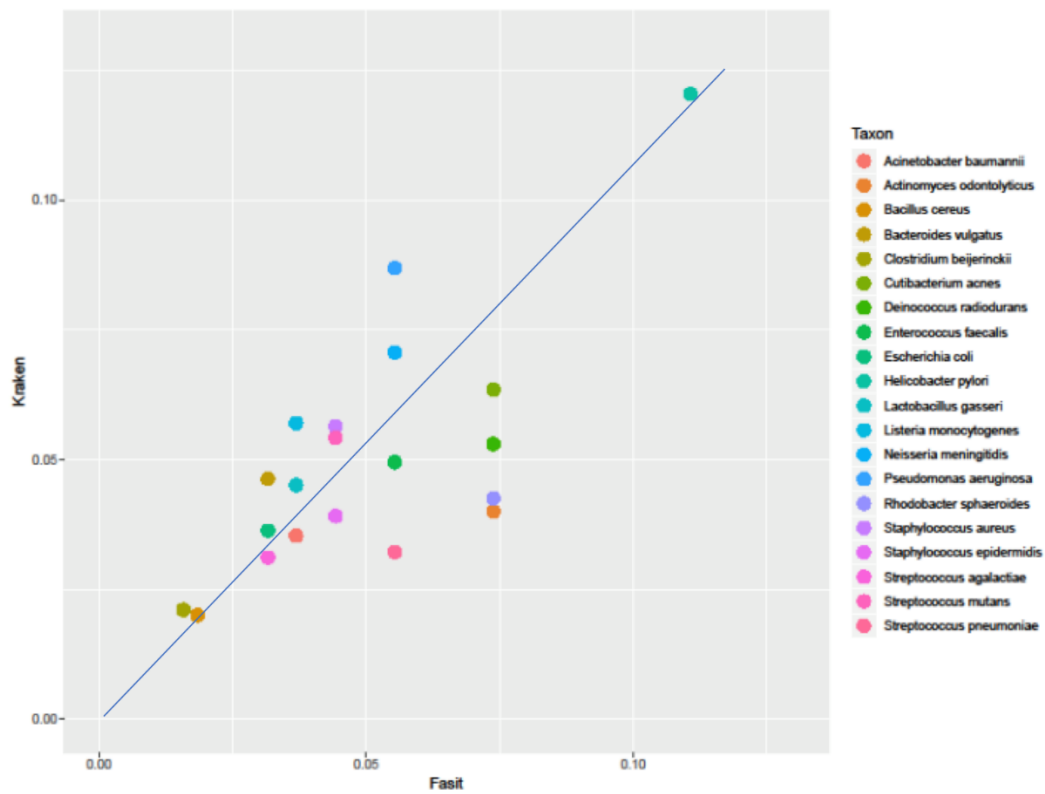


Figur 13. Vurdering av Kraken2-databaser. Illustrasjon av prosentandelen sekvenser (y-akse) som mappet til HumGut (lilla box)- og Standard (blå box) Kraken2 databasene (x-akse), fra både MiSeq og HiSeq data. Sirkler angir slengere fra 1,5-3,0 x Q1 (nedre kvartil).

### **3.4 RMS-analyse av Mock-samfunn**

For å evaluere kvaliteten til RMS-metoden i detektering av kjente bakteriearter, ble det gjennomført analyser av Mock-samfunn. Imidlertid klarte ikke Kraken2 klassifiseringen å trekke ut informasjon fra hver av de syv enkeltstammene som ble tilsatt til Mock20even (resultater ikke vist). Resultatene vist i figur 14 er dermed kun basert på analyser fra Mock20even.

For å teste hvorvidt det er korrelasjon mellom forventet og bestemt bakteriell sammensetning i Mock20even, ble det utført en Pearson-korrelasjonsanalyse (figur 14). Her vises den forventede relative andelen (fasit) av de 20 kjente bakterieartene tilstede i Mock20even, hvor deres antall 16S rRNA genkopier er tatt i betraktning, i forhold til den bestemte relative andelen av bakterier som et resultat av sekvenser produsert ved RMS-metoden, med bruk av HiSeq 3000 som sekvenseringsplattform. Det observeres en relativt god positiv korrelasjon mellom forventet og bestemt bakteriell sammensetning i Mock20even, med en Pearson-korrelasjonskoeffisient  $r = 0,75$ . En  $p$ -verdi  $< 0,001$  tilsier at korrelasjonen er signifikant. Bakteriearter som ligger under linjen er forholdsmessig mindre rikelig som et resultat av RMS-metoden enn det som er forventet, og motsatt.



Figur 14. **Korrelasjonsplott av Mock-samfunn (Mock20even).** Her vises den relative andelen av de 20 kjente bakterieartene tilstede i Mock20even. Her er den forventede relative andelen (fasit) av bakteriene hvor deres antall av 16S rRNA genkopier er tatt i betraktning (x-akse) plottet mot den bestemte relative andelen av bakterier som et resultat av sekvenser produsert ved RMS-metoden, med HiSeq 3000 som sekvenseringsplattform (y-akse).  $R = 0,75$ ,  $p < 0,001$ . Taksonomisk klassifisering av Mock-data ble utført ved bruk av Kraken2 Standard database.

### 3.5 RMS-analyse av mor-barn prøver

Det ble gjennomført preliminare statistiske analyser, Kruskal-Wallis test og Spearmans korrelasjonsanalyse, av det totale antallet på 269 bakteriearter som ble identifisert med Kraken2 HumGut databasen. Kruskal-Wallis testet for mor-barn assosierte tarmbakterier, mens Spearmans korrelasjonsanalyse testet for tarmbakterier som er korrelert med alder. På bakgrunn av statistisk signifikante resultat generert fra Kruskal-Wallis test, og Spearmans korrelasjonsanalyse, ble ulike bakteriearter selektert ut for videre analyse.

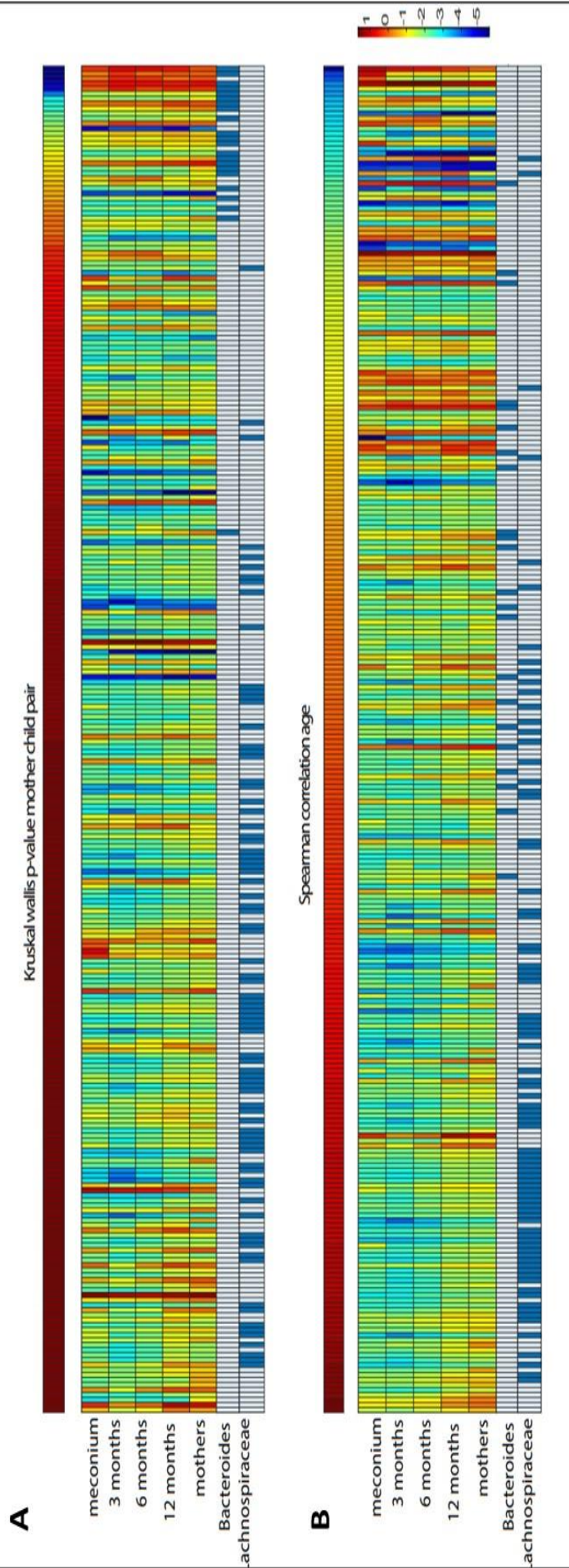
#### 3.5.1 Seleksjon av bakteriearter for videre analyse

Et antall på tolv bakteriearter (*Escherichia coli*, *Pelomonas puraquae*, *Acinetobacter parvus*, *Dorea longicatena*, *Coprococcus catus*, *Coprococcus comes*, *Bacteroides fragilis*, *Bacteroides thetaiotamicron*, *Parabacteroides distasonis*, *Bacteroides vulgatus*, *Bacteroides dorei*, *Bifidobacterium longum*) ble selektert ut for videre analyser. Fra figur 15 illustreres gjennomsnittsverdien for mengden av ulike bakterier for alle alderskategorier basert på p-verdier (Kruskal-Wallis) og korrelasjonskoeffisientens verdier (Spearmans korrelasjon).

Figurens panel A illustrerer resultatene fra Kruskal-Wallis test som ble utført for hver av de 269 bakteriene opp mot 100 mor-barn par for hver alderskategori (nyfødt/mekonium, tre mnd, seks mnd, tolv mnd). Resultater som ga lave p-verdier ble sortert ut for videre analyse, her *Bacteroides*, som dermed antas å være mor-barn assosiert. Motsatt observeres det en mye høyere p-verdi for *Lachnospiraceae*, noe som indikerer at den ikke er mor-barn assosiert.

Figurens panel B illustrerer resultatene fra Spearmans korrelasjonsanalyse som også ble utført for hver av de 269 bakteriene opp mot 100 mor-barn par for hver alderskategori (nyfødt/mekonium, tre mnd, seks mnd, tolv mnd). Det observeres at mengden av *Bacteroides* er her jevnere fordelt blant de ulike alderskategoriene, mens *Lachnospiraceae* viser en sterk positiv korrelasjon med alder, noe som indikerer at den er høyest tilstede hos mødre.

Den taksonomiske rangeringen (rekke, klasse, familie, slekt og art) av de ulike bakteriene er gitt i vedlegg 7.



Figur 15. **Kruskal-Wallis test (A) - og Spearman's korrelasjonsanalyser (B) av artsdata generert fra Kraken2 HumGut databasen.** Panel A: Det ble først utført Kruskal-Wallis test for hver bakterie opp mot 100 mor-barn par for hver alderskategori (nyfødt/meconium, tre mnd, seks mnd, tolv mnd). Resultater som ga lave p-verdier ble så sortert ut (mor-barn assosierede tarmbakterier). I panel A illustreres gjennomsnittsverdien for mengden av ulike bakterier (hver strek representerer en spesifikk bakterie, mens fargen representerer mengden hvor mørkeblå representerer mest). Skalaen på toppen, derimot, representerer styrken på p-verdien (log10) (mørkeblå er lavest p-verdi og høyere p-verdi jo lenger mot venstre i skalaen). Fra «Bacteroides» skalaen vises den laveste p-verdi (mørkeblå helt til høyre) noe som indikerer at den er høyt tilstede hos både mor og barn (mor-barn assosiert). Lachnospiraceae er jevnt fordelt og p-verdien her er høyere (mer mot venstre på skalaen) noe som indikerer at den ikke er mor-barn assosiert. Panel B: Det ble først utført Spearman's korrelasjonsanalyse for hver bakterie opp mot 100 mor-barn par for hver alderskategori (nyfødt/meconium, tre mnd, seks mnd, tolv mnd). Skalaen på toppen representerer korrelasjonskoeffisientens verdi (mørkeblå er lavest verdi og høyere verdi jo lenger mot venstre). Bacteroides er jevnt fordelt blant de ulike alderskategoriene, mens Lachnospiraceae viser en sterk positiv korrelasjon med alder (plassert godt til venstre i skalaen) noe som indikerer at den er høyt tilstede hos mødre. Analysene og figuren basert på det er utført av professor Knut Rudi (veileder).



### 3.5.2 Den relative mengden av bakteriearter hos spedbarn og mødre

Den relative mengden av bakteriearter hos spedbarn og mødre er vist i figur 16, figur 17 og figur 18 som en log<sub>10</sub> transformert prosentandel. Dataene er presentert som median og interkvartilområde.

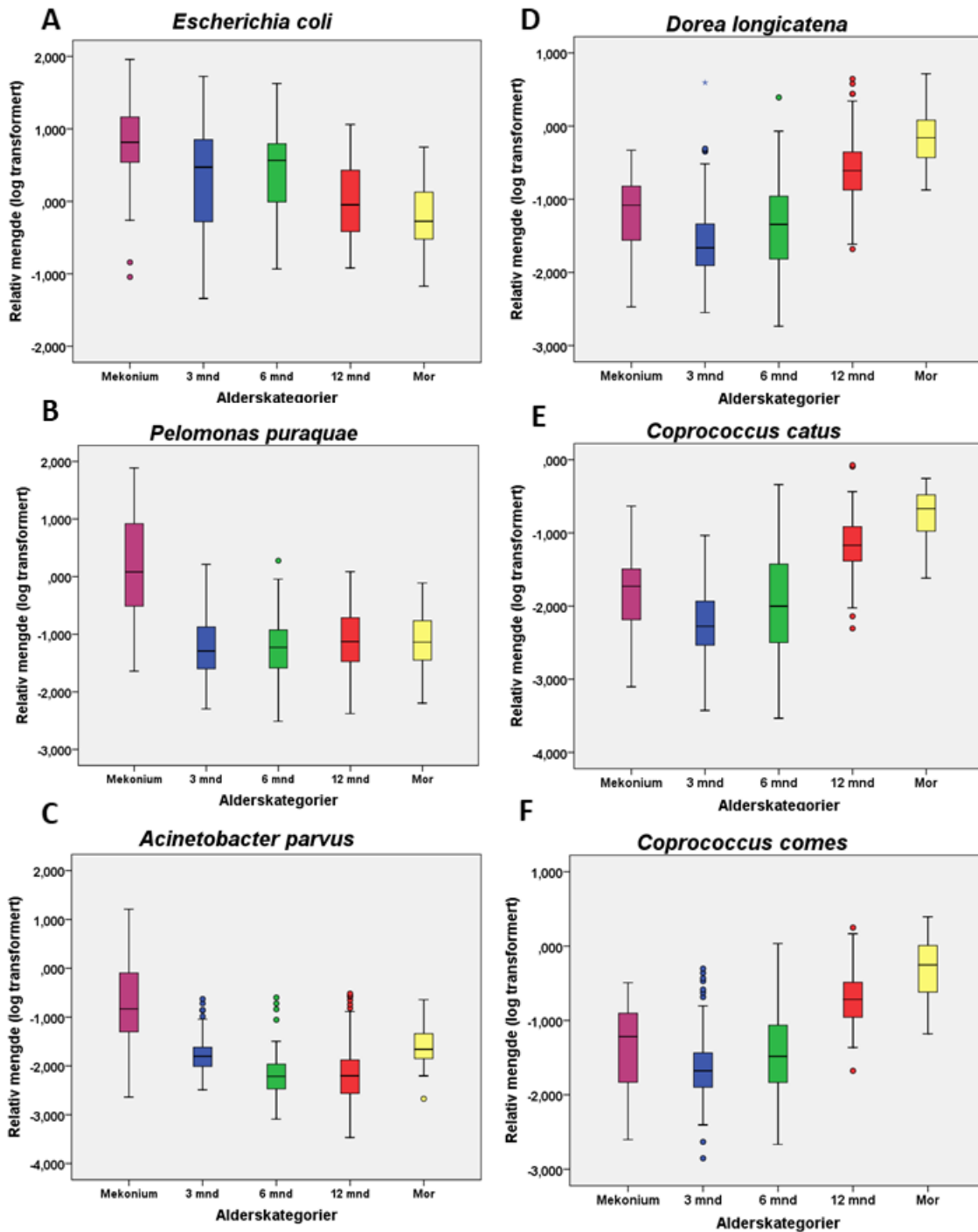
For å undersøke om forskjellene i den relative mengden av bakteriearter mellom mødre og barn i ulike alderskategorier er statistisk signifikante, ble det gjennomført Kruskal-Wallis test. Siden Kruskal-Wallis test ikke oppdager mellom hvilke mor-barn alderskategorier de potensielle forskjellene i mengden bakterien forekommer, ble det videre utført Dunn's Post Hoc test for parvise sammenligninger, med Bonferroni korreksjon av p-verdier.

#### Bakteriearter assosiert med alder

Fra figur 16 observeres det tydelig at tre av bakterieartene (*E.coli*, *P.puraquae* og *A.parvus*) er negativt assosiert med alder (synker med alder), mens resterende bakteriearter (*D.longicatena*, *C.catus* og *C.comes*) er positivt assosiert med alder (stiger med alder). Den relative mengden av for eksempel *E.coli* er lavest hos mødre (median -0,28, interkvartilområde -0,52-0,13), til forskjell fra mekonium hvor den relative mengden er størst (median 0,81, interkvartilområde 0,54-1,17). Motsatt er den relative mengden av for eksempel *D.longicatena* størst hos mødre (median -0,16, interkvartilområde -0,44-0,08) til forskjell fra tre måneder hvor den er lavest (median -1,66, interkvartilområde -1,91- -1,33), og interessant nok lavere enn mekonium (median -1,08, interkvartilområde -1,57- -0,82). Tolkningen er tilsvarende for resterende bakteriearter som visuelt synker med alder (*P.puraquae* og *A.parvus*) og øker med alder (*C.catus* og *C.comes*).

Alle Bonferroni-korrigerede p-verdier < 0,05, bortsett fra den parvise sammenligningen av *D. longicatena*, *C. catus* og *C. comes* mellom 3 mnd og 6 mnd, i tillegg til utvalgte parvise sammenligninger av *E.coli*, *P. puraquae* og *A. parvus*.

Alle Bonferroni korrigerede p-verdier for de parvise sammenligningene av den relative mengden av de ulike bakterieartene er vist i vedlegg 8.

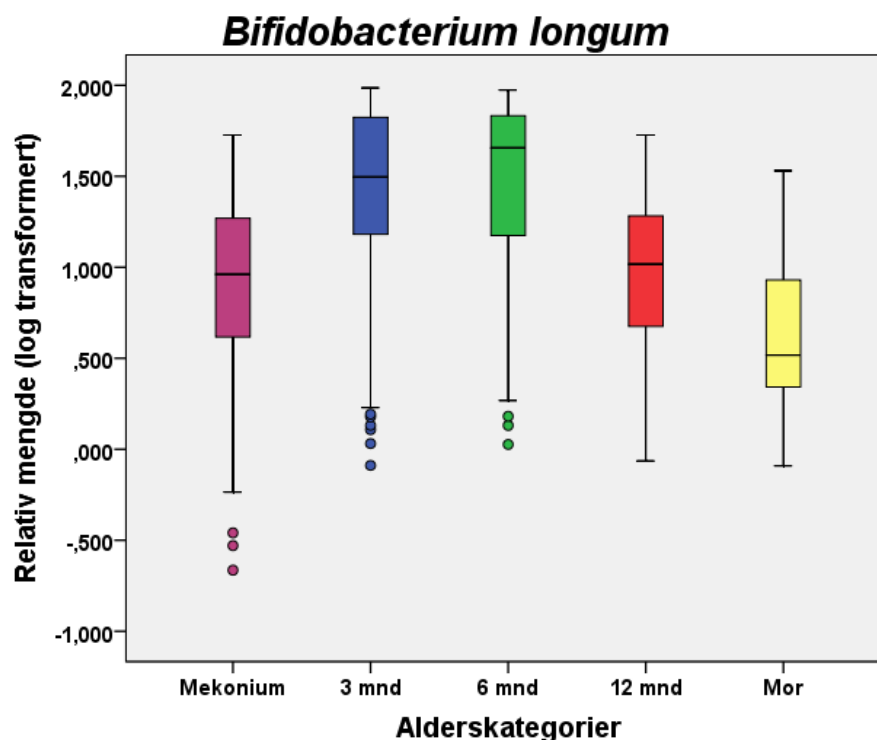


Figur 16. Relativ mengde ( $\log_{10}$ ) av ulike bakteriearter (*E. coli*, *P. puraquae*, *A. parvus*, *D. longicatena*, *C. catus* og *C. comes*) som synker (A, B, C) og stiger (D, E, F) med alder (nyfødt/mekonium, tre mnd, seks mnd, tolv mnd og mødre, illustrert med ulike farger). Dataene er presentert som median og interkvartilområde ( $Q1 \rightarrow Q3$ ). Sirkler og stjerner angir slengere fra henholdsvis  $1,5\text{--}3,0 \times Q1$  (nedre kvartil) eller  $Q3$  (øvre kvartil), og over  $3,0 \times Q1$  (nedre kvartil) eller  $Q3$  (øvre kvartil). Signifikans ble estimert ved bruk av Kruskal-Wallis test med Dunn's PostHoc test for parvise sammenligninger, og Bonferroni-korreksjon av p-verdier. Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

### ***Bifidobacterium longum* høyt representert i spedbarns tarmmikrobiota**

Fra figur 17 observeres det at den relative mengden av *Bifidobacterium longum* øker signifikant fra mekonium (median 0,96, interkvartilområde 0,61-1,27) til tre måneder (median 1,50, interkvartilområde 1,18-1,83,  $p < 0,05$ ) og seks måneder (median 1,66, interkvartilområde 1,17-1,83,  $p < 0,05$ ). Den relative mengden blir så signifikant redusert fra seks måneder til tolv måneder (median 1,02, interkvartilområde 0,67-1,30,  $p < 0,05$ ). *Bifidobacterium longum* er tydelig lavest representert hos mødre (median 0,52, interkvartilområde 0,34-0,94).

Alle Bonferroni korrigerede p-verdier for de parvise sammenligningene av den relative mengden av *Bifidobacterium longum* er vist i vedlegg 8.



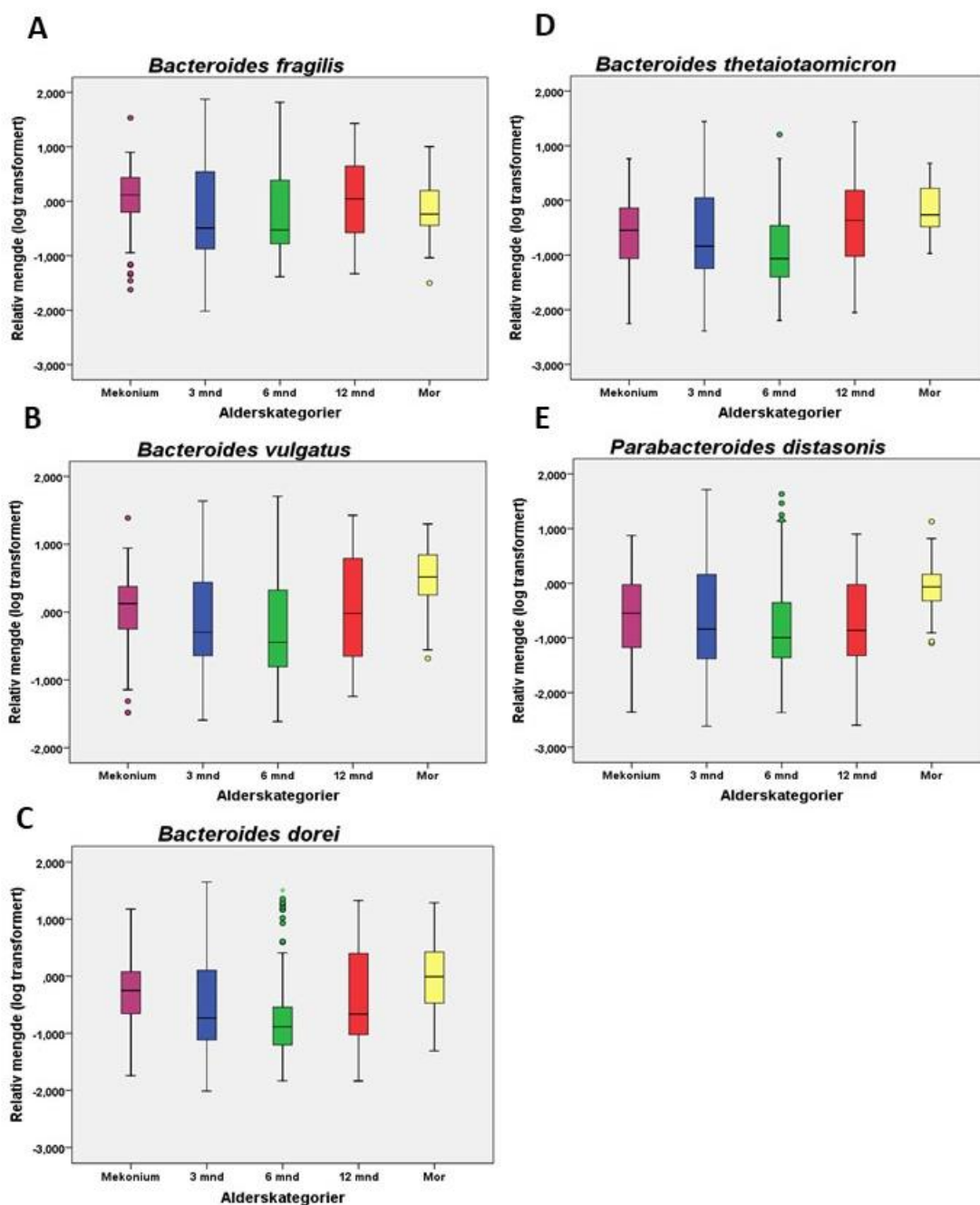
Figur 17. Relativ mengde ( $\log_{10}$ ) av *Bifidobacterium longum* mellom mor og barn i ulike alderskategorier (nyfødt/mekonium, tre mnd, seks mnd, tolv mnd, illustrert med ulike farger). Dataene er presentert som median og interkvartilområde (Q1- $\rightarrow$ -Q3). Sirkler angir slengere fra 1,5-3,0 x Q1 (nedre kvartil). Signifikans ble estimert ved bruk av Kruskal-Wallis test med Dunn's PostHoc test for parvise sammenligninger, og Bonferroni-korreksjon av p-verdier. Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

### **Mor-barn assosierte bakterier**

Fra figur 18 observeres det at *B.fragilis*, *B.thetaiotaomicron*, *P.distasonis*, *B.vulgatus* og *B.dorei* er rikelig tilstede hos både mor og barn. Imidlertid observeres det at bakteriene er høyest representert i mekonium/nyfødt blant de ulike barns alderskategorier. Den relative mengden av for eksempel *B. fragilis* er høyest i mekonium (median 0,18, interkvartilområde -0,22-0,44) og lavest i seks måneder (median -0,53, interkvartilområde -0,78-0,42). Det observeres også en økning av den relative mengden av alle arter bortsett fra *P. distasonis* fra seks mnd til tolv mnd. I tillegg observeres det en økning av den relative mengden av alle arter bortsett fra *B.fragilis* fra 12 mnd til mødre.

Visuelt sett ser det ut til at den relative mengden av bakteriene her er mer stabil mellom mor og barn i ulike alderskategorier. Det er likevel et relativt kompleks mønster av signifikante og ikke-signifikante resultater fra de parvise sammenligningene av den relative mengden av bakterier mellom mor og barn (se vedlegg 8). For eksempel var alle mor-barn parvise sammenligninger av den relative mengden *B. vulgatus* og *P. distasonis* signifikant ( $p < 0,05$ ), mens alle mor-barn parvise sammenligninger av den relative mengden av *B. fragilis* var ikke signifikant ( $p > 0,05$ ). Når det gjelder parvise sammenligninger av den relative mengden av bakterier mellom barn i ulike alderskategorier, var majoriteten av resultatene ikke-signifikante ( $p > 0,05$ ).

Alle Bonferroni korrigerede p-verdier for de parvise sammenligningene av den relative mengden av de ulike artene fra slekten *Bacteroides* er vist i vedlegg 8.



Figur 18. Relativ mengde (log<sub>10</sub>) av ulike bakteriearter (*B. fragilis*, *B. vulgatus*, *B. dorei*, *B. thetaiotaomicron*, *P. distasonis* og *B. longum*) mellom mor og barn i ulike alderskategorier (nyfødt/mekonium, tre mnd, seks mnd, tolv mnd, illustrert med ulike farger). Dataene er presentert som median og interkvartilområde (Q1->Q3). Sirkler og stjerner angir slengere fra henholdsvis 1,5-3,0 x Q1 (nedre kvartil) eller Q3 (øvre kvartil), og over 3,0 x Q1 (nedre kvartil) eller Q3 (øvre kvartil). Signifikans ble estimert ved bruk av Kruskal-Wallis test med Dunn's PostHoc test for parvise sammenligninger, og Bonferroni-korreksjon av p-verdier. Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

## 4. Diskusjon

### 4.1 RMS-analyse av mor-barn prøver

#### 4.1.1 *Proteobacteria* og *Actinobacteria* dominerer tarmmikrobiota signifikant i ung alder

Det som karakteriserer et spedbarns tarmmikrobiota er typisk et begrenset mangfold av bakterier som hovedsakelig domineres av *Proteobacteria* og *Actinobacteria*, med etterfølgende fremvekst og dominans av *Firmicutes* og *Bacteroidetes* (5). Dette ble også observert i studien. *Proteobacteria* var signifikant negativt assosiert med alder, mens *Firmicutes* var signifikant positivt assosiert med alder. Dette er en indikasjon på at tarmkoloniseringen med *Proteobacteria* er typisk vanlig i ung alder, til forskjell fra tarmkoloniseringen med *Firmicutes* som dermed er mer vanlig på et senere tidspunkt, noe som er i overensstemmelse med tidligere studier (26, 30). En ytterligere grunn til at den relative mengden av medlemmer av *Proteobacteria* var signifikant høyere hos spedbarn enn hos mødre, kan skyldes et lavere nivå av artsdiversitet hos førstnevnte, og dermed mindre «konkurransen» i tarmen i forhold til nisjeokkupasjon (5).

En tidligere dyrkningsbasert studie (37) identifiserte arten *E.coli* fra rekken *Proteobacteria* til å være en av de først representerte bakteriene i tarmen hos friske spedbarn. På den måten kan *E.coli*, som er en fakultativ anaerob bakterie, produsere et mer oksygenfattig miljø som er gunstig for etablering av de seinere forekommende bakteriene (*Bacteroidetes*, *Actinobacteria* og *Firmicutes*), som karakteriseres som strikt anaerobe (37). Det er dermed ikke overraskende med den signifikante reduksjonen i den relative mengden av *Proteobacteria* og den signifikante økningen av *Firmicutes* og *Actinobacteria*.

#### **4.1.2 Direkte mor-barn overføring av *E. coli*?**

*E. coli* var signifikant høyest representert i mekonium på likhet med det som tidligere har blitt observert (75). Dette kan tyde på mor-barn overføring av fekale bakterier, ettersom avføring under fødsel ikke er uvanlig. Tidligere studier (30, 76) har ved bruk av dyrkningsuavhengige metoder rapportert stammer av *E. coli* som deles mellom mor og barn, noe som styrker påstanden om at bakterien kan ha blitt overført fra mor til barn. Imidlertid betyr ikke det at bakterien er mor-barn assosiert, noe som blant annet kan begrunnes med den signifikante reduksjonen av *E. coli* med alder. Det kan tenkes at mor-barn assosierte bakterier vil kunne være rikelig tilstede hos både mor og barn over lenger tid (77). Istedenfor ser det ut til at *E. coli* blir utkonkurrert av andre bakterier. Imidlertid er det ingen kjennskap til mødrenes diett, hvilket kan ha en påvirkning på mor-barn overføringen av bakterier (78).

#### **4.1.3 *Bifidobacterium longum* høyest representert ved seks måneders alder**

*Bifidobacterium* er kjent for å være den dominerende anaerobe tarmbakterien hos ammende spedbarn (5, 79). I denne studien var arten *Bifidobacterium longum* høyest representert ved seks måneders alder. Den signifikante økningen av *Bifidobacterium longum* fra mekonium og opp til seks måneders alder fremhever den positive rollen til morsmelk i utforming av det *Bifidobakterielle* samfunnet, som blant annet spiller en viktig rolle i utviklingen og modningen av spedbarnets immunsystem (80). Denne signifikante økningen støttes også av tilstedeværelsen av de prebiotiske humane melkeoligosakkaridene (HMO'er), som stimulerer veksten av *Bifidobacterium* (81). En høy tilstedeværelse av *Bifidobacterium longum* hos spedbarn har også blitt rapportert av tidligere studier (79, 82, 83). I likhet med andre studier (79, 82, 83), har denne studien fremhevet den viktige rollen til *Bifidobacterium* i utformingen av tarmmikrobiota på et tidlig stadium i livet.

I overgangen fra morsmelk til fast føde er det kjent at mengden *Bifidobacterium* reduseres (84). Dette kan være en årsak til den signifikante reduksjonen av *Bifidobacterium longum* fra seks til tolv måneders alder. På dette stadiet var det en tilsvarende signifikant økning av *Firmicutes* og *Bacteroidetes* fra seks til tolv måneders alder, noe som er i overensstemmelse med tidligere studier (79). Mange av de seinere forekommende bakteriene er effektive i nedbrytning av kostfibre, og er produsenter av kortkjedede fettsyrer (8, 9). Det kan derfor tyde på at det skjer et skifte mot et mer voksenassosiert tarmmiljø, som har mer kapasitet til nedbrytning av karbohydrater.

Når tarmen introduseres for nye næringsstoffer, kan det derfor tyde på at den intraindividuelle diversiteten øker som følge av et mikrobielt skift, mens den interindividuelle diversiteten

reduseres, ettersom tarmen blir mer voksenassosiert og mindre heterogen, noe som også har blitt observert tidligere (34, 85). Imidlertid forblir det ukjent om det mikrobielle skiftet skyldes opphør av amming, innføring av fast føde eller en kombinasjon av dem. Fra en tidligere studie (30) er det bevis på at opphør av amming har størst medvirkning til seleksjonen av bakterier. Fra studien (30) hadde innføring av fast føde en medvirkning til det mikrobielle skiftet, men skiftet oppsto ikke før amming ble fraværende. Det kan dermed være en indikasjon på at det er fravær av amming, fremfor innføring av fast føde, som er årsaken til den signifikante reduksjonen av *Bifidobacterium longum*, og dermed skiftet mot en mer voksenassosiert tarmmikrobiota. Samlet tyder resultatene på at tarmmikrobiotas sammensetning utvikler seg med alder som respons på miljøfaktorer, i dette tilfellet diettendring.

#### **4.1.4 Mulig kolonisering av spedbarnstarm før fødsel**

Interessant nok var alle de tre artene (*D. longicatena*, *C. catus* og *C. comes*) fra rekken *Firmicutes*, signifikant høyere i mekonium sammenlignet med tre måneder. Mekonium antas å reflektere det intrauterine miljøet før fødsel (29). Det kan dermed være en indikasjon på at det intrauterine miljøet inneholder bakterielle samfunn som kan tilrettelegge for videre tarmkolonisering. Før fødsel kan det være vanlig at spedbarnet blir utsatt for anaerobe forhold i utero, mens ved fødselen er spedbarnets tarmmikrobiota for det meste fakultativt anaerobt (5). Dette kan være en årsak til at *Firmicutes* blir signifikant redusert fra mekonium til tre måneders alder. Det kan tenkes at tilstedeværelsen av *Firmicutes* forblir slik inntil vekstforholdene blir optimale (79), typisk mellom seks og tolv måneders alder hvor det ble observert en signifikant økning.

Kilden til bakteriene i mekonium forblir uklar, men kan være av intrauterin opprinnelse som et resultat av svelget fostervann (86) . Imidlertid trengs det ytterlige prøver fra placenta og fostervann for å kunne bekrefte tilstedeværelsen av de tilsvarende bakteriene detektert i mekonium. Med andre ord er det ikke nok bevis for å støtte ideen om «i utero» kolonisering. En tidligere studie (86) fant spesifikke de novo OTU'er, tildelt klassen *Clostridia* fra rekken *Firmicutes*, i mekonium og tilsvarende i placenta og fostervann, fra mor-spedbarns par, noe som støtter hypotesen om «i utero» kolonisering. En annen nyere studie rapporterte at den første koloniseringen av spedbarnstarmen inntreffer så snart fostermembranen brytes ned (28).

Det har også tidligere blitt observert at medlemmer av *Firmicutes* overføres fra morsmelk og til spedbarnets tarm (87). Det kan derfor være en mulighet at nyfødte har blitt ammet på et tidspunkt før mekoniumsprøvene ble tatt, men dette forlir uklart ettersom informasjon tilknyttet



tidspunkt for prøvetakning og amming er ukjent i denne studien. For eventuelt å bekrefte en slik overføring kreves det ytterligere analyser på stammenivå og analyser av morsmelk.

#### **4.1.5 *Bacteroides* - mor-barn assosiert**

I denne studien ble det observert en jevn relativ mengde av *Bacteroides* blant spedbarn i ulike alderskategorier, inkludert nyfødt/mekonium, noe som indikerer at *Bacteroides* stabilt koloniserer spedbarnets tarm. *Bacteroides* var også rikelig tilstede hos mødre, noe som tyder på at bakterien er mor-barn assosiert, på likhet med det som tidligere har blitt observert (38). Det kan tenkes at mor-barn assosierte bakterier tilpasser seg bedre i tarmen hos spedbarn over lengre tid, sammenlignet med kolonisering av bakterier avledet fra miljøet og andre områder (77). Dette kan være en mulig forklaring på den stabile koloniseringen av *Bacteroides* over tid.

Tidligere studier (88, 89) har rapportert en begrenset mengde av *Bacteroides* i avføringsprøver fra spedbarn født ved keisersnitt, noe som er en ytterligere indikasjon på at det er en sterk assosiasjon av *Bacteroides* mellom mødre og barn. Imidlertid forblir det uvisst om bakterien er direkte overført fra mor til barn, men tilstedeværelsen av *Bacteroides* i mekonium kan være en indikasjon på dette. Det har tidligere blitt vist at artene *Bacteroides fragilis* og *Bacteroides thetaiotamicron* blir direkte overført fra mor til barn (30). I tilfeller hvor ulike individer deler samme arter, har det blitt vist at artene består av ulike stammer (44, 90). Det er derfor behov for ytterligere analyser på stammenivå for å bekrefte en eventuell mor-barn bakterieoverføring.

#### **4.1.6 Utfører *Bifidobacterium* en nisjeokkupasjon?**

Det som også er interessant er at når mengden av *Bifidobacterium* etter seks måneder signifikant synker, så øker signifikant mengden av *Bacteroides*, noe som har også blitt observert tidligere (91). En årsak til det kan være at *Bifidobacterium* utfører en nisjeokkupasjon som fører til en signifikant nedgang i mengden av *Bacteroides*. *Bacteroides* er kjent for å være en ytterligere viktig bidragsyter til nedbrytningen av HMO'er (92) tilsvarende *Bifidobacterium*.

## **4.2 Metodiske betraktninger**

### **4.2.1 Tillaging av RMS-bibliotek**

Den ulike andelen av mor-barn prøver som fikk bånd på gel kan først og fremst være en indikasjon på at det i utgangspunktet er en varierende mengde av DNA hos mødre og spedbarn i ulike alderskategorier, spesielt nyfødt/mekonium (93), som er kjent for å ha en redusert mengde av DNA. Det var dermed forventet at en lavere andel av prøver fra mekonium skulle få bånd til forskjell fra for eksempel prøver fra tolv måneder. Imidlertid kan man heller ikke forvente å få like mange eller like gode utslag fra prøvene på gel da prøvene i utgangspunktet ikke ble normalisert for mengden av genomisk DNA etter kvantifiseringen på Qubit.

Det ble også observert en reduksjon i andelen av mor-barn prøver som fikk bånd etter index-PCR. Da det var et rensetrinn mellom første PCR og index-PCR kan det trolig skyldes at en del fragmenter har blitt rensset vekk. Dette var også grunnen til at rensingen av RMS-biblioteket seinere ble utført manuelt ved bruk av AMPure® XP kuler i forhold 2:1 av volum kuler og DNA, istedenfor forhold 1:1 som i utgangspunktet ble benyttet. En større mengde av AMPure® XP kuler vil sannsynligvis kun rense vekk de minste fragmentene. Grunnen til at rensingen etter hvert ble utført manuelt var på grunn av problemer med instrumentet Biomek 3000, noe som også kan være en ytterligere grunn til den observerte reduksjonen i andelen prøver som fikk bånd etter index-PCR. Samlet var utvalget av mor-barn prøver med utslag på gel likevel representativt nok for videre analyser. Fragmenter av forventet størrelse på 500-550 bp ble bekreftet og det var ingen utslag på negativ kontroll - på den måten kunne forurensning fra blant annet reagenser utelukkes. Tilsvarende observasjoner med hensyn til fragment av forventet størrelse og ingen utslag på negativ kontroll gjaldt også Mock-prøvene.

### **4.2.2 Validering av RMS-bibliotek og sekvensering**

Sekvenseringen av RMS-biblioteket på Illumina MiSeq bekreftet en relativt god kvalitet av RMS-biblioteket. Det ble observert at 65,9% av sekvensene hadde en Qscore > 30. Det vil si at det er mer enn 99,9% sannsynlighet for at riktig base har blitt avlest i forhold til dens signalintensitet som detekteres under prosessen med sekvensering.

Fra den ytterligere dypere sekvenseringen av RMS-biblioteket på Illumina HiSeq 3000 ble det her også observert en svært god kvalitet. Majoriteten av målingene hadde en PhredScore/Qscore > 35, men den eksakte andelen av sekvenser med tilhørende Qscore var ukjent.

Det ble observert at kvaliteten var lavere helt i starten av sekvenseringen på både MiSeq og HiSeq 3000. Det kan muligens skyldes bobler som kan passere gjennom flowcellen og føre til bakgrunnsstøy. Det viktigste er at kvaliteten blir forbedret og holder seg stabil utover

sekvenseringen, noe som i dette tilfellet ble observert for både sekvenseringen på MiSeq og HiSeq. Likevel ble også kvaliteten redusert mot slutten av sekvenseringen som forventet, da det blant annet blir begrenset med reagenser, DNA-polymerase etc., som dermed kan ha en påvirkning på kvaliteten av videre sekvensering.

#### **4.2.3 Vurdering av Kraken2-databaser**

Den gjennomgående høyere andelen av sekvenser som mappet til Kraken2 HumGut databasen til forskjell fra Kraken2 Standard databasen fra både MiSeq og HiSeq data indikerer at førstnevnte har et bedre utvalg av genomer. I og med at Kraken2 HumGut database inneholder taxa som er vanlige å finne i human tarm, til forskjell fra Kraken2 Standard database som har mer bredde og dermed litt uspesifikk, ble antall ulike detekterte taxa redusert fra 3100 til 269. Det som også er viktig å bemerke seg er at de 269 detekterte artene er bare et resultat av sekvensene som faktisk matchet til sekvensene tilgjengelige i Kraken2 HumGut databasen. Det betyr ikke nødvendigvis at det ikke finnes ytterligere arter i prøven, men heller at artene ikke er tilgjengelige i Kraken2 HumGut databasen.

Til tross for at Kraken2 HumGut databasen er i stand til å klassifisere bakteriene helt ned på stammenivå, ble det i denne studien valgt å operere på artsnivå. Det er mange grunner til at det er vanskelig med klassifisering på stammenivå. To eller flere stammer fra samme art vil sannsynligvis dele mange fragmenter. Hvis den ene stammen er tilstede i prøven, kan man få signal på de like fragmentene fra resterende stammer, noe som kan være vanskelig å skille de ulike stammene fra hverandre. Det forblir dermed vanskelig å si hvilken stamme som faktisk er tilstede i prøven. Et annet og viktig poeng er at Kraken2 bruker NCBI-taksonomi databasen for angivelse av taksonomi (67). Det gis ikke lenger taksonomisk identifikator til stammer, men bare ned til art. Grunnen til det er at det er for mange stammer til at alle skal få sin egen identifikator. En nylig studie (94) som undersøkte ulike system for klassifisering av stammer, rapporterte at blant annet Kraken2 ikke var i stand til å identifisere den korrekte stammen.

For å få taksonomisk informasjon på stammenivå er det dermed nødvendig med bruk av et annet klassifiseringssystem enn Kraken2. Imidlertid har Krakens2 store fordeler med hensyn til dens kommandoer i opprettelsen av egne databaser tydelig blitt belyst i studien, og dens evne til å klassifisere bakterier på artsnivå.

#### 4.2.4 RMS-analyse av Mock-samfunn

Det var en signifikant positiv korrelasjon mellom forventet og bestemt bakteriell sammensetning i Mock-samfunn (Mock20even). En Pearson-korrelasjonskoeffisient,  $r = 0,75$ , tilsier at korrelasjonen eller sammenhengen er relativt bra. Dette tyder på at RMS-metoden er relativt pålitelig i detektering av forventede bakteriearter.

Ideelt sett så skulle andelen mellom forventet og bestemt bakteriell sammensetning vært den samme, noe som ikke er tilfellet her da punktene i grafen (figur 14) ikke ligger langs linjen diagonalt. Årsakene til avvikene kan være mange. Fasiten/forventet er fortsatt noe usikker da antall 16S rRNA genkopier i en bakteriecelle kan variere mellom bakteriearter- og stammer (95). Med andre ord ville man utelukket bruk av 16S rRNA gensekvensering alene for blant annet artsidentifikasjon på grunn av intragenomisk heterogenitet av 16S rRNA genet (95).

I tillegg har selve RMS-metoden en svakhet i det ulike antallet sekvenser som produseres fra de ulike RMS-fragmentene som forekommer fra samme genom. Noen RMS-fragmenter fra samme genom har alltid veldig få sekvenser, mens andre har alltid svært mange sekvenser. Den eksakte grunnen til dette er ukjent, men kan eventuelt skyldes restriksjonskuttingen. I og med at en slik observasjon ble gjentatt med små endringer fra prøve til prøve, er det ikke en grunn til å tro at det er en tilfeldig variasjon. Likevel var kvaliteten av sekvensdata relativt god, noe som skyldes et stort antall genererte RMS-fragmenter fra hvert genom. Til tross for at noen RMS-fragmenter ga mange sekvenser, mens andre ga få sekvenser, var gjennomsnittet likevel akseptabelt.

I studien var det også ønskelig å se om RMS-metoden klarte å skille stammer fra samme art fra hverandre. I dette tilfellet gjelder det de ulike blandningene som beskrevet i 2.2.2. Imidlertid klarte ikke Kraken2 å hente ut informasjon fra hver av de syv enkeltstammene som ble tilsatt til Mock20even på bakgrunn av det som tidligere er nevnt om Kraken2. Til tross for at selve RMS-metoden i seg selv er i stand til å generere et pålitelig bibliotek som er representativt for sporing av bakterier på stammenivå, er det likevel avgjørende med et passende klassifiseringssystem. Det er dermed nødvendig med en ytterligere testing av RMS-metoden i detektering av lignende taxa.

## **4.3 Studiedesign**

### **4.3.1 Styrker ved studien**

En av hovedstyrkene ved studien er det store antallet av prøver som ble analysert, og standardiseringen av prøveinnsamling. På den måten har det bidratt til mer statistisk robuste resultater. Alle p-verdier ble også Bonferroni-korrigert, og på den måten ble sjansen for falske positive resultater redusert, ettersom Bonferroni er en streng korreksjon av p-verdier (96). I tillegg ble analysene utført på prøver fra mødre og deres spedbarn i ulike alderskategorier, noe som har gjort det mulig å overvåke utviklingen av tarmmikrobiota over tid og mellom individer. En ytterligere styrke ved studien er Mock-analysene som har kontrollert påliteligheten til RMS-metoden i detektering av kjente bakteriearter. På denne måten har Mock-analysene troverdiggjort resultatene fra mor-barn analysene, i tillegg til støttet reproduksjonen av RMS-metoden ved eventuelle seinere anledninger. Ikke minst er det en styrke ved sekvenseringen på Illumina MiSeq som sikret kvaliteten på selve RMS-biblioteket før den ytterligere dypere sekvenseringen på Illumina HiSeq 3000. Det var ikke bare kvaliteten på RMS-biblioteket som ble sikret, men også selve kostnaden for å utføre sekvenseringen på Illumina HiSeq 3000, som ellers er høyere sammenlignet med sekvenseringen på Illumina MiSeq.

### **4.3.2 Begrensninger ved studien**

En av begrensningene ved studien er dens manglende evne til direkte å evaluere innflytelsen av faktorer som påvirker den første mikrobielle koloniseringen av tarmen, blant annet diett (morsmelk versus morsmelkerstatning) (30), antibiotika (31) og fødselsmåte (vaginalt versus keisersnitt) (24). Det var også mangel på undersøkelse av mødrenes hud- og vaginale mikrobiota, som ellers kan ha en viktig rolle i koloniseringen av tarmen hos spedbarn (25).

Det kan heller ikke utelukkes at prøvene fra mekonium var blitt forurenset grunnet manglende informasjon tilknyttet intervallet for tidspunkt på prøvetakningen. Varierende tidspunkt kan ha hatt innflytelse på bakteriepreparatet hos spedbarn. Dessuten var DNA fra avføringsprøvene på forhånd ekstrahert før denne studien, og dermed kan man ikke utelukke forurensning som kan ha oppstått under denne forbehandlingen.

Når det gjelder valget av den ikke-parametriske statistiske metoden Kruskal-Wallis, utfører den, på likhet med andre ikke-parametriske tester, en analyse på rangerte data. Det vil si at testen konverterer de opprinnelige verdiene til deres rang i stigende rekkefølge hvor den laveste verdien får rang nummer 1 (97). På den måten kan det mistes informasjon fra de opprinnelige verdiene, noe som gjør testen mindre robust sammenlignet med parametriske tester. Imidlertid er mengden av bakterier mellom mødre og spedbarn i ulike alderskategorier i utgangspunktet

svært ulik, noe som er en hovedgrunnene til at valget falt på bruk av en ikke-parametrisk metode som ikke antar normalfordelte data. I tillegg er ikke-parametriske tester mer robuste mot slengere, som ble observert i studien.

## **5. Konklusjon og videre forskning**

Hovedmålet med denne studien var å undersøke mor-barn assosiasjon av tarmbakterier, samt tarmbakterier assosiert med alder, ved bruk av RMS for sporing av bakteriearter. Det ble observert en jevn relativ mengde av de ulike artene fra slekten *Bacteroides* blant spedbarn i ulike alderskategorier, noe som indikerer at *Bacteroides* stabilt koloniserer spedbarnets tarm. *Bacteroides* var også rikelig tilstede hos mødre, noe som tyder på at bakterien er mor-barn assosiert, på likhet med det som tidligere har vært observert. I tillegg tyder resultatene fra denne studien og som er i overensstemmelse med tidligere studier på at artene fra rekken *Proteobacteria* og *Actinobacteria* dominerer tarmmikrobiota signifikant i ung alder, mens artene fra rekken *Firmicutes* dominerer signifikant i en høy alder.

Tilstedeværelsen av artene i mekonium indikerer at mødrene er en potensiell kilde til den tidlige koloniseringen av spedbarnets tarm. Imidlertid trengs det ytterligere analyser på stammenivå for å bekrefte at mor-barn assosierte tarmbakterier også er tilknyttet direkte mor-barn overføringer av tarmbakterier.

## 6. Kilder

1. Walker RW, Clemente JC, Peter I, Loos RJF. The prenatal gut microbiome: are we colonized with bacteria in utero? *Pediatric obesity*. 2017;12(1):3-17.
2. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the human microbiome. *Nutrition reviews*. 2012;70 (1):S38-44.
3. Milani C, Duranti S, Bottacini F, Casey E, Turrone F, Mahony J, et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews*. 2017;81(4):e00036-17.
4. Quigley EM. Gut bacteria in health and disease. *Gastroenterology & hepatology*. 2013;9(9):560-9.
5. Rodriguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, et al. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial ecology in health and disease*. 2015;26(2):26050.
6. Sommer F, Backhed F. The gut microbiota--masters of host development and physiology. *Nature Reviews Microbiology*. 2013;11(4):227-38.
7. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21.
8. Koppel N, Maini Rekdal V, Balskus EP. Chemical transformation of xenobiotics by the human gut microbiota. *Science*. 2017;356(6344):1-11.
9. Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell*. 2016;165(6):1332-45.
10. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology & Hepatology*. 2012;9(10):577-89.
11. Rinaldi E, Consonni A, Guidesi E, Elli M, Mantegazza R, Baggi F. Gut microbiota and probiotics: novel immune system modulators in myasthenia gravis? *The New York Academy of Sciences*. 2018;1413(1):49-58.
12. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59-65.

13. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, et al. The NIH Human Microbiome Project. *Genome research*. 2009;19(12):2317-23.
14. Parekh PJ, Balart LA, Johnson DA. The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease. *Clinical and translational gastroenterology*. 2015;6(6):e91.
15. Upadhyaya S, Banerjee G. Type 2 diabetes and gut microbiome: at the intersection of known and unknown. *Gut microbes*. 2015;6(2):85-92.
16. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, et al. Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PloS one*. 2016;11(5):e0155362.
17. Kolho KL, Korpela K, Jaakkola T, Pichai MV, Zoetendal EG, Salonen A, et al. Fecal Microbiota in Pediatric Inflammatory Bowel Disease and Its Relation to Inflammation. *American Journal of Gastroenterology*. 2015;110(6):921-30.
18. Wang F, Kaplan JL, Gold BD, Bhasin MK, Ward NL, Kellermayer R, et al. Detecting Microbial Dysbiosis Associated with Pediatric Crohn Disease Despite the High Variability of the Gut Microbiota. *Cell reports*. 2016;14(4):945-55.
19. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyotylainen T, Hamalainen AM, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell host & microbe*. 2015;17(2):260-73.
20. Elgin TG, Kern SL, McElroy SJ. Development of the Neonatal Intestinal Microbiome and Its Association With Necrotizing Enterocolitis. *Clinical therapeutics*. 2016;38(4):706-15.
21. Bager P, Wohlfahrt J, Westergaard T. Caesarean delivery and risk of atopy and allergic disease: meta-analyses. *Clinical & Experimental Allergy*. 2008;38(4):634-42.
22. Thavagnanam S, Fleming J, Bromley A, Shields MD, Cardwell CR. A meta-analysis of the association between Caesarean section and childhood asthma. *Clinical & Experimental Allergy*. 2008;38(4):629-33.
23. Milani C, Mancabelli L, Lugli GA, Duranti S, Turrone F, Ferrario C, et al. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. *Applied and Environmental Microbiology*. 2015;81(20):7078-87.
24. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(26):11971-5.



25. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host & Microbe*. 2018;24(1):133-45.e5.
26. Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *American Society for Microbiology*. 2017;2(1):00164-16.
27. Korpela K, Costea P, Coelho LP, Kandels-Lewis S, Willemsen G, Boomsma DI, et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Research*. 2018;28(4):561-8.
28. Rehbinder EM, Lodrup Carlsen KC, Staff AC, Angell IL, Landro L, Hilde K, et al. Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? *American Journal of Obstetrics Gynecology*. 2018;219(3):289.e1-.e12.
29. Houghteling PD, Walker WA. Why is initial bacterial colonization of the intestine important to infants' and children's health? *Journal of Pediatric Gastroenterology and Nutrition*. 2015;60(3):294-307.
30. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*. 2015;17(5):690-703.
31. Lemas DJ, Yee S, Cacho N, Miller D, Cardel M, Gurka M, et al. Exploring the contribution of maternal antibiotics and breastfeeding to development of the infant microbiome and pediatric obesity. *Seminars in fetal & neonatal medicine*. 2016;21(6):406-9.
32. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, et al. Patterned progression of bacterial populations in the premature infant gut. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(34):12522-7.
33. Ottman N, Smidt H, de Vos W, Belzer C. The function of our microbiota: who is out there and what do they do? *Frontiers in cellular and infection microbiology*. 2012;2(104):1–11.
34. Yatsunencko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7.
35. Cheng J, Ringel-Kulka T, Heikamp-de Jong I, Ringel Y, Carroll I, de Vos WM, et al. Discordant temporal development of bacterial phyla and the emergence of core in the fecal microbiota of young children. *International Society for Microbial Ecology*. 2015;10(4):1002–14.

36. Rodríguez JM, Murphy K, Stanton CJMEHD. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial ecology in health and disease*. 2015;26(2):26050.
37. Favier CF, de Vos WM, Akkermans AD. Development of bacterial and bifidobacterial communities in feces of newborn babies. *Anaerobe*. 2003;9(5):219-29.
38. Bäckhed F, Roswall J, Peng YJCHM. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe*. 2015;17(5):690-703.
39. Salminen S, Isolauri E. Intestinal colonization, microbiota, and probiotics. *The Journal of Pediatrics*. 2006;149(5):S115-S20.
40. Barile D, Rastall RA. Human milk and related oligosaccharides as prebiotics. *Current Opinion in Biotechnology*. 2013;24(2):214-9.
41. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108 (1):4578-85.
42. Janda JM, Abbott SL. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology*. 2007;45(9):2761-4.
43. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J. Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*. 2009;55(5):856-66.
44. Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nature methods*. 2016;13(5):435-8.
45. Avershina E, Angell IL, Simpson M, Storro O, Oien T, Johnsen R, et al. Low Maternal Microbiota Sharing across Gut, Breast Milk and Vagina, as Revealed by 16S rRNA Gene and Reduced Metagenomic Sequencing. *Genes*. 2018;9(5):231.
46. Ravi A, Avershina E, Angell IL, Ludvigsen J, Manohar P, Padmanaban S, et al. 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*. 2018;149(18):44-52.
47. Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*. 2014;15(1):96.

48. Oulas A, Pavludi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*. 2015;9(5):75-88.
49. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. *PloS one*. 2014;9(4):e94249.
50. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. *BMC bioinformatics*. 2011;12(1):38.
51. Konstantinidis KT, Ramette A, Tiedje JM. The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2006;361(1475):1929-40.
52. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*. 2013;31(9):814.
53. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*. 2014;5(209):209.
54. Nagarajan N, Pop M. Sequence assembly demystified. *Nature reviews Genetics*. 2013;14(3):157-67.
55. Mende DR, Waller AS, Sunagawa S, Jarvelin AI, Chan MM, Arumugam M, et al. Assessment of metagenomic assembly using simulated next generation sequencing data. *PloS one*. 2012;7(2):e31386.
56. Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pacific Symposium on Biocomputing*. 2011:165-76.
57. Janssen P, Coopman R, Huys G, Swings J, Bleeker M, Vos P, et al. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology*. 1996;142 (7):1881-93.
58. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic acids research*. 1995;23(21):4407-14.
59. Liu MY, Worden P, Monahan LG, DeMaere MZ, Burke CM, Djordjevic SP, et al. Evaluation of ddRADseq for reduced representation metagenome sequencing. *PeerJ*. 2017;5(19):e3837.

60. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977;74(12):5463-7.
61. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*. 2011;38(3):95-109.
62. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prevention Research*. 2012;5(7):887-900.
63. Balzer S, Malde K, Lanzen A, Sharma A, Jonassen I. Characteristics of 454 pyrosequencing data--enabling realistic simulation with flowsim. *Bioinformatics*. 2010;26(18):i420-5.
64. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nature Biotechnology*. 2008;26(10):1146-53.
65. Ferrarini M, Moretto M, Ward JA, Surbanovski N, Stevanovic V, Giongo L, et al. An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BioMed Central Genomics*. 2013;14:670.
66. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990;215(3):403-10.
67. Wood DE, Salzberg SLJGB. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*. 2014;15(3):R46.
68. Chen EZ, Bushman FD, Li H. A Model-Based Approach For Species Abundance Quantification Based On Shotgun Metagenomic Data. *Statistics in biosciences*. 2017;9(1):13-27.
69. Qamar W, Khan MR, Arafah A. Optimization of conditions to extract high quality DNA for PCR analysis from whole blood using SDS-proteinase K method. *Saudi journal of biological sciences*. 2017;24(7):1465-9.
70. Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J. Rapid and simple method for purification of nucleic acids. *Journal of clinical microbiology*. 1990;28(3):495-503.
71. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical chemistry*. 2012;84(2):1003-11.
72. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments*. 2012;20(62):e3923.

73. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International journal of systematic and evolutionary microbiology*. 2017;67(5):1613-7.
74. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*. 2016;17(1):132.
75. Gosalbes MJ, Llop S, Valles Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical & Experimental Allergy*. 2013;43(2):198-211.
76. de Muinck EJ, Øien T, Storrø O, Johnsen R, Stenseth NC, Rønningen KS, et al. Diversity, transmission and persistence of *Escherichia coli* in a cohort of mothers and their infants. *Environmental Microbiology Reports*. 2011;3(3):352-9.
77. Korpela K, de Vos WM. Early life colonization of the human gut: microbes matter everywhere. *Current Opinion in Microbiology*. 2018;44:70-8.
78. Lundgren SN, Madan JC, Emond JA, Morrison HG, Christensen BC, Karagas MR, et al. Maternal diet during pregnancy is related with the infant stool microbiome in a delivery mode-dependent manner. *Microbiome*. 2018;6(1):109.
79. Avershina E, Lundgård K, Sekelja M, Dotterud C, Storrø O, Øien T, et al. Transition from infant- to adult-like gut microbiota. *Environmental Microbiology*. 2016;18(7):2226-36.
80. Turrone F, Ribbera A, Foroni E, van Sinderen D, Ventura M. Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek*. 2008;94(1):35-50.
81. Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG. The infant microbiome development: mom matters. *Trends in Molecular Medicine*. 2015;21(2):109-17.
82. Turrone F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, et al. Diversity of bifidobacteria within the infant gut microbiota. *PloS one*. 2012;7(5):e36957.
83. Milani C, Mancabelli L, Lugli GA, Duranti S, Turrone F, Ferrario C, et al. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. *Applied and Environmental Microbiology*. 2015;81(20):7078-87.
84. Laforest-Lapointe I, Arrieta MC. Patterns of Early-Life Gut Microbial Colonization during Human Immune Development: An Ecological Perspective. *Frontiers in Immunology*. 2017;8(10):788.

85. Pannaraj PS, Li F, Cerini C, Bender JM, Yang S, Rollie A, et al. Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatrics*. 2017;171(7):647-54.
86. Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Scientific Reports*. 2016;6:23129.
87. Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environmental microbiology*. 2014;16(9):2891-904.
88. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut*. 2014;63(4):559-66.
89. Rudi K, Storro O, Oien T, Johnsen R. Modelling bacterial transmission in human allergen-specific IgE sensitization. *Letters in applied microbiology*. 2012;54(5):447-54.
90. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. *Nature*. 2013;493(7430):45-50.
91. Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science translational medicine*. 2016;8(343):343ra81.
92. Marcobal A, Barboza M, Sonnenburg EDJCHM. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host & Microbe*. 2011;10(5):507-14.
93. Hansen R, Scott KP, Khan S, Martin JC, Berry SH, Stevenson M, et al. First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota. *PloS one*. 2015;10(7):e0133320.
94. Yang W, Huang L, Shi C, Wang L, Yu R. UltraStrain: An NGS-Based Ultra Sensitive Strain Typing Method for Salmonella enterica. *Frontiers in Genetics*. 2019;10:276.
95. Vetrovsky T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS one*. 2013;8(2):e57923.
96. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? *Korean journal of anesthesiology*. 2018;71(5):353-60.
97. Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean journal of anesthesiology*. 2016;69(1):8-14.

# Vedlegg

## Vedleggsliste

Vedlegg 1. REK-godkjenning

Vedlegg 2. Oversikt over Illumina index-primere

Vedlegg 3. «QScore Heatmap» fra Illumina MiSeq sekvenseringen

Vedlegg 4. «QScore Distribution» fra Illumina MiSeq sekvenseringen

Vedlegg 5. «Per Sequence Quality Scores» fra Illumina HiSeq sekvenseringen

Vedlegg 6. «Mean Quality Scores» fra Illumina HiSeq sekvenseringen

Vedlegg 7. Taksonomisk rangering av utvalgte bakteriearter

Vedlegg 8. Kruskal-Wallis test med Dunn's PostHoc og Bonferroni korrigerede p-verdier

# Vedlegg 1. REK-godkjenning



<b>Region:</b> REK sør-øst	<b>Saksbehandler:</b> Hege Holde Andersson	<b>Telefon:</b> 22845514	<b>Vår dato:</b> 08.12.2014	<b>Vår referanse:</b> 2014/518/REK sør-øst B
			<b>Deres dato:</b> 29.10.2014	<b>Deres referanse:</b>

Vår referanse må oppgis ved alle henvendelser

Karin C. Lødrup Carlsen  
Oslo universitetssykehus HF

## 2014/518 PreventADALL

Vi viser til tilbakemelding på komiteens merknader til ovennevnte prosjekt, slik de fremkom i vedtaksbrev av 17.10.2014. Søknaden ble behandlet av Regional komité for medisinsk og helsefaglig forskningsetikk (REK sør-øst) i møtet 26.11.2014. Vurderingen er gjort med hjemmel i helseforskningsloven § 10, jf. forskningsetikklovens § 4.

### Prosjektomtale

«Allergiske sykdommer rammer ca 25-40 % av befolkningen med stor individuell og samfunnsbelastning. I dag finnes ingen primærforebyggende strategier. Sykdomsårsaker er ukjente, men miljøfaktorer som redusert mikrobiologisk mangfold og økt bruk av miljøgifter har sannsynlig påvirkning av immun-utvikling ved interaksjon med gener (epigenetikk). Allergiske sykdommer starter ofte med atopisk eksem og fødemiddelallergi i første leveår etterfulgt av astma og allergisk rinitt. PreventADALL er en internasjonal innovativ randomisert faktorielt designet prospektiv fødselskohortstudie som skal teste om allergisk sykdomsutvikling kan forebygges ved lavkostnadsintervensjoner, og identifisere faktorer som tidlig i livet kan føre til sykdom med data fra 18 ukers svangerskap, biologiske tester og langtidsoppfølging av barn. Tidlige utfallsvariabler er fødemiddelintoleranse og atopisk eksem.

De gravide kvinnene vil bli bedt om å fylle ut spørreskjema som inkluderer generell helse, sosioøkonomi, livsstil (alkohol, tobakk, medisiner), kosthold, fysisk aktivitet, stress og livskvalitet. Oppfølging av barna inkluderer spørreskjema om generell helse, sykdommer, medisiner, sosioøkonomi, boforhold, kosthold, stress og livskvalitet. Gravide kvinner skal rekrutteres på svangerskapskontroll uke 18. Barna inkluderes ved fødsel. Målet er å rekruttere 3500 kvinner i løpet av 24 mnd. Det vil bli sendt ut brev til alle kvinner som skal til slik kontroll med invitasjon. De inkluderte randomiseres til 4 ulike grupper.

*Gruppe 1: Norske nasjonale anbefalinger følges om fullamming til 6 måneder og matvareintroduksjon etter dette tidspunktet. Ingen anbefalinger i forhold til hudbehandling.*

*Gruppe 2: Deltagerne introduserer kumelk, egg, torsk, peanøtt og hvete som smaksprøver mellom 3 og 4 måneders alder. Ingen anbefalinger i forhold til hudbehandling.*

*Gruppe 3: Norske nasjonale anbefalinger følges om fullamming til 6 måneder og matvareintroduksjon etter dette tidspunktet. Barnet bades i oljebad/smøres med fuktighetskrem i ansiktet minst 5 dager/uke mellom 2 uker og 9 måneders alder.*

*Gruppe 4: Deltagerne introduserer melk, egg, torsk, peanøtt og hvete som smaksprøver mellom 3 og 4 måneder. Barnet bades i oljebad/smøres med fuktighetskrem i ansiktet minst 5 dager/uke mellom 2 uker og 9*

**Besøksadresse:**  
Gullhaugveien 1-3, 0484 Oslo

**Telefon:** 22845511  
**E-post:** post@helseforskning.etikk.no  
**Web:** <http://helseforskning.etikk.no/>

All post og e-post som inngår i saksbehandlingen, bes adressert til REK sør-øst og ikke til enkelte personer

Kindly address all mail and e-mails to the Regional Ethics Committee, REK sør-øst, not to individual staff



*måneders alder.»*

### **Forskningsbiobank**

Det søkes om opprettelse av ny spesifikk forskningsbiobank «*PreventADALL*» med ansvarshavende Karin Lødrup Carlsen.

### **Saksgang**

Komiteen behandlet første gang prosjektet på møtet 26.03.2014. Vedtak i saken ble utsatt. Komiteen ønsket en nærmere redegjørelse for forsvarligheten av studien, da særlig med tanke på den delen av studien som omhandler tidlig introduksjon av mat. Komiteen ønsket å sende prosjektet til vurdering hos en ekstern konsulent for å få en vurdering av risiko og forsvarlighet særlig knyttet til denne delen av studien.

Studien er omfattende og krever mye av den enkelte deltager. Komiteen ønsket derfor en nærmere redegjørelse fra prosjektleder knyttet til prosjektets gjennomførbarhet.

I tillegg ba komiteen for en nærmere redegjørelse fra prosjektleder om studiens kobling til andre studier samt koblinger til journal og helseregistre.

### *Prosjektleders tilbakemelding*

Komiteen mottok prosjektleders tilbakemelding 09.05.2014.

Når det gjelder studiens forsvarlighet viser prosjektleder til oppdaterte internasjonale retningslinjer. Det finnes ingen vitenskapelig dokumentasjon for å utsette/ unngå introduksjon av matvarer etter 4 måneder. Det presiseres at det skal kun gis små mengder av mat, og at mat skal ikke erstatte morsmelk.

Når det gjelder forskning på barn skal risiko og ulempe for barnet være ubetydelig. I tilbakemeldingen skriver prosjektleder at hudintervensjonen kan bedre hudbarrieren for deltagende barn. Risiko for økt allergi antas å være liten basert på oppdatert forskning og retningslinjer. Prosjektleder skriver i tilbakemeldingen at det er ingen garanti for at intervensjonene ikke fører til ulempe, men det er lite trolig at små smaksprøver av matvarer (sutte på finger/smokk/teskje med litt puree) vil oppleves som ulempe i forbindelse med matvareintroduksjonen. Det er etablert ekstern overvåkningskomite for å observere ulikheter i gruppene.

Prosjektleder redegjør for kobling til andre studier ved OUS. Det presiseres at de deltagende kvinnene vil kunne delta i andre studier og at de ønsker en kobling til studier på morkaker. Informasjonsskrivet er korrigert for tydeligere å forklare dette.

I e-post mottatt 15.09.2014 skriver prosjektleder at når det gjelder kobling mot ulike helseregistre år det dette noe man ser vil være ønskelig fremover, men ikke det første året. Det vil således være ønskelig å kunne sende endringsmelding til REK, uten å måtte innhente nytt samtykke fra kvinnene /foreldrene, til kobling mot registre i fremtiden.

### *Konsulentuttalelse*

Komiteen ønsket å innhente konsulentuttalelse fra en ekstern konsulent for å få en vurdering av risiko og forsvarlighet særlig knyttet til den delen av studien som omhandler tidlig introduksjon av mat. Komiteen mottok 11.09.2014 konsulentuttalelse fra Nicholas Brodzski, Senior Consultant, Pediatric Allergy/Immunology, ved Lund universitetssykehus, Sverige. Konsulenten konkluderer med at studien er forsvarlig å gjennomføre.

### *Komiteehandling*

Komiteen behandlet søknaden på ny i møtet 17.09.2014. I sitt vedtak skriver komiteen at helseforskningsloven stiller strenge krav knyttet til forskning på barn. Etter prosjektleders tilbakemelding og konsulentuttalelsen er komiteen fortsatt usikre på forsvarligheten ved å innføre enkelte av matvarene. Matvarene skal introduseres til barn på 3-4 måneder. Komiteen mener dette vil være en sårbar alder for introduksjon da barn har et dårlig immunforsvar i denne alderen.

Komiteen ønsker derfor en uttalelse fra en allergolog/ pediater primært knyttet til introduksjon av de ulike

matvarene og tidspunktet for introduksjon.

#### *Prosjektleders tilbakemelding*

Komiteen mottok 29.okt 2014 tilbakemelding fra prosjektleder. I tilbakemeldingen skriver prosjektleder at hun ønsker å oppklare en faglig feil. Et dårlig immunforsvar referer vanligvis til nedsatt eller manglende evne til å takle infeksjoner, og henspiller på et komplekst samarbeid mellom det adaptive og det medfødte immunforsvaret. Amming er en av faktorene som bidrar til å beskytte barnet mot infeksjoner, blant annen via immunglobuliner overført fra mor. I den innsendte tilbakemeldingen redegjør prosjektleder for immunsystemets rolle i allergiutviklingen. I tillegg gis det en nærmere redegjørelse for hvilke tiltak som kan begrense risiko ved fødemiddelintroduksjon.

Prosjektleder møtte i komiteens møte 26.11.2014 for å drøfte prosjektet med komiteen.

Komiteen ønsket primært en nærmere redegjørelse fra prosjektleder knyttet til introduksjon av de ulike matvarene og tidspunktet for introduksjon, samt en vurdering av hvilke eventuelle risikoreduserende tiltak/beredskap man kan gjøre i forbindelse med studien.

#### **Komiteens vurdering**

I sitt brev datert 17.10.2014 skrev komiteen at den ønsket en uttalelse fra en allergolog/ pediater primært knyttet til introduksjon av de ulike matvarene og tidspunktet for introduksjon. Det har imidlertid vist seg vanskelig å innhente konsulentuttalelse. Komiteen besluttet derfor å invitere prosjektleder i møtet før konsulentuttalelsen foreligger.

I komitemøtet gjorde prosjektleder rede for de ulike matvarene som skal introduseres, tidspunktet for introduksjon og beredskapen i prosjektet. Prosjektleder redegjorde for at egg, melk og hvete er basisvarer som mange barn introduseres for uavhengig av studien. I følge prosjektleder er ikke peanøtter en basisvare, men fordi det finnes spor av peanøtter i mange varer, og fordi det nesten er umulig å skape toleranse på et senere tidspunkt, er det viktig å ha det med i studien.

Etter prosjektleders tilbakemelding mener komiteen at en tidlig introduksjon av egg, melk og hvete vil være forsvarlig. Komiteen finner derfor på det nåværende tidspunkt kunne godkjenne de fleste delene av studien, med unntak av tidlig introduksjon av peanøtter. Imidlertid ønsker komiteen fortsatt en vurdering av en ekstern konsulent knyttet til forsvarligheten av introduksjon av peanøtter og hvilke bedreskap man bør ha rundt dette. Komiteen utsetter derfor sluttbehandling av spørsmålet om inklusjon av peanøtter som en av fødevarerne til konsulentrapport foreligger.

#### *Om tidlig introduksjon av peanøtter.*

Etter tilbakemelding fra komiteen har man nå endret prosjektet slik at førstegangsintroduksjon av peanøtter vil tilbys ved 3 måneders kontrollen på sykehus. I følge prosjektleder vil det være svært lite sannsynlig at man vil reagere på peanøtter ved andregangs introduksjon dersom man har tålt det ved førstegangsintroduksjon på sykehuset. Komiteen betviler ikke at det vil være en svært liten sannsynlighet for alvorlige allergiske reaksjoner ved andregangsintroduksjon eller senere i studieløpet. Imidlertid skal det inkluderes et stort antall barn i studien, og ca 1750 barn vil randomiseres til en av de to gruppene med tidlig introduksjon av fødevarer. Selv med en svært liten sannsynlighet for at den enkelte deltaker vil få alvorlige allergiske reaksjoner, må man med et så høyt deltakerantall, ta høyde for at noen barn kan få alvorlige allergiske reaksjoner. Det er derfor viktig at foreldrene informeres godt om at denne type reaksjoner også kan oppstå og om hvordan man eventuelt skal håndtere en slik situasjon. Det er også viktig at det gjøres en grundig vurdering av om risiko for de barna dette eventuelt måtte gjelde er akseptabel i lys av helseforskningslovens krav til forskning på mindreårige, jf. § 18. Forskning på barn kan bare tillates hvis risiko og ulempe for barna er «ubetydelig».

Det er opprettet en ekstern overvåkningskomite for å observere mulige store forskjeller i hovedutfallene i de fire gruppene. Denne overvåkningskomiteen vil fortløpende evaluere studien. Komiteen mener at dette er et godt og tryggende tiltak, men som likevel ikke gjør en grundig forhåndsvurdering av risiko overflødig.

Det er lagt opp til en tett kontakt mellom deltagere og prosjektledelse, og ved mistanke om allergiske

symptomer har foreldrene mulighet til å kontakte de lokale barneavdelingene for en vurdering av symptomene. Komiteen mener dette er positivt, men at denne muligheten primært vil være aktuell for de barna som får for eksempel hudreaksjoner, og ikke for de barna som eventuelt får alvorlige akutte reaksjoner som luftveisproblemer. Komiteen mener at det må etableres en bedre beredskap i prosjektet.

Komiteen mener det er svært viktig at foreldrene informeres på en god måte om hvilke reaksjoner barna kan få. Prosjektleder anfører at det er viktig at man ikke sender et signal om at mat er farlig. Komiteen har forståelse for dette, men mener imidlertid det er viktig at man ikke bagatelliserer de reaksjonene som kan oppstå hos barna.

Før introduksjon av peanøtter eventuelt kan godkjennes, må det derfor utarbeides en beredskapsplan for prosjektet, særlig knyttet til de første 2 ukene med introduksjon av mat. Her bør det tydelig fremkomme hvilke rutiner prosjektet har dersom et av barna får en alvorlig allergisk reaksjon. I tillegg bør det gis en beskrivelse av hvilke informasjon foreldrene skal få. Det må utarbeides et eget informasjonsskriv til de som randomiseres til gruppene for tidlig introduksjon av mat. Beredskapsplan og informasjonsskriv må sendes komiteen for godkjenning.

Komiteen mener det er positivt at man tilbyr foreldrene at førstegangsintroduksjon av peanøtter skjer ved 3 måneders kontrollen på sykehus. Komiteen mener imidlertid at dette ikke bare bør være et tilbud; komiteen anser at førstegangsintroduksjon *må* foregå på sykehus. Komiteen mener videre at også for de påfølgende introduksjoner av peanøtter, bør det være en ramme som sikrer rask helsehjelp om nødvendig.

#### *Konklusjon*

Etter tilbakemelding fra prosjektleder skriftlig og i møtet mener komiteen at prosjektleder har svart tilfredsstillende på komiteens spørsmål knyttet til introduksjon og beredskap knyttet til introduksjon av basisfødevarer. Komiteen ønsker imidlertid å innhente en uavhengig konsulentvurdering av risiko knyttet til tidlig introduksjon av peanøtter, for å sikre at helseforskningslovens krav til forskning på barn er tilfredsstillt. Komiteen ber videre prosjektleder utarbeide og sende komiteen en beredskapsplan og informasjonsskriv knyttet til introduksjon av peanøtter, i tråd med ovenstående merknader. Komiteen eller komiteens leder vil ta stilling til introduksjon av peanøtter når konsulentuttalelsen foreligger, og når den har mottatt beredskapsplan og informasjonsskriv fra prosjektleder.

Komiteen finner derfor å kunne godkjenne prosjektet på følgende vilkår:

1) Introduksjon av peanøtter kan ikke skje før komiteen har godkjent dette. Komiteen eller komiteens leder tar stilling til introduksjon av peanøtter når konsulentuttalelse og tilbakemelding fra prosjektleder foreligger.

#### **Vedtak**

Komiteen godkjenner prosjektet i henhold til helseforskningsloven § 9 og § 33 under forutsetning av at ovennevnte vilkår oppfylles.

I tillegg til ovennevnte vilkår, er godkjenningen gitt under forutsetning av at prosjektet gjennomføres slik det er beskrevet i søknaden.

Tillatelsen gjelder til 31.12.2044. Av dokumentasjonshensyn skal opplysningene likevel bevares inntil 31.12.2051. Opplysningene skal lagres avidentifisert, dvs. atskilt i en nøkkel- og en opplysningsfil. Opplysningene skal deretter slettes eller anonymiseres, senest innen et halvt år fra denne dato.

Komiteen godkjenner også oppførelsen av en spesifikk forskningsbiobank som beskrevet i søknaden.

Biobankregisteret blir underrettet ved kopi av dette brev.

Hvis forskningsbiobanken opphører, nedlegges eller overtas av andre, skal det søkes REK om tillatelse, jf. helseforskningsloven § 30.

Komiteen godkjenner overføring av humant biologisk materiale til og fra utlandet for denne studien, jf. helseforskningsloven § 29.

Forskningsprosjektets data skal oppbevares forsvarlig, se personopplysningsforskriften kapittel 2, og Helsedirektoratets veileder ”Personvern og informasjonssikkerhet i forskningsprosjekter innenfor helse- og omsorgssektoren”

*Klageadgang*

Du kan klage på komiteens vedtak, jf. forvaltningslovens § 28 flg. Klagen sendes til REK sør-øst B. Klagefristen er tre uker fra du mottar dette brevet. Dersom vedtaket opprettholdes av REK sør-øst B, sendes klagen videre til Den nasjonale forskningsetiske komité for medisin og helsefag for endelig vurdering.

Komiteens avgjørelse var enstemmig.

Vi ber om at alle henvendelser sendes inn via vår saksportal: <http://helseforskning.etikkom.no> eller på e-post til: [post@helseforskning.etikkom.no](mailto:post@helseforskning.etikkom.no)

Vennligst oppgi vårt referansennummer i korrespondansen.

Med vennlig hilsen

Geir Olav Hjortland  
nestleder REK sør-øst b

Hege Holde Andersson  
komitésekretær

**Kopi til:** Oslo universitetssykehus, [oushfdlgodkjenning@ous-hf.no](mailto:oushfdlgodkjenning@ous-hf.no)

## Vedlegg 2. Oversikt over Illumina index-primere

### EcoRI forward (5'-3'):

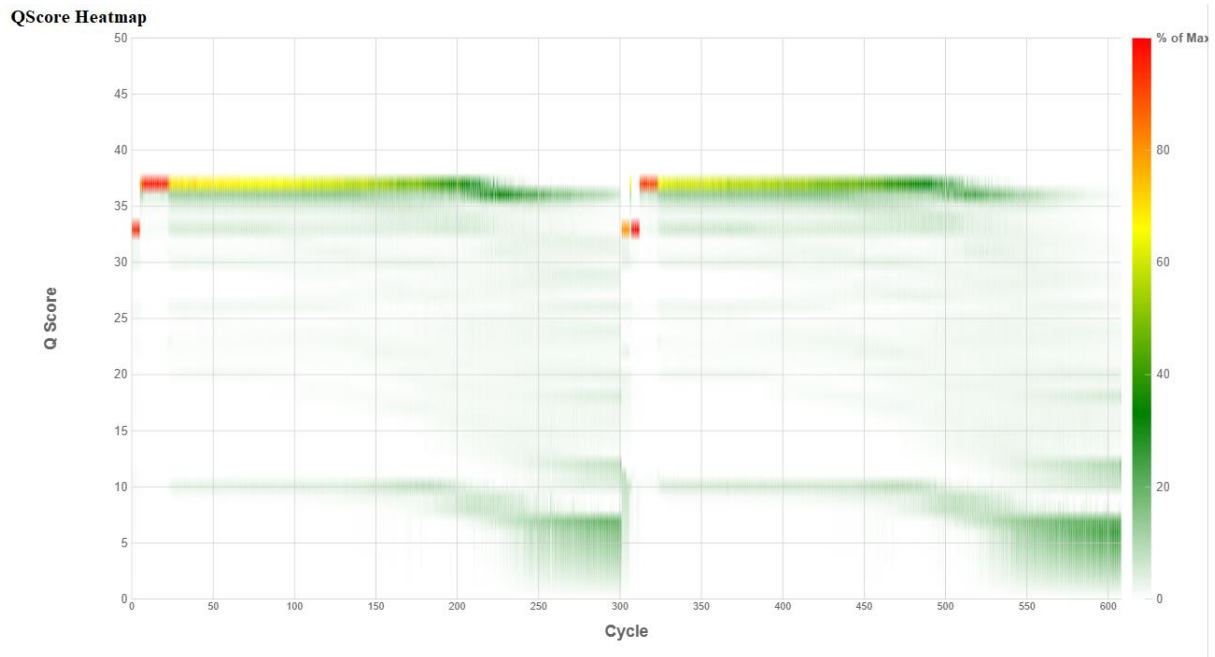
1. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtaaGACTGCGTACCAATTC
2. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtccGACTGCGTACCAATTC
3. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgcaGACTGCGTACCAATTC
4. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccGACTGCGTACCAATTC
5. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagGACTGCGTACCAATTC
6. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcGACTGCGTACCAATTC
7. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaGACTGCGTACCAATTC
8. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccGACTGCGTACCAATTC
9. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgGACTGCGTACCAATTC
10. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgGACTGCGTACCAATTC
11. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggGACTGCGTACCAATTC
12. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcGACTGCGTACCAATTC
13. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatGACTGCGTACCAATTC
14. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcGACTGCGTACCAATTC
15. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattctGACTGCGTACCAATTC
16. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagGACTGCGTACCAATTC

### MseI reverse (5'-3'):

1. caagcagaagacggcatacagatCGTGATgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
2. caagcagaagacggcatacagatACATCGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
3. caagcagaagacggcatacagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
4. caagcagaagacggcatacagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
5. caagcagaagacggcatacagatCACTCTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
6. caagcagaagacggcatacagatATTGGCgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
7. caagcagaagacggcatacagatGATCTGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
8. caagcagaagacggcatacagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA
9. caagcagaagacggcatacagatCTGATCgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

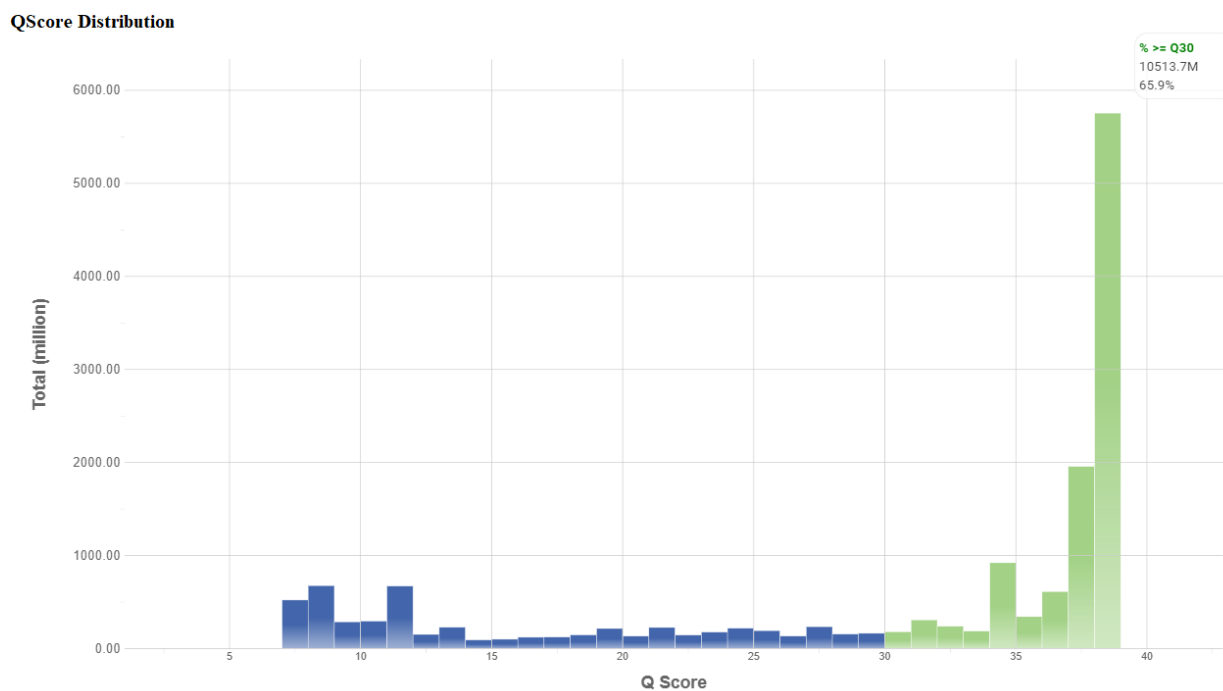
10. caagcagaagacggcatacagagatAAGCTAgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
11. caagcagaagacggcatacagagatGTAGCCgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
12. caagcagaagacggcatacagagatTACAAGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
13. caagcagaagacggcatacagagatTTGACTgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
14. caagcagaagacggcatacagagatGGAACTgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
15. caagcagaagacggcatacagagatTGACATgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
16. caagcagaagacggcatacagagatGGACGGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
17. caagcagaagacggcatacagagatCTCTACgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
18. caagcagaagacggcatacagagatGCGGACgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
19. caagcagaagacggcatacagagatTTTCACgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
20. caagcagaagacggcatacagagatGGCCACgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
21. caagcagaagacggcatacagagatCGAAACgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
22. caagcagaagacggcatacagagatCGTACGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
23. caagcagaagacggcatacagagatCCACTCgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
24. caagcagaagacggcatacagagatGCTACCgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
25. caagcagaagacggcatacagagatATCAGTgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
26. caagcagaagacggcatacagagatGCTCATgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
27. caagcagaagacggcatacagagatAGGAATgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
28. caagcagaagacggcatacagagatCTTTTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
29. caagcagaagacggcatacagagatTAGTTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
30. caagcagaagacggcatacagagatCCGGTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
31. caagcagaagacggcatacagagatATCGTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
32. caagcagaagacggcatacagagatTGAGTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
33. caagcagaagacggcatacagagatCGCCTGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
34. caagcagaagacggcatacagagatGCCATGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
35. caagcagaagacggcatacagagatAAAATGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA
36. caagcagaagacggcatacagagatTGTTGGgtgactggagttcagacgtgtgctctccgatctGATGAGTCCTGAGTAA

### Vedlegg 3. «QScore Heatmap» fra Illumina MiSeq sekvenseringen



Tilleggsfigur 1. «**QScore Heatmap**». Illustrasjon av kvaliteten (*Qscore*, y-akse) til hver enkelt syklus (totalt 300 sykluser per DNA-tråd) (x-akse) fra Illumina MiSeq sekvenseringen. Andelen av sekvenser fordelt blant de ulike *Qscores* er assosiert med ulike fargekoder (grønn, gul og rød), med henholdsvis lav, middels og høy andel. Det observeres en høyere andel av sekvenser med en høyere *Qscore* for «read 1» (til venstre, fram til syklus 300) til forskjell fra «read 2» (fra syklus 300). Hentet fra BaseSpace Illumina.

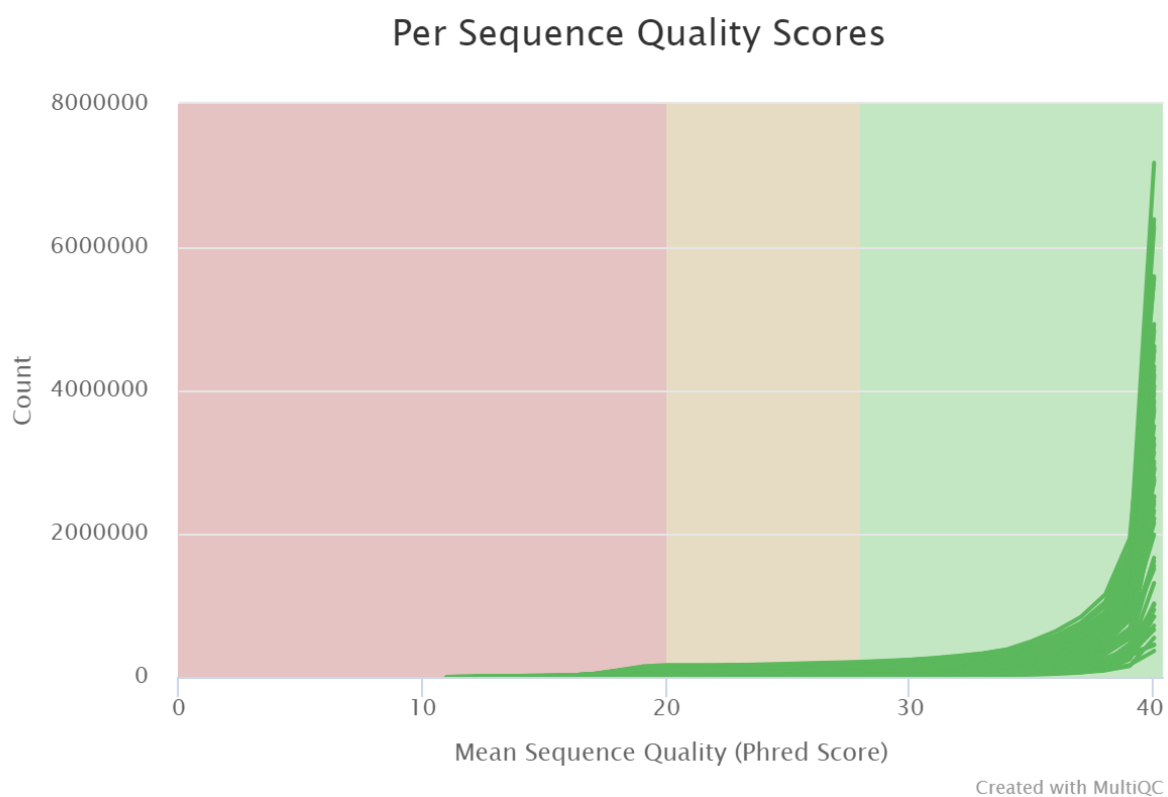
## Vedlegg 4. «QScore Distribution» fra Illumina MiSeq sekvenseringen



Tilleggsfigur 2. «**QScore Distribution**». Illustrasjon av fordelingen av *Qscore* (x-akse) blant sekvensene angitt i million (y-akse) fra Illumina MiSeq sekvenseringen. Grønne stolper indikerer høy *QScore*, mens blå stolper indikerer lav *QScore*. 65,9% av sekvensene har en *QScore* > 30. Hentet fra BaseSpace Illumina.

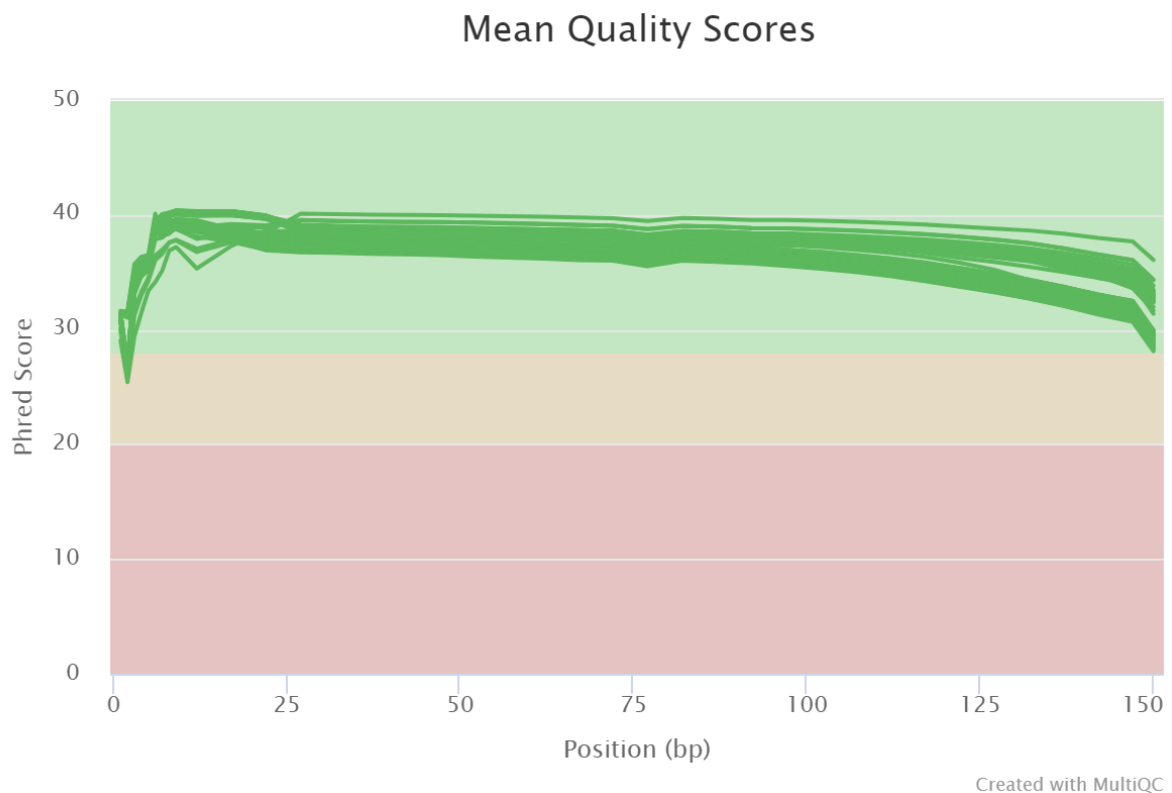


## Vedlegg 5. «Per Sequence Quality Scores» fra Illumina HiSeq sekvenseringen



Tilleggsfigur 3. «**Per Sequence Quality Scores**». Illustrasjon av fordelingen av kvalitetsscore (Phred Score) (x-akse) blant sekvensene angitt i million (y-akse) fra Illumina HiSeq3000 sekvenseringen. De ulike bakgrunnsfargene lilla, rosa og grønn er assosiert med henholdsvis lav, middels og høy Phred Score. Majoriteten av sekvensene har en Phred Score > 35. Hentet fra «MultiQC».

## Vedlegg 6. «Mean Quality Scores» fra Illumina HiSeq sekvenseringen



Tilleggsfigur 4. «Mean Quality Scores». Illustrasjon av fordelingen av gjennomsnittlig kvalitetsscore (Phred Score) (y-akse) for «read1 og read2» blant syklusene/bp (totalt 150 sykluser/bp) (x-akse) fra Illumina HiSeq3000 sekvenseringen. Bakgrunnen til grafen deler y-aksen i tre deler med ulike fargekoder som grønn, rosa og lilla assosiert med ulik kvalitet, med henholdsvis høy, middels og lav kvalitet (Phred Score). Hentet fra «MultiQC».

## Vedlegg 7. Taksonomisk rangering av utvalgte bakteriearter

Tilleggstabell 1. Taksonomisk rangering (rekke, klasse, familie, slekt) av utvalgte bakteriearter (*E. coli*, *P. puraquae*, *A. parvus*, *D. longicatena*, *C. catus*, *C. comes*, *B. fragilis*, *B. thetaiotamicron*, *P. distasonis*, *B. vulgatus*, *B. dorei*, *B. longum*).

<b>Rekke</b>	<b>Klasse</b>	<b>Familie</b>	<b>Slekt</b>	<b>Art</b>
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriaceae</i>	<i>Escherichia</i>	<i>Escherichia coli</i>
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Comamonadaceae</i>	<i>Pelomonas</i>	<i>Pelomonas puraquae</i>
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Moraxellaceae</i>	<i>Acinetobacter</i>	<i>Acinetobacter parvus</i>
<i>Firmicutes</i>	<i>Clostridia</i>	<i>Lachnospiraceae</i>	<i>Dorea</i>	<i>Dorea longicatena</i>
<i>Firmicutes</i>	<i>Clostridia</i>	<i>Lachnospiraceae</i>	<i>Coprococcus</i>	<i>Coprococcus catus</i>
<i>Firmicutes</i>	<i>Clostridia</i>	<i>Lachnospiraceae</i>	<i>Coprococcus</i>	<i>Coprococcus comes</i>
<i>Bacteroidetes</i>	<i>Gammaproteobacteria</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>Bacteroides fragilis</i>
<i>Bacteroidetes</i>	<i>Gammaproteobacteria</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>Bacteroides vulgatus</i>
<i>Bacteroidetes</i>	<i>Gammaproteobacteria</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>Bacteroides dorei</i>
<i>Bacteroidetes</i>	<i>Gammaproteobacteria</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>Bacteroides thetaiotamicron</i>
<i>Bacteroidetes</i>	<u><i>Bacteroidia</i></u>	<i>Porphyromonadaceae</i>	<i>Parabacteroides</i>	<i>Parabacteroides distasonis</i>
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Bifidobacteriaceae</i>	<i>Bifidobacterium</i>	<i>Bifidobacterium longum</i>

## Vedlegg 8. Kruskal-Wallis test med *Dunn's* PostHoc og Bonferroni korrigererte p-verdier

Tilleggstabell 2. Resultater fra sammenligningen av den relative mengden av bakteriearter (*D. longicatena*, *C.catus* og *C.comes*) mellom mor og barn i ulike alderskategorier ved bruk av Kruskal-Wallis test med *Dunn's* PostHoc for parvise sammenligninger og Bonferroni korrigererte p-verdier. Gul farge indikerer statistisk signifikante resultat ( $p < 0,05$ ). Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

Sammenligning av mor-barn prøver	<i>Dorea longicatena</i> (Adj.Sig.)	<i>Coprococcus catus</i> (Adj.Sig.)	<i>Coprococcus comes</i> (Adj.Sig.)
<b>Kruskal-Wallis test med PostHoc og Bonferroni korreksjon</b>			
Mor-12 mnd	0,001	0,001	0,001
Mor-3 mnd	0,000	0,000	0,000
Mor-6 mnd	0,000	0,000	0,000
Mor- Mekonium	0,000	0,000	0,000
12 mnd – 3 mnd	0,000	0,000	0,000
12 mnd – 6 mnd	0,000	0,000	0,000
12 mnd - Mekonium	0,000	0,000	0,000
3 mnd – 6 mnd	0,544	0,055	0,410
3 mnd - Mekonium	0,014	0,002	0,030
6 mnd - Mekonium	1,000	1,000	1,000

Tilleggstabell 3. Resultater fra sammenligningen av den relative mengden av bakteriearter (*E. coli*, *P. puraquae* og *A. parvus*) mellom mor og barn i ulike alderskategorier ved bruk av Kruskal-Wallis test med Dunn's PostHoc for parvise sammenligninger og Bonferroni korrigererte p-verdier. Gul farge indikerer statistisk signifikante resultat ( $p < 0,05$ ). Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

Sammenligning av mor-barn prøver	<i>Escherichia coli</i> (Adj.Sig.)	<i>Pelomonas puraquae</i> (Adj.Sig.)	<i>Acinetobacter parvus</i> (Adj.Sig.)
<b>Kruskal-Wallis test med Dunn's PostHoc og Bonferroni korreksjon</b>			
Mor-12 mnd	0,532	1,000	0,000
Mor-3 mnd	0,000	1,000	0,168
Mor-6 mnd	0,000	1,000	0,000
Mor- Mekonium	0,000	0,000	0,000
12 mnd – 3 mnd	0,002	1,000	0,000
12 mnd – 6 mnd	0,000	1,000	1,000
12 mnd - Mekonium	0,000	0,000	0,000
3 mnd – 6 mnd	1,000	1,000	0,000
3 mnd - Mekonium	0,000	0,000	0,000
6 mnd - Mekonium	0,000	0,000	0,000

Tilleggstabell 4. Resultater fra sammenligningen av den relative mengden av bakteriearter (*B. fragilis*, *B. vulgatus*, *B. dorei*, *B. thetaiotamicron* og *P. distasonis*) mellom mor og barn i ulike alderskategorier ved bruk av Kruskal-Wallis test med Dunn's PostHoc for parvise sammenligninger og Bonferroni korrigererte p-verdier. Gul farge indikerer statistisk signifikante resultat ( $p < 0,05$ ). Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

Sammenligning av mor-barn prøver	<i>Bacteroides fragilis</i> (Adj.Sig.)	<i>Bacteroides vulgatus</i> (Adj.Sig.)	<i>Bacteroides dorei</i> (Adj.Sig.)	<i>Bacteroides thetaiotamicron</i> (Adj.Sig.)	<i>Parabacteroides distasonis</i> (Adj.Sig.)
<b>Kruskal-Wallis test med Dunn's PostHoc og Bonferroni korreksjon</b>					
Mor-12 mnd	1,000	0,000	0,000	0,051	0,000
Mor-3 mnd	0,276	0,000	0,000	0,000	0,000
Mor-6 mnd	0,318	0,000	0,000	0,000	0,000
Mor- Mekonium	1,000	0,000	0,060	0,000	0,000
12 mnd – 3 mnd	0,016	1,000	1,000	0,520	1,000
12 mnd – 6 mnd	0,019	0,042	0,012	0,000	1,000
12 mnd - Mekonium	1,000	1,000	1,000	0,858	1,000
3 mnd – 6 mnd	1,000	1,000	0,149	0,202	1,000
3 mnd - Mekonium	0,003	1,000	0,201	1,000	1,000
6 mnd - Mekonium	0,003	1,145	0,000	0,108	0,696

Tilleggstabell 5. Resultater fra sammenligningen av den relative mengden av bakteriearten *B. longum* mellom mor og barn i ulike alderskategorier ved bruk av Kruskal-Wallis test med Dunn's PostHoc for parvise sammenligninger og Bonferroni korrigerede p-verdier. Gul farge indikerer statistisk signifikante resultat ( $p < 0,05$ ). Statistikk er utført i Statistical Package for the Social Sciences (SPSS).

<b>Sammenligning av mor-barn prøver</b>	<i>Bifidobacterium longum</i> (Adj.Sig.)
<b>Kruskal-Wallis test med Dunn's PostHoc og Bonferroni korreksjon</b>	
Mor-12 mnd	0,001
Mor-3 mnd	0,000
Mor-6 mnd	0,000
Mor- Mekonium	0,016
12 mnd – 3 mnd	0,000
12 mnd – 6 mnd	0,000
12 mnd - Mekonium	1,000
3 mnd – 6 mnd	1,000
3 mnd - Mekonium	0,000
6 mnd - Mekonium	0,000