Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Chemistry, Biotechnology and
Food Science

Master Thesis 2015
60 credits

Effect of maternal probiotic consumption on childrens microbiota and allergy related diseases

Acknowledgements

This master thesis was carried out at the Norwegian university of life sciences (NMBU), department of chemistry under supervision of Professor Knut Rudi and Ph.d student Ekaterina Avershina.

Thanks for the opportunity to take part in new and exciting fields of science and letting me write my thesis at your lab. Also thanks for planning and pointing out structural strategies to realize the thesis. Also a great thanks to the other ph.d students at MiDiv (jane and Anu) for assistance and always being helpful when needed. It has been a pleasure working at MiDiv. You are all passionate about your work.

A big thanks to family and friends for support through thick and thin

Ås, August 2015

Krister Lundgård

Sammendrag

Omfanget av allergirelaterte sykdommer i den vestlige sivilisasjon har eskalert i løpet av de siste tiårene. En positiv korrelasjon mellom sterile miljøer, mindre husholdninger, færre søsken og allergiutvikling har gitt opphav til den foreslåtte hygienehypotesen.

Tarmens bakterieflora har gjennom studier vist seg å ha en innvirkning på allergirelaterte sykdommer. Forslag som administrering av gunstige probiotiske bakterier kan bidra til å forhindre utvikling av allergi, selv om de underliggende mekanismene ennå ikke er forstått. Det finnes lite data om mulig modulasjon av kommensale bakterier ved konsum av probiotika. En viktig begrensning i tidligere studier er oppløsningen i analysen. Ingen nøkkeltaxa som kan være relatert til probiotisk administrasjon og atopisk utvikling er foreløpig funnet. Det har tidligere blitt vist at mors probiotiske forbruk reduserer forekomsten av atopisk eksem hos barn. For å undersøke bakterieprofilen på artsnivå, ble et komplett datasett avledet fra ProPACT studie utført i Trondheim, Norge anskaffet. Par av mor og barn med avføring samplet på flere tidsrammer ble valgt. For å teste hypotesen om at forbruket av en kommersielt probiotisk produkt endrer tarmens microbielle sammensetningen hos barn ble en dybdesekvense tilnærming valgt. 16S rRNA-genet ble valgt som markør og sekvensering ble utført på Illuminaplattformen ved bruk av en Miseq enhet.

En tydelig temporalt mønster ble funnet hos barn i alle aldre. Kompositoriske endringer blant høyt forekommende bakteriearter ble observert. Spesielt blant barn med utviklet atopi i den probiotiske gruppen, selv om tydelige profiler ble observert i alle kategorier. Disse funnene strider i mot tidligere studier på dette materialet, konkluderende med at ingen slike profiler kunne observeres. Observasjoner gjort i denne studien, kan åpne opp et vindu for å forstå samspillet mellom administrert probiotika på kommensal bakterieflora. Styrken i denne studien ligger i det store utvalget av prøver samlet inn over en tidsramme fra ett stort utvalg av mødre og deres barn.

Abstract

Prevalance of allergy related diseases in western civilization has escalated over the last few decades. A positive correlation between more sterile environments, smaller households, fewer siblings and allergy development has given rise to the proposed hygiene hypothesis.

The gut microbiota has through numeral studies been shown to have an impact on allergy related diseases. Suggestions that administrated beneficial bacteria may help prevent development of allergy has been demonstrated, albeit the underlying mechanisms are not yet understood. There are little data on possible modulation of the commensal bacteria by probiotics. A major limitation in previous studies are the resolution of the analysis. To best of knowledge there has not been found any key taxa that can be related to probiotic consumption and atopic development. It has previously been demonstrated that maternal probiotic consumption reduces the incidence of atopic dermatitis in children. To investigate the bacterial profile at species level, a complete dataset derived from the ProPACT study conducted in Trondheim Norway was acquired. Pairs of mother and children with faeces sampled at multiple timeframes were selected. To test the hypothesis that consumption of a commercial probiotic brand alters the gut microbiotic composition in children a deep sequencing approach were chosen. The 16s rRNA gene was chosen and sequencing were performed on the Illumina platform by Miseq bench top device.

A distinct temporal pattern was found in children at all ages. Compositional changes amongst the abundant bacterial species was observed. Most notably were shifts among children with developed atopy in the probiotic group, though categorial profiles categories. These findings contradict previous studies on this material concluding that no such profiles could be observed. Observations made in this study, may open up a window to understanding interaction of administrated probiotics on the commensal gut microbiota. The strength of this study lies in the vast array of samples collected over a timeframe from a large cohort of mothers and their children.

Table of contents

Abbrevations

AD	Atopic dermatitis
Bp	Base pairs
DNA	Deoxyribonucleic acid
Dntp	Deoxyribonucleotide triphosphate
GIT	Gastro intestinal tract
IEC	Intestinal epithelial cell
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
ProPACT	Probiotics in the PACT study
rRNA	Ribosomal ribonucleic acid
SCFA	Short chain fatty acid
igE	Allergen specific immunoglobulin E
T-cell	Thymus cell
TLR	Toll like receptor
NMBU	Norges miljø og biovitenskapelige universitet
GA	Genetic analysis AS

1. Introduction

The beneficial effect of many microorganisms like bacteria and fungi has been exploited for thousands of years, albeit the exact mechanism of beneficial effects has not been well understood. Focus on the role of bacterial symbioses in the human gut has the recent couple of decades escalated. Even though it´s certain that bacteria play an important role in gut health maintenance, the mechanisms remain to be elucidated. Since the middle of the 20.th century, allergy related diseases such as AD, asthma and allergic rhinoconjunctivitis has seen a significant rise among western populations. According to the hygiene hypothesis (Strachan 1989) there are strong indications that atopic symptoms can be directly linked to the human microbiota (Martinez & Holt 1999). Methods to modulate the microbiota through consumption of probiotics for instance is a great field of interest, even though the exact allergy preventive effects remains unclear (Martín et al. 2013). Efforts to relate disease and microbiota have previously been limited by resolution of the analysis, which in turn might lead to neglect important players. New sequencing technologies, facilitates screening of the microbiome at far more extensive levels than practically possible just a few years ago (Loman et al. 2012). This opens up for studies on large cohorts with great analytical depth as well as the opportunity to profile how the dynamics of the gut microbes might correlate to ailments. Monitoring the microbial dynamics of the gut through colonization, composition, interactions and metabolism might hold the key to creating microbiota profiles that can be related to the gut health status.

Bacteria of the human intestinal tract

1.1.1 Development of the normal flora

The commensal bacteria are responsible for a range of tasks such as fermentation of complex polysaccharides (e.g cellulose) to available forms of nutrition such as SCFA (butyrate, acetate, butyrate) and vitamin synthesis including B_{12} (cobalamin, strictly synthesized by bacteria), and vitamin K (menaquinone) (LeBlanc et al. 2011; Topping & Clifton 2001). Humans are thought to be born germ free, although recent studies on *meconium* indicate that

bacteria might be acquired *in utero* (Valles et al. 2014). Either way, it is generally accepted that a crucial part in colonization of the gut microbiota takes place during child delivery and the immediate subsequent time-period, even though understanding establishment of bacterial abundance is yet to be clarified. The mature human gut microbiota consists of $10^{13} – 10^{14}$ bacterial cells; equal to ten times the amount of host cells and is highly distinguishable between individuals. Factors that influence the bacterial composition can be host specific like pH, bile acids, pancreatic enzymes, mucus composition, or host independent including diet, medication, and environmental factors, as well as adhesion capacity, enzymes, and metabolic capacities of the bacteria (Penders et al. 2013). Bacterial load along the gastrointestinal tract varies. In the lower part of the gut the flux is slower, with less secretion of antimicrobial compounds compared with the upper part. Therefore the colon has the highest density of bacteria. All though there is inconsistency in literature, it is predicted that the gut harbours hundreds of taxa (Dethlefsen et al. 2008), but is dominated by a few species (Kraal et al. 2014). The colonization of the infant gut microbiota evolves in stages, reflected through selection based on nutrient availability, genetic and environmental factors. Starting off with primary exposure from the maternal birth canal and sources in the immediate vicinity, delivery mode seems to play a crucial role for selection initially. Children delivered by caesarean section harbours a slightly different and less beneficial bacterial pattern, and a healthy microbiotic basis appears to be inherited from the mothers (Penders et al. 2006). The infant colon is partly aerobic at an early stage and typical early colonizers consists of aerobic and facultative anaerobic bacteria belonging to taxa like *Staphylococcus, Streptococcus, lactobacilli* and *Enterobacteria*, with a rapid skew towards late colonizing, strictly anaerobic bacteria (*Bifidobacterium, Bacteroides, Eubacteria, and Clostridia*) (Palmer et al. 2007). Nutrition and bacterial transfer through breast milk or formula feeding is an important factor for establishment of early colonizers (Salminen & Isolauri 2006). Next after delivery mode and breast-feeding, birth order has also been shown to be an important contributor of bacterial colonization (Penders et al. 2013). *bifidobacterium* rapidly becomes a dominant colonizer as it is introduced through mother's milk or formula as well as selected for by nutrients in the milk. A rapid deacrease in *Staphylococci* from 10 days with *bifidobacteria* and *bacteroides* peaking at 4 months is observed (Vebo et al. 2011). Achieving a microbiota that resembles the mothers is supported by the introduction of solid foods and is observed maturing around two years of age, even though the alfa diversity at this age is shown to be much lower than in comparison with the mothers (Avershina et al. 2014; Isolauri et al. 2004). Efforts have been taken to map the human gut microbiota through microbiome and functional analysis based on

metagenomic data. The Human Microbiome Project found the interspecific microbiota to be resilient but functional stabile while the Metagenomics of the Human Intestinal Tract study discovered that the gut microbiota clustered into three distinct but stable “enterotypes” of interacting microbes (Storro 2013).

1.1.2 Potential role of the gut microbiota in allergy

Bacteria play an important role in stimulating the development and maintenance of the immune system. A dysbiosis in the microbiota is associated with deviations in immune responses such as development of atopic disorders (Storro 2013). The prevalence of immunologic atopic diseases (e.g asthma, AD /atopic eczema) has over the last few decades seen a significant rise in western populations (Von 1998). Children with a family history of atopy are particularly vulnerable with regards to allergy development. However, the rapid increase in atopic incidences over a narrow timeline rules out genetic divergence. Number of siblings, hygiene and general exposure to infections during the first years of life has been shown to give a negative correlation in development of allergy disorders such as AD. In contrast, "sterile" environments and smaller households seem to correlate positively with certain immune deficiency symptoms, which in turn has led to the proposed "hygiene hypothesis" (Martinez & Holt 1999; Strachan 1989). Although atopy is defined inconsistent, in general it is recognized by the presence of elevated levels of total and allergen-specific T and B cell produced immunoglobulin E in the serum, leading to positive skin-prick tests to common allergens. AD is a chronic inflammatory skin disease, where symptoms are often characterized by itchy red rashes, and or lesions in the skin, and it supplements a triade of atopic diseases including allergic rhinitis, asthma and eczema (Hari et al. 2010). Atopic diseases are states of chronic inflammation caused by aberrant T-cell differentiated TLR responses against antigens (Penders et al. 2007). Pattern recognition receptors on the IEC´s including TLR´s recognize molecular patterns conserved among both pathogenic and commensal bacteria. A cascade of immunological responses to eliminate the microbe is initiated. In general the immunological response is restricted to penetration of the IECs, but in the absence of commensal bacteria the mucus layer which protects the IECs from microbial penetration is weakened, and lateral access induce inflammatory response (Kubinak & Round 2012). It has been demonstrated that commensal bacteria are able to induce T-regulatory cell response through symbiosis factors where signalling with secreted polysaccharide A from bacteria down-regulates immunologic T-cell response, and in that way helps preventing inflammation caused by commensal symbiosis (Round et al. 2011). Therefore the presence of

certain commensal bacteria might explain a higher tolerance against inflammation through a certain microbiotic composition. Experiments on gnotobiotic mice showed that a *Clostridia*-containing microbiota modulates the intestinal permeability through mucosa, reducing the uptake of allergens, and thus keeping up the integrity of the mucosa layer can in turn intercept allergens from entering the cell barrier (Stefka et al. 2014). Several studies show that microbiota differs between subjects suffering from allergic conditions and healthy individuals. The idea that there are small or single bacterial groups that promote allergy is however manifested in conflicting results. Reduced intestinal microbial diversity in infancy is hypothesized to correlate with allergy, and exploring dynamic establishment of the infant gut might give insight in to the development of such immunological conditions (Penders et al. 2007; Storro 2013).

In a cohort of Finnish infants, gut microflora differed between healthy subjects and atopic development. Patterns showed early colonization with *Clostridia* and less *Bifidobacterium* clustering together with atopic sensitization (Kalliomaki et al. 2001). Other studies have found negative correlation between levels of *Clostridium difficile,* a common pathogen, and allergy development in small children (Sjogren et al. 2009).

1.1.3 Probiotics and potentials to modulate the gut microbiota

Probiotics are defined as viable microorganisms which when administrated in adequate amounts may provide beneficial health factors to the host (WHO/FAO 2002). The potential of probiotics to prevent development or reduce the severity of certain diseases such as atopy, antibiotic related, acute, and travellers diarrhoea are well documented (de Vrese & Marteau 2007; Dotterud et al. 2010). Probiotics are commonly administrated orally through foods, and especially fermented diary products, although other variants such as capsules exist (Ouwehand et al. 2002). In general it is assumed that a probiotic effect is only achieved while the bacteria reach the target alive, and it is therefore recommended with a cell density of 10^7 per gram or ml. Essential properties of probiotics include resistance to pancreatic enzymes, acid and bile adhesion (Ouwehand et al. 2002). However confusion with regards to the exact mechanism this is accomplished exists. There are several commercially produced probiotic strains and probiotic combinations are common. A number of genera have the potential of probiotic effects, although bacteria belonging to the lactic acid bacteria *Lactobacilla spp.* and the *Bifidobacteria spp*. are best described, also other species such as strains of *Escherichia*

coli and the probiotic yeast *Saccharomyces boulardii* have been reported with beneficial effects (de Vrese & Marteau 2007). *Lactobacillus rhamnosus GG (LGG)* is probably the most frequently studied and widely used probiotic strain. Reported health promoting effects of *LGG* include prevention and treatment of GIT infections and diarrhea as well as stimulation of immune responses that might reduce or prevent certain allergic symptoms (Segers & Lebeer 2014). Administration of the probiotic strain *LGG* to small children suffering from AD has shown elevated serum levels of Interleukin 10, an anti-inflammatory mediator which is known to down-regulate pro-inflammatory IgE (Pessi et al. 2000). There are generally three suggestions for molecular mechanisms of potential microbiota- and immuno modulation by probiotics. Microbe to microbe interactions can through exclusion reject pathogens by competition or secreted substances, but also promote the endogenous commensal microbiota by cooperation for nutrients and cell communication. Second, certain probiotics may enhance the epithelial cell barrier by production of proteins, peptides or other metabolites. Thirdly, induction of Treg cells, or modulation of dendritic cell function through lipoteichoic acid and other cell surface factors as well as certain metabolites might achieve immunomodulation (Lebeer et al. 2008; Segers & Lebeer 2014). The latter point is mainly thought to be effective in the small intestine, as this part of the GIT has a lower bacterial load than the large intestine and thus probiotic strains are more susceptible to bypass exclusion by commensals.

Other possible microbiotic modulators

Fecal transplantation has been suggested as a therapeutic method to restore aberrant microbiota. Bacterial suspension from healthy donors is introduced either through nasogastric tube, enema, colonoscopy or capsules (Hamilton et al. 2013). Reconstituting the normal flora through fecal transplantation has shown to be a promising alternative to antibiotic treatment in patients with relapsing *clostridium difficile* infections (Hamilton et al. 2013; Youngster et al. 2014).

Besides consuming live strains of bacteria for promoting a healthy gut microbiota, more indirect approaches are also possible. Prebiotics are generally non-digestible food ingredients, usually oligo- or polysaccharides that stimulate growth and/or activity of a limited number of beneficial bacteria. Resistance to gastric acidity, hydrolysis by the host, and to gastrointestinal absorption as well as being available as a fermentation substrate by the beneficial intestinal microflora are essential prebiotic properties (Roberfroid 2007). Human milk oligosaccharides are known to harbor prebiotic effects, and in case of formula fed infants, efforts are made to find suitable prebiotic substitutions (Hsieh et al. 2015). Formula fed infants with prebiotics

galacto- and long chain fructo-oligosaccharide supplementation has been shown to decrease the incident of AD compared with placebo. The study conducted by (Moro et al. 2006) also showed a significant increase in *Bifidobacterium* compared to the control group.

1.2 Culture independent methods to analyse the gut microbiota

Early efforts to describe the human gut composition relied on time-consuming techniques often requiring viable bacteria. This could lead to major underestimations, and false conclusions since the vast majority of the gut bacteria are strictly anaerobic and thus difficult to grow *in vitro*, or considered non-culturable. Screening the microbiota without the need of culturing has promising prospects, not only for clinical studies, but also as a diagnostic tool (Loman et al. 2012). Even though the incorporation of high throughput DNA sequencing for detection of bacteria, opens up for a much more un-biased exploration of the microbiome, it is however not without challenges.

Several nucleic acids technologies exists. Common for all is the need for isolation of the organism`s DNA from cell constituents as well as other particles and substances. A qualitative PCR of desired area within the genome is sometimes necessary prior to sequencing as the initial DNA concentration after cell lysis usually is not adequate Amplification of genomic sequences generates a massive amount of identical DNA copies in a 2 fold per cycle manner. A PCR reaction is performed by temperature cycling, where denaturing of the double stranded DNA at high temperature is followed by lowering the temperature allowing primer annealing. Increasing the temperature initiates primer extension by the polymerase and the complementary DNA strand is synthesized (*Figure 1A*). Besides polymerase and oligonucleotide primers, the four dntp´s and buffers are needed for the reaction to succeed (Kubista et al. 2006).

1.2.1 Quantitative polymerase chain reaction

Quantitative PCR (qPCR), sometimes referred to as real time PCR, is a refinement of the classic qualitative PCR method. qPCR allows for simultaneous amplification and detection of desired DNA target without introducing amplification bias known to be a limitation with analysis following PCR. Essentially a fluorescent dye is added to the reaction mixture that can

then be measured during amplification. The fluorescent dye is either target specific or unspecific. The emission is proportional to the amount of formed product, and the amount of DNA molecules present in the initial sample (Kubista et al. 2006). A widely used specific dye is the taq-man probe, which is designed to hybridize to specific oligomers within the desired target. Each time the polymerase encounters the probe, a quencher molecule is cleaved off in turn generating a fluorescent signal (Yu et al. 2005). Unspecific dyes such as Eva-green binds to all DNA, which is an obvious drawback in situations where the detection of specific targets within mixed samples is desirable. The output of qPCR yields an exponential amplification curve, where a threshold value (Ct-value) is set as the number of cycles required to reach a certain fluorescence level (*Figure 1B)*. By including a standard curve, typically a 10 fold dilution series, with known concentration or DNA copy number, calculation of initial products in unknown samples is possible (Kubista et al. 2006).

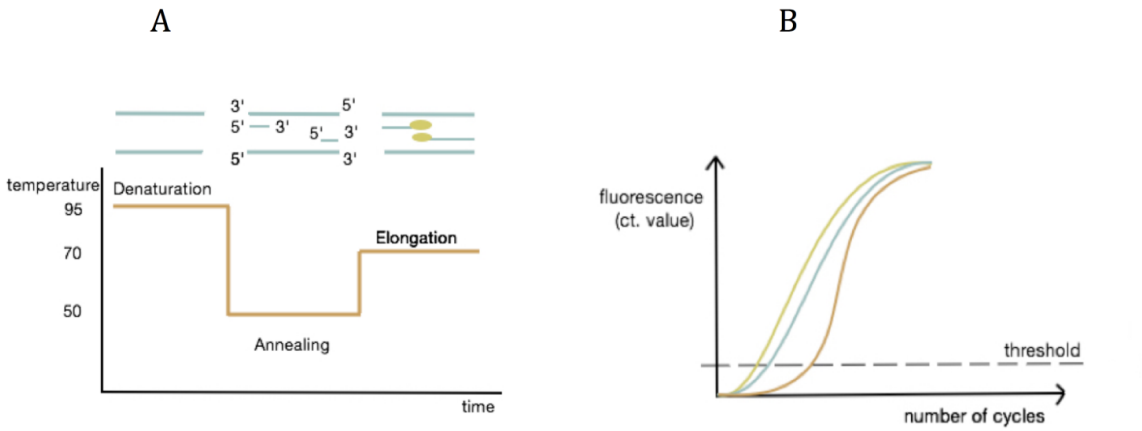


Figure 1: A three step PCR is shown, where DNA is denatured at high temperature, primers annealed by lowering the temperature and finally a heat stable polymerase elongates the primers when increasing the temperature (A). Example of amplification curves from a qpcr. Number of DNA copies is growing exponentially as PCR cycling succeeds, reflected by fluorescence intensity. DNA concentration can be calculated by the threshold ct-value of respective samples (B) (Lundgård, this thesis adapted from (Kubista et al. 2006).

1.2.2 Sequencing technologies

Sanger sequencing was developed in the mid 1970`s, and was the first commercial successful principal of reading DNA (Sanger et al. 1977). As the demand for less labour intensive, higher outputs and lower cost has emerged, so has the progression of new technologies. DNA sequencing is in general the analysis of the order of bases in a string of the genome and has many applications. Some of the fields sequencing is applied to include identification of organisms and taxonomy, gene mapping and mutation analysis.

Next generation sequencing is a huge leap forwards compared to traditional Sanger sequencing with regards to labour and cost. The first commercial available next generation sequencing technology was the 454 pyrosequencing principle produced by Roche. Today there are several competing technologies available, however due to lower cost per sequence and higher sequencing quality, the Illumina platform has recently gained popularity (Nelson et al. 2014). The next generation sequencing platforms differ in the sequencing chemistry, but relies on the principal of "sequencing by synthesis" and the ability to handle millions of reads originating from a diversity of organisms and multiple samples simultaneously (Loman et al. 2012). In general a shared workflow is made up of constructing sequencing libraries from DNA fragments by PCR, clonally amplification to generate clusters of DNA copies and finally the fragments are massively paralleled sequenced by base-calling from the generation of a signal (Mardis 2011).

Sanger sequencing

In automated Sanger sequencing, four different fluorescent dideoxynucleotides (ddntp´s) are added together with the dntp`s normally used for PCR. On ddntp`s the 3` hydroxyl group are removed, and will therefore occasionally terminate the synthesis reaction as it substitutes the dntp´s. The result is many DNA fragments of different lengths with a reporter DNA base on the 3`end. The Fragmented DNA strands are then size-separated on a highly sensitive capillary gel where a laser emits light from the terminal fluorochrome and in turn registered by a detector yielding the DNA sequence. The advance of Sanger sequencing is it´s relatively long reading lengths. The disadvantage is the lack of possibility to do multiplex reading, making it a one sequence per sample tool not suitable for analysis of complex communities (Metzker 2005).

Illumina sequencing

On the illumina, platform the DNA is barcoded with indexprimers that allows a multiple of samples to be sequenced in one reaction. The barcoded primers contain a unique string of typically of a few nucleotides. This serves as a tag that is utilized for sample identification during analysis. Illumina supports paired end reading allowing multiple barcode combinations for sequencing of many samples simultaneously. These primers also include a gene specific region as well as oligomers for bridge amplification and a region for sequencing primer

hybridization. Indexed PCR products are added to a flow cell that contains a solid phase with DNA oligomers complementary to the colony amplification region of the indexed PCR products. Bridge amplification is necessary to achieve sufficient amount of signal in the sequencing step, and results in millions of clusters each consisting of identical DNA strands. After Bridge amplification is performed, sequencing primers are hybridizing to the complementary region of the DNA strands and a massive parallel sequencing reaction is initialized. Fluorescent and complementary reversibly blocked 3`OH termination nucleotides are incorporated by the polymerase. Due to a lack of free 3`OH group, the synthesis is temporarily terminated. A laser emits the fluorescence at wavelength with respect to one of the four incorporated nucleotides, and the signal is imaged. After the nucleotides are base-called, the signal is quenched, 3`OH protection group chemically removed and the process is repeated. The Illumina sequencing technology allows read length of 300 bp paired end read totaling DNA read lengths of total 600 bp (http://www.illumina.com/systems/sequencing.html, 14.04.15). The fact that Illumina sequencing is based upon natural competition between the four bases and error correction by the polymerase yields a low error ratio (Loman et al. 2012; Mardis 2011; Metzker 2010). A generalized overview of the Illumina sequencing is shown in *Figure 2*

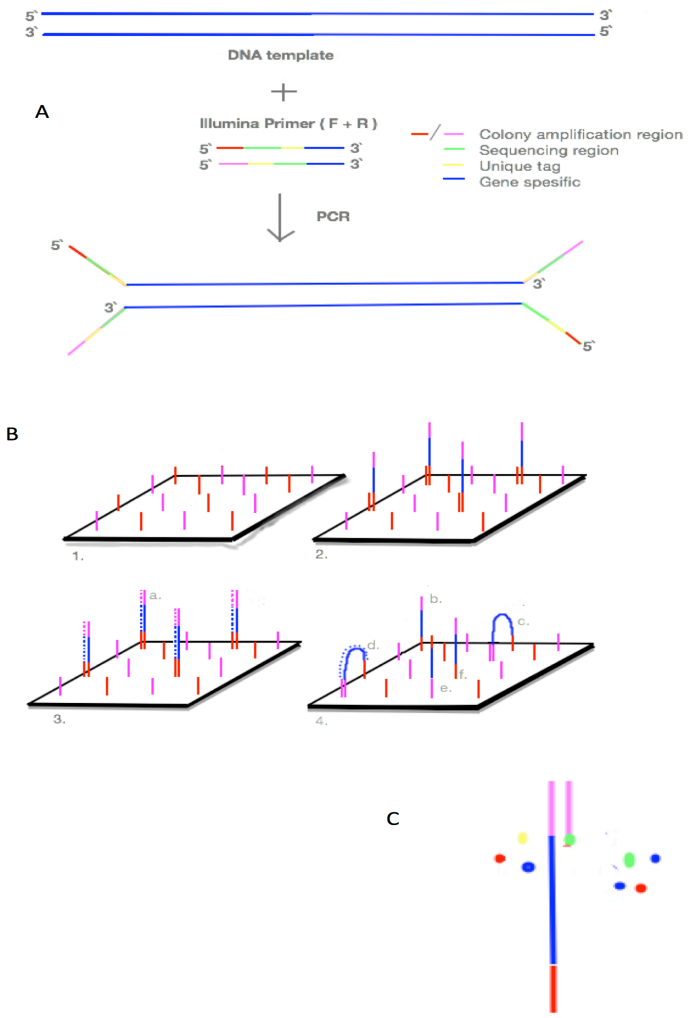


Figure 2: Illumina PCR indexing by barcoded primers (A). Solid phase oligomers on flow cell (B1), indexed PCR products hybridizes to oligomers on the flow cell (B2), synthesis of strand (B3). Original template is washed away (3a - 4b). Bridge amplification generating millions of identical clusters (4 c-f). Reverse strand is removed, barcode read and sequencing starts (C). In paired end read, the process is repeated from step B2 reading the reverse strand.

Other next generation sequencing platforms

There are several other next generation platforms on the marked, although the most important Illumina competitor is probably the 454 pyrosequencing by Roche. The principle of

pyrosequencing is detection pyrophosphate. Prior to sequencing, PCR clustering is performed with the DNA fragments attached to beads in emulsions, spatially separating the DNA clones. The beads are then transferred to micro-wells for sequencing. In the sequencing reaction, one of the four nucleotides is incorporated at a time. As the polymerase incorporates nucleotides, a molecule of pyrophosphate is released. Pyrophosphate is converted to ATP, which is utilized by luciferase to oxidize luciferin in turn generating a light signal imaged by a camera (Margulies et al. 2005). On the Ion torrent platform the release of hydrogen ions after incorporation of nucleotides are detected. A probe hybridization approach using fluorescence makes up the sequencing principle on the SOLiD platform (Loman et al. 2012). Apart from the above mentioned next generation sequencing platforms, single molecule sequencing technologies, generally referred to as third generation sequencing are in the works. As the name implies, these systems are capable of analyzing data from single molecules of DNA in real time rather than by analysis after stepwise nucleotide incorporation (Mardis 2011).

1.2.3 The 16s rRNA gene

Selected targets within the genomes are chosen based on heterogeneity distinguishable between organisms and conservation within groups suitable for priming during PCR. For these reasons the 16s rRNA gene has for several years been utilized to characterize bacterial populations. The bacterial 16s rRNA is located on the ribosomes as a part of the small subunit, and the gene averages about 1500 bp, is strictly found in all bacteria and archaea, is susceptible to mutations in multiple hyper variable regions in the order of >50 to several hundreds bp suitable for barcoding bacteria. Evolutionary conserved regions suitable for primer hybridization flank the variable regions during PCR (Hartmann et al. 2010). Bacterial taxonomic classification by the 16s rRNA gene is not without it`s limitations. Heterogeneity in the variable regions within species is a known problem, and is believed to cause overestimation of the diversity (Sun et al. 2013). Commonly a cutoff at 97-99% similarity at the species level is used. This is however not a universal standard, and not true in all cases. Bias due to multiple gene copies is another that factor might lead to overestimations. The 16s rRNA gene within a species may exhibit 1,5 % divergence, while other species may share as much as 99,5% similarity of the 16s rRNA gene but will only hybridize <50% of the genome, and thus be phenotypically and/or biochemically distinguishable at higher taxonomic levels. (Janda & Abbott 2007; Sun et al. 2013).

Other targets for taxonomic classification

Due to in limitations in divergence of the 16s rRNA gene in closely related populations, protein encoding genes with low rate of horizontal gene transfer as well as a high evolutionary rate has been suggested as alternatives for predicting phylogenetic analysis. Candidates includes *gyrB* (gyrase subunit B), *infB* (translation initiation factor 2), as well as the tRNA encoding 16s – 23s intergenic spacer region (ITS), or combinations of these genes (Namsolleck et al. 2004).

1.3 DNA based analysis

As the capacity to analyze larger sample sets through rapid development of technologies such as the next generation sequencing platforms emerges, so do adequate routines for handling large data sets. On the analytic side, the generation of massive amount of data highlights the need of efficient pipelines as well as computer power that can handle such inputs (Loman et al. 2012). Phylogenetic analysis by second generation sequencing platforms typically handles read lengths ranging from 100 to several hundred bp, and thus exceeding this limit necessitates shotgun sequencing followed by *De Novo* assembly (Loman et al. 2012).

1.3.1 Analytical pipelines

However limitations regarding the 16s rRNA gene, more precise prediction tools, databases and general analytic pipelines are progressively improving to encounter for such uncertainties, (Janda & Abbott 2007). Multiple open source databases containing previously sequenced organisms for bacterial classification based on the 16s rRNA gene exists today. The ribosomal database project (RDP) and greengenes are such online curated classification tools frequently used for analysis of high throughput sequencing data (Larsen et al. 1993). Qiime is an open source pipeline developed for analysis of sequencing data generated on e.g the Illumina platform. Qiime contains default scripts for demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction as well as tools for diversity analysis (J Gregory Caporaso 2010).

1.4 Allergy related studies

The prevention of allergy among children in Trondheim (PACT) is a life-style interventional trial aimed to reduce indoor dampness and tobacco exposure as well as increase the intake of n-3 polyunsaturated fatty acids in a non-randomized control study during pregnancy and the first 2 years after birth. The trial lasted from 2000 through 2004, and encouraged anticipation of all pregnant women and parents in Trondheim of children up to 2 years of age as an effort to reduce development of allergic diseases (Storro et al. 2010).

The proPACT (probiotic in the PACT study) study is targeted as a sub study of the PACT study where a cohort of mothers was given probiotics during last trimester of pregnancy (36. week) throughout the 3 first months after given birth. The aim was to investigate the impact of maternal probiotic consumption on allergy development in children during the first two years of life. The administrated probiotics consisted of *LGG, L. acidophilus La-5* and *B. animalis* subsp. *lactis Bb-12.* The incidence of AD was almost halved in children born from mothers who consumed probiotic milk (Dotterud et al. 2010). ProPACT was followed up by investigating whether administration of probiotics to mothers during pregnancy and breastfeeding period altered the gut microbiota of the mothers and their children. The prevalence of the administrated probiotic strains was significantly higher in the probiotic group albeit only *LGG* was shown to colonize the children at age 10 days and 3 months. No difference between the groups at 1 and 2 years were found. There was also not found any proof that probiotics altered the general microbial composition (Dotterud et al. 2015).

The IMPACT (immunology and microbiology in PACT) was a nested case-cohort study of the PACT study involving children with atopic disease, and randomly selected sensitized and non-sensitized children. The aim was to investigate microbiotic impact on cytokine profiles (Øien et al. 2006). Colonization patterns of *E.coli, B.longum* and *B.fragilis* were detected. At different time points, *B.longum, Enterococcus* and the phyla *Firmicutes* and *Actinobacter* correlated positively with elevated SigE levels. There was however found no correlation in colonization patterns from children with AD (Storro et al. 2011; Vebo et al. 2011).

1.5 Purpose of this thesis

Atopy development in western industrialized countries has seen a significant rise the last years. Apart from affliction caused to subjects suffering from such conditions, states of socioeconomic and medical concerns should be encountered for. It is well documented that the microbiota is an important mediator of immune development. Exploring temporal bacterial gut development as well as potentials to modulate bacterial composition might be a mean of diagnosis as well as for therapeutic purposes.

Maternal impact on the child microbiota is important, and thus utilizing probiotics to potentially modulate the mother-to-child transmission might be a mediator for establishment of a healthy gut. However, there are limited numbers of intervention studies performed on large human cohorts showing the temporal development at high resolution.

The aim of this thesis is therefore to investigate potential bacterial patterns in children diagnosed with AD.

The question to be addressed was if maternal consumption of probiotics could alter the childrens temporal gut microbiota composition where atopy did or did not develop.

Achieving this goal, an approach using deep sequencing of the gut microbiota on the Illumina platform was chosen. Feces samples originated from the ProPACT study, and the 16s rRNA gene was selected as the taxonomic predictor.

Handling large sample sets for sequencing introduces several challenges. Therefore a sub goal was to evaluate and optimize the process.

2. Materials and methods

2.1 Study design and experimental setup

All samples in this thesis originate from mother/children pairs of the ProPACT study. This study was designed as randomized double blinded where subjects were selected by chance, split in two groups; probiotics and a control group. Neither researchers nor participants knew which group was given probiotics or placebo. The children were to be breastfed during the trial period. The probiotics were given to the mothers in form of skimmed fermented milk produced by Tine meierier BA, branded as Biola ®, and contained the commercially manufactured bacterial strains *L. rhamnosus GG* (LGG), *B. animalis subsp. lactis Bb-12* (Bb-12) *and L.acidophilus La-5 (La-5)*. Placebo milk was produced according to Biola but sterilized by heat treatment. Both milks had equal looks and taste, and packed in neutral packaging.

A number of other parameters (family history of atopy, gender, birth weight, breastfeeding, prematurity, parity, maternal age, parental smoking behaviour and pet exposure) were collected as mothers filled out questionnaires on time points 6 weeks, 1 year and two years. Stool samples were collected from the children at the ages of 10 days, 3 months and 1 and 2 years, while maternal stool samples where collected in the last trimester and 3 months post labour. Mother-child pairs assessed to the probiotic group were 138 and equivalent 140 for the control group. For more details on sampling and study design see (Dotterud et al. 2010).

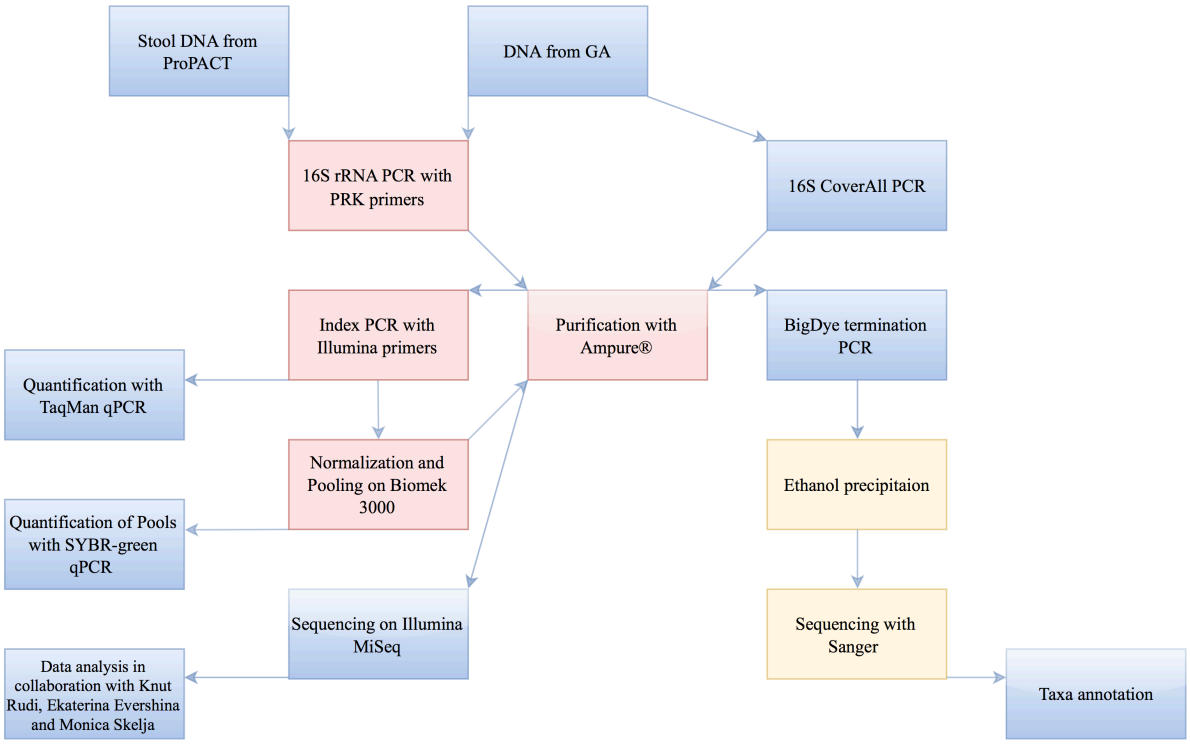


Figure 3: A schematic overview of the sample preparation process is shown. A total of 1516 samples originated from the ProPACT material, while 96 samples came from GA. Red shading indicates that samples were verified on gel electrophoresis and by Qubit® measurements. Yellow shading indicates steps performed by Knut Rudi at Høgskolen i Hedmark.

2.2 Polymerase chain reaction PCR

Two approaches of PCR were utilized in this thesis. Qualitative PCR for generating sufficient amount of targeted amplicons, and quantitative PCR for calculation of product amount. For all PCR setups, reagents where mixed and kept on ice during set-up and either refrigerated or stored at -20 °C after amplification. 96 well PCR plates where used for setup, and a PCR plate centrifuge were used for mixing and collecting droplets.

2.2.1 Qualitative PCR

Prokaryote PCR (PRK PCR)

PRK primers (Appendix 1) targeting the V3-V4 area of the 16s rRNA gene was chosen, resulting in amplicons of 466 bp (Yu et al. 2005). Each reaction contained 1 µl of extracted DNA (unknown concentration), 1,25U HOT FIREpol® DNA polymerase, 1x HOT FIREpol®

buffer B2, 2,5 mM Magnesium-dichloride ($Mgcl_2$), 0,2mM dNTP mix, 0,2µM Forward/Reverse primers (PRK341F, PRK806R *Appendix A*) and PCR grade H_2O to a final volume of 25µl. A three cycling step PCR was run under following conditions: Initialization at 95°C for 15 min, 25 cycles for 30 sec at 95°C and 55°C, 72°C for 1 min denaturation, annealing and elongation respectively. Samples were purified with Ampure®

Index PCR

Construction of the Illumina library pools was carried out with modified PRK primers containing indexes and adapters for Illumina sequencing (*Appendix A*). Purified DNA product from the prokaryote PCR was used. Setup corresponded to the prokaryote PCR protocol, except a FirePol® polymerase was used instead. Each sample was given it´s unique pair of primers as an identification tag. A total of 16 forward and 36 reverse = 576 combinations exists in the library used. The three cycling steps PCR conditions were as follows: Initiation at 95°C for 5 minutes, 10 cycles at 95°C for 30 sec, at 50°C for 1 min, and at 72°C for 45 sec with one cycle at 72°C for 7 min to ensure elongation of all fragments. The resulting amplicons were 594 bp.

Temperature gradient PCR

Evaluation of efficient annealing temperature for uracil modified PRK primers (*Appendix A*) was performed by a temperature gradient PCR. Positive DNA templates as well as negative controls of water were used for the reaction. The annealing temperatures were set to 50 ± 10 °C. Samples were prepared according to the PRK PCR protocol.

CoverAll® PCR

Preparing DNA fragments for Sanger sequencing was done by PCR with CoverAll™ primers (Mangala F-1 forward, 16S 1015U R reverse) targeting the 16S rRNA gene. Each reaction included 1,25U HOT FIREpol® DNA polymerase, 1x HOT FIREpol® buffer B2, 2,5mM $MgCl_2$, 0,2mM dNTPs, 0,2 mM primers, 1µL DNA template (unknown concentrations) and PCR grade H_2O to final volumes of 25µL. PCR conditions were as follows: initiation at 95°C for 15 min, 30 cycles of denaturing at 95°C for 30s, annealing at 55°C for 30s, elongation at 72°C for 1min 20s, finishing with one cycle additional elongation at 72°C for 7min to ensure synthesis of all fragments. The amplified products were cleaned up by the use of Ampure.

2.2.2 Quantitative PCR – Normalization of PCR products

All qPCR reactions were performed in 96-wells LightCycler plates with appropriate sealing on a Lightcycler® 480 system (*Roche*). Ct-values were calculated on the LightCycler 480 software, LinRegPCR and processed in Excel (*Microsoft co.*).

Eva-green

Quantification by Eva-green binding dye was carried out with final concentrations per reaction of 1X Hot FirePol® EvaGreen qPCR Mix 0,2 µM forward and reverse primers, 1 µl (unknown concentrations) DNA template and PCR grade H_2O to final volumes of 25 µL. Plates were spun down and sealed. The forward and reverse primers targeted the colony amplification regions of the indexed PCR products. Thermocycling was performed 95 °C initiating for 15 minutes, followed by 40 PCR cycles at 95 °C denaturing for 15s, 60 °C annealing for 30s and extension at 72 °C for 45s. A high-resolution melting curve was included at the end.

Taq-man

Quantification by Taq-man specific probe was carried out with a 16s rRNA gene targeting probe as described by (Yu et al. 2005). The forward and reverse primers targeted the colony amplification regions of the indexed PCR products. Each reaction contained 1X HotfirePol probe qPCR mix, 0,2 µM primers, 0,1 µM Taq-man probe, 1 µL (unknown concentration) 1:200 diluted DNA template and PCR grade H_2O to final volumes of 25 µL. Thermocycling was conducted with initiation at 95 °C for 15 minutes, followed by 40 PCR cycles of denaturing at 95 °C, and annealing/elongation at 60 °C for 1 min. To evaluate reproducibility between plates, standards were included in each run. A 10 fold dilution series was included in each pool for calculation of copy numbers in unknown samples. Three samples were chosen as standards, measured by Qubit and diluted to concentrations ranging from 10^4 – 10^9 copies/µL. Three parallels were run for each standard. The average Ct-values were used for calculation of the standard curve.

SYBR-green

Quantification of the pooled index PCR products where accomplished according to the Quanta Biosciences™ protocol for PerfeCta® NGS library quantification kit for Illumina sequencing platform. Fragment length of the included standards deviated from the indexed

samples, and size was therefore corrected for when calculating the concentration. Calculated DNA concentrations from standard curve were compared with Qubit measurements.

2.3 Purification of amplified PCR products

Cleaning up amplified DNA templates following PCR products is crucial. Unincorporated nucleotides and primers, short fragments of primerdimers as well as other contaminants must be removed from the samples. Contaminants might lead to inflate quantification, as well as inhibit preceding PCR and sequencing reactions. In large sample sets this step is quite labor intensive. Template recovery by different methods was therefore evaluated. For all methods the purified products were assessed by gelelectrophoresis and by Qbit measurements

Ampure®

Magnetic separation of DNA fragments after PCR was performed by a modifications to the original Ampure® purification protocol (Beckman Coulter). A script was made on a Biomek 3000 (Beckman Coulter) laboratory automation robot. This script allowed for automating all the steps of the Ampure protocol. According to manufactory recommendations the ratio of Ampure® to template was 1:1. This would exclude DNA fragments of length shorter than 200 bp. 10 µl of Ampure® bead solution was added to each sample containing 10µl of DNA template, and mixed by pipetting. After 5 minute incubation, the plate was moved to the magnet and incubated for another 2 minutes for binding of the magnetic beads. The supernatant is then removed and thrown away. Adding 100µl of 80% ethanol followed by 30 seconds incubation before removed and thrown away performed washing steps. The wash step was repeated twice. After the ethanol throwaway the samples were air dried for 30 minutes before removed from the magnet and added 20 µl of PCR graded H2O used as elution liquid. The samples were incubated for 2 minutes and placed back on the magnet, where they were incubated for 5 minutes for separation of eluate and beads. The purified DNA was finally transferred to a new PCR plate and stored at -20 °C. Calculating the ratio pre and post clean up based on Quit measurements as well as fragment size confirmation by gel electrophoresis did verification of purified products. Purification after library pooling was carried out adjusting the ratio of Ampure to DNA concentration to 0,8:1 to compensate for increased fragment length. DNA input was 150 µL, while elution was done in 40 µL tris buffer (10mM, pH 8,5).

Uracil glykosylase

Modified PRK primers with uracil pyrimidine substitution for thymine or cytosine on the 3`ends (*Appendix A*) were implemented to the DNA fragments by PRK PCR. Master mix contained 0,1 U uracil glycosylase, 1X Reaction buffer, 1 µL PCR template (unknown concentration) and PCR grade H_2O for volume adjustment to a total of 10 µL/reaction was prepared. A two-step heat reaction with 30 minutes at 37 °C enzyme activation and cleavage of the N-glycosylic bond of the uracil at 95 °C for 15 minutes was initiated. Also a catalyst in form of spermine tetrahydrocloride was added in a concentration of of 0,3 mM/reaction to evaluate increasing cleavage efficiency.

Dilution

The effect of diluting the PCR products prior to index-PCR was compared with cleaning up by Ampure®. Two parallels of 48 samples were either diluted 1:100 with PCR graded H_2O, or purified according to the Ampure® protocol. Index PCR with one index pair was used to amplify the products. Verification was done by qPCR and gel electrophoresis.

2.4 Qualitative confirmation of amplified PCR products

Gelelectrophoresis

Verification of DNA fragments were carried out on 1% agarose gel with 1X TAE running buffer. Gelred loading dye was used for visualization with 1:2 gelred to loading buffer were prepared for working solutions. 5 µL DNA was added to 3 µL of working solution, and 5 µL applied to the gel. Small gels migrated ran for 20 minutes, while large gels ran for 50 minutes. Current was 100V (small gels) and 150V (large gels). 100 bp size-ladder template as well as controls were included. The fragments where visualized by UV-light on Gel Doc™ XR Imaging (Bio-Rad laboratories, USA).

2. Qubit™

DNA concentrations were done by fluorometric Qubit™ measurements using the Quant-iT™ assay. Working solutions were prepared according to manufactures´ recommendation where 2 µL DNA was added to a mix of 1:200 Quant-iT™ reagent and Quant-iT™ buffer respectively, to a total volume of 200 µL. Measurements were executed in a Qubit™

fluorometer. For calibration of the instrument, 10 µL of two standards were added to 1:200 working solution.

2.5 Normalization and pooling indexed PCR products

Pooling samples in equal concentrations is crucial prior to high throughput sequencing. Normalization is necessary to achieve a uniform cluster density when samples are loaded on the chip. Copy-number of the indexed PCR products were calculated in excel from the given Ct-values based on the standard curve from the included dilution series during Taq-man qPCR. The upper part of the Ct-values was set to correspond to 10 µL template. Generally this was expressed as *Equation 1:* $Volume = \frac{Upper\ threshold\ copy\ number * 10\ \mu L}{Copy\ number\ in\ sample}$

Maximum/minimum volume per sample was set to 1 and 10 µL. Samples that failed to fall within this area, was added in volumes corresponding to the nearest limit. A script for automatically pooling equal sample concentrations were made on a Biomek 3000 (Beckman Coulter) laboratory automation robot. The robot was fed the excel sheet that contained volumes as well as information on target and destination.

2.6 Sequencing

Illumina

Preparation of the indexed and pooled samples for sequencing was carried out according to the "Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system" protocol (Illumina 2011). Modifications with regards of spiking the samples with PhiX, was increased to 15% to ensure appropriate cluster separation. Final DNA chip loading concentration was 6pM. Sequencing was carried out on a MiSeq located at NMBU by the use of Nextera® MiSeq Reagent Kit v3 (Illumina).

Sanger

A collection of DNA templates used for control purposes were provided by Genetic analysis (GA) and sequenced using the Sanger method. A BigDye labeling reaction was performed on the CoverAll amplified products. Forward and reverse strands were sequenced separately and thus two reactions were run in parallel with forward and reverse primers respectively. Each

reaction contained 5x BigDye sequencing buffer, BigDye Terminator v1.1 (1µL), 3,2µM sequencing primers (see CoverAll PCR), 1µL DNA template, and H_2O to a total volume of 10µL. The termination PCR conditions were as follows: Activation at 95°C for 1 min, 25 cycles denaturation at 96°C for 1 min, annealing at 50°C for 5s and annealing/elongation at 60°C for 4 min. Precipitation by ethanol used for DNA clean up as well as sequencing was carried out at Høgskolen i Hedmark.

2.7 Data analysis

Qiime

Sequences generated from Illumina sequencing of the 16S rRNA gene was extracted and annotated using the Qiime pipeline (conducted by Monika Skelja). The raw sequences were quality filtered, paired end reads joined, combined with metadata and de-multiplexed based on barcode information. Operational taxonomic units (OTUs) were generated based on 99% homology with taxonomy assigned from the greengenes database. Diversity analysis was calculated from the samples based on OTUs and metadata.

CLC main workbench (Qiagen) was used to filter the sequences as well as joining the forward and reverse reads after Sanger sequencing. Taxonomic annotation re-annotation was done by through the Ribosomal Database Project ab score >99,5 % identity (RDP (Wang et al. 2007)), and Blast at > 99% identity (*NCBI*)

Statistical testing was executed in Excel (*Microsoft Co.*) unless stated otherwise. Generally, all P-values <0,05 rejected the null hypothesis. Testing of variance between the respective groups was executed by the use of the two-sided independent t-test assuming unequal variances.

3. Results

3.1.1 16s rRNA gene analysis

The samples were sequenced on the Illumina Miseq in three batches, where each pool contained 480 – 576 uniquely barcoded samples. Clustering densities were 456 ± 70 K/mm^2, with Q30 scores of 88.5 ± 0.5 % percent respectively. A total of 11.2 ± 1.7 million reads were generated. Normalized data yielded a total of 21 158 715 sequences ranging between 9 and 16 9024, with a mean value of 13 316 ± 19 751. 1605 samples total were sequenced including controls and propact samples. A total of 1336 OTUs were generated. Number of successfully sequenced propact samples are summed up in *table 1*.

Table 1: Number of successfully sequenced samples (76 %) in probiotic and atopy groups are summed up in table 1.

	Biola and atopy	Biola, no atopy	Placebo and atopy	placebo, no atopy	
Mothers pregnant	22	83	56	66	
Mothers 3 months	18	61	39	57	
Child 10 days	17	65	30	64	
Child 3 montths	22	74	29	66	
Child 1 year	24	82	40	70	
Child 2 years	29	65	18	57	N = 1154

3.1.2 Diversity of probiotic and atopic groups

Diversity of the the samples were explored by the use of α-diversity calculation through Qiime. These data were used to make rarefraction plots. Number of observed species were plotted against number of sequences. The results are given in *Figure 3*.

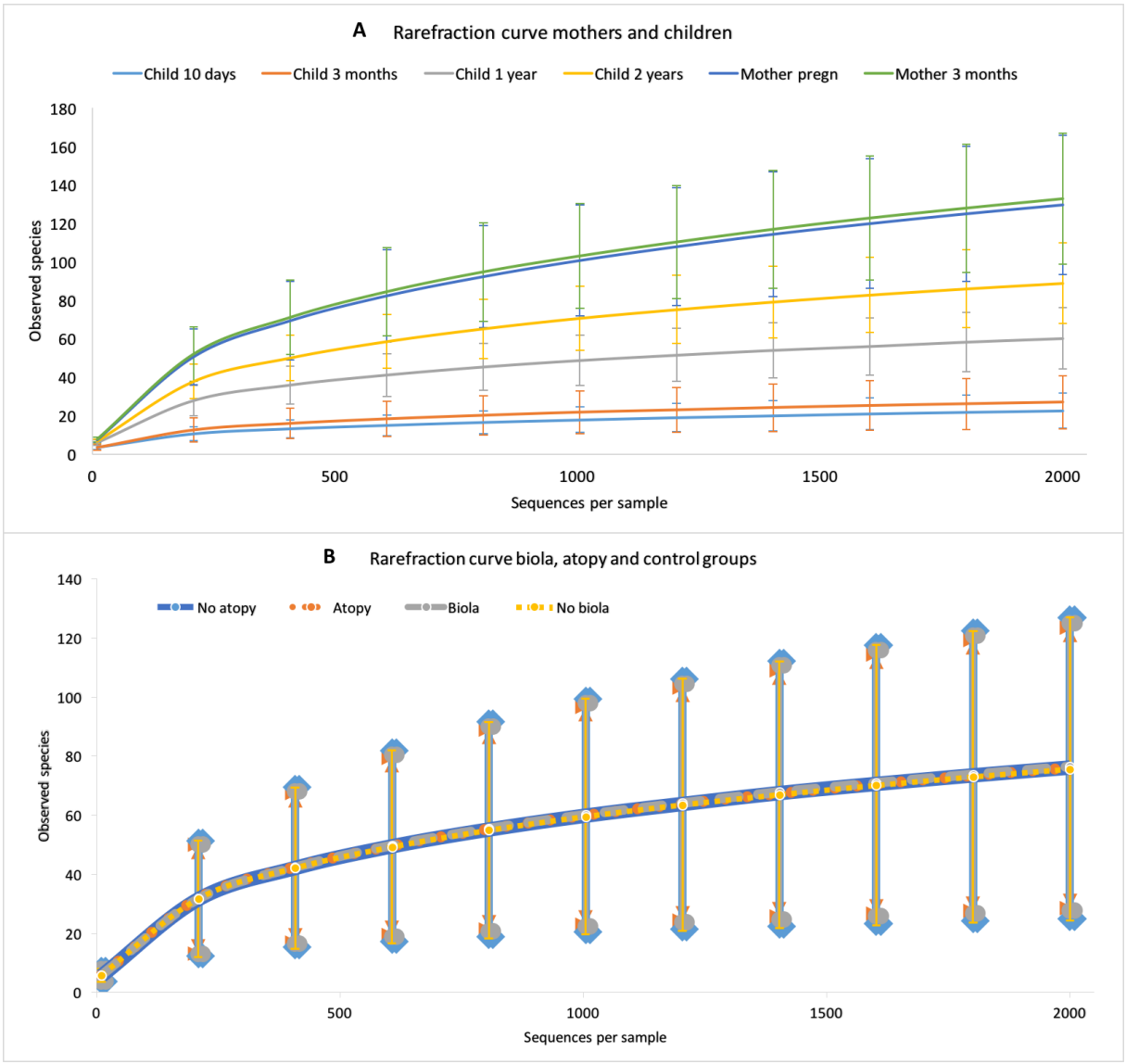


Figure 4: The rarefraction curves show observed observed species in response to sequences per sample (mean values ± sd), between age groups (A), and between biola and atopy groups (B).

Figure 3A shows that species richness is much lower for newborns, but increases over the timeline. At 2 years of age, a diversity resembling the maternal is not reached. No difference could be observed for the atopic and biola groups as shown in *Figure 3B*.

The beta diversity data was collected from the Qiime weighted unifrac analysis and analysed using the emperor package yielding a principal component analysis plot. Clustering was uniform biola- placebo and atopy- no atopy groups. Distinct patterns appeared for the age groups where children at 2 years clustered together with maternal samples. Results are given in Appendix B

Investigating dominant class of bacteria within the biola- placebo (*conducted by Ekaterina Avershina in* MATLAB® 2014a software (MathWorks Inc., USA). Atopy- no atopy groups were done by extracting data from the Qiime output and processed in Excel (*Microsoft, USA*) gave a general overview of taxa within the respective groups. A non parametric Kruskal-Wallis test approach was used to calculate p.values by the use of R (*R development core team*) with the NMBU R commander plugin (*Kristian Hovde Liland*). Bar charts yielding the 10 most abundant bacterial classes were made. Results are given in *Figure 4.*

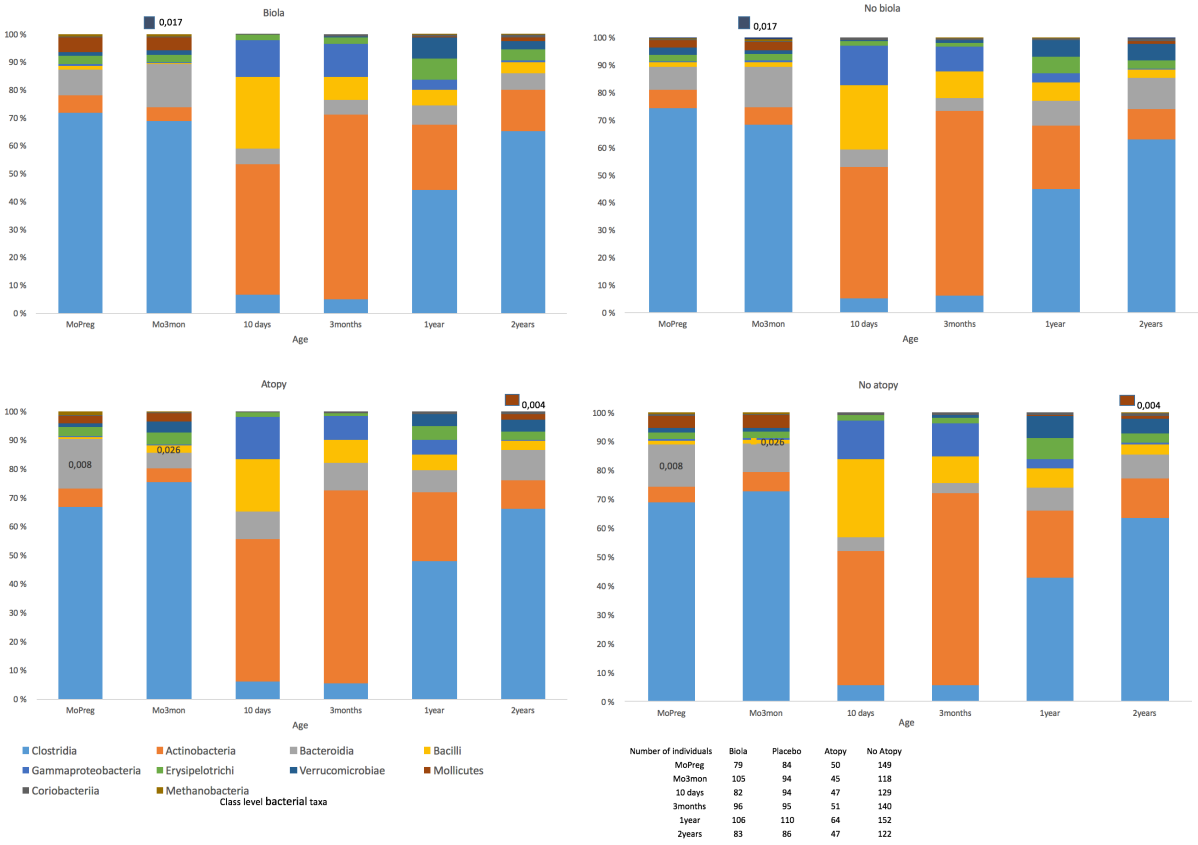


Figure 5: The bar charts shows 10 dominating class of bacteria in biola and placebo groups as well as atopy and no atopy groups. P.values are given for the taxa that are significantly different (<0,05) of the respective categories.

The microbial communities are highly distinguishable at different stages with regards to age. The Actinobacteria class was dominated by genus of bifidobacteria comprising a about half the gut microbial composition at 10 days and 3 month old children (Appendix B). Two year olds have, at class level, a composition that resembles the maternal (*Figure 4*). One distinct class were significantly different in the biola-placebo group while several were detected in the atopy- no atopy group (p.values <0,05), though only molicutes were found significantly

different in children regardless of category. A complete table of bacterial classes with corresponding p.values are given in *Appendix C.*

3.1.2 Diversity at species level

To research differences at species level, testing of significant different OTUs in the biola and atopy groups were conducted by comparing the variance using the two-sided t.test, assuming unequal variance. Groups of children were sorted on age and categorized on biola or placebo, and further sub categorized on atopic development or not. Subtracting the mean values between the groups gave the direction of OTU abundance. The relative number of OTUs was not considered. Species with significant p.values <0,05 and within the upper limit of the rarefraction curves (ranging from 32 – 167 species) within the respective age groups were collected and. Most OTUs had to be reassigned using the *rdp* and blast databases. Results are given in *Table 1.*

Table 2: Significant different species in respective categories are shown. Categories biola and placebo are sub categorized in atopy positive (+) or negative (-). P. values (<0,05) of respective categories are given. Rank gives the order of specie abundancy in the categories.

Rank	P.value	Biola Atopy +/-	Placebo Atopy +/-	OTU	Average difference sequences	Identity	Species (Phyla; Class; Order; Family; Specie)
7	0,001	+		46	1176		p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; s__Bifidobacterium dentum
19	0,016		+	302	152	Child 10 days	p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; s__Streptococcus mitis
27	0,011	+		109*	123		p__Firmicutes; c__Bacteroidetes; o__Bacteroidales; f__Bacteroidaceae; s__Parabacteroides distasonis
30	0,039	+		40	6		p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; s__Bacteroides uniformis
2	0,048	-		2	949		p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; s__ Bifidobacterium breve
22	0,046	+		11	267	Child 3 months	p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; s__ Veillonella ratti
44	0,047		+	1071*	43		p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; s__Parabacteroides distasonis
2	0,012		-	20	184		p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; s__Clostridium ramosum
4	0,023		+	3	310	Child 1 Year	p__Actinobacteria; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; s__Bifidobacterium catenelatum
68	0,040	+		158	26		p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; s__Roseburia_faecis
30	0,011	+		22	59		p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; s__Alistipes putredinis
64	0,020	+		49	30		p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; s__Bacteroides caccae
78	0,046	+		122	131	Child 2 years	p__Tenericutes; c__Mollicutes; o__RF39; f__; g__; s__ *
98	0,036		+	187	18		p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; s__Blautia obeum
99	0,042		+	81	11		p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; s__Clostridium_xylanolyticum

With two exceptions, all species given in *Table 1* were positively associated with atopy. Children in the biola group had more significant species compared to placebo. One OTU

(*Mollicutes* RF 39) could not be identified with high confidence at specie level. Closest Blast hit yielded at identy 83% *Spiroplasma apis* (accession NR_121708.1) within the Mollicutes class. *Parabacteroides distasonis* were assigned to two OTUs (*) and the only specie observed over multiple timepoints*;* positively related with atopy in the biola group at 10 days and placebo group at 3 months.

To investigate whether certain species would show patterns throughout the timeline, significant different OTUs in the groups were matched with regards to age. Analysis was performed on the complete dataset. The results are given in *Figure 6.*

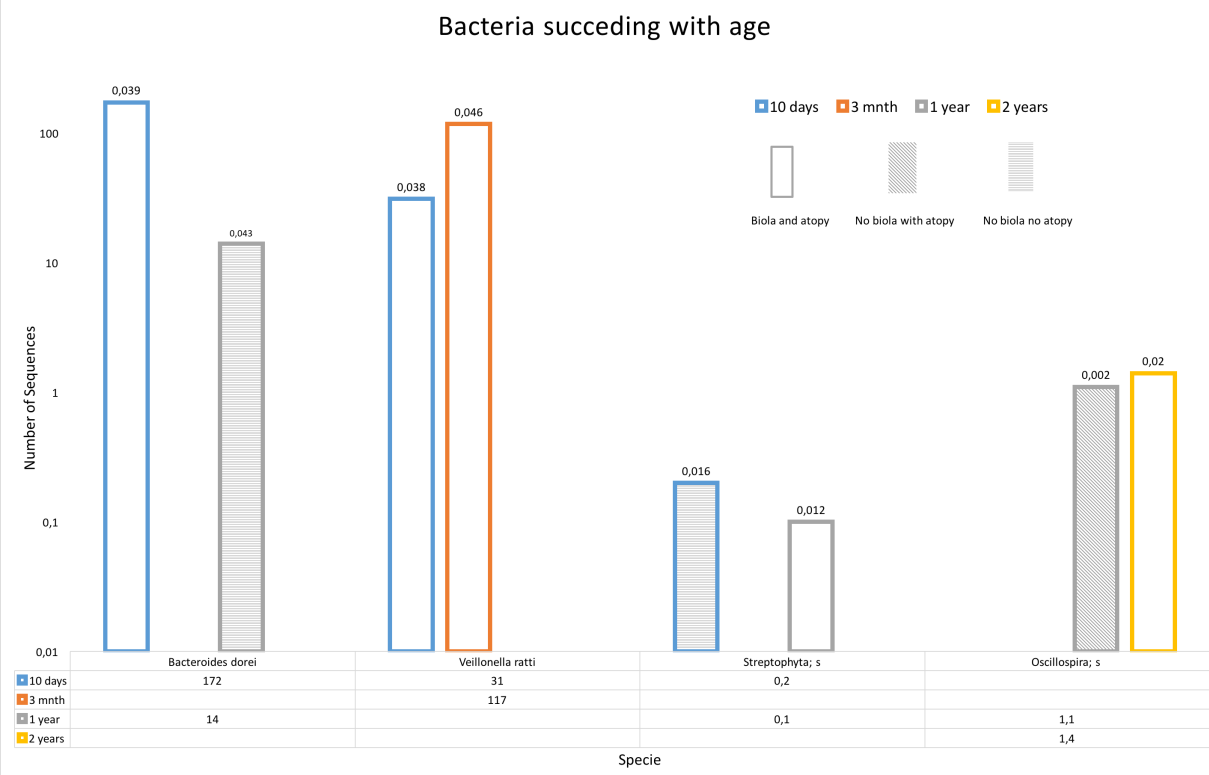


Figure 6: The boxplot shows bacterial species with a p.value <0,05 that gives a pattern between age groups. Mean difference in sequences are given between the groups. P.values are placed outside the bars. The Y-axis is logarithmic to compensate for divergent data. The number of sequences are shown with the x labels.

The plot in *Figure 6* shows that certain species is prevalent at different ages. Except for *Veillonella ratti,* there seems to be little systematics in direction with regards to atopy/biola categories.

To investigate any higher order systematics, all significant OTUs, regardless of abundancy were summed up and categorized at phylum level. The result is given in *Figure 7*.

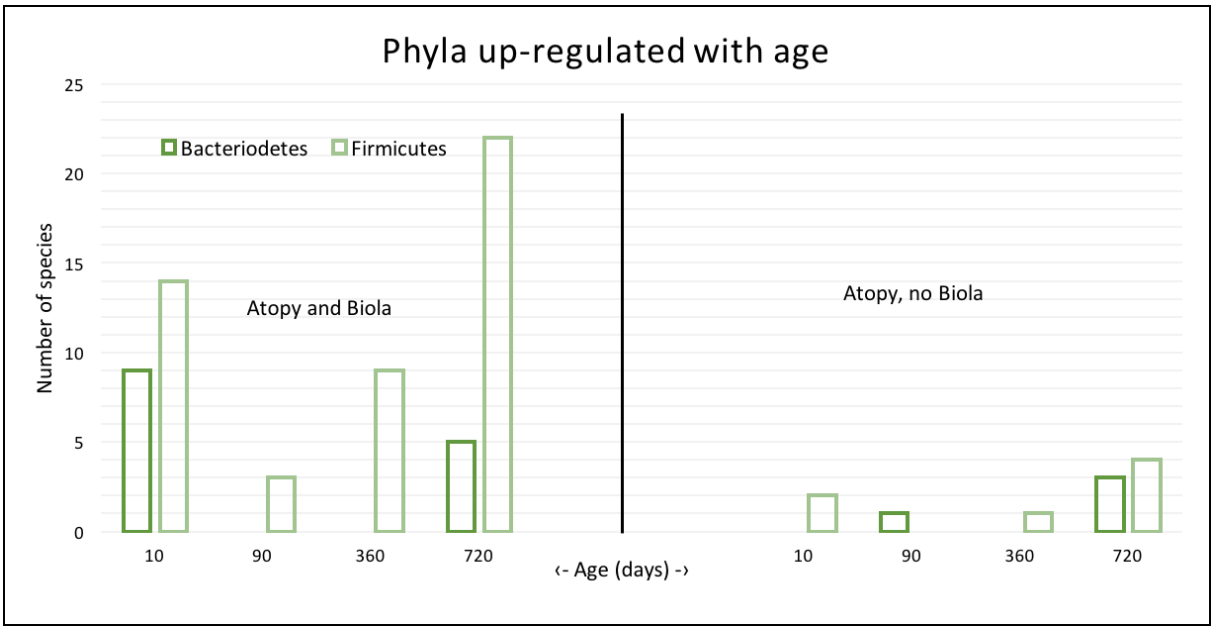


Figure 7: The table shows number of OTU`s within the phyla *Firmicutes* and *Bacteriodetes* that are significant different on p.values <0.05. The numbers represent up regulation of species within the respective phyla at the given age.

Higher prevalence of Firmicutes was observed for children at most ages.

3.2 Methodological optimization and technical evaluation

3.2.1 Automation of sample preparation

To minimize the hands on sample preparation, automation of purifying samples were considered prior to index PCR. Attempt to cleave primer contamination using uracil substituted primers incorporated during the first amplification step did not yield satisfactory results (results not shown). A method for performing automatic purification by the use of a Biomek 3000 laboratory automation robot was therefore developed. The general steps can be seen in Appendix D. Gelelectrophoresis of a test run consisting of 48 parallel PCR amplified samples (purified versus non purified) was performed. Scoring matrices were made based on band intensity of either product or generation of undesired primerdimers. A graphical illustration is shown in Figure 8A.

Normalization of indexed PCR products was performed by comparing two qPCR methods. Parallels of purified and non purified PCR products were quantified by Evagreen and TaqMan real time PCR. It was desirable to evaluate the specificity of EvaGreen versus TaqMan qPCR methods.

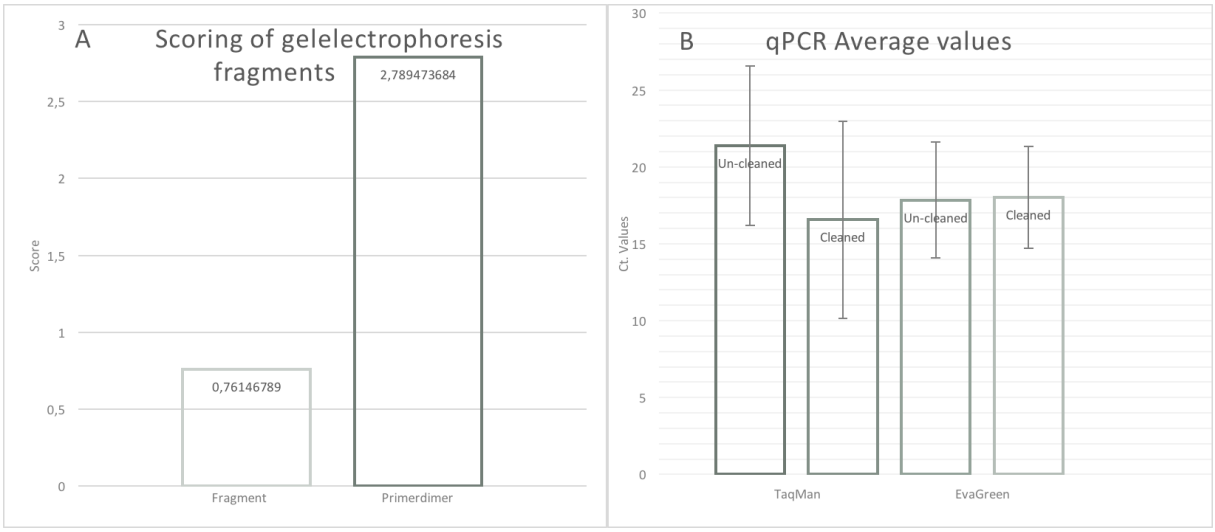


Figure 8: The plots shows evaluation of 48 parallel, ampure purified or not purified respectively, samples after indexed PCR. Subjective scoring of band intensity from gelelectrophoresis is given as ratios in figure A. Mean Ct. values (± sd) are given for the two qPCR methods (B).

Generation of primerdimers shown in *Figure 8A* from purified samples were due to failed amplification of some products. For successful amplified products, primerdimers were basically non-existent. Divergent reproducibility as shown through the standard deviation in *Figure 8B* were to be expected, as samples were not normalized prior to quantification. For the EvaGreen method (*Figure 5B*) there was due to the unspecific binding dye basically no difference in Ct.values between purified and non-purified samples. The Taq man method was chosen for calculation of sample concentration prior to normalization.

Furthermore reproducibility of the sequencing was evaluated by constructing libraries from a selection of bacterial species DNA controls. Three different libraries were constructed from a collection of 88 samples (4 single species, 88, 30 and 10 samples respectively). Three parallels were mixed in a ratio of 1:1 based on Qbit measurements after purification and prior to index PCR. Results are given in *Figure 10 and 11*.

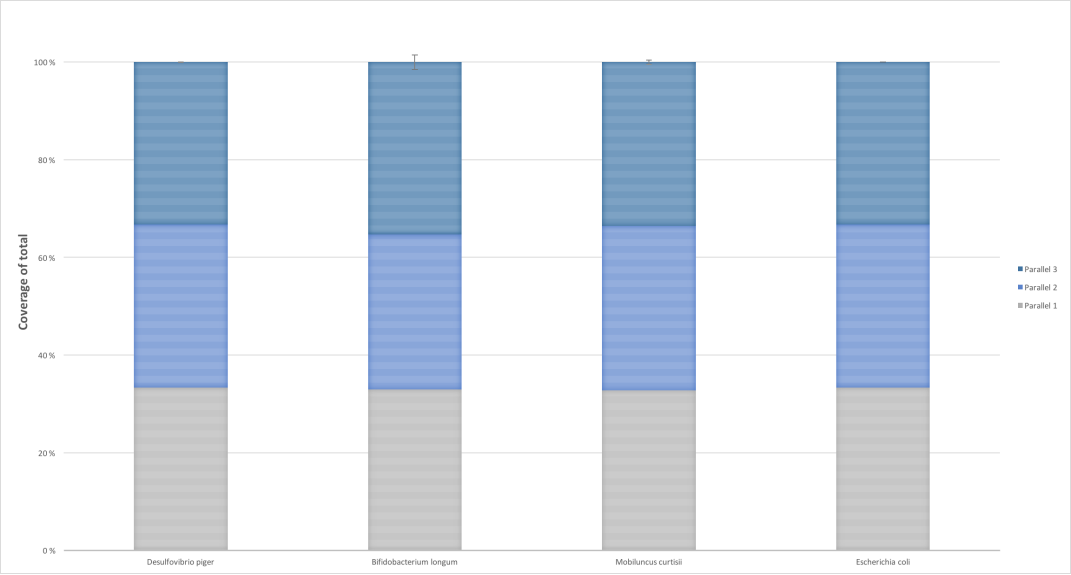


Figure 9: Recover of four species sequenced in triplicates. Mean (± sd) percentage are given.

All four bacterial species were highly reproducible after Illumina sequencing (*Figure 10).*

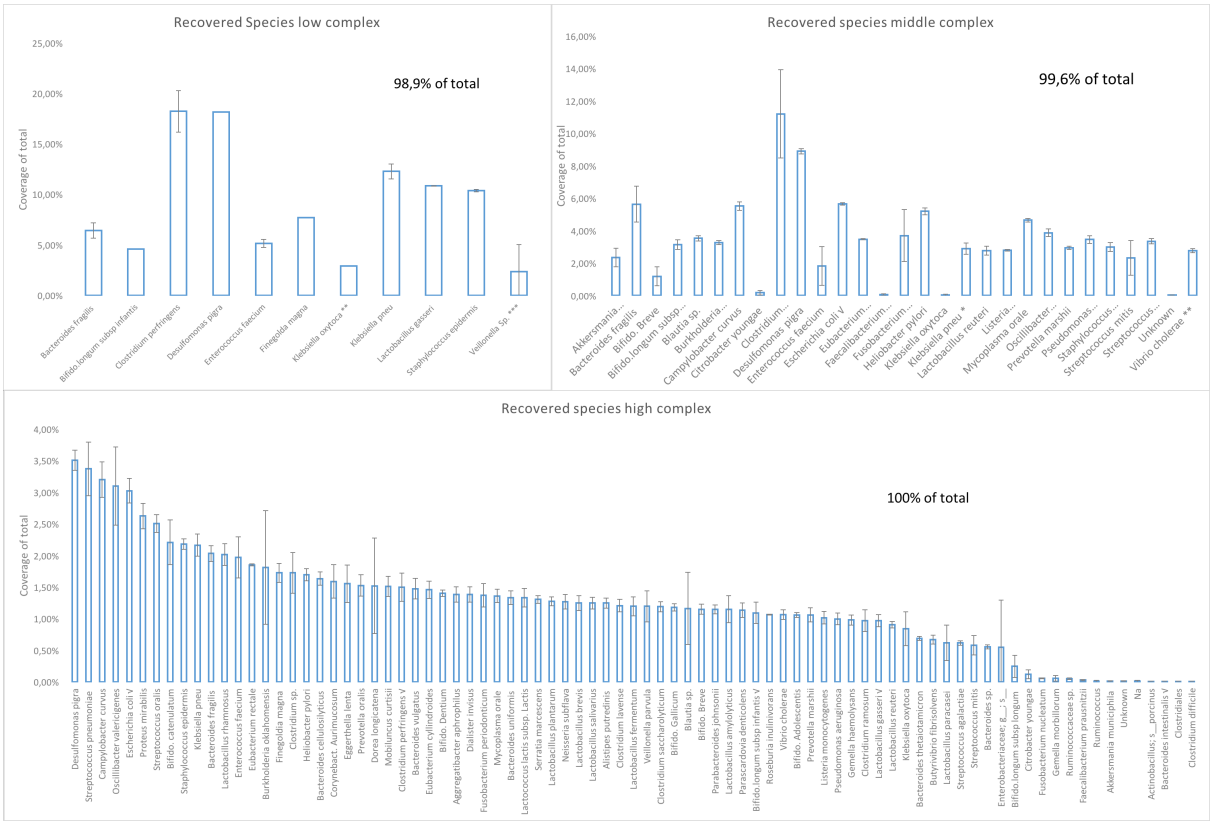


Figure 10: Graphs showing replicates of included controls during Illumina sequencing. Average percentage seqeunces per OTU of each control out of total OTU´s are given as well as the ratio of all control sequences to the total number of sequences.

The graphs in *Figure 11* shows different amplification efficiency for different bacterial species. For some species there seems to be a lower reproducibility than others as reflected through the standard error of the mean. Not all species could be reproduced from the Illumina sequencing. 90% of the low abundance sequences making up totally 98,9% were found. Recovery rate of 93% making up totally 99,6% percent of total sequences were found in the middle complex samples. 82% recovery rate from of the high complex samples making up 100% of total sequences. A complete table is given in Appendix G.

4. Discussion

4.1 Diversity of gut microbiota related to age

It is observed that the complexity of the gut microbiota increases over time. Acquired genuine species seems to have a greater leap between the age of three months to one year and two years than between 10 days and three months. Species richness is made up of an average between 20 to 120 unique taxa over the timeline. This supports recent suggestions that reports on individual bacterial species of the gut has previously been overestimated {Avershina, #104}. An overrepresentation of bifidobacterium is apparent in 10 days and 3 month old children, consistent with previous findings {Avershina, 2014 #48}. Interindividual gut microbiota composition seems to be cluster together at a greater degree for children at 10 days and 3 months as well as the maternal, than at one and two year of age. A more individual specific bacterial profile at late maturation therefore seems to be the case. Despite that the class level composition at two years closely resembles the maternal gut microbiota, the microbiotic richness observed in the maternal microbiota is not met. This suggests that the children's complex microbiota is not fully developed at two years and that sub populations are yet to be recruited.

4.4 Effect of probiotics on the microbiota and atopy

No difference in species richness could be observed for biola or placebo groups regardless of atopic development. Individual outliers were observed, but generally clustering was uniform at any given age. For the 10 most abundant bacterial class level taxa, mollicutes, comprising a total bacterial load of 2% in the atopic group, and 0,1% in the non-atopic group were significantly different for children at 2 years of age between healthy and affected individuals. Comparing childrens gut microbiota at species level indeed revealed significant difference between children in the respective groups. Eight unique species were observed with increased abundance within the atopy-biola group, whereas *Bifidobacterium breve* was the only bacteria showing negative association with atopy and biola consumption. Interestingly two species of *Bifidobacterium* correlated positively with and atopy development. *B.dentium* at 10 days of age were found to be the seventh most abundant bacteria in the biola- while *B.catenulatum* in one year old children from the placebo-atopy group ranked as the fourth most abundant bacteria. The fact that these species are overrepresentated in children with atopy development

might suggest a counter beneficial effect of these certain species. It is also feasible that due to their dominance that they might have a negative effect on recruitment of other beneficial species. On the contrary, ranking as the second most abundant specie in children at 3 months of age, *B. breve* was detected to be negatively associated with atopy in the biola group. It has previously been demonstrated that *B.breve* and *B. dentium* are among the dominating Bifidobacterial species of the human gut microbiota {Avershina, 2013 #105}. The pattern of *Bifidobacterium* colonization in this dataset seems to be temporal, and variations may have modulatory effects on atopy development. The observed class of mollicutes at two years of age in atopic children, was identified in the biola group. It was not successful to identify organism at specie level with high confidence. However, along more dissimilar sequences, blast search aligns it´s OTU to *Spiroplasma apis.*
The only observation made at multiple time points amongst the highly abundant species, *Parabacteroides distasonis* was associated with atopy and biola at 10 days and atopy and placebo at 3 months. It therefore seems that bacterial profiles at the respective ages in individuals susceptible for atopic development is quite distinct.

Since it cannot be ruled out that certain niche species might play an important role bacterial succession with regards to age was investigated on a full dataset. Four species in the low abundance microbiota was observed, however no species occurred at all ages. Only *Veillonella ratti* persisted within the same category (Biola and atopy) at 10 days and 3 months. There was found no observations for the Biola no atopy category. Together this gives further strong indications that age is an important factor considering how the microbiota responds to biola and the development of atopy.

To investigate succession at higher taxonomic level, all significant OTUs was compared and gathered in the two dominating phyla. The overall dominance of number of species within the two phyla were in the atopic group regardless of treatment, although more significant species were found in the biola group. In general Firmicutes dominated over Bacteriodetes regardless of age among children with atopy, with more evident differences in the biola group.

4.2 Technical validation

Included single species triplicates showed good reproducibility from the Illumina sequencing. Deviations in more complex samples were observed. This might be due to competition during the amplification step. As a note, the samples should probably have been normalized after index PCR not before.

Triplicates during Illumina sequencing showed acceptable results. Most species were identified, but most of the OTUs had to be re-assigned taxonomy. Some of the samples had too poor quality to be aligned to reference or was not identified at lower taxonomic level.

Conclusion

Maternal administrated probiotics seems to have an impact on childrens microbiota. At specie level, composition of core bacteria in children is altered. The microbiotic pattern is most evident in children with atopic development. The fact that biola is correlated positively with altered microbiota and atopy might suggest that biola are non-beneficial for these subjects. The impact of the altered microbiota is here shown to be highly temporal. The high abundance of altered key taxa suggests that bacterial profiles at species level might predict atopic development.

Since this dataset do not contain absolute quantification of bacterial species, following up these findings by the use of a target specific qPCR might reveal a more correct composition with regards to the total bacterial load of the gut microbiota. It would also be interesting to do a functional metagenomic profile on these species to map a metabolic pattern.

References

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P. & Rudi, K. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol*, 87 (1): 280-90.

de Vrese, M. & Marteau, P. R. (2007). Probiotics and Prebiotics: Effects on Diarrhea. *The Journal of Nutrition*, 137 (3): 803S-811S.

Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol*, 6 (11): e280.

Dotterud, C. K., Storro, O., Johnsen, R. & Oien, T. (2010). Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial. *Br J Dermatol*, 163 (3): 616-23.

Dotterud, C. K., Avershina, E., Sekelja, M., Simpson, M. R., Rudi, K., Storro, O., Johnsen, R. & Oien, T. (2015). Does Maternal Perinatal Probiotic Supplementation Alter the Intestinal Microbiota of Mother and Child? A Randomized Controlled Trial. *J Pediatr Gastroenterol Nutr*.

Hamilton, M. J., Weingarden, A. R., Unno, T., Khoruts, A. & Sadowsky, M. J. (2013). High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. *Gut Microbes*, 4 (2): 125-35.

Hari, A., Flach, T. L., Shi, Y. & Mydlarski, P. R. (2010). Toll-like receptors: role in dermatological disease. *Mediators Inflamm*, 2010: 437246.

Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W. & Nilsson, R. H. (2010). V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *J Microbiol Methods*, 83 (2): 250-3.

Hsieh, C.-C., Hernández-Ledesma, B., Fernández-Tomé, S., Weinborn, V., Barile, D. & de Moura Bell, J. M. L. N. (2015). Milk Proteins, Peptides, and Oligosaccharides: Effects against the 21st Century Disorders. *BioMed Research International*, 2015: 146840.

Illumina. (2011). *Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system*. http://www.illumina.com. Available at: (http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf (accessed: 27.11).

Isolauri, E., Salminen, S. & Ouwehand, A. C. (2004). Microbial-gut interactions in health and disease. Probiotics. *Best Pract Res Clin Gastroenterol*, 18 (2): 299-313.

J Gregory Caporaso, J. K., Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Pena, Julia K Goodrich, Jeffrey I Gordon, Gavin A Huttley, Scott T Kelley, Dan Knights, Jeremy E Koenig, Ruth E Ley, Catherine A Lozupone, Daniel McDonald, Brian D Muegge, Meg Pirrung, Jens Reeder, Joel R Sevinsky,

Peter J Turnbaugh, William A Walters, Jeremy Widmann, Tanya Yatsunenko, Jesse Zaneveld and Rob Knigh. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 2010*, 7: 2.

Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*, 45 (9): 2761-4.

Kalliomaki, M., Kirjavainen, P., Eerola, E., Kero, P., Salminen, S. & Isolauri, E. (2001). Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing. *J Allergy Clin Immunol*, 107 (1): 129-34.

Kraal, L., Abubucker, S., Kota, K., Fischbach, M. A. & Mitreva, M. (2014). The prevalence of species and strains in the human microbiome: a resource for experimental efforts. *PLoS One*, 9 (5): e97279.

Kubinak, J. L. & Round, J. L. (2012). Toll-like receptors promote mutually beneficial commensal-host interactions. *PLoS Pathog*, 8 (7): e1002785.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L., et al. (2006). The real-time polymerase chain reaction. *Mol Aspects Med*, 27 (2-3): 95-125.

Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R., Macke, T. J., Marsh, T. L. & Woese, C. R. (1993). The ribosomal database project. *Nucleic Acids Research*, 21 (13): 3021-3023.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev*, 72 (4): 728-64, Table of Contents.

LeBlanc, J. G., Laino, J. E., del Valle, M. J., Vannini, V., van Sinderen, D., Taranto, M. P., de Valdez, G. F., de Giori, G. S. & Sesma, F. (2011). B-group vitamin production by lactic acid bacteria--current knowledge and potential applications. *J Appl Microbiol*, 111 (6): 1297-309.

Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., Robinson, E. R. & Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol*, 10 (9): 599-606.

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470 (7333): 198-203.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437 (7057): 376-80.

Martín, R., Miquel, S., Ulmer, J., Kechaou, N., Langella, P. & Bermúdez-Humarán, L. G. (2013). Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. *Microbial Cell Factories*, 12: 71-71.

Martinez, F. D. & Holt, P. G. (1999). Role of microbial burden in aetiology of allergy and asthma. *The Lancet*, 354: SII12-SII15.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Res*, 15 (12): 1767-76.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet*, 11 (1): 31-46.

Moro, G., Arslanoglu, S., Stahl, B., Jelinek, J., Wahn, U. & Boehm, G. (2006). A mixture of prebiotic oligosaccharides reduces the incidence of atopic dermatitis during the first six months of age. *Arch Dis Child*, 91 (10): 814-9.

Namsolleck, P., Thiel, R., Lawson, P., Holmstrøm, K., Rajilic, M., Vaughan, E. E., Rigottier-Gois, L., Collins, M. D., de Vos, W. M. & Blaut, M. (2004). Molecular methods for the analysis of gut microbiota. *Microbial Ecology in Health and Disease*, 16 (2-3): 71-85.

Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L. & Graf, J. (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One*, 9 (4): e94249.

Øien, T., Storrø, O. & Johnsen, R. (2006). Intestinal microbiota and its effect on the immune system — A nested case-cohort study on prevention of atopy among small children in Trondheim: The IMPACT study. *Contemporary Clinical Trials*, 27 (4): 389-395.

Ouwehand, A., Salminen, S. & Isolauri, E. (2002). Probiotics: an overview of beneficial effects. *Antonie van Leeuwenhoek*, 82 (1-4): 279-289.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. (2007). Development of the Human Infant Intestinal Microbiota. *PLoS Biol*, 5 (7): e177.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P. A. & Stobberingh, E. E. (2006). Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *Pediatrics*, 118 (2): 511-521.

Penders, J., Stobberingh, E. E., van den Brandt, P. A. & Thijs, C. (2007). The role of the intestinal microbiota in the development of atopic disorders. *Allergy*, 62 (11): 1223-36.

Penders, J., Gerhold, K., Stobberingh, E. E., Thijs, C., Zimmermann, K., Lau, S. & Hamelmann, E. (2013). Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood. *J Allergy Clin Immunol*, 132 (3): 601-607 e8.

Pessi, T., S\'fctas, Y., Hurme, M. & Isolauri, E. (2000). Interleukin-10 generation in atopic children following oral Lactobacillus rhamnosus GG. *Clinical & Experimental Allergy*\, 30\ (12\): 1804\-1808\.

Roberfroid, M. (2007). Prebiotics: The Concept Revisited. *The Journal of Nutrition*, 137 (3): 830S-837S.

Round, J. L., Lee, S. M., Li, J., Tran, G., Jabri, B., Chatila, T. A. & Mazmanian, S. K. (2011). The Toll-Like Receptor 2 Pathway Establishes Colonization by a Commensal of the Human Microbiota. *Science*, 332 (6032): 974-977.

Salminen, S. & Isolauri, E. (2006). Intestinal colonization, microbiota, and probiotics. *The Journal of Pediatrics*, 149 (5): S115-S120.

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

Segers, M. E. & Lebeer, S. (2014). Towards a better understanding of Lactobacillus rhamnosus GG--host interactions. *Microb Cell Fact*, 13 Suppl 1: S7.

Sjogren, Y. M., Jenmalm, M. C., Bottcher, M. F., Bjorksten, B. & Sverremark-Ekstrom, E. (2009). Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clin Exp Allergy*, 39 (4): 518-26.

Stefka, A. T., Feehley, T., Tripathi, P., Qiu, J., McCoy, K., Mazmanian, S. K., Tjota, M. Y., Seo, G. Y., Cao, S., Theriault, B. R., et al. (2014). Commensal bacteria protect against food allergen sensitization. *Proc Natl Acad Sci U S A*, 111 (36): 13145-50.

Storro, O., Oien, T., Dotterud, C. K., Jenssen, J. A. & Johnsen, R. (2010). A primary health-care intervention on pre- and postnatal risk factor behavior to prevent childhood allergy. The Prevention of Allergy among Children in Trondheim (PACT) study. *BMC Public Health*, 10: 443.

Storro, O., Oien, T., Langsrud, O., Rudi, K., Dotterud, C. & Johnsen, R. (2011). Temporal variations in early gut microbial colonization are associated with allergen-specific immunoglobulin E but not atopic eczema at 2 years of age. *Clin Exp Allergy*, 41 (11): 1545-54.

Storro, O. A., E ; Rudi, K. (2013). Diversity of intestinal microbiota in infancy and the risk of allergic disease in childhood. *Current Opinion in Allergy & Clinical Immunology.*, 13(3) (June 2013): 7.

Strachan, D. P. (1989). *Hay fever, hygiene, and household size*, vol. 299. 1259-1260 pp.

Sun, D.-L., Jiang, X., Wu, Q. L. & Zhou, N.-Y. (2013). Intragenomic Heterogeneity of 16S rRNA Genes Causes Overestimation of Prokaryotic Diversity. *Applied and Environmental Microbiology*, 79 (19): 5962-5969.

Topping, D. L. & Clifton, P. M. (2001). *Short-Chain Fatty Acids and Human Colonic Function: Roles of Resistant Starch and Nonstarch Polysaccharides*, vol. 81. 1031-1064 pp.

Valles, Y., Artacho, A., Pascual-Garcia, A., Ferrus, M. L., Gosalbes, M. J., Abellan, J. J. & Francino, M. P. (2014). Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants. *PLoS Genet*, 10 (6): e1004406.

Vebo, H. C., Sekelja, M., Nestestog, R., Storro, O., Johnsen, R., Oien, T. & Rudi, K. (2011). Temporal development of the infant gut microbiota in immunoglobulin E-sensitized and nonsensitized children determined by the GA-map infant array. *Clin Vaccine Immunol*, 18 (8): 1326-35.

Von, M. (1998). The rising trends in asthma and allergic disease. *Clinical & Experimental Allergy*, 28: 45-49.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*, 73 (16): 5261-7.

WHO/FAO. (2002). *Guidelines for evalution of probiotics in food*. FAO/WHO. London, Ontario, Canada. 11 pp.

Youngster, I., Russell, G. H., Pindar, C., Ziv-Baran, T., Sauk, J. & Hohmann, E. L. (2014). Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection. *JAMA*, 312 (17): 1772-8.

Yu, Y., Lee, C., Kim, J. & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng*, 89 (6): 670-9.

APPENDIX

Appendix A: Primers of this thesis

PRK primers targeting prokaryotic 16S reran gene:

Forward (PRK341F): 5 ` CCTACGGGRBGCASCAG 3`

Reverse (PRK806): 5 ` GGACTACYVGGGTATCTAAT 3`

PRK Illumina primers:

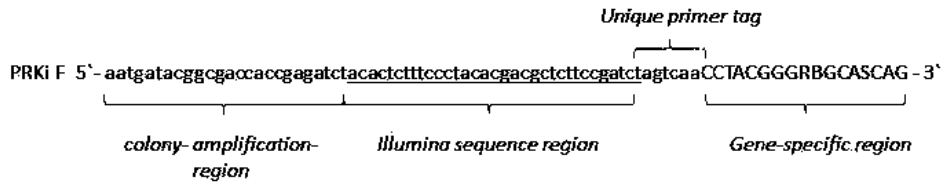
Forward (PRKi F):

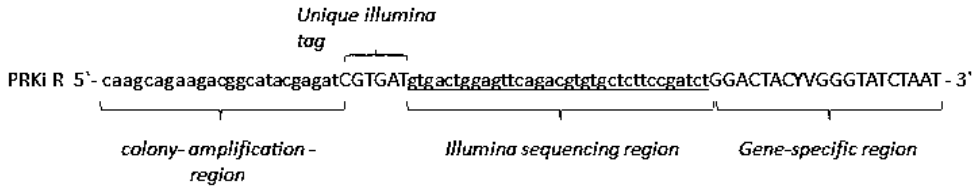
01. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCCTACGGGRBGCASCAG
02. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccCCTACGGGRBGCASCAG
03. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaCCTACGGGRBGCASCAG
04. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccCCTACGGGRBGCASCAG
05. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagCCTACGGGRBGCASCAG
06. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcCCTACGGGRBGCASCAG
07. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaCCTACGGGRBGCASCAG
08. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccCCTACGGGRBGCASCAG
09. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgCCTACGGGRBGCASCAG
10. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgCCTACGGGRBGCASCAG
11. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggCCTACGGGRBGCASCAG
12. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcCCTACGGGRBGCASCAG
13. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatCCTACGGGRBGCASCAG
14. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcCCTACGGGRBGCASCAG
15. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctCCTACGGGRBGCASCAG
16. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagCCTACGGGRBGCASCAG

Reverse (PRKi R):

01. caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
02. caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
03. caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
04. caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
05. caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
06. caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
07. caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
08. caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
09. caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
10. caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
11. caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
12. caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
13. caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
14. caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
15. caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
16. caagcagaagacggcatacgagatGGACGGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
17. caagcagaagacggcatacgagatCTCTACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

18. caagcagaagacggcatacgagatGCGGACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
19. caagcagaagacggcatacgagatTTTCACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
20. caagcagaagacggcatacgagatGGCCACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
21. caagcagaagacggcatacgagatCGAAACgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
22. caagcagaagacggcatacgagatCGTACGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
23. caagcagaagacggcatacgagatCCACTCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
24. caagcagaagacggcatacgagatGCTACCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
25. caagcagaagacggcatacgagatATCAGTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
26. caagcagaagacggcatacgagatGCTCATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
27. caagcagaagacggcatacgagatAGGAATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
28. caagcagaagacggcatacgagatCTTTTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
29. caagcagaagacggcatacgagatTAGTTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
30. caagcagaagacggcatacgagatCCGGTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
31. caagcagaagacggcatacgagatATCGTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
32. caagcagaagacggcatacgagatTGAGTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
33. caagcagaagacggcatacgagatCGCCTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
34. caagcagaagacggcatacgagatGCCATGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
35. caagcagaagacggcatacgagatAAAATGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
36. caagcagaagacggcatacgagatTGTTGGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT





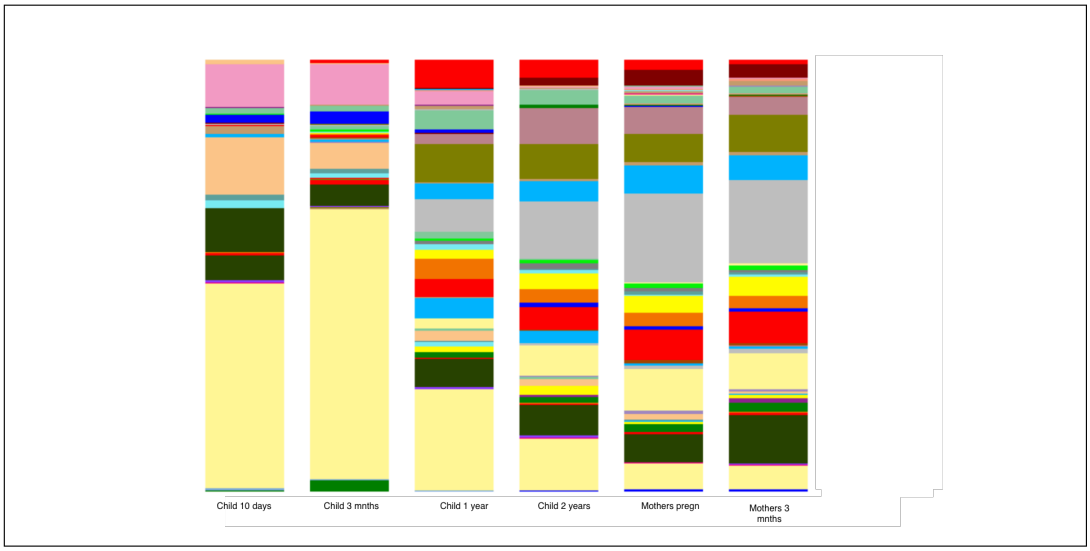
PRK primers with uracil substitution

Forward (PRK341F): 5`CCTACGGGRBGCASUAG 3`
Reverse (PRK806): 5`GGACTACYVGGGTATCUAAT 3`

Cover All™ Primers

Forward: 5`TCCTACGGGAGGCAGCAG 3`
Reverse: 5`CGGTTACCTTGTTACGACTT 3`

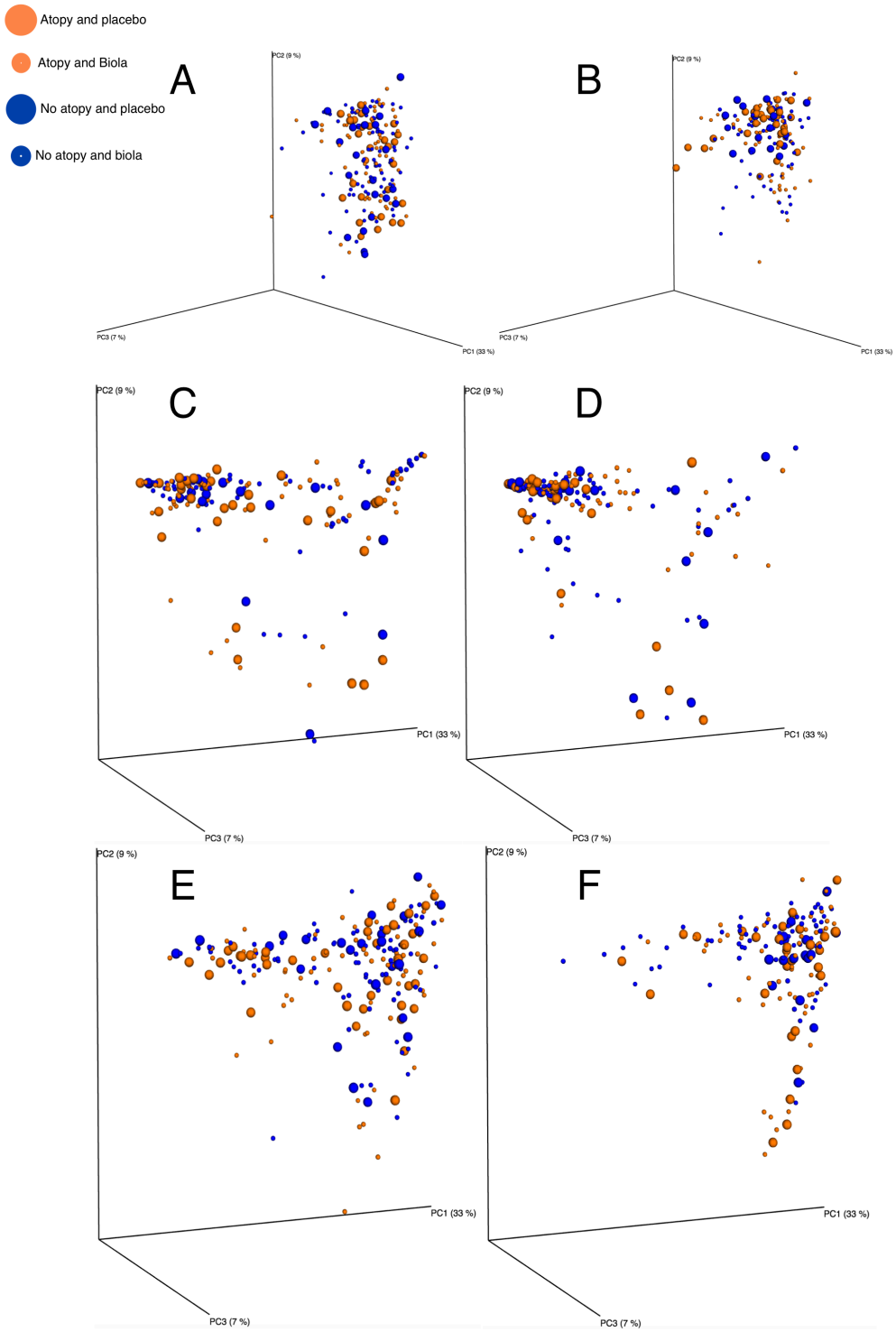
Appendix B: Genus level bacterial taxa at all ages. Only dominant genus are shown in the legend

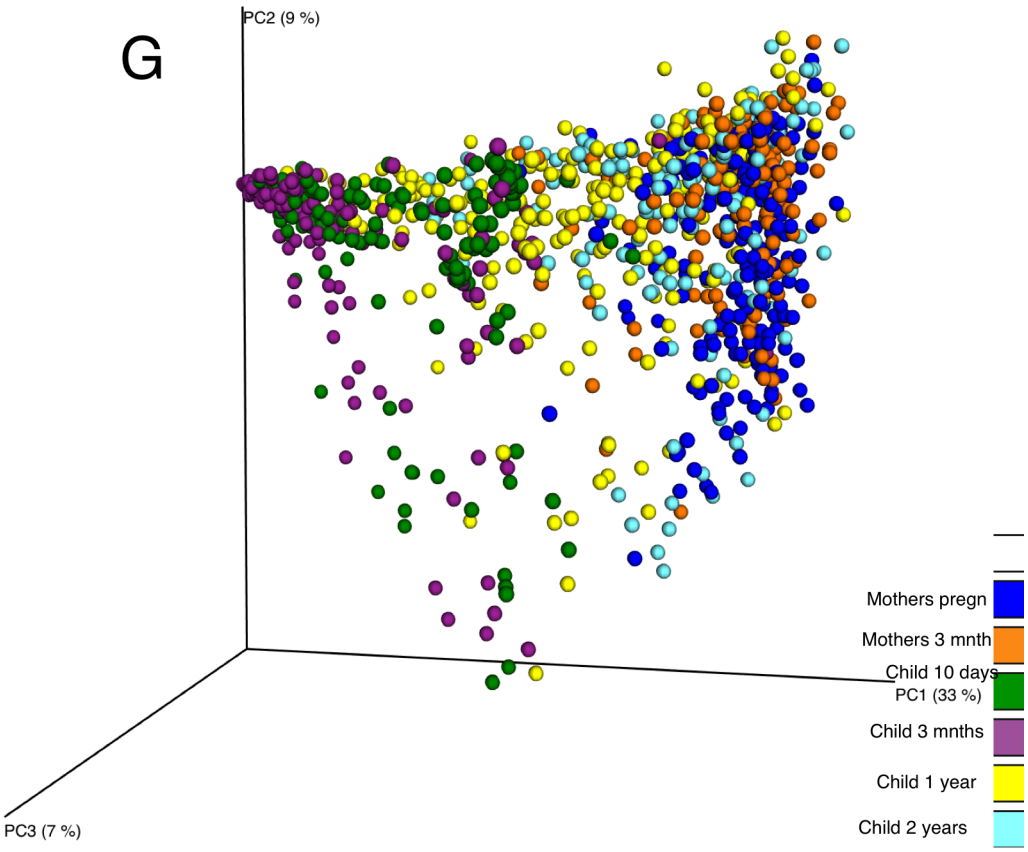


View Table (.txt)

Legend	Taxonomy	Total count	Total %	10 %	90 %	360 %	720 %	.100 %	.200 %
	Unassigned;Other;Other;Other;Other;Other	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Archaea;p__Euryarchaeota;c__Methanobacteria;o__Methanobacteriales;f__Methanobacteriaceae;g__Methanobrevibacter	0	0.2%	0.0%	0.0%	0.0%	0.2%	0.5%	0.5%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinobaculum	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces	0	0.5%	0.3%	2.5%	0.1%	0.1%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus	0	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Varibaculum	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermabacteraceae;Other	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermabacteraceae;g__Brachybacterium	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Leucobacter	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Kocuria	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Rothia	0	0.1%	0.4%	0.2%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;Other	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium	2	25.6%	47.5%	62.6%	23.5%	11.8%	5.9%	5.4%
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__	0	0.2%	0.1%	0.1%	0.1%	0.4%	0.2%	0.3%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium	0	0.0%	0.0%	0.2%	0.1%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella	0	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Eggerthella	0	0.3%	0.6%	0.3%	0.5%	0.3%	0.1%	0.1%
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Slackia	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Actinobacteria;c__Nitriliruptoria;o__Nitriliruptorales;f__Nitriliruptoraceae;g__	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides	1	7.2%	5.6%	5.1%	6.4%	7.1%	6.5%	11.2%

Appendix C: PCoA plots shows mothers pregnant (A), mothers 3 months (B), children 10 days (C), children 3 months (D), Children 1 year (E) and children 2 years (F). Diversity at all ages (G)




G
PC2 (9 %)
PC1 (33 %)
PC3 (7 %)
Mothers pregn
Mothers 3 mnth
Child 10 days
Child 3 mnths
Child 1 year
Child 2 years

Appendix D:

Atopy							No Atopy						
Class	MoPreg	Mo3mon	10 days	3months	1year	2years	Class	MoPreg	Mo3mon	10 days	3months	1year	2years
Clostridia	0,661546	0,742423	0,060508	0,052971	0,479839	0,659357	Clostridia	0,684487	0,721885	0,055995	0,055709	0,427444	0,632402
Actinobacteria	0,0642	4,96E-02	0,494069	0,671631	0,238782	0,099322	Actinobacteria	0,055681	0,067192	0,465438	0,664163	0,230058	0,138638
Bacteroidia	0,17124	0,050694	0,097406	0,096746	0,074941	0,103336	Bacteroidia	0,143562	0,099192	0,046707	0,033557	0,080194	0,079917
Bacilli	0,008035	0,026083	0,179585	0,077592	0,054772	0,030867	Bacilli	0,012831	0,012129	0,268523	0,093348	0,065254	0,036415
Gammaproteobacteria	0,001599	0,003171	0,147968	0,082751	0,049248	0,004574	Gammaproteobacteria	0,005196	0,005989	0,133148	0,112376	0,030717	0,005675
Erysipelotrichi	0,03326	0,03998	0,016491	0,00921	0,048363	0,029891	Erysipelotrichi	0,022983	0,02052	0,020132	0,021043	0,072584	0,033102
Verrucomicrobiae	0,010434	0,04019	1,52E-05	3,64E-05	0,041533	0,039322	Verrucomicrobiae	0,015244	0,014076	0,000811	0,007678	0,078035	0,04949
Mollicutes	0,02753	0,027503	4,89E-05	2,17E-07	0,001047	0,019968	Mollicutes	0,042393	0,04279	5,25E-05	4,43E-05	0,001508	0,009063
Coriobacteriia	0,002954	2,93E-03	0,002293	0,007423	0,009032	0,008778	Coriobacteriia	0,004789	0,003122	0,007732	0,008934	0,008604	0,010118
Methanobacteria	0,011272	2,71E-03	7,82E-06	0	4,62E-06	0,000373	Methanobacteria	0,005936	0,005937	3,99E-06	3,85E-06	2,82E-06	0,00191
Other	0,000488	1,60E-04	1,98E-04	2,63E-04	3,54E-04	0,000569	Other	0,000533	0,000302	0,000299	0,000507	0,000346	8,65E-03
Betaproteobacteria	0,002535	0,003485	0,000108	0,000308	0,000989	0,000401	Betaproteobacteria	0,003199	0,002531	0,000243	0,000622	0,001101	0,000584
Alphaproteobacteria	0,000795	0,0094	0,000955	0,000724	0,000399	0,001364	Alphaproteobacteria	0,001017	0,001107	0,000383	0,000958	0,00078	0,000484
Fusobacteriia	0,000736	0,000214	1,24E-05	1,8E-05	3,87E-05	2,45E-05	Fusobacteriia	9,22E-05	0,000179	4,1E-05	0,000125	0,000438	2,22E-05
Deltaproteobacteria	0,000647	0,00047	0,000158	2,29E-05	3,31E-05	0,000111	Deltaproteobacteria	0,000783	0,001353	0,000347	0,000118	3,76E-05	0,000342
Epsilonproteobacteria	0,000191	0,000175	2,22E-06	5,86E-05	9,05E-06	8,11E-06	Epsilonproteobacteria	6,5E-05	0,000571	1,15E-06	0,000323	4,08E-05	1,03E-05
Tenericutes;c__RF3	7,3E-05	8,54E-05	0	0	6,6E-07	0	Tenericutes;c__RF3	0,000619	0,000441	1,64E-06	0	0	1,19E-06
TM7-3	6,95E-05	0,000283	3,51E-05	6,5E-05	0,000584	0,001561	TM7-3	0,00012	0,00013	4,03E-06	0,000196	0,002725	0,000867
Cyanobacteria;c__4C0d-2	6,62E-05	0,00018	7,32E-05	0	4,2E-06	0,000115	Cyanobacteria;c__4C0d-2	0,000353	0,00043	1,78E-05	3,84E-05	1,04E-06	0,000169
Synergistia	5,39E-05	0	0	0	0	0	Synergistia	2,4E-06	1,03E-05	0	0	0	0
[Saprospirae]	2,69E-05	1,16E-05	8,93E-07	0	1,28E-06	0	[Saprospirae]	7,07E-06	1,42E-06	5,16E-06	0	0	0
Chloroplast	2,63E-05	9,12E-05	8,9E-06	2,35E-05	1,71E-05	1,37E-05	Chloroplast	6,93E-06	1,52E-05	1,42E-05	4,82E-06	3,55E-05	5,12E-05
Nitriliruptoria	8,26E-06	6,34E-07	8,4E-05	0,00015	8,41E-05	6,14E-05	Nitriliruptoria	9,19E-06	1,34E-05	7,52E-05	0,000174	9,63E-05	7,47E-05
Opitutae	0	2,87E-06	0	0	0	0	Opitutae	3,5E-05	3,42E-05	0	0	0	0
[Lentisphaeria]	0	0	0	0	0	4,14E-07	[Lentisphaeria]	2,23E-05	4,87E-05	0	0	0	0
Flavobacteriia	0	0,000101	0	7,5E-06	0	0	Flavobacteriia	1,23E-06	6,95E-07	2,08E-06	3,05E-06	2,5E-06	1,57E-06
[Rhodothermi]	0	0	0	0	0	0	[Rhodothermi]	0	0	0	0	9,23E-06	0
Sphingobacteriia	0	0	0	0	0	0	Sphingobacteriia	0	0	0	0	0	0
Chlamydiia	0	0	0	0	0	0	Chlamydiia	0	0	0	8,07E-05	0	0

BIOLA							PLACEBO						
class	MoPreg	Mo3mon	10 days	3months	1year	2years	class	MoPreg	Mo3mon	10 days	3months	1year	2years
'Clostridia'	71,55	68,58	6,53	4,89	43,90	65,05	'Clostridia'	73,85	68,24	5,06	6,05	44,69	63,00
'Actinobacteria'	6,09	4,73	46,81	66,14	23,45	14,78	'Actinobacteria'	6,42	6,23	47,74	67,04	23,18	10,82
'Bacteroidia'	9,01	15,44	5,45	5,29	6,69	6,02	'Bacteroidia'	8,27	14,74	6,54	4,84	8,87	11,33
'Bacilli'	1,43	0,59	25,87	8,25	5,85	3,92	'Bacilli'	1,65	1,62	23,23	9,57	6,62	3,08
'Gammaproteobacteria'	0,72	0,22	13,01	11,81	3,59	0,73	'Gammaproteobacteria'	0,35	0,62	14,30	9,20	3,55	0,36
'Erysipelotrichi'	2,76	2,61	2,00	2,33	7,44	3,77	'Erysipelotrichi'	2,39	2,47	1,88	1,27	5,69	2,80
'Verrucomicrobiae'	1,45	1,56	0,00	0,28	7,54	3,11	'Verrucomicrobiae'	2,78	1,17	0,11	0,84	5,99	6,06
'Mollicutes'	5,30	4,61	0,01	0,00	0,21	1,22	'Mollicutes'	2,54	2,98	0,01	0,01	0,08	1,20
'Coriobacteriia'	0,31	0,36	0,17	0,79	0,96	0,95	'Coriobacteriia'	0,28	0,51	1,05	0,92	0,83	0,94
'Methanobacteria'	0,78	0,82	0,00	0,00	0,00	0,16	'Betaproteobacteria'	0,31	0,48	0,01	0,08	0,14	0,06
'Betaproteobacteria'	0,22	0,15	0,03	0,04	0,07	0,04	'Methanobacteria'	0,29	0,64	0,00	0,00	0,00	0,13
'Alphaproteobacteria'	0,08	0,09	0,08	0,09	0,07	0,10	'Alphaproteobacteria'	0,60	0,10	0,03	0,09	0,06	0,04
'TM7-3'	0,01	0,02	0,00	0,03	0,17	0,12	'TM7-3'	0,02	0,01	0,00	0,01	0,26	0,09
'Deltaproteobacteria'	0,11	0,05	0,01	0,01	0,00	0,02	'Deltaproteobacteria'	0,11	0,09	0,04	0,01	0,00	0,05
'Fusobacteriia'	0,03	0,04	0,01	0,01	0,03	0,00	'RF3'	0,04	0,06	0,00	0,00	0,00	0,00
'Epsilonproteobacteria'	0,09	0,02	0,00	0,00	0,01	0,00	'4C0d-2'	0,06	0,02	0,00	0,01	0,00	0,01
'4C0d-2'	0,02	0,04	0,00	0,00	0,00	0,02	'Fusobacteriia'	0,00	0,01	0,00	0,01	0,03	0,00
'RF3'	0,02	0,05	0,00	0,00	0,00	0,00	'Epsilonproteobacteria'	0,01	0,01	0,00	0,04	0,00	0,00
'Nitriliruptoria'	0,00	0,00	0,01	0,02	0,01	0,01	'Nitriliruptoria'	0,00	0,00	0,01	0,02	0,01	0,01
'Chlamydiia'	0,00	0,00	0,00	0,01	0,00	0,00	'Chloroplast'	0,00	0,00	0,00	0,00	0,00	0,01
'Chloroplast'	0,00	0,00	0,00	0,00	0,00	0,00	'Opitutae'	0,00	0,00	0,00	0,00	0,00	0,00
'Flavobacteriia'	0,00	0,00	0,00	0,00	0,00	0,00	'Flavobacteriia'	0,01	0,00	0,00	0,00	0,00	0,00
'[Lentisphaeria]'	0,00	0,00	0,00	0,00	0,00	0,00	'[Lentisphaeria]'	0,00	0,00	0,00	0,00	0,00	0,00
'[Saprospirae]'	0,00	0,00	0,00	0,00	0,00	0,00	'Synergistia'	0,00	0,00	0,00	0,00	0,00	0,00
'Synergistia'	0,00	0,00	0,00	0,00	0,00	0,00	'[Saprospirae]'	0,00	0,00	0,00	0,00	0,00	0,00
'Opitutae'	0,00	0,00	0,00	0,00	0,00	0,00	'Chlamydiia'	0,00	0,00	0,00	0,00	0,00	0,00

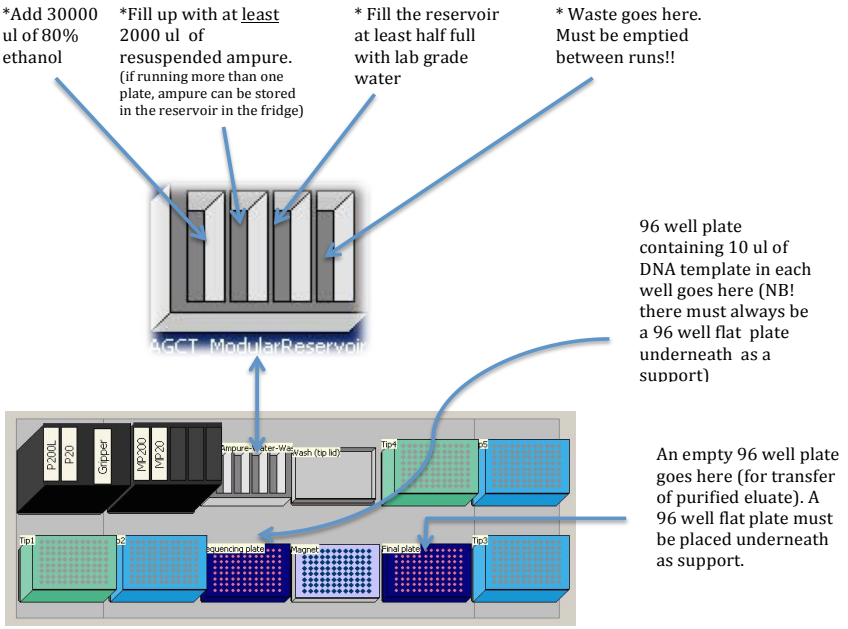
Atopy	KRUSKAL_WALLIS p_values						Biola	KRUSKAL_WALLIS p_values					
class	MoPreg	Mo3mon	10 days	3months	1year	2years	class	MoPreg	Mo3mon	10 days	3months	1year	2years
'Methanobacteria'	0,2853	0,7576	0,0887	0,5461	0,8276	0,3308	'Methanobacteria'	0,2240	0,3716	0,1853	0,3148	0,5474	0,9765
'Clostridia'	0,5722	0,1662	0,3929	0,5862	0,1631	0,5231	'Clostridia'	0,2058	0,9195	0,1052	0,7884	0,8446	0,4695
'Mollicutes'	0,1451	0,3409	0,4209	0,2661	0,7149	0,0040	'Mollicutes'	0,1911	0,4599	0,7610	0,3059	0,6993	0,8830
'Bacteroidia'	0,2344	0,0079	0,2987	0,9316	0,4217	0,1396	'Bacteroidia'	0,2161	0,6115	0,1529	0,9009	0,3975	0,1040
'Actinobacteria'	0,7334	0,4489	0,8084	0,6233	0,8618	0,2104	'Actinobacteria'	0,6769	0,0503	1,0000	0,9749	0,7729	0,1645
'Erysipelotrichi'	0,3056	0,0518	0,2690	0,4331	0,0506	0,5914	'Erysipelotrichi'	0,7013	0,2209	0,1892	0,8909	0,3435	0,1454
'Bacilli'	0,7185	0,0262	0,2521	0,3202	0,7385	0,5603	'Bacilli'	0,6020	0,3608	0,4063	0,7834	0,7788	0,8602
'Betaproteobacteria'	0,6040	0,9642	0,7964	0,1224	0,4120	0,4067	'Betaproteobacteria'	0,5122	0,8457	0,1692	0,4197	0,6956	0,6687
'Coriobacteriia'	0,1510	0,9615	0,8214	0,5848	0,8825	0,4288	'Coriobacteriia'	0,4310	0,0171	0,6796	0,8862	0,3429	0,1563
'Gammaproteobacteria'	0,5868	0,5402	0,5068	0,4577	0,2919	0,6634	'Gammaproteobacteria'	0,5652	0,1178	0,3040	0,4422	0,6015	0,3684
'Fusobacteriia'	0,7474	0,3544	0,7018	0,0737	0,6782	0,6196	'Fusobacteriia'	0,8920	0,2291	0,2459	0,1975	0,5436	0,9905
'Deltaproteobacteria'	0,0824	0,8019	0,1717	0,1191	0,5671	0,9111	'Deltaproteobacteria'	0,9911	0,3610	0,8440	0,2851	0,8317	0,5513
'Verrucomicrobiae'	0,1907	0,3419	0,8171	0,1003	0,1864	0,9888	'Verrucomicrobiae'	0,9638	0,9950	0,8551	0,2392	0,6424	0,1521
'Opitutae'	0,1907	0,4158	NA	NA	NA	NA	'Opitutae'	0,1689	0,0658	NaN	NaN	NaN	NaN
'Alphaproteobacteria'	0,1332	0,8654	0,8288	0,5941	0,2064	0,1184	'Alphaproteobacteria'	0,9620	0,9864	0,1695	0,8622	0,3735	0,3294
'RF3'	0,2289	0,2011	0,5461	NA	0,1233	0,5348	'RF3'	0,9506	0,8236	NaN	NaN	NaN	NaN
'4C0d-2'	0,7783	0,4568	0,8049	0,3921	0,3689	0,3773	'4C0d-2'	0,9451	0,2018	0,7888	0,3148	0,3083	0,3687
'TM7-3'	0,4531	0,9503	0,0627	0,5038	0,0463	0,9943	'TM7-3'	0,6525	0,5272	0,9228	0,1510	0,9836	0,6159
'[Lentisphaeria]'	0,0488	0,0208	NA	NA	NA	0,1072	'[Lentisphaeria]'	0,1089	0,6196	NaN	NaN	NaN	NaN
'Epsilonproteobacteria'	0,0662	0,2364	0,4607	0,9873	0,8094	0,7251	'Epsilonproteobacteria'	0,1260	0,6128	NaN	0,6758	0,0866	0,1635
'Chloroplast'	0,2737	0,0006	0,0932	0,6881	0,5916	0,4627	'Chloroplast'	0,2918	0,9318	0,1289	0,3198	0,2765	0,0326
'Synergistia'	0,4263	0,1621	NA	NA	NA	NA	'Synergistia'	0,3025	0,2906	NaN	NaN	NaN	NaN
'Nitriliruptoria'	0,2807	0,1155	0,7157	0,5574	0,2331	0,8824	'Nitriliruptoria'	0,1689	0,1798	0,6944	0,7519	0,3733	0,4632
'[Saprospirae]'	0,3615	0,3054	0,9226	NA	0,1233	NA	'[Saprospirae]'	0,3322	0,0995	0,9228	NaN	NaN	NaN
'Flavobacteriia'	0,5624	0,4713	0,3920	0,2884	0,5164	0,5348	'Flavobacteriia'	0,3322	NaN	0,2843	0,3198	NaN	0,3087
'Chlamydiia'	NA	NA	NA	0,5461	NA	NA	'Chlamydiia'	NaN	NaN	NaN	0,3198	NaN	NaN

Appendix E: Ampure protocol for Biomek 3000

Protocol for ampure purification on Biomek 3000

Protocol suited for automatically cleaning up large amount of 16s samples

1. Open Biomek software and open Ampure cleanup exp (under Methods< Kristian)
2. place rack tools, reservoirs, tip boxes etc. on the workspace according to picture 1.
3. Add reagents to the reservoir as illustrated below



Pic. 1: This is the setup for reagents, tips etc. on the workspace

3. Power up the robot. Remember to "Home all axes" before running the program (Instrument < home all axes).

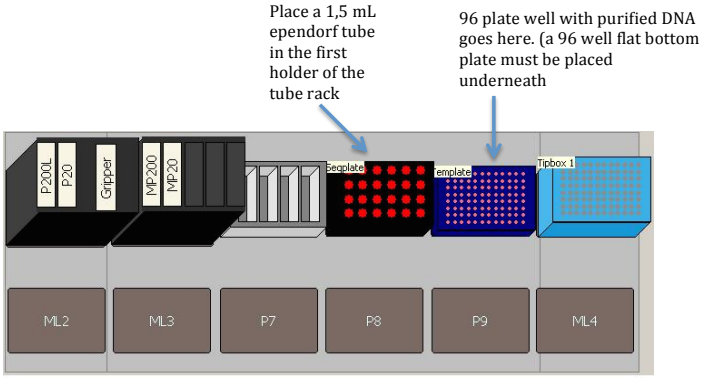
4. Press the start button, confirm that the setup matches. The robot can be paused or panic stopped (from the machine) at any time and resumed from the last point in case of failures etc.

The robot performs the following tasks:
1. Adds 10 ul of ampure (1:1) in each well.
2. Incubates for 5 minutes, moves plate to magnet, incubates for 2 minutes and removes supernatant to waste.
3. Wash step one; adds 100 ul of etoh, incubates for 30 sec and removes etoh.
4. Wash step 2 repeats step 3, but includes an additional step for removing all etoh.
5. Samples air dries for 30 minutes.
6. Eluation; Plate is removed from the magnet, and 20 ul of water is added. Incubates for 2 minutes, moves plate back to the magnet and incubates for 5 minutes.
7. Transfers 16 uL of eluate to final plate. Done

Appendix F: Protocol for normalization after DNA quantification

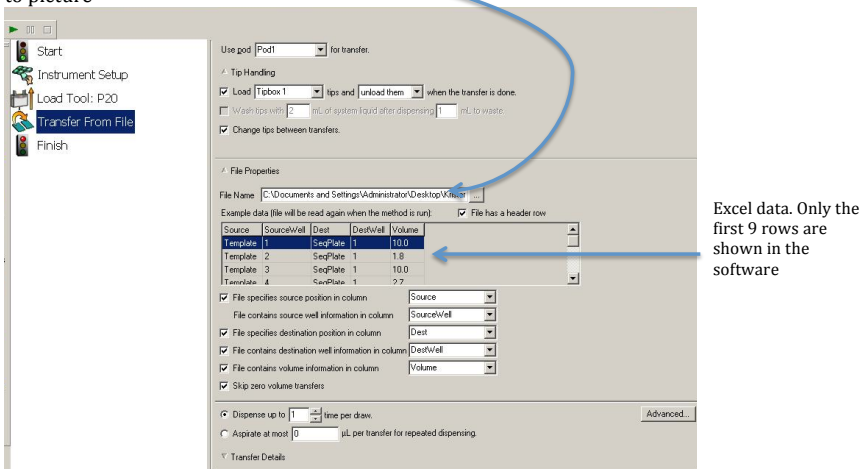
Protocol for normalization on Biomek of quantified 16s Samples

1. Calculate average DNA concentrations using quantification method of choice.
2. Put concentration data (volume) under the volume column in the excel template sheet found in the Krister biomek folder on the computer desktop (normalization template). Maximum amount is 20 ul. (NB! The software reads wells row-wise; That is from left to right). The software doesn`t read older excel file, but storing csv-files on the Biomek-computer works fine.
3. Save the file as csv (comma separated) file.
4. Open biomek software and "Normalization direct transfer" (under methods< Kristian)
5. Place rack tools, tip boxes etc. according to picture 1:



Pic. 1: Setup on the workspace

6. Point the software to the path were the csv file was stored (see picture 2), and make sure everything is set up according to picture

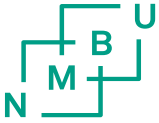


pic. 2: Window with transfer details

7. Remember to "home all axes" (Instrument < home all axes) first. Press start ▶ ; The robot now transfers liquids from the plate to the tube.

Appendix G: Table showing all controls in this thesis. (ND = not detected, Y = detected, WS = wrong specie, HO = higher order only)

Bakterie GA	Speicies sanger	Species high	MID	Species single	Bakterie GA	Speicies sanger	Species high	Species mid	Species single
municiphila	ND	Y	Y		Klebsiella oxytoca	Y	Y	Y	
Alistipes putredinis	ND	Y			Klebsiella pneu	Y	Y	Y	
cellulosilyticus	Y	Y			Lactobacillus acidophilus	Y	WS		
Bacteroides fragilis	Y	Y	Y		Lactobacillus amylolyticus	Y	Y		
intestinalis v	Y	Y			Lactobacillus brevis	Y	Y		
thetaiotamicron	Y	Y			Lactobacillus fermentum	Y	Y		
Bacteroides uniformis	Y	Y			Lactobacillus gasseri v	Y	Y		
Bacteroides vulgatus	Y	Y			Lactobacillus paracasei	Y	Y		
Bifido. Adolescentis	Y	Y			Lactobacillus plantarum	Y	Y		
Bifido. Angulatum	Y	ND			Lactobacillus reuteri	Y	Y	Y	
Bifido. Breve	Y	Y	Y		Lactobacillus rhamnosus	WS	Y		
Bifido. catenulatum	Y	Y			Lactobacillus salivarius	Y	Y		
Bifido. Dentium	Y	Y			Lactis	Y	Y		
Bifido. Gallicum	Y	Y			Listeria monocytogenes	Y	Y	Y	
infantis v	Y	Y	Y	Y	Mobiluncus curtisii	Y	Y		Y
longum	Y	Y			Mycoplasma orale	Y	Y	Y	
Blautia coccoides	Y	Y	HO		Neisseria subflava	Y	Y		
oklahomensis	Y	Y	Y		Oscillibacter valericigenes	Y	Y	Y	
fibrisolvens	Y	Y			Parabacteroides johnsonii	Y	Y		
Campylobacter curvus	Y	Y	Y		Parascardovia denticolens	Y	Y		
Citrobacter youngae	ND	Y	Y		Prevotella marshii	Y	Y	Y	
asparagiforme	HO	Y			Prevotella oralis	Y	Y		
Clostridium leptum	Y	Y			Proteus mirabilis	Y	Y		
perfringens v	Y	Y	Y		Proteus vulgaris	Y	WS		
Clostridium ramosum	Y	Y			Pseudomonas aeruginosa	Y	Y	Y	
saccharolyticum	WS	Y			Roseburia inulinivorans	Y	Y		
Aurimucosum	Y	Y			Ruminococcus albus	Y	HO		
Desulfomonas pigra	WS	Y	Y	WS	Salmonella bongori	Y	ND		
Dialister invisus	Y	Y			Salmonella enterica	Y	ND		
Dorea longicatena	Y	Y			Serratia marcescens	Y	Y		
Eggerthella lenta	Y	Y			Shigella dysenteriae	Y	ND		
aerogenes v	Y	ND			Shigella sonnei	Y	ND		
hormaechei oharae	Y	ND			Staphylococcus aureus v	WS	ND		
v	Y	WS			Staphylococcus epidermis	Y	Y	Y	
Enterococcus faecium	Y	Y	Y		Streptococcus agalactiae	Y	Y		
Escherichia coli v	Y	Y	Y	Y	Streptococcus equinus	Y	WS		
cyllindroides	Y	Y	Y		Streptococcus infantarius	WT	WS		
Eubacterium rectale	Y	Y			Streptococcus mitis	WT	Y	Y	
prausnitzii v	Y	Y	Y		Streptococcus oralis	Y	Y	Y	
Finegoldia magna	Y	Y			Streptococcus pneumoniae	Y	Y		
periodonticum	Y	Y	Y		Streptococcus pyogenes	Y	ND		
Gemella haemolysans	Y	Y			Veillonella atypica	WS	WS		
Hafnia alvei	WS	WS			Veillonella dispar v	Y	WS		
Heliobacter pylori	Y	Y	Y		pseudotuberculosis	ND	ND		
					Vibrio cholerae	Y	Y	Y	
					pseudotuberculosis	ND	ND		
						77/88	72/88	28/30	4/4



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no