UNIVERSITY OF TURKU
Faculty of Science

Optimization of microbial DNA extraction in 96-format from fecal samples for Next Generation Sequencing

Natalie Tomnikov

Master's Degree Program in Physiology and Genetics

Master's thesis

Credits: 30 op

Supervisor(s):

Teemu Kallonen, PhD

Sanja Vanhatalo, MD

Eero Vesterinen, PhD

13.06.2022

Turku

Master's thesis

Major: Biology
Author: Natalie Tomnikov
Title: Optimization of microbial DNA extraction in 96-format from stool samples for Next Generation Sequencing
Supervisors: PhD Teemu Kallonen, MD Sanja Vanhatalo and PhD Eero Vesterinen
Number of pages: 70 pages and 7 appendices
Date: 13.06.2022

Gut microbiota composition is often determined by using Next Generation Sequencing methods. The demand for rapid, efficient, and reliable microbial profiling is continuously increasing, thus making the optimization of high-throughput 96-format DNA extraction integral for downstream applications. In this study, we analyzed the impact of four pre-treatment methods to the DNA extraction and 16S metabarcoding results. More specifically, pre-treatments with and without bead-beating were compared.

Fecal DNA was extracted from infant, adult and senior fecal samples. The extraction also included negative controls and ZymoBIOMICS Gut Microbiome Standards to test the resolution of the sequencing target and assess DNA yield. The optimized kit was DNA Stool 200 Kit special H96 with Chemagic MSM I extraction robot. Four pre-treatments, including bead-beating and chemical lysis, were applied to the samples to assess the impact of the sample lysis and homogenization. The microbial composition was determined using 16S sequencing targeting V3V4 and V4 regions with Illumina Miseq platform.

The observed compositions of the Gut Microbiome Standards differed considerably from the expected theoretical compositions with all methods and sequencing targets. The fecal samples showed different degrees of microbial diversity across different pre-treatment groups; however, the inclusion of bead-beating generally lead to higher degrees of microbial diversity. The extraction in 96-format proved to be? feasible.

As the DNA extraction methods are advancing, it is important to validate new procedures with the existing NGS-methods. The application of 96-format could reduce the hands-on time as well as human error in clinical microbiology. The 96-format extraction systems proved to be suitable for fecal DNA extraction. However, the results indicate the need of a standardized methods for microbial profiling.

Key words: optimization, next generation sequencing, 16S, gut microbiome, dna extraction, feces

Pro gradu -tutkielma

Pääaine: Biologia
Tekijä: Natalie Tomnikov
Otsikko: Mikrobi DNA-eristyksen optimointi ulostenäytteistä 96-formaatissa uuden sukupolven sekvensointimenetelmiä varten.
Ohjaajat: FT Teemu Kallonen, FM Sanja Vanhatalo ja FT Eero Vesterinen
Sivumäärä: 70 sivua ja 7 liitettä
Päivämäärä: 13.06.2022

Suoliston mikrobikoostumuksen selvittämiseksi käytetään usein uuden sukupolven sekvensointimenetelmiä (*engl. NGS).* Korkean kapasiteetin 96-formaatissa olevan DNA-eristyksen optimoiminen alavirran sovelluksille on tärkeää, sillä kysyntä nopealle, tehokkaalle ja luotettavalle mikrobiologiselle profiloinnille on alati kasvamassa. Tässä tutkimuksessa analysoimme neljän esikäsittelytavan vaikutusta DNA-eristys -ja 16S-viivakoodaustuloksiin. Tarkemmin sanottuna esikäsittelyjä ilman -ja helmiravistelun kanssa vertailtiin.

Uloste-DNA:ta eristettiin aikuisen, vauvan ja seniorin ulostenäytteistä. Eristykseen kuului myös negatiivisia kontrolleja sekä ZymoBIOMICS Gut Standard -standardi sekvensointimenetelmän tarkkuuden testaamiseksi ja DNA:n saannon arvioimiseksi. Optimoitu kitti oli DNA Stool 200 Kit special H96 ja eristysrobotti Chemagic MSM I. Lyysiksen ja homogenisoinnin vaikutuksen tutkimiseksi näytteille tehtiin neljä eri esikäsittelyä, jotka sisälsivät kemiallisen -ja helmiravistelulyysiksen (*engl*. bead-beating). Mikrobikoostumus selvitettiin sekvensoimalla 16S-ribosomin V3V4 -ja V4-alueet käyttämällä Illumina Miseq-alustaa.

Havaitut Gut Standard -standardien mikrobikoostumukset erosivat huomattavasti odotetuista teoreettisista koostumuksista kaikkien esikäsittelyjen ja sekvensointikohteiden suhteen. Eri esikäsittelymenetelmät tuottivat ulostenäytteille vaihtelevia diversiteettejä. Kuitenkin helmikäsittelyn sisällyttäminen johti yleensä korkeimpiin diversiteettilukuihin. DNA:n eristys 96-formaatissa todettiin mahdolliseksi

On tärkeää validoida uusia menetelmiä vanhojen toimintatapojen valossa, koska DNA-eristysmenetelmät kehittyvät. 96-formaatin käyttöönottaminen voisi vähentää käytännön työaikaa sekä ihmisestä johtuvia virhelähteitä kliinisen mikrobiologian alalla. Tämä 96-formaatti todettiin toteutettavissa olevaksi ulostenäytteiden DNA-eristyksessä. Kuitenkin tulokset viittaavat siihen, että standardisoituja menetelmiä tarvitaan mikrobikoostumuksen selvittämisessä.

Avainsanat: optimointi, uuden sukupolven sekvensointi, 16S, suolistomikrobiomi, dna-eristys, uloste

Contents

Abbreviations 1
1 Introduction 2
1.1 Introduction to the gut microbiome 2
1.1.1 Development of the gut microbiome 2
1.1.2 Factors influencing the gut microbiota 3
1.1.3 Studying the gut microbiota 8
1.2 Next Generation Sequencing 9
1.2.1 16S sequencing 10
1.2.2 Clustering, diversity measures and rarefaction 12
1.3 Optimization of the stool processing pipeline 15
1.3.1 DNA extraction optimization 15
1.3.2 96-format 18
1.4 Aims of the thesis 20
2 Materials and methods 21
2.1 Sample collection and storage 21
2.2 DNA extraction, quantification and quality control 21
2.3 16S sequencing 23
2.4 Bioinformatic methods and data visualization 25
3 Results 26
3.1 DNA yield and integrity 26
3.2 Gut Standards 28
3.2.1 Relative abundances 28
3.2.2 Alpha diversities 31
3.2.3 Beta diversities 32
3.3 Infant fecal samples 33
3.4 Adult fecal samples 38
3.5 Senior fecal samples 43
3.6 Negative and extraction controls 48
3.7 Results summary 51
4 Discussion 52
4.1 Feasibility of the 96-format extraction 52
4.2 Bead-beating and diversity measures 55
4.3 Extraction cross-contamination 57
4.4 Challenges and limitations 58
4.4.1 Statistical power 58

4.4.2 High variability 59
4.4.3 Other contamination and sequencing contamination 59
4.5 Future outlooks 60
5 Conclusions 61
Acknowledgements 62
References 63
Appendices

Abbreviations

AIEC	Adherent Invasive *Escherichia coli*
AMP	Antimicrobial protein
GI	Gastrointestinal
HMO	Human milk oligosaccharide
IBS	Inflammatory bowel syndrome
MAC	Microbiota-accessible carbohydrate
mGWAS	Microbe genome-wide association studies
NSG	Next Generation Sequencing
OTU	Operational taxonomic unit
PAMP	Pathogen-associated molecular pattern
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
SCFA	Short-chain fatty acid
V-region	Variable region
16S rRNA	16S ribosomal RNA

1 Introduction

1.1 Introduction to the gut microbiome

The human gastrointestinal (GI) tract contains various microorganisms, including bacteria, viruses, archi and fungi. These microorganisms are termed “gut microbiota”. Although the term “gut microbiota” encompasses all of the microorganism residing in the GI tract, most often the term is used to refer to intestinal bacteria. The human GI tract is one of the largest nexuses between the host and environmental factors (Thursby & Juge 2017). Indeed, the gut microbiome has co-evolved with the host to form a complex but mutually beneficial relationship (Ley et al. 2006). It is estimated that the number of microorganisms exceeds 10^{14} cells and weights roughly 1.5-2kg. During an average lifespan, over 60 tonnes of food, along with numerous microorganisms, pass through the gastrointestinal tract. Undoubtedly, the stool samples are thought to reflect the gut microbiota so well. The overwhelming amount of information about the gut microbiota is partly due to large-scale studies, such as European Metagenomics of the Human Intestinal Tract or NIH-funded Human Microbiome Project (Shreiner et al. 2015). Despite the heaps of advances in the field of microbiology, there is much to uncover about the complex connection that humans and the gut microbiota share.

1.1.1 Development of the gut microbiome

The colonization of human GI tract is generally believed to occur at birth, although some studies report finding traces of microbiota in placenta (Aagaard et al. 2014; Rodríguez et al. 2015) and thus challenging the sterile womb paradigm. Another study found that the placenta contained only very small amounts of microbial biomass and the bacterial DNA detected in the birth tissue came from contaminating sources, such as laboratory reagents or equipment (de Goffau et al. 2019). At any rate, the microbiota composition of infants is affected by the delivery method as it is the first interaction between the newborn and the outside world. The microbiota of vaginally delivered infants has a high abundance of genus *Lactobacillus* during the first few days, reflecting the lactobacilli found in the vaginal microflora (Turroni et al. 2020). Moreover, vaginally delivered infants’ fecal microbiome resembles on average that of their

mothers'. Contrarily, infants delivered by C-section have delayed colonization of *Bacteroides* and *Bifidobacterium* but are more abundant in skin and surface derived genera such as *Staphylococcus* and *Propionibacterium* (Yang et al. 2016). During the early stages of life, the gut microbiome is typically low in diversity and is consisted of two main phyla, *Actinobacteria* and *Proteobacteria*. By the age of three, the diversity in microbiome increases and begins to resemble that of an adult in structure, diversity and functional capabilities. In adulthood, the gut microbiome is relatively stable, yet still susceptible to changes caused by life events. The microbiota of healthy adults is dominated by the phyla *Firmicutes* and *Bacteroidetes* (80-90%), with traces of *Proteobacteria*, *Fusobacteria*, *Cyanobacteria, Verrumicrobia* and *Actinobacteria* (Salazar et al. 2017). During elderly years, the diversity of microbiome decreases and becomes unstable; much like a mirror image of neonatal GI tract colonization. More specifically, lower levels of *Firmicutes*, namely *Clostridium* cluster XIVa and *Faecalibacterium prausnizii,* as well as *Actinobacteria,* mainly *Bifidobacterium*, have been found. In addition, increased abundances of *Proteobacteria* have been detected (Salazar et al. 2017).

1.1.2 Factors influencing the gut microbiota

Every individual harbors a unique gut microbiota that is influenced by many factors. The gut microbiome is susceptible to shaping by the host and alterations in environmental selective pressures (figure 1). The gut ecosystem is very dynamic. However, human gut offers relatively low number of biochemical niches, making it necessary for gut microbes to adapt (Thursby & Juge 2017). Gut microbes generally derive energy through fermentation and sulphate reduction of dietary and host carbohydrates (Ley et al. 2006). Thus, microbes that thrive in the gut are limited by their phenotypic traits.

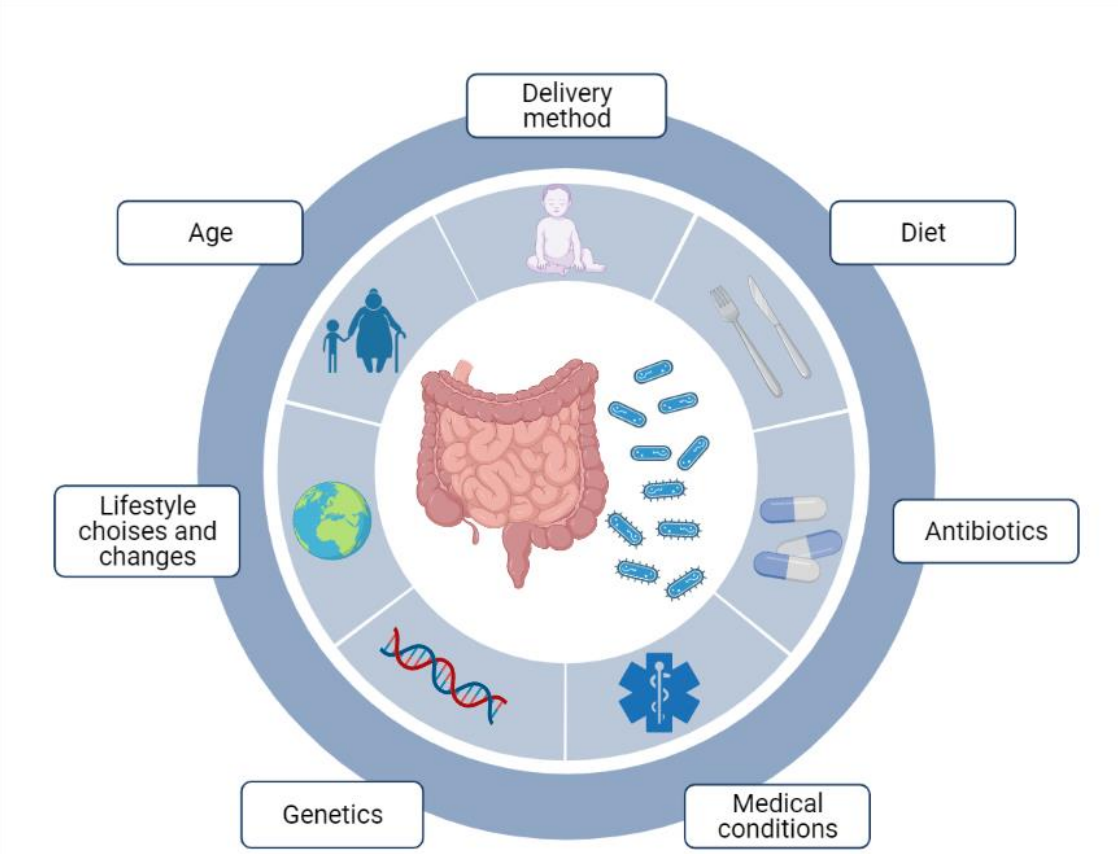


Figure 1. Factors influencing the gut microbiome. Created in BioRender.com

Diet is arguably one of the most profound factors affecting the gut microbiome. The effects of nutrition reflect on the microbial community structure early on. For example, human breast milk contains oligosaccharides (HMOs, human milk oligosaccharides) that can be utilized by *Bifidobacterium longum* and *Bacteroides* spp., allowing them to compete with other bacteria (Turroni et al. 2020). On the other hand, formula-fed infants have lower levels of *Bifidobacterium* spp. Moreover, formula-fed infants have increased microbial diversity and altered levels of other bacteria groups, such as lactobacilli and *Clostridium difficile* (Bezirtzoglou et al. 2011). Malnourished children have undeveloped and dysbiotic gut microbiota that is more abundant in enteropathogens, such as *Enterobacteriaceae* (Kau et al. 2015).

With the introduction of solid foods, the alterations in gut microbiota become more pronounced. For example, a comparative study between European and Burkina Faso (Africa) children showed significant differences in the gut microbiota between the two

groups (de Filippo et al. 2010). Children from Burkina Faso had a higher abundance of *Bacteroidetes* and depletion of *Firmicutes*. Moreover, the study found unique signatures from *Prevotella* and *Xylanibacter*, genera known to hydrolyze cellulose and xylan, absent in the European children. The diet of the European children consisted of typical "western-style" foods, whereas the Burkina Faso individuals consumed predominately vegetarian foods rich in fibers.

Strict "plant-based" and "animal-based" diets, or "Western" and "Mediterranean" style foods modulate the microbiome composition in different ways. High-fat, high-sugar and low dietary fiber "Western-style" diets promote gut inflammation and dysbiosis in mice model (Agus et al. 2016). Agus et al. (2016) found that the Western diet promotes the growth of bacteria that are capable of metabolizing simple sugars and favors Adherent Invasive *Escherichia coli* (AIEC) colonization. Moreover, the Western diet is low on microbiota-accessible carbohydrates, MACs (Sonnenburg & Sonnenburg 2014).

Nutrients from plant and animal cell vary in their availability to the gut microbiota. Nutrients from animal cells tend to be more accessible to the gut microbiota due to processing of food (Zinöcker & Lindseth, 2018). The easy accessibility to these nutrients in Western diets advances the growth of certain bacteria and facilitates the expansion of the bacteria's ecological niche. On the other hand, Mediterranean diets rich in fruits, vegetables, lean meats and whole grains promote a homeostatic gut and reduce systematic inflammation (Nagpal et al. 2019). One of the advantages of Mediterranean diet is dietary fiber. Plant cells have cells walls consisting of mainly fibre, which fiber-degrading bacteria can hydrolyze (Zinöcker & Lindseth 2018). This favors the growth of bacteria, such as *Bascteroides spp.* and *Akkermansia mucinphilia*, that produce beneficial metabolites, such as short-chain fatty acids (SCFAs). SCFAs (e.g., acetate, byrate and propionate) protect the integrity of the colon mucus layer, increase the gut motility, stimulate the production of some neurotransmitters and reduce gut inflammation (Sonnenburg & Sonnenburg 2014). For example, SCFAs are able to regulate T cell dependent immunological pathways. Contrarily, a Wester diet induced inflammation could be mediated by pathogen associated molecular patterns (PAMPs) that are produced by microbes in processed foods (Zinöcker & Lindseth 2018).

In addition to dietary habits, antimicrobials and antibiotics have a profound impact on the gut microbiota. Antibiotics are life-saving drugs for many people. For example, antibiotics have been important in improving health outcomes, clean water and sanitary conditions (Browne et al. 2021). However, antimicrobial compounds have an adverse effect on the gut microbiota. Antibiotic treatment may occur during any stage of life. During infancy, antibiotic treatment causes short-term perturbations in the development of gut microbiota in healthy (full) term infants (Turta et al. 2022). Yet the effects of antibiotic treatment can reflect up to six years in life (Uzan-Yulzari et al. 2021). Infants treated with antibiotics have less diverse and less stable bacterial communities (Yassour et al. 2016). *Lachnospiraceae* spp. and *Clostridiales* are particularly vulnerable populations (Bokulich et al. 2016). Moreover, antibiotic treatment during infancy seemed to increase the likelihood of childhood asthma (Cotten 2016), obesity in later life and other disorders, such as IBS (Bokulich et al. 2016). One of the biggest concerns with antibiotic treatments is the increase in antibiotic resistance. However, breast feeding is a protective factor against antibiotic resistance genes (Nadimpalli et al. 2020). With dietary intervention and early detection, the adverse effects of early-age antibiotic use can be mitigated.

Similarly, antibiotic usage causes acute gut microbiota dysbiosis in adults (Palleja et al. 2018). Administration of three antimicrobial compounds resulted in increase of pathobionts (e.g., *Enterobacteriaceae*) and decrease in butyrate-producing species (e.g., *Faecalibacterium prausnitzii*). Broad-range antibiotics target a wide range of bacteria, including those that are beneficial to humans. Thus, disrupting the microbial balance in the gut may lead to growth of opportunistic pathobionts, such as toxigenic *Clostridium difficile*. Palleja et al. (2018) found that the baseline microbiota in healthy adults was restored after 1.5 months of antibiotic treatment, although several common species remained undetectable after 180 days. Antibiotic usage may cause short-term consequences, such as antibiotic-associated diarrhea (ADD), or long-term consequences, such chronic dysbiosis of the gut. Dysbiosis of the gut microbiome has been linked to many medical conditions, ranging from Chron's disease to depression. For example, in irritable bowel syndrome (IBS), microbial dysbiosis is thought to facilitate the adhesion of pathogens to the bowel wall (Guinane & Cotter 2013). The onset of inflammatory bowel syndrome (IBS) is generally not thought to be caused by a

single organism, but rather by an overall dysbiosis. However, IBS is manifested with a lower microbial diversity and low-grade inflammation (Guinane & Cotter 2013). Additionally, intestinal inflammation is linked to reduced diversity in *Bacteroidetes* and *Firmicutes* phyla. Reduced bacterial diversity could affect the host's ability to defend itself from pathobionts, as the microbiome is also shaped by the host's immune system (Thursby & Juge, 2017). Paneth cells in the GI tract produce antimicrobials, such as angiotensin 4, lysozymes, histatins and lipopolysaccharide (LPS)-binding protein. Most antimicrobial proteins (AMPs) work by disrupting the bacterial cell wall or inner membrane via enzymatic attack, thus killing and limiting the growth of bacteria.

In addition to immunity, host genetics influence the gut microbiota as well. Host genetics define many attributes important to gut microbiome, such as nutrient availability in the gut, microbiome community structure and the threshold activity of the human immune system (Hall et al. 2017). Genetic variation between hosts contributes to the notable differences of the gut microbiome in individuals. Microbe genome-wide association studies (mGWAS) have found multiple connections between human genetic variants and the gut microbiota (Hall et al. 2017; Wang et al. 2016). For example, *NOD2* gene (*nucleotide-binding oligomerization domain-containing 2*) mutations are a risk factor for developing inflammatory bowel diseases (Hugot et al. 2001). Inflammatory bowel disease (IBD) encompasses two primary diagnoses: Chron's disease and ulcerative colitis, and is characterized by chronic inflammation of the gut. Nucleotide-binding oligomerization domain-containing protein 2 (NOD2) is an intracellular pattern recognition receptor for muramyl dipeptide, a cell wall component of gram negative and positive bacteria (Hugot et al. 2001). NOD2 stimulates an immune signaling cascade that leads to the production of cytokines, antimicrobial defensins and T cell-stimulating molecules. Deficiency in NOD2 results in over-activation of the inflammatory response to beneficial gut bacteria. In mouse models, wild type (not dysbiotic) mice that received fecal transplant from *NOD2*-deficient mice had an increased risk of developing colitis (Couturier-Maillard et al. 2013).

It is evident that the gut microbiota is influenced by many factors. In everyday life, our gut microbiota is shaped by the life choices we make. Gut microbiota is very dynamic. Indeed, the gut microbiome composition can change in a matter of days. For example, significant changes in gut microbiome composition and metabolism can be seen in people switching from plant-based diet to animal in just four days (David et al. 2014).

1.1.3 Studying the gut microbiota

Understanding the complex relationship between the gut microbiota and health is challenging. Utilizing in-vivo animal models have been crucial in studies manipulating the gut microbiota. Especially germ-free (GF) mice have been especially useful in microbiome experiments (Kennedy et al. 2018). The germ-free status allows the colonization of only known microbes and the creation of gnobiotic mice. Transfer of host phenotypes via human stool microbiota transplantation is essential for understanding the aetiology of microbiota-related diseases. For example, transferring endotoxin-producing *Enterobacter cloacae* B29 strain to GF and high fat diet mice induced obesity and insulin resistance in the mouse model (Fei & Zhao 2012). However, mice studies may not truly reflect the reality as mouse and human differ considerably in size, metabolic rate and dietary habits (Hugenholtz & de Vos 2018).

The aetiology of diseases cannot be studied in humans in the same way as in animal models. In other words, manipulating the gut microbiota in humans is not viable. As this is the case, birth cohort studies may help to identify disease-relevant microbial signatures at an early stage. Birth cohorts follow individuals from certain population(s) for a longer time, usually up to childhood, but potentially through adolescence, adulthood and later life (Kuh et al. 2016). Many cohorts focus on the first 1-2 years of life (perinatal period) during which the infant undergoes rapid microbial colonization and immune system development. Various biological samples are collected from the infant or the mother, including faces, breast milk, urine and DNA (García-Mantrana et al. 2019; Korpela et al. 2019). In addition, questionnaires and psychological evaluations are usually conducted. While longitudinal population studies are the preferred type of design to determine causal relationship between health and early life, variation between

individuals makes cross-sectional designs challenging to manage. Moreover, these types of studies require large amounts of time and resources, are at risk of losing follow-up samples and thus having lower statistical power as well as possible selection bias (Canova & Cantarutti 2020). The implementation of optimized protocols for sampling and analysis is needed to improve the reproducibility and suitability of large-scale microbiome studies.

Traditionally, culture-based methods have been used to determine the microbiome composition of a sample. These methods generally would focus on easy-to-culture bacteria. Bacterial culture methods have not remained completely stagnant, however. Microbial culturomics introduces novel techniques, such as new selective culture media (Bonnet et al. 2020). Cultures enable isolation of some bacterial communities and identification based on morphology. This shows that culture methods are still feasible, despite the apparent abandonment of the culture media. However, it is estimated that only 10-50% of the gut bacteria are culturable (Eckburg et al. 2005). Culture-based approaches also require more time in order to let the bacteria proliferate, adding significant amounts of hands-on-time. Thus, culture-independent approaches might be better at providing an insight into the gut microbiota. In particular, the development of fast and cheaper Next Generation Sequencing technologies has been revolutionary.

1.2 Next Generation Sequencing

Next Generation Sequencing, parallel sequencing or high-throughput sequencing are terms that describe a DNA sequencing technology that has revolutionized research in the genomic field. The term ”next generation” implies a significant improvement to the previous sequencing technologies, mainly to chain-termination-based Sanger sequencing technologies. While Sanger sequencing remains a good alternative for some applications, Next Generation technologies tend to be cheaper, quicker, more reliable and require less DNA (Thursby & Juge 2017). Arguably the most common NGS method for bacterial taxonomic and compositional profiling relies on 16S sequencing.

1.2.1 16S sequencing

16S ribosomal RNA (figure 2) is the RNA component that is a part of a larger 30S subunit of prokaryotic ribosome. The entire length of the 16S rRNA gene is roughly 1500 base pairs long. The 16S ribosomal RNA gene (rRNA gene) has highly conserved primer binding sites that can be used in phylogenetic research. In addition, the rRNA gene has nine hypervariable regions (V1-V9) surrounded by the conserved sites (C1-C9). Withing the hypervariable regions there are more conserved parts that correlate to higher level taxonomy (e.g., phylum) and less conserved parts that correspond to lower-level taxonomy (e.g., genus). The hypervariable regions are often used with PCR and sequencing technologies.

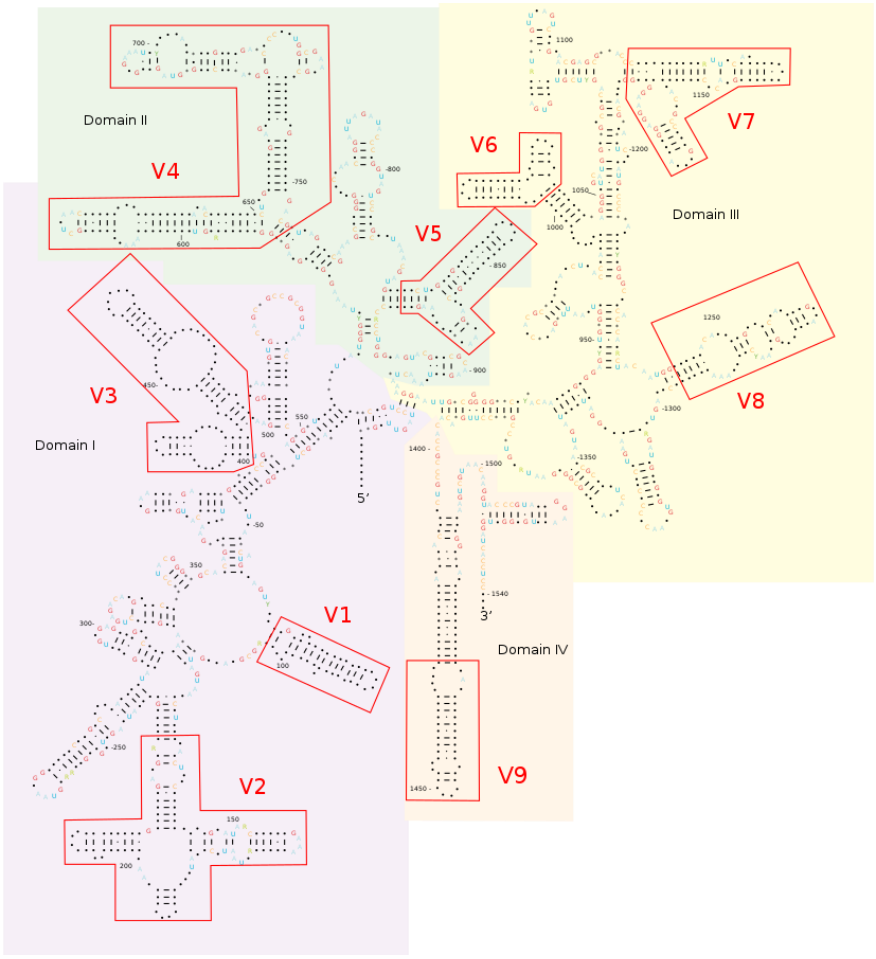


Figure 2. The structure of 16S ribosomal RNA. The 16S ribosome contains nine variable regions (V1-V9) that are outlined in red color. Source: Wikipedia Commons (2011)

16S rRNA marker gene is one of the most commonly used marker genes for bacterial profiling. Selecting the appropriate V-region (primer selection) is crucial in 16S amplicon sequencing (di Segni et al. 2018). Depending on the research goals and sequencing platforms used, microbiome studies can use either a single variable region (V-region) or a combination of V-regions. Different primer pairs are used to target different V-regions. Studies have shown that use of different sequencing platforms and targeting different sub-regions produce unique microbial compositional profiles (Fadrosh et al. 2014; Tremblay et al. 2015). The capacity of the selected sequencing platform can affect the selection of a V-region. For example, some Illumina platforms generate only short reads (around 150-200 base pairs) using fragment library sequencing protocol and cannot sequence longer V-regions (Bartram et al. 2011). For example, the V4 region is approximately 254bp long and the V3V4 region 443bp long (Bukin et al. 2019). Other sequencing platforms, such as Roche 454 or Ion Torrent, are able to sequence 2 or 3 adjacent V-regions (around 400-500 base pairs). On the other hand, longer reads can also be generated on Illumina platforms with paired-end sequencing (Fadrosh et al. 2014).

Ideally, multiple V-regions can be used to achieve better taxonomic and compositional accuracy. However, due to experimental and sequencing platform limitations, longer reads cannot always be obtained and thus research groups might use one or two V-regions. Individual V-regions have different abilities in differentiating and resolving taxonomic profiles. For example, one study found that V4 region had the highest accuracy while classifying the phylum *Bacteroides* (Pinna et al. 2019). The underlying cause of the correlation between V-regions and taxonomic resolution remains unknown (Fadrosh et al. 2014).

One of the advantages of 16S amplicon sequencing is the vast availability of reference sequences. For example, the updated SILVA 138 reference database holds roughly 9,400,000 small subunit ribosomal ribonucleic acid (SSU rRNA) sequences (SILVA 2019). Furthermore, many bioinformatic pipelines are free of charge and available worldwide. However, as any method, 16S sequencing has its limitations. Despite the low cost and wide use in bacterial taxonomic profiling, 16S sequencing is subject to

profiling bias. Primer selection has a significant impact of bacterial compositional profiles, as stated above, since the 16S rRNA gene is prone to gene copy variation. In addition, 16 sequencing provides no information on the function of the bacteria. On the contrast, shotgun sequencing (whole DNA sequencing) and metatranscriptome (whole RNA) sequencing can provide us information on bacteria's gene expression or functions as a species level (Slatko et al. 2018). 16S amplicon sequencing is also limited to detection of bacteria and cannot estimate the abundances of viruses, yeast, archea or other microorganisms in the gut.

1.2.2 Clustering, diversity measures and rarefaction

There are many units for marker gene analysis, all constructed in different ways (Chiarello et al. 2022). All of these methods aim to minimize sequencing errors within a pool of reads. Arguably the most traditional method is OTU clustering, which will be explained later in more detail. The newer methods include Amplicon Sequencing Variant (ASV) approach that aims to describe sequences that are statistically supported being in a sample. ASVs are also referred as Exact Sequence Variants (ESVs) or zero-radius OTUs (zOTUs) perhaps due to these methods not using a similarity threshold, or a reference database until assigning taxonomy (Chiarello et al. 2022). However, the method that is used in this thesis is OTU clustering. OTU clustering begins with assigning sequences a similarity threshold (usually 97%) and clustering them into Operational Taxonomic Units (OTUs). Sequence similarity is customarily computed as the percentage of sites that agree in a pairwise alignment. The similarity threshold of 97% was derived from a study that found that most procaryotic strains had 97% 16S rRNA sequence similarity (Konstantinidis & Tiedje 2005). A single sequence is then selected as a representative of an OTU. This representative sequence is annotated and the information from the annotation is applied to all the remaining sequences within that specific OTU. The sequences are mapped against a reference database, such as SILVA (used in this thesis), Greengenes, RDP or NCBI (Balvočiute & Huson 2017). One of the significant benefits of OTU clustering is that it doesn't require large computational power if the clusters are not created *de novo* (without a reference database). A 16S amplicon may have millions of reads but this could result in only thousands of OTUs. Thus, OTU clustering allows rapid analysis of the sequence data (Nguyen et al. 2016).

However, OTU creation may be subject to reference bias if the OTUs are created surrounding a closed reference database. Results from 16S rRNA gene amplicon sequencing can be applied to determine microbial composition at genus-level. Bacterial strains are not always accurately assigned at species-level with 16S sequencing due to sequence similarity of the V-regions (Rossi-Tamisier et al. 2015).

Next step is to visualize and describe the bacterial community structure in stool samples. This is done via different diversity measures. These measures do not offer us information on changes in abundance of specific genus or taxa but allow us to determine broader changes or differences in the composition of bacteria (or other microbes). Alpha and beta diversity are such measures.

Alpha diversity measures estimate the structure of a microbial community regarding its richness (number of taxonomic groups), evenness (distribution of abundances of the groups) or both. In microbiome studies, computing the alpha diversities of amplicon sequencing data is a typical first approach in assessing differences between environments or treatments (Willis 2019). Alpha diversity contains many indices, such as observed OTU number, Shannon entropy and Chao 1, that all account for different things. Shannon entropy measures both evenness and richness in a single equation (Gauthier & Derome 2021). Along with Simpson's index that measures sample richness, Shannon entropy is one of the most popular diversity indices in community ecology. Chao 1 index in a non-parametric estimator for measuring the richness of a microbial community (Kim et al. 2017). It is based upon a notion that the rarest species (or taxa) infer the most information about the number of missing species. This index gives more weight to low-abundance taxa, approximating the number of species (or taxa) that are represented only by a single individual (singleton) or by two individuals (doubletons). Thus, Chao 1 is especially useful for data sets skewed toward low-abundance taxa. Observed OTUs is another diversity index for richness. It visualizes how many Operational Taxonomic Units (OTUs) are present in a sample. All of the indices are a great tool to estimate sample richness and evenness, however direct comparison between those indices can be difficult due to indices accounting for different parameters.

While alpha diversity is a diversity measure applicable to a single sample, beta diversity metrics measure the degree to which samples differ (dissimilarity) from one another. Beta diversity measures can be grouped in a few different ways (Goodrich et al. 2014). First, the metrics can be quantitative (using sequence abundance, e.g., Bray-Curtis or weighted UniFrac), or qualitative (presence or absence or sequences, e.g., binary Jaccard or unweighted UniFrac). Second, they can be based on phylogeny (UniFrac metrics) or not (e.g., Bray-Curtis). Once the beta diversity distance matrices are calculated, Principal Coordinate Analysis (PCoA) can be performed. PCoA visualizes the microbiome data sets in two- or three-dimensional scatterplots. Principal Coordinates (PCs/PCos), which explain a certain percentage of the variability, are plotted to create a visual representation of differences among microbial community samples.

When a microbiome sample is collected from a specific niche, there is a need to evaluate how well the sample reflects the true diversity of that niche, which relates to sample richness and relative abundance. It is challenging to determine which community has the highest diversity as we compare samples with different library sizes. Rarefaction is a statistical method that allows comparison between samples of different library sizes (Willis 2019). It estimates the number of taxa expected in a random sample taken from a pool of samples. Rarefaction informs us if the sample consisted of a specific number of individuals would likely have been there. The rarefaction method is dependent upon rarefaction curve, which measures number of OTUs with a given depth of sequencing. In other words, OTU counts are normalized to match the sample with the lowest OTU count. Rarefaction is a good tool to aid comparisons of compositional profiles as well as alpha and beta diversity indices.

1.3 Optimization of the stool processing pipeline

Many clinical researchers are looking to incorporate gut microbiome data into their studies, deciphering the intricate connection between gut and health. This has led to a rapid development of Next Generation Sequencing technologies, so researchers could acquire large amounts of microbiome data and apply it to the clinical setting. However, there is still no consensus on standard stool processing protocol to guarantee the sample quality and feasibility of metagenomic analysis. The detected compositional profile of gut microbiome is influenced by various factors in the stool processing pipeline (figure 3). All the steps in the pipeline are possible targets of optimization. Therefore, it is important to consider the pipeline leading to Next Generation Sequencing.

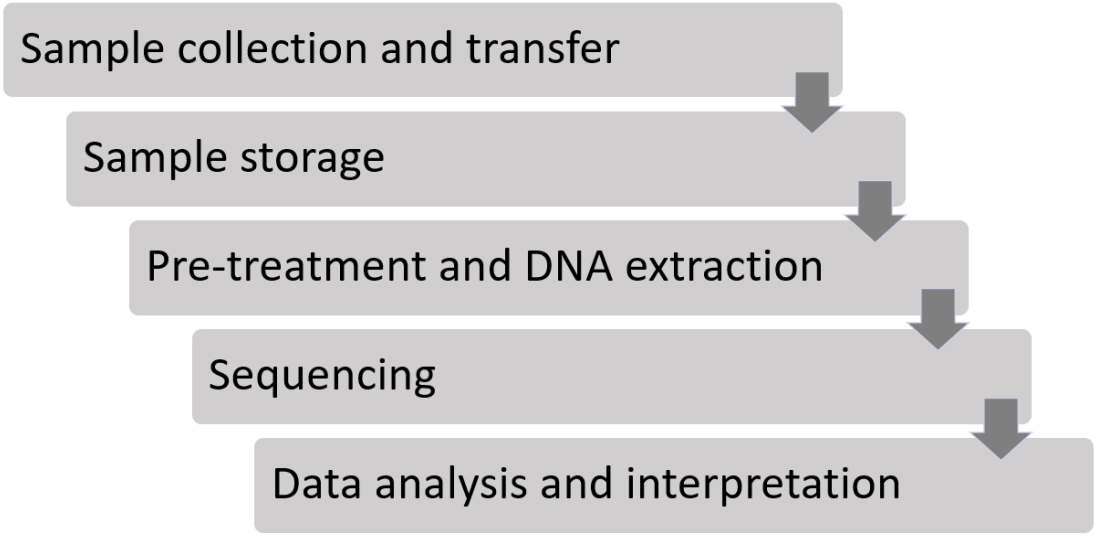


Figure 3. Simplified stool processing pipeline.

1.3.1 DNA extraction optimization

Although NGS technologies are known to have a high-throughput capacity, bacterial DNA extraction methods are often still manual and requiring significant amounts of hands-on-time. Indeed, the DNA extraction step can be considered as a bottleneck in a large-scale microbiology laboratory (Rintala et al. 2017). Furthermore, manual extraction methods are often subject to human error that may cause variability among experiments and lead to less reliable results. Manual extraction is done without the help

of an extraction robot and every extraction reagent is added to the sample individually using a pipette. On the other hand, semiautomatic extraction often includes a manual pre-treatment step, after which the DNA is further extracted with an extraction robot.

DNA extraction critically influences the bacterial compositional profiles (Videnska et al. 2019). A central step in microbial DNA isolation is to obtain sufficient amounts of high-quality DNA for downstream applications. However, DNA extraction from stool samples is not without challenges as several PCR inhibitors, such as bile salts, may be co-extracted with the sample (Lantz et al. 1997). Moreover, fecal samples contain normal proteins, that might inhibit PRC-reactions and other downstream applications (Schrader et al. 2012). Another challenge in bacterial DNA extraction is ensuring the complete lysis of the bacterial cell wall without shearing the genomic DNA. Commercial kits take different approaches to tackle these challenges mainly by differing from each other in the steps of DNA extraction. The core differences in extraction kits may stem from cell lysis methods and different extraction chemistries for contaminant removal (Rintala et al. 2017). Standard bacterial DNA isolation protocol includes cell lysis, contaminant removal, DNA binding, washing and elution. For example, cell lysis step can be mechanical or chemical. Contaminant removal can include incubation with protein degrading agents, such as proteinase K. Proteinase K is an enzyme that inactivates nucleases that might degrade DNA during extraction (Moore et al. 2008). Commercial kits also differ in the DNA capture method. Typically, DNA can be bound to a silica membrane in a filter or silica-coated magnetic beads. In the next step, DNA is washed with various ethanol-based solutions to remove residual salts, humic acids and other contaminants. Finally, DNA is eluted with water or low-salt elution buffer. A successful DNA isolation protocol should produce a high yield of pure and high integrity DNA, with good representation from all bacterial taxonomic groups (Yuan et al. 2012). Furthermore, DNA extraction protocols can be optimized to be time efficient and reproducible.

Optimization of DNA extraction procedures can be facilitated by different microbial standards. Some microbial community standards mimic the human gut microbiome, allowing to validate DNA extracting protocols. Microbiome community standards can

also act as a positive control for sample processing. Commercial mock standards may contain a mixture of microbes with cross kingdom representation. For example, ZymoBIOMICS Gut Standard (Zymo Research, USA) contains a defined composition of 18 bacterial, 2 fungal and 1 archaeal strain.

Homogenization is a process where a substance, such as stool, is reduced to small particles and distributed evenly though a liquid (e.g., buffer or a stool stabilizer). In gut microbiome context, bacterial cell wall is ruptured, and DNA is released and mixed into the surrounding liquid. Fecal samples contain hard-to-lyse gram-positive bacteria, that unlike gram-negative bacteria, have a thick peptidoglycan layer (Silhavy et al. 2010). Lysis of these gram-positive bacteria is especially important for achieving an unbiased representation of the bacterial community in the sample. Different methods of cell lysing have been developed and the chosen methods can be influenced by the ease of the chosen method, availability of the laboratory equipment and previous studies as well as the research objective. Homogenization and cell lysis can happen in multiple ways (figure 4). Cell lysis can be non-mechanical, mechanical or both (Islam et al. 2017). Non-mechanical lysis can be divided into enzymatic (e.g., lytic), chemical (e.g., detergents) and physical (e.g., osmotic shock). Mechanical methods include liquid shearing (e.g., bead-beading in buffer) and solid shearing (e.g., mortar and pestle grinding), though the latter is rarely used in gut microbiome studies

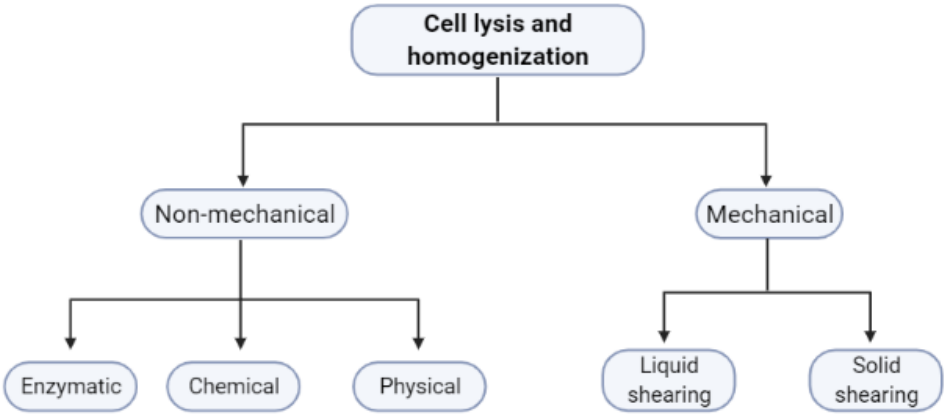


Figure 4. Cell lysis and homogenization methods. Created in BioRender.com

Bead-beating (liquid shearing) has been a “golden standard” in DNA extraction from stool samples. Even Human Microbiome Standards project (IHMS, 2011-2015) has incorporated bead-beating step into their recommended fecal DNA extraction protocol (IHMS 2015). The beads used in extraction procedures are small and are made of various materials, including ceramic, glass, silica, or zirconia (Wu et al. 2019). The bead size, material and the duration of homogenization process may affect the efficacy of bead-beating. Stool samples are placed into a bead-containing tubes or a plate and are homogenized with a bead-mill. Mechanical cell lysis and homogenization via bead-beating has been shown to increase the diversity of gut microbiota (Costea et al. 2017). This may be due to the added signatures from hard-to-lyse gram-positive bacteria. Moreover, pre-treatment prior to DNA extraction with bead-beating is generally more effective at lysing gram-positive bacteria than enzymatic or chemical methods, and it improves DNA recovery (Hsieh et al. 2016). Furthermore, bead-beating mixes the stool samples effectively and reduces intra-sample variance.

1.3.2 96-format

96-systems fit most of the modern laboratory equipment. In 96-format, the same sample size of 96 could be processed throughout the sample processing pipeline. More exactly, 96 samples could go through pre-treatment (lysis and homogenization), DNA extraction, DNA quantification, library preparation and sequencing. The first 96-plate system was introduced by a Hungarian researcher, Dr. Gyula Takátsy, in 1951 (Takatsy 1955). At that time, a serious influenza epidemic that was at loose in Hungary urged the researchers to develop a fast, reliable and economic method for identifying the influenza virus. The diagnostics were carried out on a hand-made microtiter plate; the liquid handling was performed with knitting needles. For time saving purposes, Dr. Takátsy arranged the knitting needles in a way that he could keep them in his hands without problems. This led to the development of a plate with 8x12 wells that could quickly and easily be filled. Thus, the 96-format was born, and it has prevailed through the years.

Recently, 96-format has been implemented in various laboratory settings. In the context of microbial DNA extraction, and in the context of this thesis, 96 samples go through the DNA extraction process. The introduction of 96-format could reduce hands-on time

and human error. Moreover, 96-format allows the processing of duplicate or triplicate samples on the same plate. Although 96-format allows flexible and faster diagnostics, it needs to be optimized for an optimal outcome. Furthermore, the format is not without its challenges. One of those challenges is contamination.

Contamination can be difficult to avoid in any laboratory setting. Potential sources of contamination include laboratory staff, working surfaces, PCR grade water, reagents and the samples themselves. Transfer of stool particles between adjacent wells, i.e., cross-contamination, is not to be overlooked on 96-format DNA extraction. Even careful pipetting can generate aerosols, that despite biosafety cabinets, may end up cross-contaminating stool samples. Cross-contamination between high bacterial biomass fecal samples (e.g., adult and senior) will probably not affect the compositional profiles of either sample, since both of those samples have abundant microbial community of their own. However, cross-contamination from a high-biomass to low-biomass sample (e.g., adult to infant) has a higher chance to skew the compositional profile of the low-biomass sample (Wu et al. 2019). Similarly, cross-contamination has a higher chance of affecting a sample that is otherwise of poor quality.

Negative and extraction controls are a way to monitor cross-contamination. In context of this thesis, negative controls are the storage liquid that the stool sample was in (e.g, OMNIgene fluid) and extraction controls are the DNA extraction lysis buffer. Negative and extraction controls are placed in a manner where cross-contamination could possibly occur, for example between samples. The controls go through the same processing pipeline as the stool samples. Thus, negative and extraction controls also may indicate the contamination profile of the whole processing pipeline. Unidirectional workflow, separation of pre- and post-PCR spaces and aseptic working technique are preventative measures that mitigate the effects of contamination.

1.4 Aims of the thesis

Many studies seem to focus on the differences between DNA extraction kits, largely comparing them based on compositional profiles and alpha and beta diversities. However, the role of proper cell lysis and homogenization is often overlooked. Moreover, the majority of the kits compared are manual and lack high-throughput capacity.

PerkinElmer's (PerkinElmer, Finland) Chemagic DNA Stool 200 H96 kit with Magnetic Separation Module I (MSM I) extraction system was chosen to be optimized. The Chemagic DNA stool isolation system is a semiautomatic protocol with magnetic bead-based DNA capture method and high-throughput capacity. Normally, this Chemagic stool protocol doesn't include a bead-beating step as lysis occurs chemically. However, multiple papers have shown that mechanical lysis with bead-beating increases microbial diversity and DNA recovery (Hsieh et al. 2016; Costea et al. 2017). Additionally, Human Microbiome Standards project (IHMS, 2011-2015) has recommended a bead-beating step (IHMS 2015).

This thesis aims to optimize a reliable and reproducible 96-format DNA extraction method for Next General Sequencing and other downstream applications. We will implement a bead-beating step with a bead plate in 96-format. Specifically, this study will look at DNA yield, integrity, microbial diversity measures, (cross-)contamination and easy execution as optimization criteria. Four different pre-treatment conditions will be implemented to test the lysis efficacy. We will extract bacterial DNA from infant, adult and senior individuals. ZymoBIOMICS Gut Standard (Zymo Research, USA) will be used to assess the resolution of the extraction method. This study will also include different negative controls to detect contamination and cross-contamination.

2 Materials and methods

2.1 Sample collection and storage

Fecal samples from healthy infant, adult and senior volunteers were collected in OMNIgene GUT tubes (DNA Genotek, USA) and DNA/RNA Shield Fecal Collection Tubes (Zymo Research, USA) from single individual per age group. OMNIgene collection tubes contained 2mL and DNA/RNA Shield tubes contained 9mL of the collection fluid. After three days of storage at room temperature, adult and senior fecal samples were divided into 400µl aliquots and infant samples into 600µl aliquots. The samples were frozen at -80°C until used for DNA extraction.

2.2 DNA extraction, quantification and quality control

All the samples were extracted using Chemagic DNA Stool 200 H96 kit (PerkinElmer, Finland) with Magnetic Separation Module I (MSM I) extraction robot (PerkinElmer). Three replicates of adult and senior samples and two replicates of infant sample were extracted. The extraction also included negative controls (OMNIgene fluid, DNA/RNA Shield fluid), extraction controls (Chemagic Lysis Buffer 1) and ZymoBIOMICS Gut Standard D6331 (Zymo Research, USA). Negative and extraction controls were used to detect cross-contamination, and Gut Standard to mimic the human gut microbiome and assess the resolution of the extraction method.

Pre-treatment procedures were modified from "Purification Protocol for Human Feces Material Using the Chemagic Magnetic Separation Module I" protocol (appendix 1) and preliminary testing was conducted prior to this extraction (appendix 5). The following volumes of reagents and samples were used in every pre-treatment group. Lysis Buffer 1 (800µL) was added into 200µL of fecal sample. Subsequently, 925µL of Lysis Buffer 1 was added into 75 µL of ZymoBIOMICS Gut Standard. For negative controls (OMNIgene and DNA/RNA Shield fluid) the volume of 800µL of Lysis Buffer 1 was added into 200µL of each collection fluid. Finally for Lysis Buffer 1 extraction control, 1mL of buffer was used. The pre-treatment was divided into four groups (table 1).

Table 1. Pre-treatment groups of DNA extraction. Abbr.=abbreviation. In abbreviation column the letter "C" indicates chemical lysis, "prot" proteinase K and "M" mechanical lysis.

Group	Abbreviation	Lysis	Pre-treatment
1	Cprot	Chemical	Manufacturer's protocol; incubation with proteinase K
2	C	Chemical	Manufacturer's protocol; incubation *without* proteinase K
3	CM	Chemical+Mechanical	Bead plate+TissueLyser 15 Hz; 2 x 5 min
4	CMprot	Chemical+Mechanical	Bead plate+TissueLyser 15 Hz; 2 x 5 min+proteinase K incubation

Group 1 or hence in text "Cprot". MSM I manufacturer's modified protocol with proteinase K incubations. The aforementioned amounts of fecal samples, Gut Standard and negative controls were added into 2mL screw cap tubes. The tubes were vortexed and 15µL of proteinase K was added. The tubes were incubated in thermo shaker at 70ºC for 10 min, followed by incubation at 95ºC for 5 min. Samples were centrifuged at 13,447 xg for 5 min. The lysate (800µL) was then transferred unto the sample plate and the extraction proceeded according to the manufacturer's protocol using MSMI.

Group 2 or hence in text "C". MSM I manufacturer's modified protocol without proteinase K incubations. The aforementioned amounts of fecal samples, Gut Standard and negative controls were added into 2mL screw cap tubes. The tubes were vortexed and incubated in thermo shaker at 70ºC for 10 min, followed by incubation at 95ºC for 5 min. The Samples were centrifuged at 13,447 xg for 5 min. The lysate (800µL) was then transferred unto sample plate and the extraction proceeded according to the manufacturer's protocol using MSMI.

Group 3 or hence in text "CM". Bead-beating with bead plate and TissueLyser II. The aforementioned amounts of fecal samples, Gut Standard and negative controls were added into PowerBead Pro Plates (Qiagen, USA). The plate was placed in TissueLyser II (Qiagen, USA) and shaken at 15 Hz for 5 min. The plate was reorientated so that the side that had been closest to the machine was now the farthest, and the plate was shaken again with the same settings. After shaking, the plate was centrifuged at 4,500 xg for 6 min and 800µL of lysate was transferred unto the sample plate and the extraction proceeded according to the manufacturer's protocol using MSMI.

Group 4 or hence in text “CMprot”. Bead-beating with bead plate and TissueLyser II and proteinase K incubations. The aforementioned amounts of feacal samples, Gut Standard and negative controls were added into PowerBead Pro Plates (Qiagen, USA). Homogenization was performed as in group CM (see above) and 800µL of lysate was transferred into 2mL screw cap tubes with 15µL of proteinase K. The tubes were vortexed and incubated in thermo shaker at 70ºC for 10 min, followed by incubation at 95ºC for 5 min. The tubes were briefly spinned and the lysate was transferred unto the sample plate and the extraction proceeded according to the manufacturer’s protocol using MSMI.

After extraction, DNA concentration was measured with Qubit 2.0 Fluorometer (ThermoFisher Scientific, USA) using Qubit dsDNA High Sensitivity Assay kit. DNA concentration measurement was tested in 96-format using Quant-iT dsDNA 1x HS kit (ThermoFisher Scientific, USA), however it was proven to be unreliable (see appendix 4). DNA integrity was determined by 1% TBE agarose gel. The DNA was divided into two 100µL aliquots and stored at -80ºC

2.3 16S sequencing

Microbial composition was determined by sequencing V3V4 and V4 regions of 16S ribosomal gene using Illumina MiSeq platform. The sequence library was constructed according to the Illumina library preparation protocol (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf) with minor differences to V3V4 protocol and V4 library with in-house protocol (see below). For V4, two out of three extraction replicates of adult and senior samples and all two extraction replicates of the infant samples were sequenced. V3V4 sequencing included all extraction replicates of fecal samples. ZymoBIOMICS Microbial Community DNA Standard (ZymoBIOMICS, USA) and PCR grade water were included in the sequencing as controls. Although the DNA extraction included DNA/RNA Shield samples, for the purpose of this thesis only OmniGENE fecal samples were sequenced and analyzed. Sequencing also excluded DNA/RNA Shield fluid negative controls.

The V3V4 protocol differed from Illumina's recommendation mainly regarding PCR reaction final volumes and DNA visualization procedures. Prior to PCR, the DNA samples were diluted to 2,5 ng/µL in PCR grade water. Briefly, amplicon PCR included 2x KAPA HiFi HotStart ReadyMix (Roche, USA), Illumina amplicon forward and reverse primers (6,6µM), PCR grade water and microbial DNA (16,5ng). The final volume of the amplicon PCR reaction was 33µL. After PCR, 8µL of the product was analyzed with 1,5% TAE agarose gel (120V, 1h). Index PCR was performed according to Illumina's instructions. After amplicon and index PCR, the products were purified with AMPure XP magnetic beads (Becman Coulter, USA). The concentration of the library samples was measured with Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA) using Qubit dsDNA High Sensitivity Assay kit. The purified samples were mixed in equimolar concentration to 4nM library pool. The library pool was denatured, diluted to concentration of 4pM and 8% denaturated PhiX control was added. The library samples were sequenced with Illumina MiSeq Reagent kit v3 (600 cycles) on Miseq system with 2x 300 bp paired ends following the manufacturer's instructions and using Miseq V3 reagent kits (Illumina, USA).

In V4 library preparation, amplicon PCR and index PCR were combined. The DNA samples were diluted in PCR grade water to 10ng/µL concentration prior to PCR. PCR was performed with KAPA HiFi High Fidelity PCR kit with dNTPs (Roche, USA). The desired concentration of each component was the following: 1x for 5x KAPA HiFi Fidelity Buffer, 0,3mM for dNTP mix, 0,5U for KAPA HiFi DNA polymerase and PCR grade water. Reverse and forward primers included in-house modifications validated by Rintala et al. (2017). The forward and reverse primer sequences were 5'-AATGATACGGCGACCACCGAGATCTACAC -i5- TATGGTAATT-GT-GTGCCAGCMGCCGCGGTAA-3'(forward) and 5'-CAAGCAGAAGACGGCATACGAGAT -i7- AGTCAGTCAG-GC-GGACTACHVGGGTWTCTAAT-3' (reverse), respectively, where i5 and i7 indicate the sample specific indices (see appendix 3 for index sequences). Primers were added in the concentration of 0,3µM. The concentration of template DNA was 50ng. The final volume of the reaction was 25µL. Combined amplicon and index PCR was performed

under following conditions; initial denaturation at 98ºC for 4 min, followed by 30 cycles consisting of denaturation at 98ºC for 20s, annealing at 65ºC for 20s and extension at 72ºC for 35s, and with a final extension at 72ºC for 10min. After PCR, 5µl of the product was analyzed with 1,5% TBE agarose gel (100V, 1h15min). PCR products were purified with in-house purification protocol (appendix 2) with AMPure XP magnetic beads (Becman Coulter, USA). The concentration of the library samples was measured with Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA) using Qubit dsDNA High Sensitivity Assay kit. The purified samples were mixed in equimolar concentration to 4nM library pool. The library pool was denatured, diluted to concentration of 4pM and 2% denaturated PhiX control was added. The library samples were sequenced with Illumina MiSeq Reagent kit v3 (600 cycles) on Miseq system with 2x 250 bp paired ends following the manufacturer's instructions and using Miseq V3 reagent kits (Illumina, USA).

2.4 Bioinformatic methods and data visualization

Raw sequence quality was visually checked using FastQC (Babrahan Bioinfromatics) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw sequence data for both libraries was processed and analyzed with CLC Microbial Genomics Module (CLC Genomics Workbench 21.0.3, Qiagen, Denmark). Workflows "Data QC and OTU clustering" and "Alpha and beta diversities" were used in analyzing the data with default settings. 16S read pairs were demultiplexed based on unique sequence indices and merged. The cut-off for fecal samples for the number of reads was 200,000 sequences, however negative and extraction controls were not filtered based on number of reads. One infant sample from V3V4 sequencing was excluded due to a low number of reads. Index and adapter trimming from 5' end was performed for both libraries. Sequences were mapped using SILVA 16S version 132 with 97% similarity for OTU clustering. Low abundance OTUs were filtered. Neighbour-Joining (NJ) tree was used for calculating alpha diversity and Shannon entropy, Chao 1 as well as total OTU number (observed OTUs) were selected to represent alpha diversity. Rarefaction value of 78,948 was used. Beta diversity calculations were performed with weighted UniFrac and Bray-Curtis indices and visualized with Principal Coordinate Analysis (PCoA). OTU tables, alpha and beta diversities were exported to GraphPad Prism 9.0.1 (151) for

data visualization. Statistical analyses were not appropriate to be performed due to a low sample size.

3 Results

3.1 DNA yield and integrity

The average concentrations of adult, senior and infant fecal samples as well as ZymoBIOMICS Gut Standards and their standard deviations across different pre-treatment groups are shown in table 2.

Table 2. The average DNA concentrations and standard deviations (ng/µl) of adult, senior and infant fecal samples and ZymoBIOMICS Gut Standards across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. No data is available for Senior DNA/RNA Shield samples for groups 1 (Cprot) and 2 (C) due to a mishap in DNA extraction. (n=2)

	1	2	3	4	Average all	Stdev all
Zymobiomics Gut Standard	6.445	6.03	4.93	4.005	5.35	1.1
Infant OMNIgene	32.05	33.5	11.65	8.11	21.33	13.31
Infant DNA/RNA Shield	25.05	18.61	8.115	7.13	14.73	8.62
Adult OMNIgene	28.8	31.1	26.17	23.2	27.32	3.4
Adult DNA/RNA Shield	29.87	33.17	24.67	27.33	28.76	3.63
Senior OMNIgene	46.57	53.7	47.43	55.33	50.76	4.4
Senior DNA/RNA Shield	36.07	15.04			25.56	14.87

The average concentrations across different pre-treatment groups for adult and senior samples were relatively similar (table 2). The highest concentration for adult samples was in non-bead-beating group Cprot with both stabilizer liquids. For senior OMNIgene samples, the highest concentration was in bead-beading with proteinase K incubation (group CMprot). For infants, the highest concentration was achieved without bead beading (groups Cprot and C for OMNIgene and DNA/RNA Shield respectively). Adult fecal samples seemed to have the least deviation in concentrations across pre-treatment groups, whereas infant samples had the highest standard deviation.

Gel electrophoresis of the stool and extraction can be seen from figure 4. The genomic DNA is well over 1000bp in length in all stool samples. Adult and senior OMNIgene

samples have a smear in pre-treatmnet groups Cprot (1) and C (2). However, the smear is fainter in groups CM (3) and CMprot (4). Extraction controls are not contaminated.

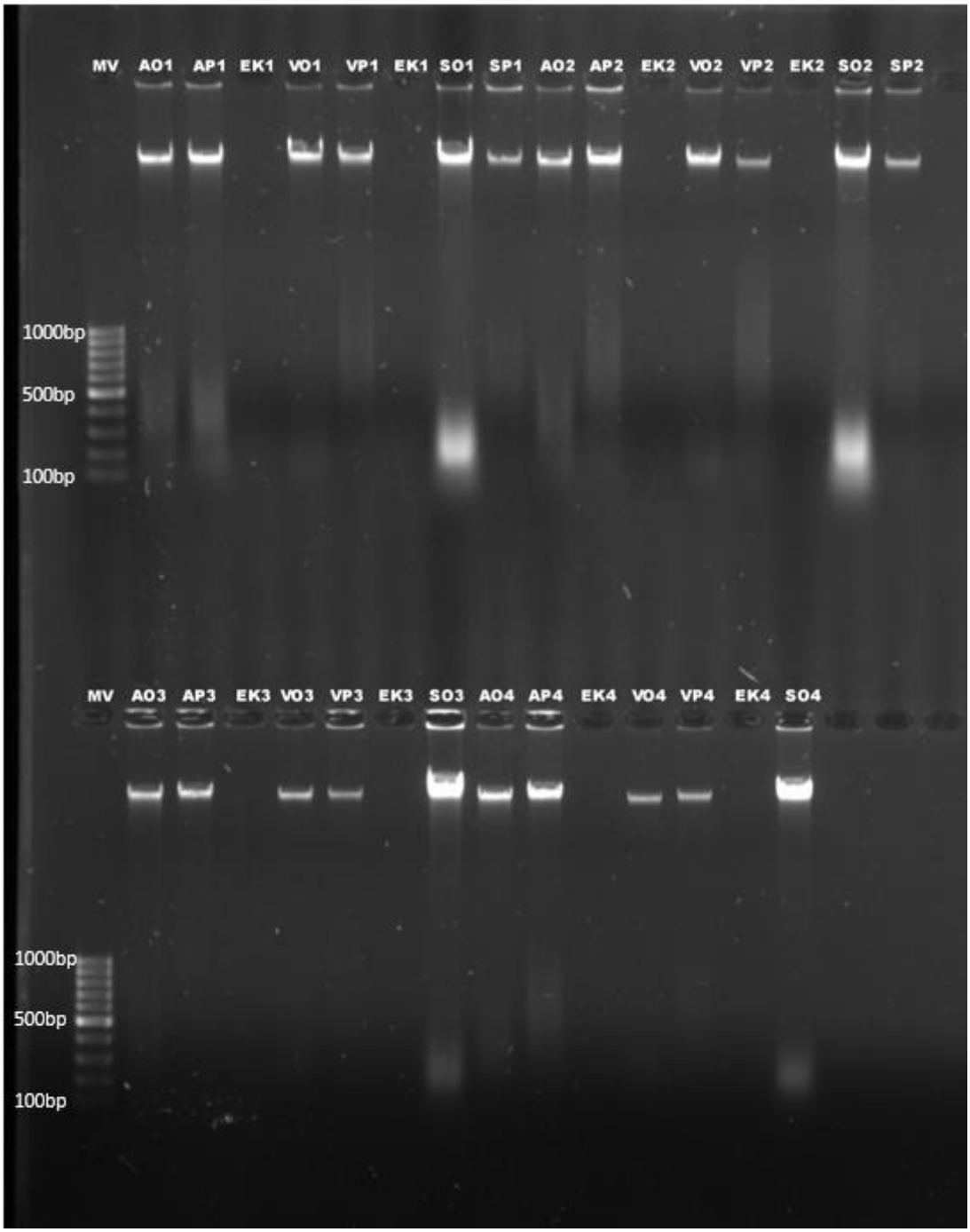


Figure 4. Agarose gel of stool samples and extraction controls across different pre-treatment groups after genomic DNA extraction. MV=molecular weight (DNA ladder), AO=adult OMNIgene, AP=adult DNA/RNA Shield, EK=extraction control, VO=infant OMNIgene, VP=infant DNA/RNA Shield, SO=senior OMNIgene, SP=senior DNA/RNA Shield. Number after the sample names indicate the number of the pre-treatment group (1,2,3 and 4). Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. DNA ladder GeneRuler 100bp. Gel 1% TBE. Gel run 110V, 1h.

3.2 Gut Standards

3.2.1 Relative abundances

Relative abundances of ZymoBIOMICS Gut Standards across different pre-treatment groups are shown in figure 5 (see appendix 6 for full list of detected genera). The manufacturer's (Zymo) expected abundances are shown in the figure on left. In relation to ZymoBIOMICS Gut standard, V3V4 sequencing produced more similar results to ZymoBIOMICS with V4 visibly differing from the manufacturer's expected abundances. However, the results were relatively similar across pre-treatment groups within the same sequencing target.

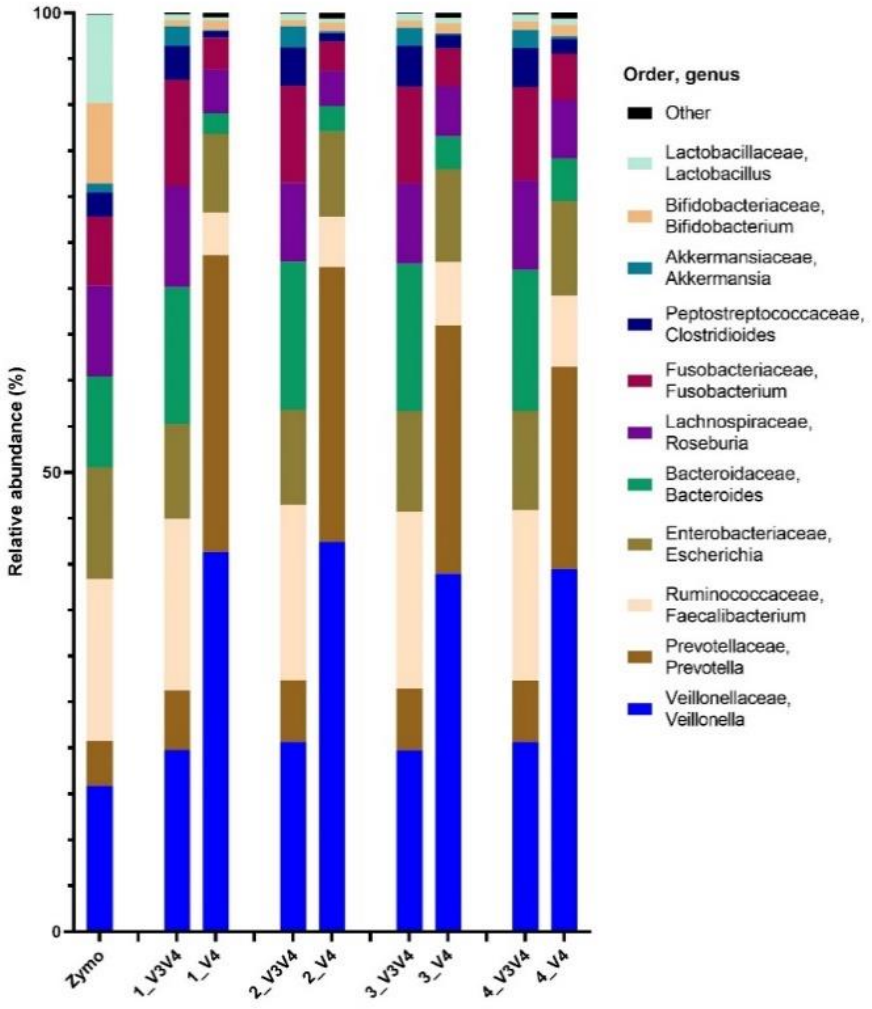


Figure 5. Relative abundances of ZymoBIOMICS Gut Standards across different pre-treatment groups with V3V4 and V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The manufacturer's expected relative abundances are shown on the left (Zymo). The bar chart shows 11 genera that are above ca. 1% abundance. (n=2)

V3V4 sequencing produced more similar results to the manufacturer's expected abundances, as mentioned above (figure 5). However, the V3V4 sequencing target had higher relative abundances of the genera *Fusobacterium, Clostridioides, Akkermansia, Bacteroides, Veillonella* and *Prevotella* than the manufacturer. On the other hand, the genera *Bifidobacterium* and L*actobacillus* were more abundant with ZymoBIOMICS' sequencing result. Visible changes with the pre-treatment are hard to see as the changes in relative abundances fluctuate between 1-2%.

V4 sequencing favored the genera *Veillonella and Prevotella* in comparison with the manufacturer's abundances (figure 5). V4 sequencing detected 5-6 times higher relative abundances with *Prevotella* and two times higher abundance with *Veillonella* in relation to the manufacturer. The genera *Faecalibacterium, Roseburia, Bactreriodes, Fusobacterium, Clostridioides, Akkermansia, Bifidobacterium* and *Lactobacillus* were lower in abundance in comparison with ZymoBIOMICS. V4 sequencing results had more variation between the pre-treatments groups than V3V4 sequencing. The genera *Veillonella* and *Prevotella* decreased in abundance from non-bead-beating groups (C and Cprot) to bead-beating groups (CM and CMprot). Subsequently, the genera *Bacteroides, Faecalibacterium, Roseburia, Lactobacillus* and *Bifidobacterium* increased in relative abundance with bead-beating (groups CM and CMprot). The relative abundance of *Esherichia* stayed consistent across the pre-treatment groups.

Both sequencing targets favored the genera *Veillonella* and *Prevotella* in relation to the manufacturer's result. The genera *Bifidobacterium* and *Lactobacillus* were lower in abundance, whereas *Esherichia* stayed consistent in abundance with both sequencing results.

Genera below 1% relative abundance ("other" in figure 5) are shown in figure 6. Pre-treatment groups Cprot and C in V4 sequencing failed to detect the genus *Enterococcus* (figure 6, appendix 6). Other low-abundance genera were detected with both sequencing targets. V3V4 had more similar results in low abundance genera across pre-treatment groups, whereas V4 sequencing had more variation across pre-treatment groups. Both V3V4 and V4 sequencing also detected other genera (ca. 0,1-0,6%) that were not present in the ZymoBIOMICS gut standard (appendix 6).

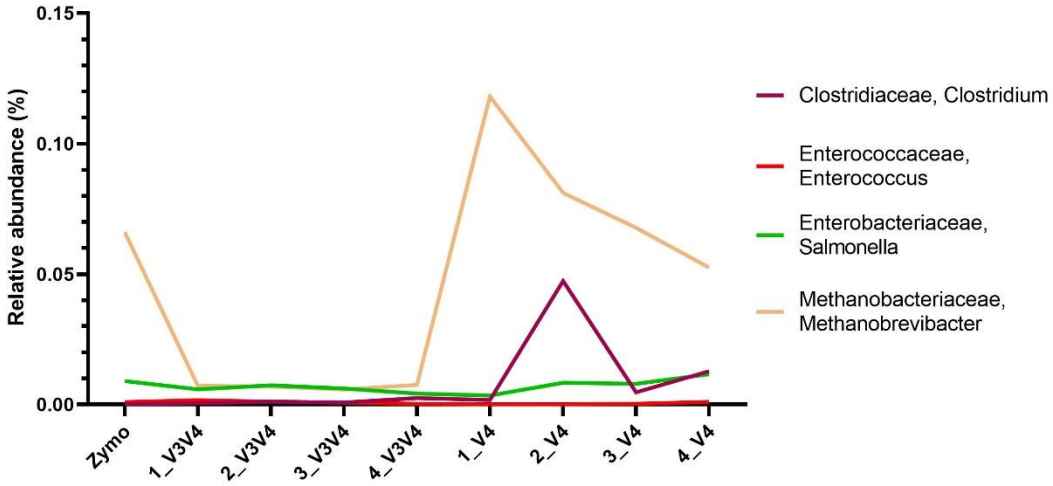


Figure 6. ZymoBIOMICS Gut Standard relative abundances below 1% abundance across different pre-treatment groups. Groups: Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. Manufacturer's (Zymo) expected abundances are shown on the left. (n=2)

3.2.2 Alpha diversities

Bead-beating groups CM and CMprot produced the highest alpha diversities with both sequencing targets, with the exception of Shannon entropy in V3V4 sequencing in group CM (figure 7). However, the deviation between replicates seems to be generally larger in beat-beating groups. V4 sequencing produced higher alpha diversity measures, apart from Shannon entropy.

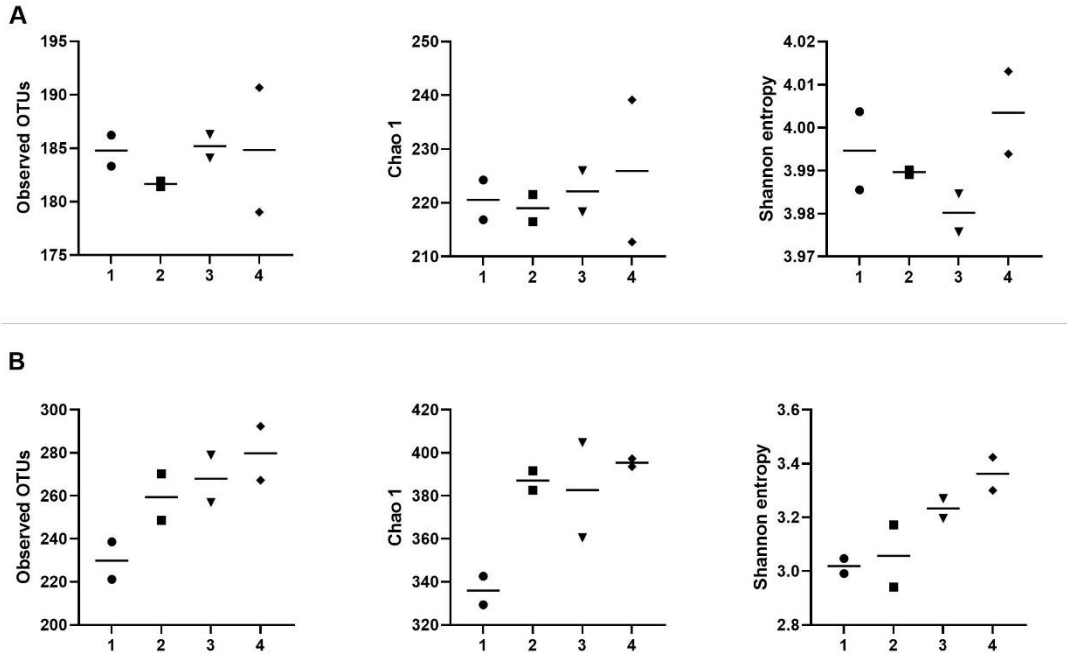


Figure 7. Alpha diversity indices of ZymoBIOMICS Gut Standards across different pre-treatment groups. A) V3V4 sequencing B) V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The different symbols indicate individual replicates within a pre-treatment group. The horizontal lines indicate the average in a pre-treatment group. (n=2)

3.2.3 Beta diversities

Beta diversities of ZymoBIOMICS Gut Standards are shown in figure 8. Diversity measures show variation across the pre-treatment groups. Group Cprot (chemical lysis, proteinase K) replicates grouped together with both sequencing targets and diversity measures.

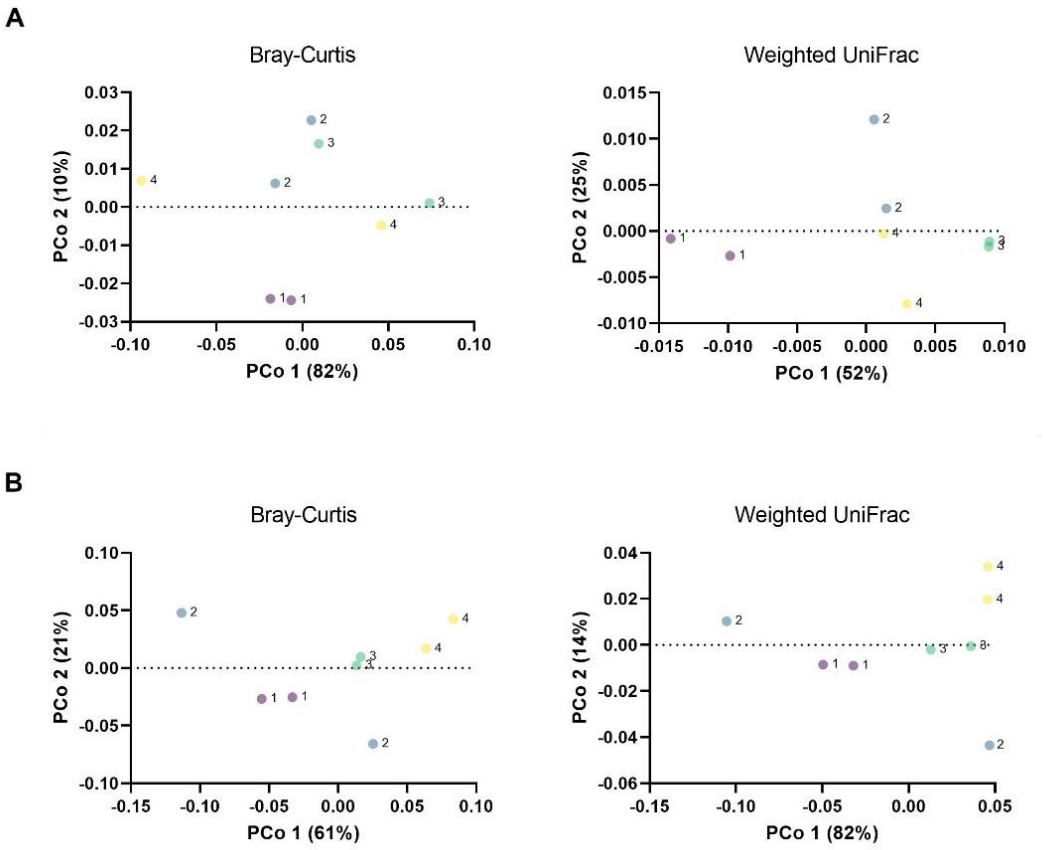


Figure 8. Beta diversity of ZymoBIOMICS Gut Standards across different pre-treatment groups with Bray-Curtis and Weighted UniFrac measures. A)V3V4 sequencing B)V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. Pre-treatments groups are indicated with differently colored dots and numbers adjacent to the dots. Purple=Cprot, blue=C, green=CM and yellow=CMprot (n=2)

3.3 Infant fecal samples

The most abundant phyla for infant samples were *Bacteroidetes*, *Firmicutes* and *Proteobacteria* (appendix 7A). Relative abundances of infant samples at genus level are shown in figure 9. Compositional profiles with the same sequencing target were relatively similar within the same sequencing target, however V3V4 and V4 differed from each other.

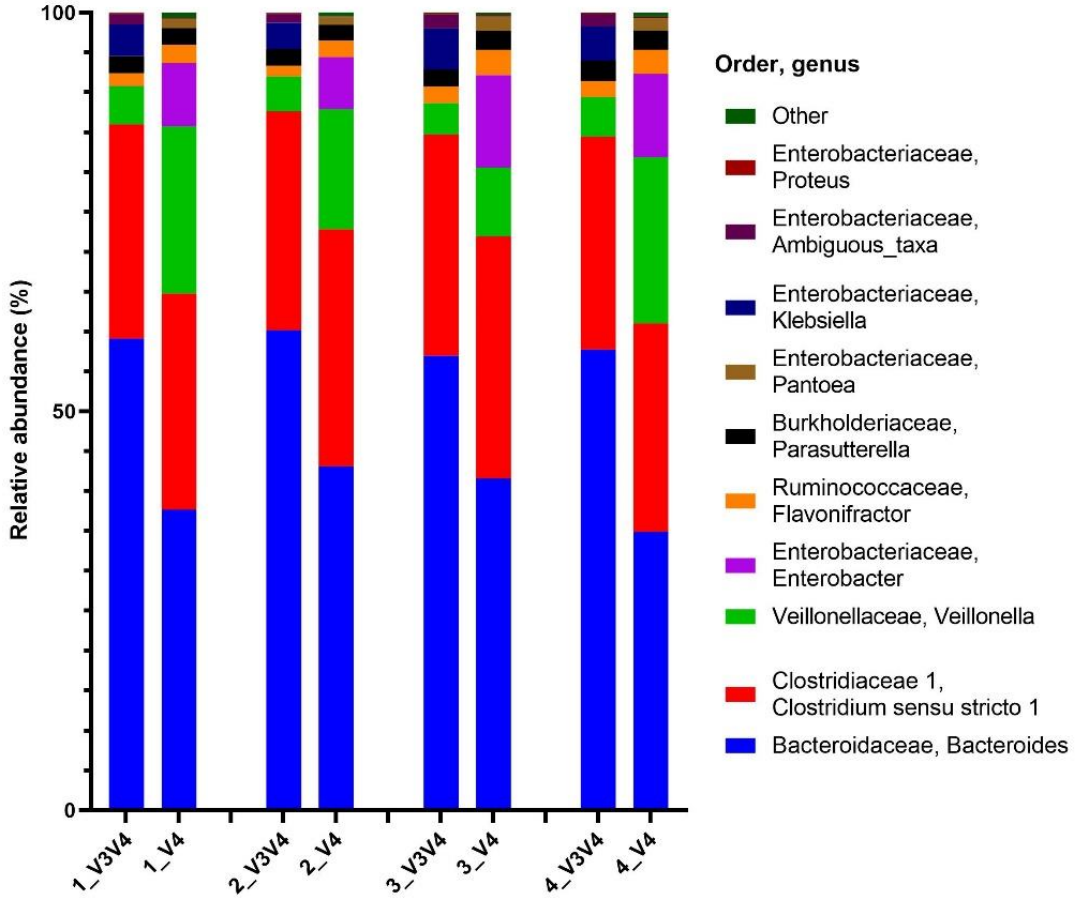


Figure 9. Relative abundance of V3V4 and V4 sequenced infant fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The bar chart shows 10 most abundant genera. (n=2, n=1 in V3V4 sequenced group CM)

With V3V4 sequencing, the genera *Flavonifactor* and *[Enterobactericeae] Ambiguous_taxa* increased in abundance with bead-beating. The genus *Bacteroides* decreased slightly in abundance with bead beating. *Clostridium sensu stricto 1* and *Parasutterella* stayed consistent across the pre-treatment groups. The genus *Veillonella* stayed relatively similar in abundance across pre-treatment groups, however the genus is

lowest in abundance in group CM (bead-beating without proteinase K). *Klebsiella* peaked in abundance in groups C (manufacturer's protocol with protK) and CM (bead-beating). The genera *Proteus* and *Enterobacter* stayed similar in abundance across the pre-treatment groups, however their abundances are below 1%. V3V4 sequencing did not detect *Pantoea.*

With V4 sequencing, the genera *Enterobacter, Flavonifactor, Parasutterella* and *Pantoea* increase in abundance with bead-beating. The genus *Bacteroides* peaked in abundance in groups C and CM and was lowest in abundance in group CMprot. *Veillonella* was lowest in abundance in groups C and CM, however the genus stayed consistent in abundance in groups Cprot and CMprot. The genera *Klebsiella, [Enterobactericeae] Ambiguous_taxa* and *Proteus* were below 1% in abundance, however these genera stayed similar in abundance across the pre-treatment groups. Finally, C*lostridium sensu stricto 1* stayed consistent across the groups.

With both sequencing targets, C*lostridium sensu stricto 1, Flavonifactor* and *Proteus* had similar trends for decreasing or increasing in abundance across pre-treatment groups. Both V3V4 and V4 detected low abundances (below 1%) of *Proteus.* V3V4 did not detect *Pantoea* and only very low abundances of *Enterobacter,* whereas V4 detected both *Pantoea* and *Enterobacter*. Similarly, V4 detected only low abundances of *Klebsiella*, whereas with V3V4 the genus was more abundant.

The 5 least abundant genera visible in the figure 9 are shown in greater detail in figure 10. *Klebsiella* is barely detected with V4 sequencing, whereas *Pantoea* increases with V4 sequencing.

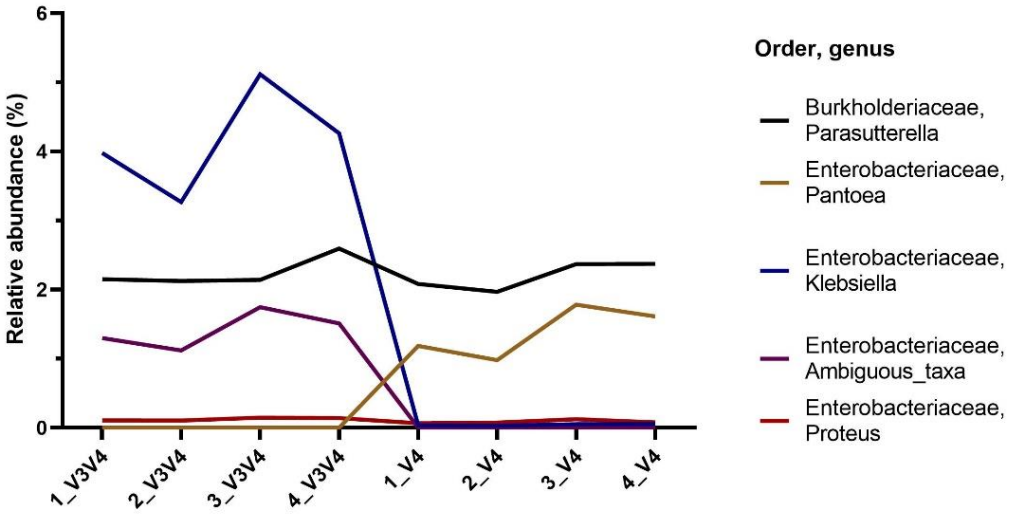


Figure 10. Line chart of 5-10 least abundant genera of infant fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. (n=2)

Alpha diversity indices of infant fecal samples are shown in figure 11. Group CM with V3V4 sequencing contains only one replicate due to the other replicate not having enough sequence reads. V3V4 sequencing produced the highest diversity indexes with bead-beating groups (CM and CMprot). Contrarily, V4 sequencing yielded the highest diversity indices with group Cprot (chemical lysis with proteinase K) with the exception of Shannon entropy. The deviation between the replicates is relatively large and diversity indices with V4 sequencing are higher, except with Shannon entropy index.

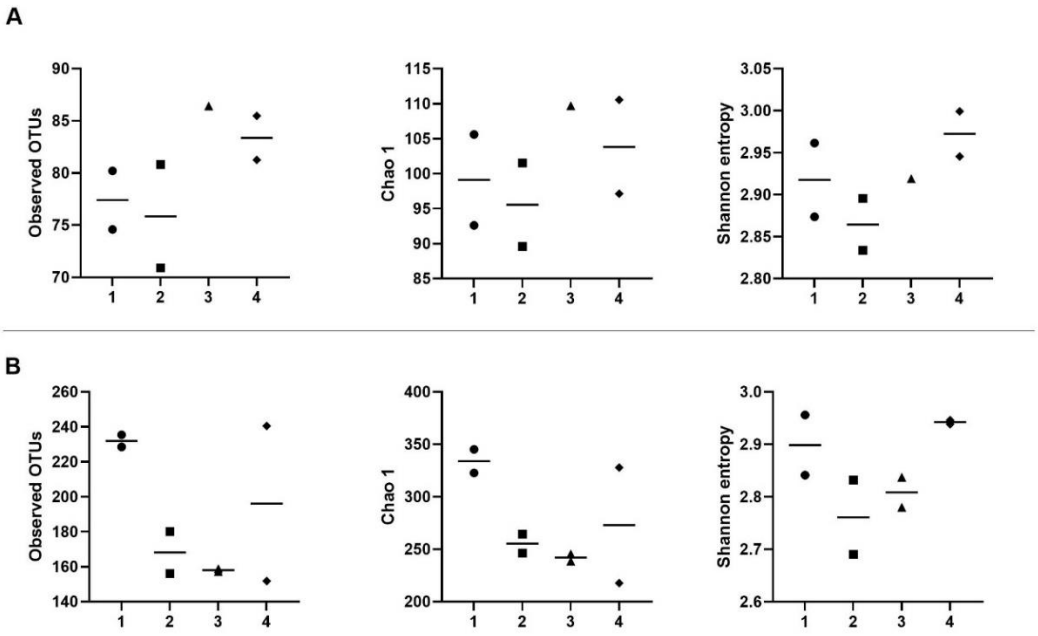


Figure 11. Alpha diversity indices of infant fecal samples across different pre-treatment groups. A) V3V4 sequencing B) V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The symbols indicate individual replicates within a pre-treatment group. The horizontal lines indicate the average in a pre-treatment group. (n=2, n=1 in V3V4 sequenced group 3/CM)

As for beta diversities, infant fecal samples did not exhibit specific grouping among the pre-treatment groups (figure 12).

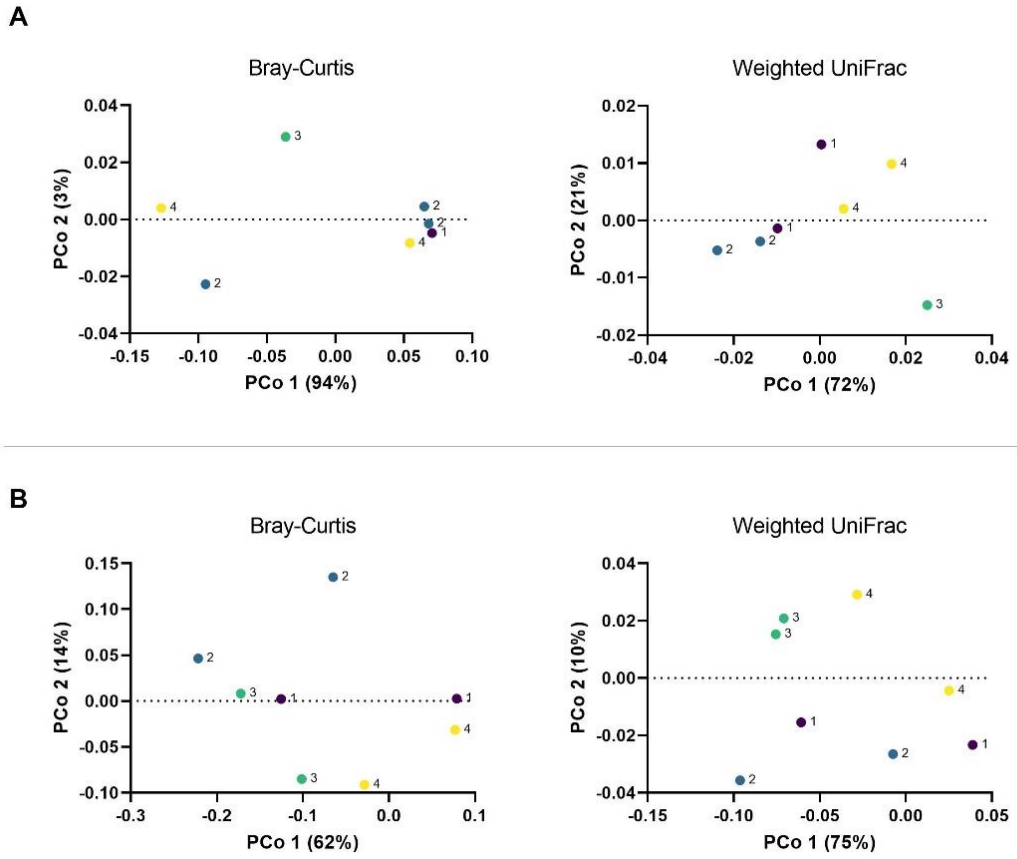


Figure 12. Beta diversity of infant fecal samples across different pre-treatment groups with Bray-Curtis and Weighted UniFrac measures. A)V3V4 sequencing B)V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. Pre-treatments groups are indicated with differently colored dots and numbers adjacent to the dots. Purple=Cprot, blue=C, green=CM and yellow=CMprot.(n=2, n=1 in group 3/CM with V3V4 sequencing

3.4 Adult fecal samples

Adult fecal samples are dominated by the phyla *Firmicutes*, followed by *Bacteroidetes* and *Actinobacteria* (appendix 7B). Relative abundances of adult samples at genus level are shown in figure 13.

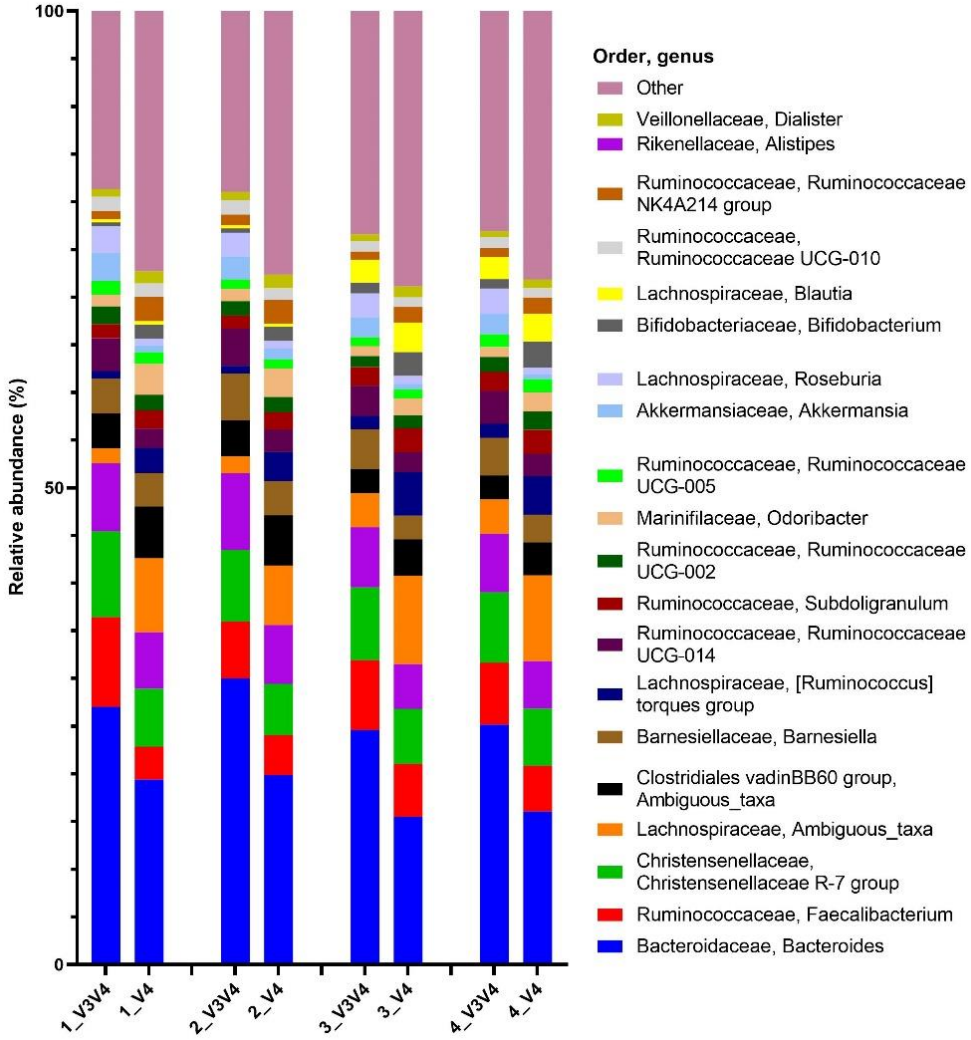


Figure 13. Relative abundance of V3V4 and V4 sequenced adult fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The bar chart shows 20 most abundant genera (n=2)

With V3V4 sequencing, the genera *[Lachnospiraceae] Ambiguous_taxa, [Ruminococcus] torques group, Subdoligranulanum, Bifidobacterium* and *Blautia* increase in relative abundance with bead-beating (figure 13). On the other hand, *Bacteroides, Faecalibacterium, [Clostridiales vadinBB60 group] Ambiguous_taxa, Ruminococcaceae UCG-002, Odoribacter, Ruminococcaceae UCG-005, Akkermansia, Ruminococcaceae UCG-010, Alistipes, Ruminococcaceae NK4A214 group* and Dialister decrease in abundance with bead-beating groups (CM and CMprot). The genera *Christensenellaceae R-7 group, Barnesiella, Ruminococcaceae UCG-014* and *Roseburia* stayed consistent in abundance across different pre-treatment groups.

With V4 sequencing the genera *Faecalibacterium, [Lachnospiraceae] Ambiguous_taxa, [Ruminococcus] torques group, Subdoligranulanum, Bifidobacterium* and *Blautia* increase in abundance with bead-beating (figure 13). The genera *Bacteroides, [Clostridiales vadinBB60 group] Ambiguous_taxa, Barnesiella, Odoribacter, Ruminococcaceae UCG-005, Akkermansia, Ruminococcaceae NK4A214 group, Alistipes* and *Dialister* decreased in abundance with bead-beating. Finally, *Christensenellaceae R-7 group, Ruminococcaceae UCG-014, Ruminococcaceae UCG-002, Roseburia and Ruminococcaceae UCG-010* stayed consistent across the pre-treatment groups.

With both sequencing results, the genera *Christensenellaceae R-7 group, [Lachnospiraceae] Ambiguous_taxa, [Clostridiales vadinBB60 group] Ambiguous_taxa, [Ruminococcus] torques group, Ruminococcaceae UCG-014, Subdoligranulanum, Odoribacter, Akkermansia, Roseburia, Bifidobacterium, Blautia, Alistipes* and Dialister had similar trends for decreasing or increasing in abundance across pre-treatment groups. For example, *Christensenellaceae R-7 group* stayed consistent in abundance across pre-treatments with both V3V4 and V4 sequencing.

Figure 14 shows 5 least abundant genera that are shown in figure 13 in more detail. As mentioned above, gram-positive *Blautia and Bifidobacterium* increased in abundance with bead-beating in both sequencing results. Contrarily, gram-positive *Ruminococcaceae UCG-010* and *Ruminococcaceae NK4A214 group,* as well as *Dialister,* decreased in abundance with bead-beating groups CM and CMprot.

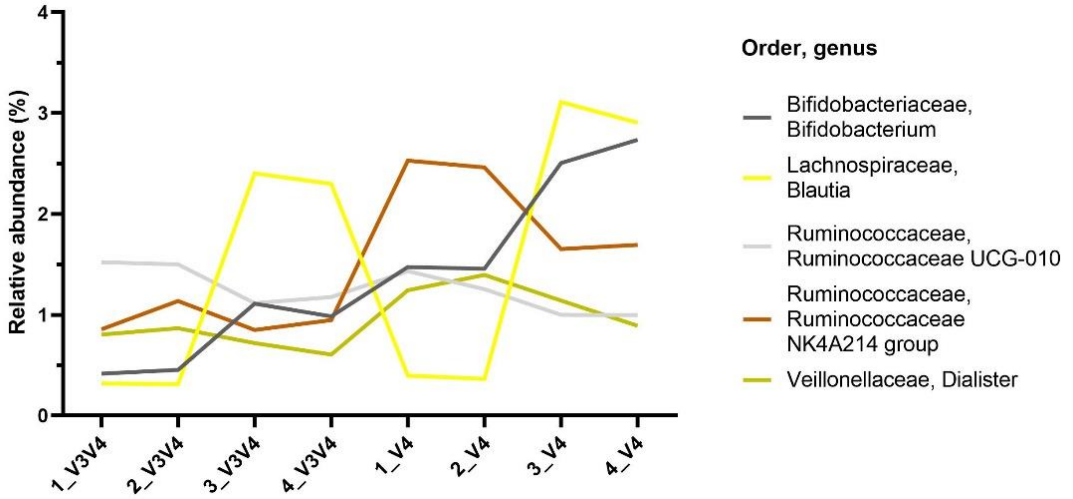


Figure 14. Line chart of 15-20 least abundant genera of adult fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. (n=2)

In adult samples, bead-beating groups C and CMprot produced higher alpha diversity measures than without bead-beating with both sequencing targets (figure 15). V3V4 sequencing with both bead-beating groups (CM and CMprot) produced similar diversity measures with every diversity index. However, bead-beating without proteinase K (group CM) yielded the highest diversity measures with V4 sequencing.

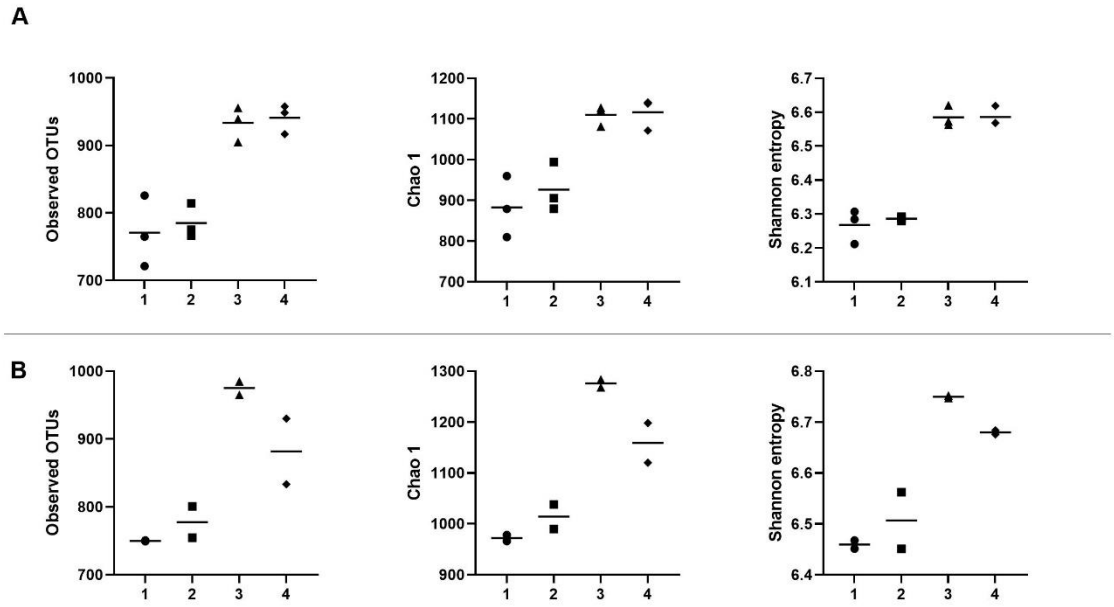


Figure 15. Alpha diversity indices of adult fecal samples across different pre-treatment groups. A) V3V4 sequencing B) V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The symbols indicate individual replicates within a pre-treatment group. The horizontal lines indicate the average in a pre-treatment group. (n=3 in V3V4 sequenced samples, n=2 in V4 sequenced samples)

For beta diversities, bead-beating samples without proteinase K incubation (CM) and with proteinase K incubation (CMprot) are loosely grouped in adult fecal samples (figure 16). Similarly, chemical lysis samples with proteinase K incubation (Cprot) and without proteinase K (C) are loosely grouped. Both sequencing results exhibit similar trends in pre-treatment grouping.

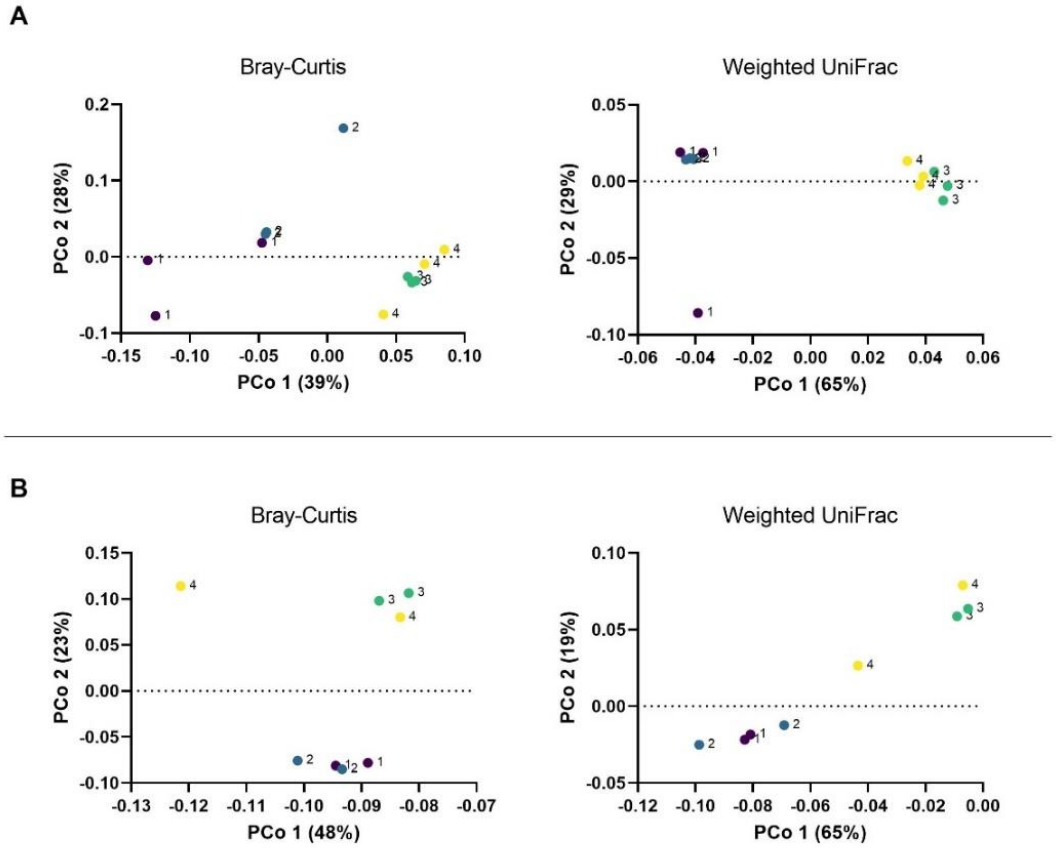


Figure 16. Beta diversity of adult fecal samples across different pre-treatment groups with Bray-Curtis and Weighted UniFrac measures. A)V3V4 sequencing B)V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. Pre-treatments groups are indicated with differently colored dots and numbers adjacent to the dots. Purple=Cprot, blue=C, green=CM and yellow=CMprot. (n=3 with V3V4 sequencing, n=2 with V4 sequencing)

3.5 Senior fecal samples

For senior fecal samples, the most abundant phyla were *Firmicutes*, *Bacteroidetes* and *Proteobacteria* (appendix 7C). Relative abundances of senior samples at genus level are shown in figure 17. Compositional profiles with the same sequencing target were relatively similar to each other, however V3V4 and V4 sequencing targets led to different compositional profiles.

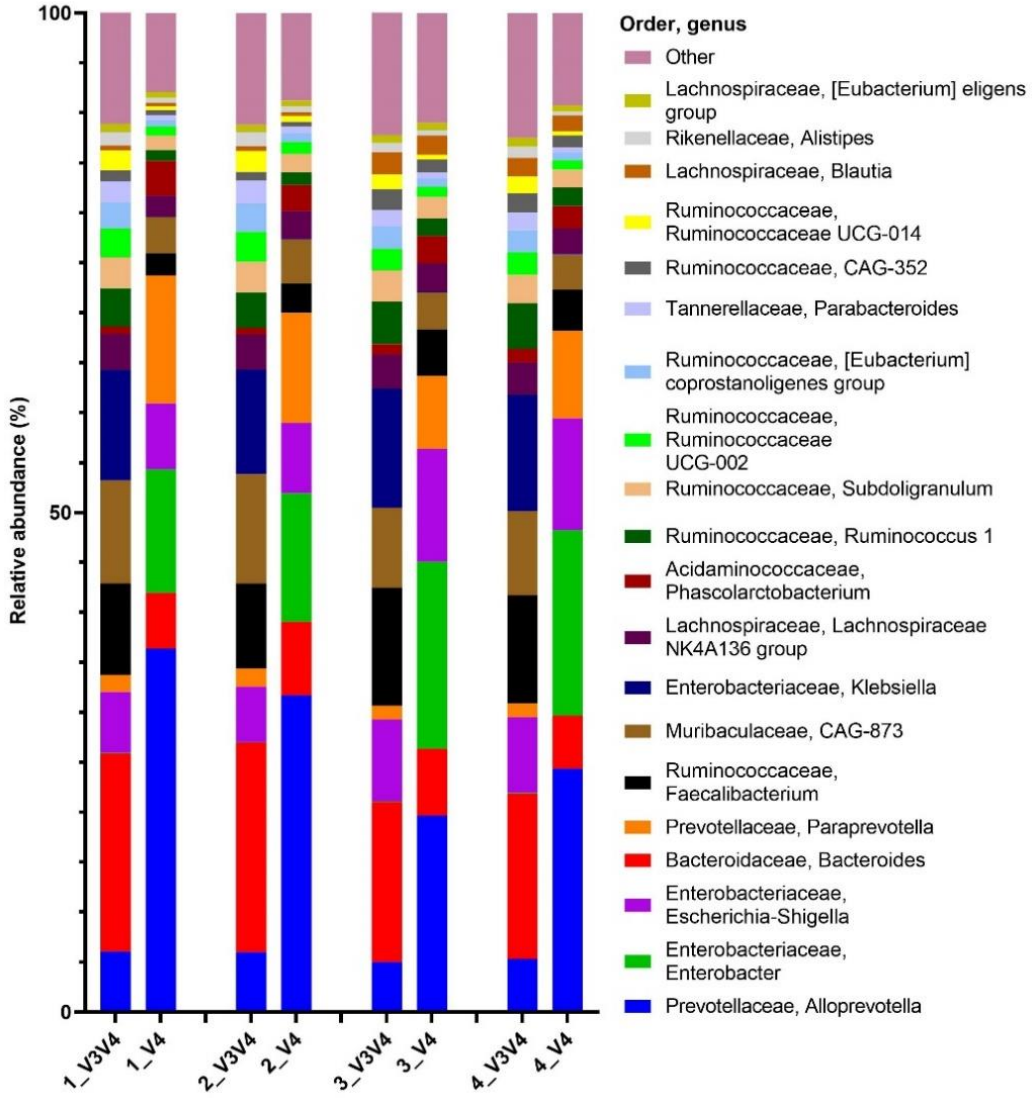


Figure 17. Relative abundance of V3V4 and V4 sequenced senior fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The bar chart shows 20 most abundant genera. (n=2)

With V3V4 sequencing, the genera *Esherichia-Shigella, Faecalibacterium, Phascolarctobacterium, Ruminococcus 1, Ruminococcus CAG-352* and *Blautia* increased in abundance with bead-beating (figure 17). On the other hand, *Bacteroides, Paraprevotella, Muribaculaceae CAG-873, Ruminococcus UCG-002, [Eubacterium] coprostanolignes group, Parabacteroides, Ruminococcus UCG-014* and *Alistipes* decrease in abundance with bead-beating groups (C and CMprot). *Alloprevotella, Klebsiella, Lachnospiraceae NK4A136, Subdoligranulanum* and *[Eubacterium] eligens group* stayed consistent across the pre-treatment groups.

With V4 sequencing, the genera *Enterobacter, Esherichia-Shigella, Faecalibacterium, Lachnospiraceae NK4A136, Ruminococcus 1, Subdoligranulanum, Ruminococcus CAG-352* and *Blautia* increased in abundance with bead-beating (figure 17). The genera *Alloprevotella, Bacteroides, Paraprevotella and Phascolarctobacterium* decreased in abundance with bead-beating. *Muribaculaceae CAG-873, Ruminococcus UCG-002, [Eubacterium] coprostanolignes group, Parabacteroides, Ruminococcus UCG-014, Alistipes and [Eubacterium] eligens group* stayed consistent across the pre-treatment groups.

With both sequencing results, the genera *Esherichia-Shigella, Bacteroides, Paraprevotella, Faecalibacterium, Ruminococcus 1, Ruminococcus CAG-352, Blautia* and *[Eubacterium] eligens group* had similar trends for decreasing or increasing in abundance across pre-treatment groups. However, depending on the pre-treatment group, V4 sequencing detected 6-3 times higher abundances of *Alloprevotella* than V3V4 sequencing. V4 sequencing also favored the genus *Enterobacter,* where V3V4 sequencing detected only minimal abundances (ca. 0,04-0,06%). Subsequently, V4 sequencing detected only low abundances of *Klebsiella (*ca. 0,04-0,06%), whereas V3V4 sequencing favored the genera.

Figure 18. shows 5 least abundant genera visible in figure 17. These five genera are shown in greater detail in the line chart. The genera *Blautia* and *Ruminococcaceae CAG-352,* both gram-positive bacteria, increase in abundance with bead-beating groups (CM and CMprot) with both sequencing targets.

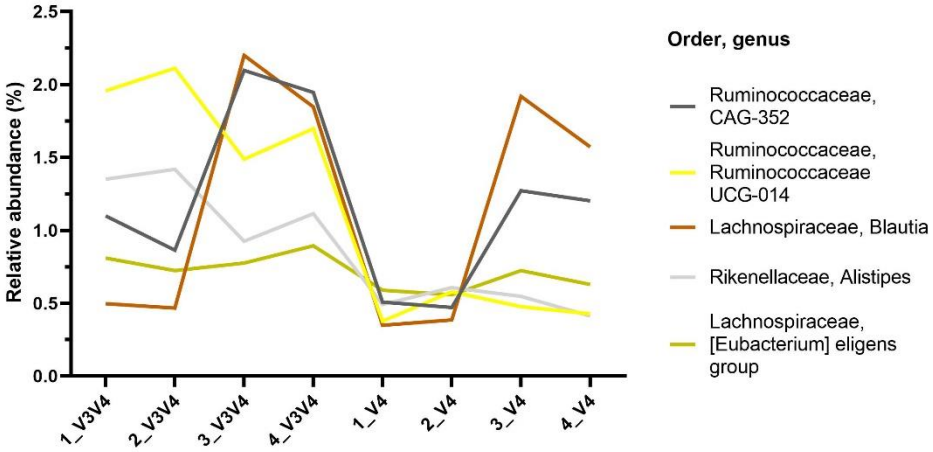


Figure 18. Line chart of 15-20 least abundant genera of senior fecal samples across different pre-treatment groups. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. (n=2)

With senior, bead-beating groups CM and CMprot produced higher alpha diversity measures than without bead-beating with both sequencing targets (figure 19). Bead-beating with proteinase K (group CMprot) produced the highest alpha diversity measures with V3V4 sequencing. Contrarily, bead-beating without proteinase K (group CM) yielded highest diversity measures with V4 sequencing.

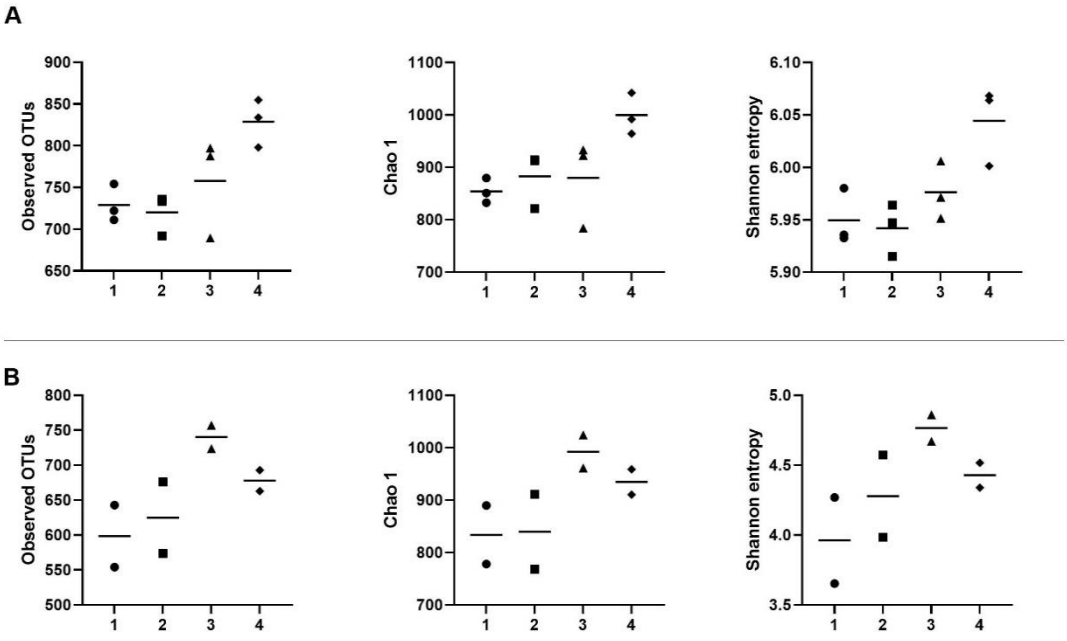


Figure 19. Alpha diversity indece of senior fecal samples across different pre-treatment groups. A) V3V4 sequencing B) V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The symbols indicate individual replicates within a pre-treatment group. The horizontal lines indicate the average in a pre-treatment group. (n=3 in V3V4 sequenced samples, n=2 in V4 sequenced samples)

Bead-beating samples without proteinase K incubation (CM) and with proteinase K incubation (CMprot) are loosely grouped in senior fecal samples (figure 20). Similarly, chemical lysis samples with proteinase K incubation (Cprot) and without proteinase K (C) are loosely grouped. Both sequencing results exhibit similar trends in pre-treatment grouping. However, V4 sequenced samples suggest having more variation within groups Cprot and C.

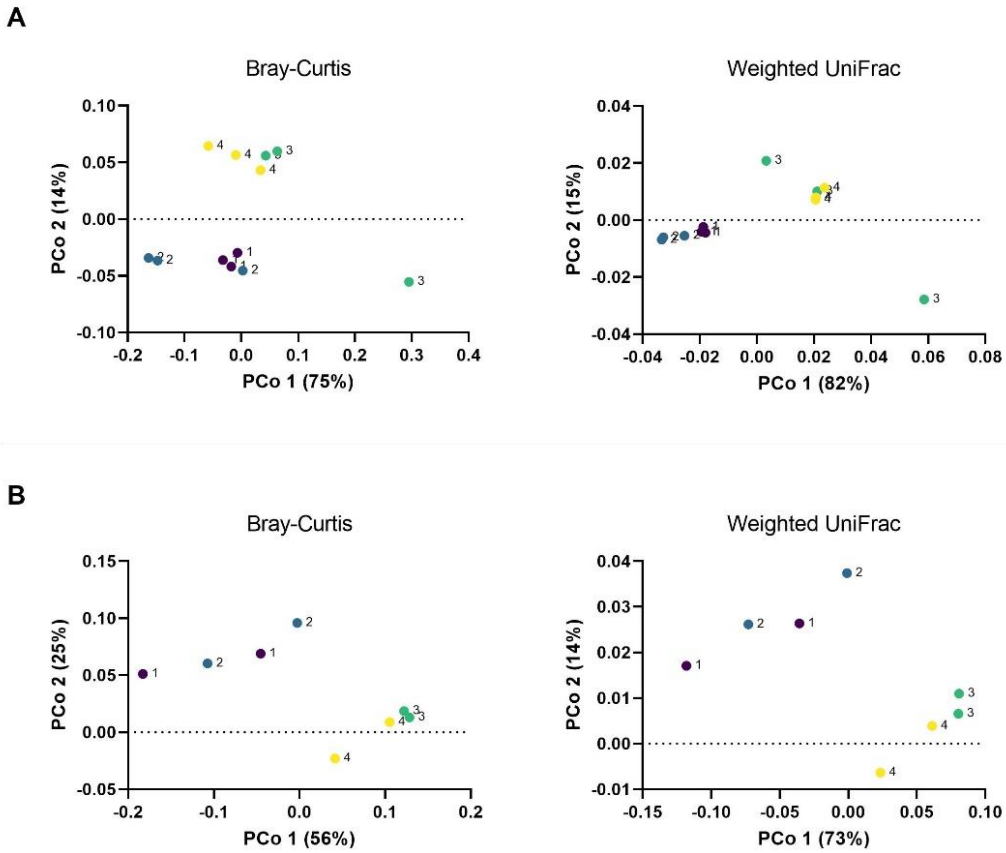


Figure 20. Beta diversity of senior fecal samples across different pre-treatment groups with Bray-Curtis and Weighted UniFrac measures. A)V3V4 sequencing B)V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. Pre-treatments groups are indicated with differently colored dots and numbers adjacent to the dots. Purple=Cprot, blue=C, green=CM and yellow=CMprot.(n=3 with V3V4 sequencing, n=2 with V4 sequencing)

3.6 Negative and extraction controls

Negative and extraction controls with V3V4 sequencing are dominated by genera present in fecal samples, such as *Bacteroides, Faecalibacterium* and *Klebsiella* (figure 21). For the most part, the compositional profiles resemble those of fecal samples.

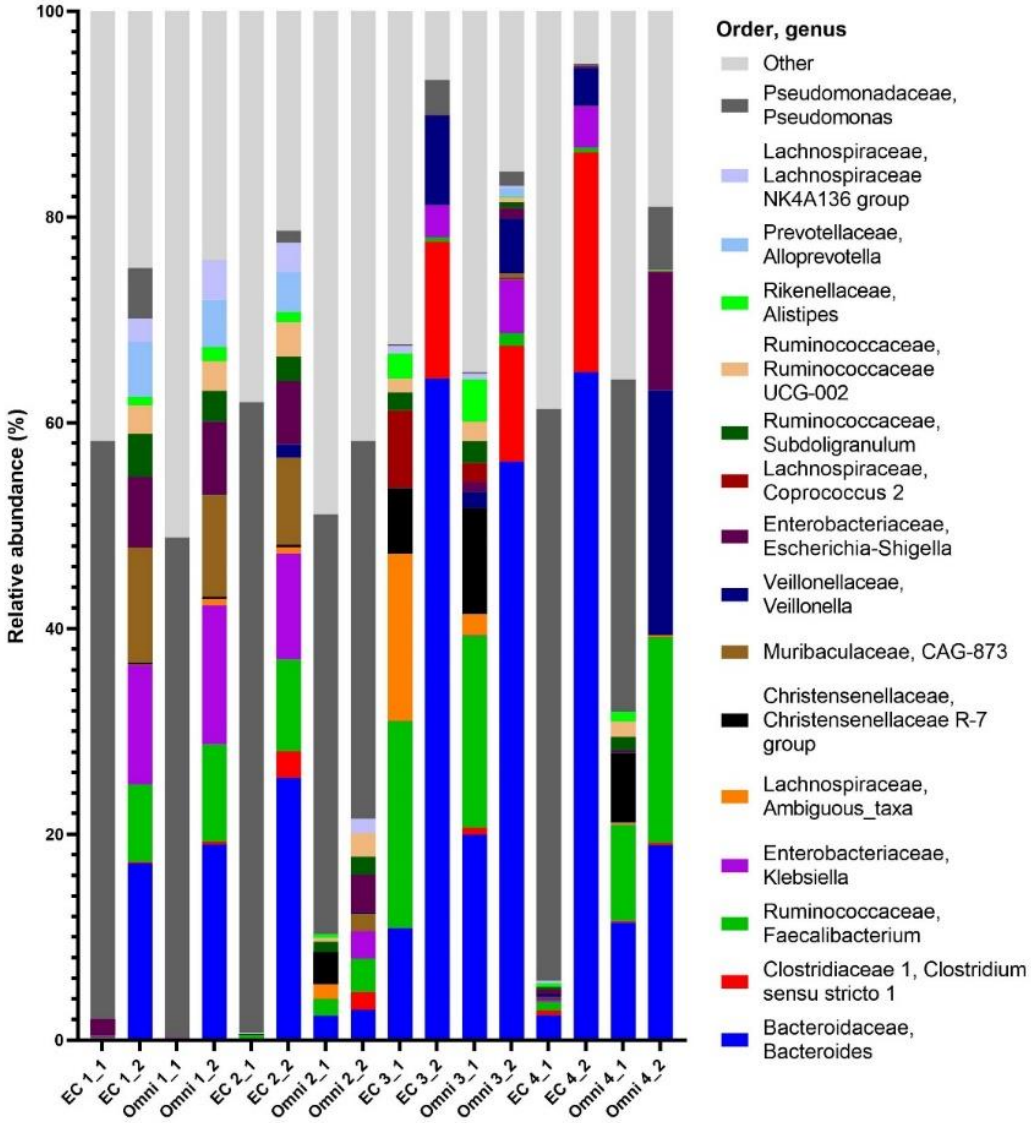


Figure 21. V3V4 sequenced Lysis Buffer 1 extraction controls (EC) and OMNIgene fluid (Omni) negative controls. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The extraction group is indicated before the underscore and the replicate number after the underscore. The bar chart shows 16 most abundant genera.

Negative and extraction controls with V4 sequencing are dominated by genera present in fecal samples, such as *Bacteroides, Veillonella* and *Alloprevotella* (figure 22). It can be seen that the negative controls contained same genera as the fecal samples (figure 22).

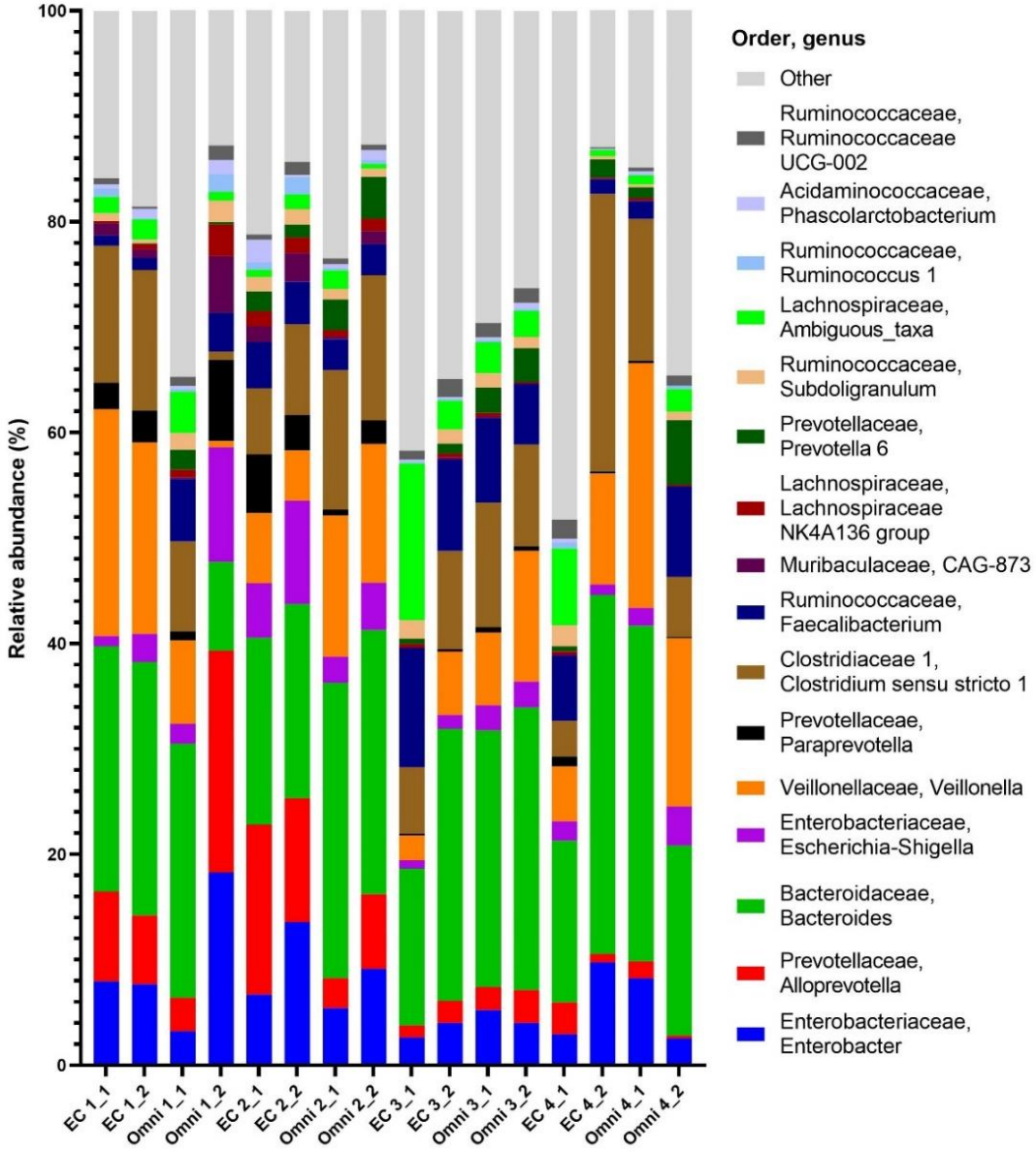


Figure 22. V4 sequenced Lysis Buffer 1 extraction controls (EC) and OMNIgene fluid (Omni) negative controls. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The extraction group is indicated before the underscore and the replicate number after the underscore. The bar chart shows 16 most abundant genera.

Both V3V4 and V4 detected genera present in the fecal samples. V3V4 sequenced controls had more variation in the compositional profiles, whereas V4 results were more uniform.

V3V4 sequenced controls had on average 203,857 reads before trimming and 198,598 reads after trimming (table 4). PCR-control contained 2,894 reads (not in table). The individual read counts of negative and extraction controls are shown in table 4. Bead-beating did not increase the read count (t-test=0.28).

Table 3. Read counts, average lengths and trim percentages of V3V4 Lysis Buffer 1 extraction controls (EC) and OMNIgene fluid (Omni) negative controls. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The extraction group is indicated before the underscore and the replicate number after the underscore.

Name	Number of reads	Avg.length (bp)	Number of reads after trim	Percentage trimmed (%)	Avg.length (bp) after trim
EC 1_1	8,764	300	8,385	95.68	275
EC 1_2	82,124	300	7,9437	96.73	281
Omni 1_1	11,764	299	11,259	95.71	274
Omni 1_2	522,972	300	507,531	97.05	280
EC 2_1	5,938	299	5,562	93.67	273
EC 2_2	296,300	300	284,690	96.08	280
Omni 2_1	9,070	300	8,650	95.37	275
Omni 2_2	5,188	299	4,872	93.91	275
EC 3_1	479,774	300	469,214	97.8	281
EC 3_2	58,428	300	57,014	97.58	283
Omni 3_1	476,310	300	466,584	97.96	281
Omni 3_2	155,446	300	151,840	97.68	282
EC 4_1	5,066	298	4,622	91.24	272
EC 4_2	1,106,800	300	1,081,719	97.73	283
Omni 4_1	8,612	299	8,154	94.68	277
Omni 4_2	29,158	300	2,8035	96.15	279

V4 sequenced controls had on average 98 419 reads before trimming and 28 422 reads after trimming (table 5). The PCR-control contained 50,888 reads (not in table). The individual read counts of negative and extraction controls are shown in table 5. Bead-beating did not increase the read count (t-test=0.36).

Table 4. Read counts, average lengths and trim percentages of V4 sequenced Lysis Buffer 1 extraction controls (EC) and OMNIgene fluid (Omni) negative controls. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. The extraction group is indicated before the underscore and the replicate number after the underscore.

Name	Number of reads	Avg.length (bp)	Number of reads after trim	Percentage trimmed (%)	Avg.length (bp) after trim
EC 1_1	24,342	251	2,133	8.76	234
EC 1_2	19,214	251	6,211	32.33	239
Omni 1_1	52,022	251	14,789	28.43	232
Omni 1_2	263,548	251	226,014	85.76	240
EC 2_1	146,224	251	16,493	11.28	230
EC 2_2	122,550	251	28,992	23.66	236
Omni 2_1	45,104	251	16,973	37.63	233
Omni 2_2	97,970	251	19,603	20.01	234
EC 3_1	277,822	251	19,962	7.19	231
EC 3_2	50,732	251	14,373	28.33	233
Omni 3_1	168,428	251	20,162	11.97	232
Omni 3_2	161,336	251	18,150	11.25	231
EC 4_1	6,278	251	3,304	52.63	238
EC 4_2	43,926	251	24,039	54.73	238
Omni 4_1	77,468	251	14,737	19.02	232
Omni 4_2	17,742	251	8,814	49.68	240

3.7 Results summary

No notable difference in DNA yields could be seen between pre-treatment methods, except infant samples showing higher yields in groups with chemical lysis. Genomic DNA was visually assessed with gel electrophoresis and DNA was over 1000bp in length. Beat-beating led to higher diversity measures in Gut Standards, and adult and senior fecal samples. Infant fecal samples did not exhibit group-specific increase in abundance. Negative and extraction controls showed contamination from fecal samples. With all samples, the sequencing target (V3V4 or V4) had a high impact on compositional profiles.

4 Discussion

4.1 Feasibility of the 96-format extraction

As stated before, obtaining sufficient amounts of genomic DNA is integral in DNA extraction. We used Gut Standards to assess the DNA yield. Zymo Research states that the expected yield from 75µl of the Gut Standard is 1µg of DNA. (Zymo Research 2022). The average concentration of the Gut Standards across pre-treatment groups is 5,35 ng/µl (table 2). Hence, the absolute concentration in the elute (200µL) is 1 070ng or ~1µg. Therefore, it can be concluded that the lysis efficiency was in accordance with the manufacturer's instructions with all pre-treatment methods. With fecal samples, isolation kits affect the DNA yield, so there are no standard concentrations for fecal samples (Videnska et al. 2019). As the 16S rRNA library preparation was successful, with the exception of one infant fecal sample, it can be concluded that DNA quantity and quality were sufficient with all of the pre-treatment methods.

In terms of whole genome sequencing (WGS), different sequencing platforms have different requirements for DNA yields. The expected DNA yields could range from 100ng to 1000ng of genomic DNA (Nouws et al. 2020). For example, PacBio systems require 1µg of genomic DNA (PacBio 2022). In clinical setting, it is often important to obtain species-level metagenomic information as bacterial functionality and metabolic activity vary greatly within bacterial genera, making genus-level taxonomic resolution lacking. Whole genome sequencing typically has higher requirements for DNA quantity and quality than 16S rRNA sequencing. In this study, all of the pre-treatment methods for fecal samples produced over 1µg of genomic DNA. Thus, it can be concluded that DNA quantity with all of the pre-treatment methods would be sufficient for more demanding microbiome analysis as well.

The Gut Standard proved to be helpful in assessing the lysis accuracy of pre-treatment groups. For example, V4 sequencing in pre-treatment groups Cprot and C did not detect the gram-positive genus *Enterococcus* (figure 6, appendix 6). The lysis could be insufficient in those two first groups, since the genus was detected in bead-beating

groups CM and CMprot. Low-abundance species below 1% were not detected in the abundances according to the Zymo Research. A study by Seol et al. (2022) showed similar results where species abundances below 1% were not detected or deviated from the manufacturer's abundances. Although, some bioinformatic pipelines might exclude such low-abundance genera from the analysis. Indeed, using ZymoBIOMICS Gut Standard as an indicator for taxonomic resolution can be challenging. Direct comparisons between the manufacturer's abundances and the abundances of this thesis are likely not practicle, as the two have different extraction methods. Moreover, ZymoBIOMICS has the abundances in species-level, making it necessary to aggregate them to genus level to match the taxonomic resolution of this thesis. Despite aggregating to genus level, expected manufacturer's abundances and observed Gut Standard abundances seemed to differ considerably (figure 5). Additionally, ZymoBIOMICS did not disclose the bioinformatic pipeline that was used in analyzing the sequence data. Based on previous research, all the steps in the stool microbiome processing pipeline influence the compositional profiles of microbial samples (Wu et al. 2019). Considering this, the manufacturer's expected relative abundance percentages might not be entirely applicable in the context of this thesis. Alternatively, an internal control, such as spike-in-control, could be added to the samples. In this thesis, internal controls were not used, however they could provide additional reassurance of the success of DNA extraction and sequencing.

In addition to the quality check from the Gut Standards, DNA integrity was assessed visually with 1% TBE agarose gel. A faint smear is visible in adult and infant samples in groups Cprot and C, however the smear is more pronounced in senior samples in the same groups (figure 4). The majority of the smear falls upon 100-500bp range. On the other hand, in bead-beating groups CM and CMprot, the smear is not as apparent in senior fecal samples, and in infant as well as adult samples, the smear is barely visible. In conclusion, DNA integrity was approvable.

With the consideration of hands-on-time, bead-beating groups CM and CMprot rise above the other two groups. In groups Cprot and C, the pre-treatment was performed in 2mL screw cap tubes. Screwing and unscrewing all 96 caps takes a considerable amount

of time. Moreover, there is a risk of mixing the caps and cross-contaminating samples. Additionally, tube centrifuges typically hold 24 samples at a time. Thus, all 96 samples contained in the screw cap tubes could not be centrifuged at the same time. On the other hand, pre-treatment was performed in 96-format with groups CM and CMprot. This results in the reduction of hands-on-time, mainly with homogenization step. Homogenization with 96 bead plate allows more user-friendly experience, as all the samples can be centrifuged and homogenized at the same time, and a multichannel pipette can be used to transfer the samples. With the right equipment, both homogenization and proteinase K treatment can be done in 96-format, further chipping form the hands-on-time. The pre-treatment of 96 samples with proteinase K took roughly 1.5-2 hours. Then, the extraction with MSM I took 1h. Lastly, the measuring of DNA concentration and dividing the DNA into aliquots took 1h more. In conclusion, DNA extraction and quantity measurements can be done on the same day.

Despite the user-friendliness and time saving aspects of 96-format extractions, there are challenges that need to be considered. Studies with a high-throughput capacity are more prone to batch effect. Batch effect occurs when non-biological factors influence the sequencing data of the experiment (Leek et al. 2010). Possible factors include laboratory personnel differences, reagent lot or batch, time of day when the experiment was conducted and laboratory conditions. One of the concerns with this phenomenon is that the outcomes of experiments are confounded with batch effect factors, leading to an incorrect conclusion. Careful study design, randomization and certain bioinformatic tools can mitigate the batch effect.

Another challenge is contamination and cross-contamination in 96-format extractions. As can be seen from figures 21 and 22, cross-contamination from fecal samples is evident in extraction and negative controls. With 96 extraction plate, cross-contamination can spread from one well to all directions, possibly even further than the adjacent wells. This makes cross-contamination hard to trace or locate to a specific well or sample. It looks unlikely that cross-contamination can be fully eliminated from 96-format extraction. Thus, a tolerable level of cross-contamination should be discussed,

and whether it is based on read counts, abundances, diversity measures or something else.

4.2 Bead-beating and diversity measures

All bacterial genera were identified out of ZymoBIOMICS Gut Standard, except for V4 sequenced groups Cprot and C did not identify the genus *Enterococcus* (figure 6, appendix 6). The groups Cprot and C followed the modified manufacturer's protocol with pre-treatment group Cprot including an incubation step with proteinase K. Both V3V4 and V4 sequenced Gut Standard compositional profiles differed for ZymoBIOMICS expected compositional profiles. However, V3V4 sequenced standards resembled ZymoBIOMICS expected compositional profiles more. Correspondence via email with Zymo Research revealed that they had used V3V4 region to sequence their Gut Standard, explaining this similarity. Bead-beating generally led to a higher degree of alpha diversity in Gut Standards (figure 7). The only deviant alpha diversity measure was in V3V4 sequenced group Cprot with Shannon diversity index. The "outlier" diversity measure may be due to pipetting errors. With beta diversities, the Gut Standard pre-treatment groups seemed to group loosely together.

Bead-beating generally led to a higher degree of microbial alpha diversity in fecal samples as well. Bead-beating brought signatures from hard-to-lyse gram-positive bacteria (figures 14 and 18). Senior and adult fecal samples showed the highest alpha diversities with beat-beating (figures 15 and 19). Infant fecal samples showed variable microbial alpha diversities across pre-treatment groups (figure 11). Perhaps infant samples do not benefit from bead-beating as much as senior and adult samples since infant gut microbiota is not as diverse. The variation in infant alpha diversities could be attributed to early shearing of genomic DNA. Too vigorous bead-beating could shear the DNA into small fragments (Yuan et al. 2012). This can lead to the formation of chimeric products during PCR amplification of gene targets. However, the chimeras were filtered out during bioinformatic analysis, making this possibility unlikely.

Infant fecal samples may need a gentler cell lysis and homogenization procedures. Interestingly, V4 sequenced infant samples had higher alpha diversity indices than

V3V4 sequenced samples, except for Shannon entropy indices that had more equal values. As V-regions have different taxonomic resolution (Fadrosh et al. 2014; Palleja et al. 2018), this discrepancy could be due to V4 sequencing favoring the genera that are naturally present in the infant sample, making the V4 alpha diversity indices seem higher. Chao 1 and observed OTUs are both diversity indices that account for community richness, however Shannon entropy index measures both richness and evenness (Goodrich et al. 2014). This could explain why Shannon entropy index did not follow the trends of other diversity indices. It is evident that infant samples would benefit from further optimization in 96-format.

With beta diversities, adult and senior samples with bead-beating samples are grouped closer together than non-bead-beating samples with both sequencing targets (figures 16 and 20). This would seem to indicate that the pre-treatment method does make a difference. However, infant fecal samples showed higher dispersion in their beta diversities (figure 12). This may be due to the difficulties in pipetting the sample or reasons discussed above.

In conclusion, bead-beating seems to bring signatures from low-abundance bacteria. For example, genera *Bifidobacterium* and *Blautia* increased in abundance in adult fecal samples (figure 14). The detection of low-abundance, but clinically significant bacterial genera is important. The genus *Bifidobacterium* has been shown to be health promoting to humans, producing beneficial metabolites, such as short-chain fatty acids (Xiao et al. 2014). *Bifidobacterium* plays a big role in early life microbe colonization and the maturation of the nascent immune system. Moreover, some studies show lower abundance of *Bifidobacterium* in obese patients (Dewulf et al. 2013; Xiao et al. 2014), making the genus a possible target for studies regarding diet intervention. On the other hand, *Blautia* has been associated with high visceral mass and obesity (Beaumont et al. 2016). Twin studies have found that microbiome signatures from *Blautia* that are associated with obesity are heritable (Goodrich et al. 2016). More specifically, *Blautia* influences *CD36*, a gene involved in many functions, including long-chain fatty acid tasting on the tongue. Both *Bifidobacterium* and *Blautia* are gram-positive genera. The higher abundance of these genera in adult groups CM and CMprot underlines the importance of bead-beating for capturing signatures from “troublesome” bacteria.

Due to the low effect size, it is challenging to infer whether the proteinase K treatment had significant impact on fecal or Gut Standard compositional profiles. Proteinase K incubation was implemented on a modified manufacturer's protocol (group Cprot) and on the other bead-beating pre-treatment (group CMprot). As proteinase K can disrupt bacterial cells and solubilize human tissues (Moore et al. 2008), the addition of this enzyme could help to detect low-abundance and hard-to-lyse bacteria. However, some bacterial species could be more or less susceptible to proteinase K treatment as the combination of component amino acids in the cell wall have inherently different resistances to cleavage.

The target variable region had a high impact on microbial compositional profiles. V3V4 and V4 sequencing on fecal samples seems to have opposite results regarding relative abundances (figures 9, 13 and 17). Different sequencing targets seemed to favor different bacterial genera. Indeed, the sequencing region seemed to have more impact on the compositional profiles than pre-treatment. Various studies show that the selection of the V-region had more profound impact on the compositional profiles than the extraction method (Mancabelli et al. 2020; Abellan-Schneyder et al. 2021; Palkova et al. 2021). However, both sequencing targets showed similar trends in alpha and beta diversity measures.

4.3 Extraction cross-contamination

With 96 extraction systems, cross-contamination can be difficult to avoid. In this study, negative controls (OMNIgene fluid) and extraction controls (PerkinElmer Lysis Buffer 1) were placed on the extraction plate in a manner that cross-contamination from fecal and Gut Standard samples could be detected. More specifically, negative controls were placed between fecal samples and the Gut Standards. Cross-contamination was observed through negative and extraction control read counts and relative abundances of the controls. V3V4 sequencing contained more reads than V4 for sequencing (table 3 and 4). This may be due to the differences in library preparation protocols. V3V4 protocol contained two PCR steps, allowing more possibility for cross-contamination. However, the length of the reads from both sequencing targets did not match the

expected read length from fecal samples (Illumina, 2013), indicating that the DNA contamination in negative controls is fragmented.

Relative abundance chart of V4 targeted negative and extraction controls (figure 22) show cross-contamination from extracted fecal samples (figures 9, 13 and 17), largely from adult and senior samples. Same trend can be seen in V3V4 sequenced negative and extraction controls. Genera present in fecal samples show in relative abundances of negative controls, implying that cross-contamination from fecal samples has occurred. The relative abundances of V3V4 sequenced negative controls reflected fecal samples sequenced with the same sequencing target, and respectively V4 sequenced controls resembled fecal samples sequenced with the same V-region.

4.4 Challenges and limitations

4.4.1 Statistical power

This study consisted of adult, infant and senior fecal samples from single individual per age group. Only three replicates of fecal samples for adult and senior groups and two replicates for infant group were extracted. Moreover, not all of the adult and senior replicates were sequenced with V4 sequencing target. DNA/RNA Shield fluid samples were not sequenced with V3V4 target, thus further limiting the sample size. These V4 sequenced Shield fluid samples were left out of the thesis since the focus of the thesis was on the OMNIgene fecal samples. Furthermore, the fecal samples of the project for which this study was optimizing the extraction protocol for contains only OMNIgene collection fluid stored samples. Due to the low sample size, statistical analyses could not be performed. The population size was limited for several reasons.

First, fecal samples for this optimization thesis were challenging to obtain.
In further studies, a larger sample size is recommended so the statistical power that is needed to observe the effect that the pre-treatment has could be obtained. Second, the effect size in microbiome studies –such as this thesis- can be difficult to calculate (Debelius et al. 2016). In microbiome studies, true effect size is seldom known, and the compositional profiles of samples are generally unknown. Powers analyses are

typically based on assumptions that rarely turn out to be true in the context of microbiome analyses. Diversity based analyses are challenging because diversity measures like alpha and beta diversities require permutative analyses. Permutative analyses, like other non-parametric tests, do not have a specific distribution for the test statistic, making power calculations difficult.

4.4.2 High variability

Some fecal samples were hard to pipette. For example, infant samples were of thicker consistency and "stickier" than the other fecal samples. Due to the variable consistency of the fecal samples, the desired amount of 200µL of sample material was challenging to obtain for all replicates. Fecal matter is proven to be of heterogenous consistency (Wu et al. 2019) and together with challenges pipetting the samples, high variation within replicates could be explained.

4.4.3 Other contamination and sequencing contamination

Contamination can come from several different sources. Studies show (Salter et al. 2014; Weiss et al. 2014) that contamination from DNA extraction kits, molecular grade water and PCR reagents can skew sequence-based results especially in low-biomass samples. Contaminating bacterial sequences frequently match water- and soil-associated genera (Salter et al. 2014). In high-biomass environments, such as feces, the effect of contaminating agents from extraction kit is arguably not as strong as in low-biomass environments. It can be concluded that the kit biome did not have an impactful contaminating effect on the samples. However, the contamination from either from the extraction kit or other sources, such as PCR reactions or sequencing method associated sources, cannot be completely disregarded.

A phenomenon that has impacted NGS platforms is known as index hopping. Index hoping occurs when index sequences assigned to one sample are incorrectly assigned to other sample(s) in the pool of samples (Schnell et al. 2015). Index hopping can cause deleterious artifacts in the experiment, manifesting in high number of reads or artificially deflated diversity measures. In this study, index hopping could be detected with PCR-controls. V3V4 sequenced PCR control had 2,894 sequences and V4 PCR

control had 50,888 sequences. Index hopping is known to occur in small probabilities in Illumina sequencing platforms (Farouni et al. 2020; van der Valk et al. 2020), therefore the possibility of such contamination occurring remains. However, the sequences of PCR controls were lower than the average read number of different negative controls, indicating that the contamination to negative and extraction controls most likely came from the fecal samples. Moreover, the library preparation workflow included index hopping mitigating steps recommended by Illumina, namely pooling just briefly before sequencing, double indexing, adapter clean-up and storage at -20ºC (Illumina, 2018). The samples with poor quality of DNA and only few sequenced reads receive the majority of misaligned reads (van der Valk et al. 2020). Luckily, fecal samples in this thesis had a high number of reads (over 200,000) and good quality DNA, thus most likely avoiding the effects of index hopping.

4.5 Future outlooks

From a clinical perspective, stool samples are a great sample type. Stool samples are non-invasive and sampling can be done in the comfort of one's own home. Furthermore, stool sampling could be a better option for small children or babies, if invasive procedures such as blood sampling can be avoided. Stool analyses can detect a myriad of things, including leucocytes, occult blood, calprotectin, fat, sugars, pancreatic enzymes and infectious microorganisms (Kasırga, 2019). With the increase of metabolic diseases, early detection of possible risk factors remains crucial. For this reason, standardized sample processing pipelines are the backbone of reliable clinical results.

Future method-centered studies are important in order to optimize stool processing pipeline and produce reliable and reproducible protocols. Future studies on DNA extraction should focus on the limitations of this study, mainly the effect size and cross--contamination, and focus on providing a reliable high-throughput protocol. As the need for rapid microbial profiling is rising, DNA extraction methods need to meet the demands of downstream applications in terms of DNA quality, quantity and integrity. Moreover, many clinicians might want relatively fast DNA extraction and downstream processing. As such, DNA isolation methods will increasingly rely on high-capacity

protocols, like 96-format extractions. Decreasing human-error and hands-on-time in semiautomated systems will be important. Automated pipetting robots could ameliorate the intra-sample variance and reduce human labor.

Next Generation Sequencing methods will play a big role in microbiome diagnostics. In clinical setting it is often important to obtain species or even strain-level taxonomic resolution, as the microbes vary in functional and metabolic properties with the genera. Thus, microbiome studies will arguably favor whole genome sequencing or metagenome sequencing. In case whole genome sequencing technologies are not available, the use of non-contiguous V-regions could achieve optimal taxonomic resolution (Pinna et al. 2019).

5 Conclusions

This study aimed to optimize a reliable and reproducible 96-format DNA extraction procedure suitable for Next Generation Sequencing and other downstream applications. The optimization included four pre-treatment groups that included bead-beating and chemical lysis, and was performed with ZymoBIOMICS Gut Standard and infant, adult and senior fecal samples as well as different negative controls. High-throughput extraction in 96-format proved to be feasible, with added bead-beating step increasing the bacterial diversity of fecal samples. This study argued that 96-format semiautomated extraction reduced hands-on-time and human error. However, further studies are needed to create and optimal standardized pipeline for stool samples.

Acknowledgements

I want to thank my supervisors Teemu Kallonen, Sanja Vanhatalo and Eero Vesterinen for endless patience and plentiful advice during this process. A special thank you goes to Anna Musku and Minna Lamppu who introduced me to the inner workings of the lab and taught me the basics of the microbial work. I'd also like to thank Turku University Hospital for collaboration and Microbiome and Resistance group for the warm working environment. Lastly, a huge thank you goes to my friends and family for the support and for reminding me that there is a balance between university and personal life.

References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science Translational Medicine*, *6*(237), 237ra65.

Abellan-Schneyder, I., Matchado, M. S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M., & Neuhaus, K. (2021). Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *MSphere*, *6*(1).

Agus, A., Denizot, J., Thévenot, J., Martinez-Medina, M., Massier, S., Sauvanet, P., Bernalier-Donadille, A., Denis, S., Hofman, P., Bonnet, R., Billard, E., & Barnich, N. (2016). Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. *Scientific Reports 2016 6:1*, *6*(1), 1–14.

Babraham Bioinformatics. Fast QC. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ >[Retrieved 13.05.2022]

Balvočiute, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics*, *18*(2), 1–8.

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Applied and Environmental Microbiology*, *77*(11), 3846.

Beaumont, M., Goodrich, J. K., Jackson, M. A., Yet, I., Davenport, E. R., Vieira-Silva, S., Debelius, J., Pallister, T., Mangino, M., Raes, J., Knight, R., Clark, A. G., Ley, R. E., Spector, T. D., & Bell, J. T. (2016). Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biology*, *17*(1).

Bezirtzoglou, E., Tsiotsias, A., & Welling, G. W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*, *17*(6), 478–482.

Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., Lieber, A. D., Wu, F., Perez-Perez, G. I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M. G., & Blaser, M. J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science Translational Medicine*, *8*(343), 343ra82.

Bonnet, M., Lagier, J. C., Raoult, D., & Khelaifia, S. (2020). Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. *New Microbes and New Infections*, *34*, 100622.

Browne, A. J., Chipeta, M. G., Haines-Woodhouse, G., Kumaran, E. P. A., Hamadani, B. H. K., Zaraa, S., Henry, N. J., Deshpande, A., Reiner, R. C., Day, N. P. J., Lopez, A. D., Dunachie, S., Moore, C. E., Stergachis, A., Hay, S. I., & Dolecek, C. (2021). Global antibiotic consumption and usage in humans, 2000–18: a spatial modelling study. *The Lancet Planetary Health*, *5*(12), e893–e904.

Bukin, Y. S., Galachyants, Y. P., Morozov, I. v., Bukin, S. v., Zakharenko, A. S., & Zemskaya, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data 2019 6:1*, *6*(1), 1–14.

Canova, C., & Cantarutti, A. (2020). Population-Based Birth Cohort Studies in Epidemiology. *International Journal of Environmental Research and Public Health*, *17*(15), 1–6.

Chiarello, M., McCauley, M., Villéger, S., & Jackson, C. R. (2022). Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLOS ONE*, *17*(2), e0264443.

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F. E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., … Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology 2017 35:11*, *35*(11), 1069–1076.

Cotten, C. M. (2016). Adverse Consequences of Neonatal Antibiotic Exposure. *Current Opinion in Pediatrics*, *28*(2), 141.

Couturier-Maillard, A., Secher, T., Rehman, A., Normand, S., de Arcangelis, A., Haesler, R., Huot, L., Grandjean, T., Bressenot, A., Delanoye-Crespin, A., Gaillot, O., Schreiber, S., Lemoine, Y., Ryffel, B., Hot, D., Nùñez, G., Chen, G., Rosenstiel, P., & Chamaillard, M. (2013). NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *The Journal of Clinical Investigation*, *123*(2), 700.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. v., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559.

de Filippo, C., Cavalieri, D., di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14691–14696.

de Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J., Parkhill, J., Charnock-Jones, D. S., & Smith, G. C. S. (2019). Human placenta has no microbiome but can harbour potential pathogens. *Nature*, *572*(7769), 329.

Debelius, J., Song, S. J., Vazquez-Baeza, Y., Xu, Z. Z., Gonzalez, A., & Knight, R. (2016). Tiny microbes, enormous impacts: What matters in gut microbiome studies? In *Genome Biology* (Vol. 17, Issue 1). BioMed Central Ltd.

Dewulf, E. M., Cani, P. D., Claus, S. P., Fuentes, S., Puylaert, P. G. B., Neyrinck, A. M., Bindels, L. B., de Vos, W. M., Gibson, G. R., Thissen, J. P., & Delzenne, N. M. (2013). Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. *Gut*, *62*(8), 1112.

di Segni, A., Braun, T., Benshoshan, M., Barhom, S. F., Saar, E. G., Cesarkas, K., Squires, J. E., Keller, N., & Haberman, Y. (2018). Guided Protocol for Fecal Microbial Characterization by 16S rRNA-Amplicon Sequencing. *Journal of Visualized Experiments : JoVE*, *2018*(133), 56845.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science (New York, N.Y.)*, *308*(5728), 1635.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*(1).

Farouni, R., Djambazian, H., Ferri, L. E., Ragoussis, J., & Najafabadi, H. S. (2020). Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nature Communications 2020 11:1*, *11*(1), 1–8.

Fei, N., & Zhao, L. (2012). An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *The ISME Journal 2013 7:4*, *7*(4), 880–884.

García-Mantrana, I., Alcántara, C., Selma-Royo, M., Boix-Amorós, A., Dzidic, M., Gimeno-Alcañiz, J., Úbeda-Sansano, I., Sorribes-Monrabal, I., Escuriet, R., Gil-Raga, F., Parra-Llorca, A., Martínez-Costa, C., Collado, M. C., Baüerl, C., Villoldo, E., Zafra, C., Olivares, L., Pérez-Martínez, G., Mira, A., … Atero, A. (2019). MAMI: A birth cohort focused on maternal-infant microbiota during early life. *BMC Pediatrics*, *19*(1), 1–8.

Gauthier, J., & Derome, N. (2021). Evenness-Richness Scatter Plots: a Visual and Insightful Representation of Shannon Entropy Measurements for Ecological Community Analysis. *MSphere*, *6*(2).

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., & Ley, R. E. (2016). Genetic determinants of the gut microbiome in UK Twins. *Cell Host & Microbe*, *19*(5), 731.

Goodrich, J. K., di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., Knight, R., & Ley, R. E. (2014). Conducting a Microbiome Study. *Cell*, *158*(2), 250.

Guinane, C. M., & Cotter, P. D. (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therapeutic Advances in Gastroenterology*, *6*(4), 295.

Hall, A. B., Tolonen, A. C., & Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics 2017 18:11*, *18*(11), 690–699.

Human Microbiome Project (2015). IHMS_SOP 006 V2, Standard operating procedure for fecal samples DNA Extraction, protocol Q INRA <http://www.human-microbiome.org/index.php?id=Sop&num=006> [Retrieved 08.05.2022]

Hsieh, Y. H., Peterson, C. M., Raggio, A., Keenan, M. J., Martin, R. J., Ravussin, E., & Marco, M. L. (2016). Impact of Different Fecal Processing Methods on Assessments of Bacterial Diversity in the Human Intestine. *Frontiers in Microbiology*, *7*(10), 1643.

Hugenholtz, F., & de Vos, W. M. (2018). Mouse models for human intestinal microbiota research: a critical evaluation. *Cellular and Molecular Life Sciences*, *75*(1), 149.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J. P., Belaiche, J., Almer, S., Tysk, C., O'morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M., & Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature 2001 411:6837*, *411*(6837), 599–603.

Islam, M. S., Aryasomayajula, A., & Selvaganapathy, P. R. (2017). A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*, *8*(3).

Illumina (2013). 16S Metagenomic Library Sequencing Preparation. <https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf> [Retrieved 26.03.2022]

Illumina (2018). Effects of Index Misassignment on Multiplexing and Downstream Analysis. <https://emea.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862> [Retrieved 13.06.2022]

Kasırga, E. (2019). The importance of stool tests in diagnosis and follow-up of gastrointestinal disorders in children. *Turkish Archives of Pediatrics/Türk Pediatri Arşivi*, *54*(3), 141.

Kau, A. L., Planer, J. D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., Manary, M. J., Liu, T. C., Stappenbeck, T. S., Maleta, K. M., Ashorn, P., Dewey, K. G., Houpt, E. R., Hsieh, C. S., & Gordon, J. I. (2015). Functional characterization of IgA-targeted bacterial taxa from malnourished Malawian children that produce diet-dependent enteropathy. *Science Translational Medicine*, *7*(276), 276ra24.

Kennedy, E. A., King, K. Y., & Baldridge, M. T. (2018). Mouse microbiota models: Comparing germ-free mice and antibiotics treatment as tools for modifying gut bacteria. *Frontiers in Physiology*, *9*(10), 1534.

Kim, B.-R., Shin, J., Guevarra, R. B., Lee, J. H., Kim, D. W., Seol, K.-H., Lee, J.-H., Kim, H. B., & Isaacson, R. E. (2017). Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J. Microbiol. Biotechnol*, *27*(12), 2089–2093.

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, *102*(7), 2567–2572.

Korpela, K., Dikareva, E., Hanski, E., Kolho, K. L., de Vos, W. M., & Salonen, A. (2019). Cohort profile: Finnish Health and Early Life Microbiota (HELMi) longitudinal birth cohort. *BMJ Open*, *9*(6), e028500.

Kuh, D., Wong, A., Shah, I., Moore, A., Popham, M., Curran, P., Davis, D., Sharma, N., Richards, M., Stafford, M., Hardy, R., & Cooper, R. (2016). The MRC National Survey of Health and Development reaches age 70: maintaining participation at older ages in a birth cohort study. *European Journal of Epidemiology*, *31*(11), 1135.

Lantz, P. G., Matsson, M., Wadström, T., & Rådström, P. (1997). Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR. *Journal of Microbiological Methods*, *28*(3), 159–167.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*, *11*(10), 733–739.

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, *124*(4), 837–848.

Mancabelli, L., Milani, C., Lugli, G. A., Fontana, F., Turroni, F., van Sinderen, D., & Ventura, M. (2020). The Impact of Primer Design on Amplicon-Based Metagenomic Profiling Accuracy: Detailed Insights into Bifidobacterial Community Structure. *Microorganisms*, *8*(1).

Moore, E., Arnscheidt, A., KrÜger, A., StrÖmpl, C., & Mau, M. (2008). Simplified protocols for the preparation of genomic DNA from bacterial cultures. *Molecular Microbial Ecology Manual*, 1905–1919.

Nadimpalli, M. L., Bourke, C. D., Robertson, R. C., Delarocque-Astagneau, E., Manges, A. R., & Pickering, A. J. (2020). Can breastfeeding protect against antimicrobial resistance? *BMC Medicine 2020 18:1*, *18*(1), 1–11.

Nagpal, R., Shively, C. A., Register, T. C., Craft, S., & Yadav, H. (2019). Gut microbiome-Mediterranean diet interactions in improving host health. *F1000Research*, *8*.

Nguyen, N. P., Warnow, T., Pop, M., & White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms and Microbiomes 2016 2:1*, *2*(1), 1–8.

Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Piérard, D., Marchal, K., Roosens, N. H. C., Vanneste, K., & de Keersmaecker, S. C. J. (2020). Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing Escherichia coli isolates. *Scientific Reports 2020 10:1*, *10*(1), 1–16.

PacBio (2022). Microbial whole genome sequencing. <https://www.pacb.com/products-and-services/applications/whole-genome-sequencing/microbial/> [Retrieved 04.04.2022]

Palkova, L., Tomova, A., Repiska, G., Babinska, K., Bokor, B., Mikula, I., Minarik, G., Ostatnikova, D., & Soltys, K. (2021). Evaluation of 16S rRNA primer sets for characterisation of microbiota in paediatric patients with autism spectrum disorder. *Scientific Reports 2021 11:1*, *11*(1), 1–13.

Palleja, A., Mikkelsen, K. H., Forslund, S. K., Kashani, A., Allin, K. H., Nielsen, T., Hansen, T. H., Liang, S., Feng, Q., Zhang, C., Pyl, P. T., Coelho, L. P., Yang, H., Wang, J., Typas, A., Nielsen, M. F., Nielsen, H. B., Bork, P., Wang, J., … Pedersen, O. (2018). Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology 2018 3:11*, *3*(11), 1255–1265.

Pinna, N. K., Dutta, A., Haque, M. M., & Mande, S. S. (2019). Can targeting non-contiguous V-regions with paired-end sequencing improve 16S rRNA-based taxonomic resolution of microbiomes?: An in silico evaluation. *Frontiers in Genetics*, *10*(JUL).

Rintala, A., Pietilä, S., Munukka, E., Eerola, E., Pursiheimo, J. P., Laiho, A., Pekkala, S., & Huovinen, P. (2017). Gut microbiota analysis results are highly dependent on the 16s rRNA gene target region, whereas the impact of DNA extraction is minor. *Journal of Biomolecular Techniques*, *28*(1).

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M. C., Marchesi, J. R., & Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health and Disease*, *26*(0).

Rossi-Tamisier, M., Benamar, S., Raoult, D., & Fournier, P. E. (2015). Cautionary tale of using 16s rRNA gene sequence similarity values in identification of human-associated bacterial species. *International Journal of Systematic and Evolutionary Microbiology*, *65*(6), 1929–1934.

Salazar, N., Valdés-Varela, L., González, S., Gueimonde, M., & de los Reyes-Gavilán, C. G. (2017). Nutrition and the gut microbiome in the elderly. *Gut Microbes*, *8*(2), 82.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 1–12.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014–1026.

Seol, D., Lim, J. S., Sung, S., Lee, Y. H., Jeong, M., Cho, S., Kwak, W., & Kim, H. (2022). Microbial Identification Using rRNA Operon Region: Database and Tool for Metataxonomics with Long-Read Sequence. *Microbiology Spectrum*.

Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, *31*(1), 69.

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, *2*(5).

SILVA (2019). Release information: SILVA 138 SSU. <https://www.arb-silva.de/documentation/release-138/> [Retrieved 14.04.2022]

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59.

Sonnenburg, E. D., & Sonnenburg, J. L. (2014). Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metabolism*, *20*(5), 779.

Takatsy, G. (1955). The use of spiral loops in serological and virological micro-methods. *Acta Microbiol. Acad. Sci. Hung*, *3*(191).

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

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J. L., & Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, *6*(AUG), 771.

Turroni, F., Milani, C., Duranti, S., Lugli, G. A., Bernasconi, S., Margolles, A., di Pierro, F., van Sinderen, D., & Ventura, M. (2020). The infant gut microbiome as a microbial organ influencing host well-being. *Italian Journal of Pediatrics 2020 46:1*, *46*(1), 1–13.

Turta, O., Selma-Royo, M., Kumar, H., Collado, M. C., Isolauri, E., Salminen, S., & Rautava, S. (2022). Maternal Intrapartum Antibiotic Treatment and Gut Microbiota Development in Healthy Term Infants. *Neonatology*, *119*(1), 93–102.

Uzan-Yulzari, A., Turta, O., Belogolovski, A., Ziv, O., Kunz, C., Perschbacher, S., Neuman, H., Pasolli, E., Oz, A., Ben-Amram, H., Kumar, H., Ollila, H., Kaljonen, A., Isolauri, E., Salminen, S., Lagström, H., Segata, N., Sharon, I., Louzoun, Y., … Koren, O. (2021). Neonatal antibiotic exposure impairs child growth during the first six years of life by perturbing intestinal microbial colonization. *Nature Communications 2021 12:1*, *12*(1), 1–12.

van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L., & Guschanski, K. (2020). Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Molecular Ecology Resources*, *20*(5), 1171–1181.

Videnska, P., Smerkova, K., Zwinsova, B., Popovici, V., Micenkova, L., Sedlar, K., & Budinska, E. (2019). Stool sampling and DNA isolation kits affect DNA quality and bacterial composition following 16S rRNA gene sequencing using MiSeq Illumina platform. *Scientific Reports*, *9*(1).

Wang, J., Thingholm, L. B., Skiecevičie, J., Rausch, P., Kummen, M., Hov, J. R., Degenhardt, F., Heinsen, F. A., Rühlemann, M. C., Szymczak, S., Holm, K., Esko, T., Sun, J., Pricop-Jeckstadt, M., Al-Dury, S., Bohov, P., Bethune, J., Sommer, F., Ellinghaus, D., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics 2016 48:11*, *48*(11), 1396–1406.

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, *15*(12), 1–3.

Wikipedia Commons (2011). 16S. < https://commons.wikimedia.org/wiki/File:16S.svg> [Retrieved 01.01.2022]

Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, *10*(10), 2407.

Wu, W. K., Chen, C. C., Panyod, S., Chen, R. A., Wu, M. S., Sheen, L. Y., & Chang, S. C. (2019). Optimization of fecal sample processing for microbiome study — The journey from bathroom to bench. *Journal of the Formosan Medical Association*, *118*(2), 545–555.

Xiao, S., Fei, N., Pang, X., Shen, J., Wang, L., Zhang, B., Zhang, M., Zhang, X., Zhang, C., Li, M., Sun, L., Xue, Z., Wang, J., Feng, J., Yan, F., Zhao, N., Liu, J., Long, W., & Zhao, L. (2014). A gut microbiota-targeted dietary intervention for amelioration of chronic inflammation underlying metabolic syndrome. *Fems Microbiology Ecology*, *87*(2), 357.

Yang, I., Corwin, E. J., Brennan, P. A., Jordan, S., Murphy, J. R., & Dunlop, A. (2016). The Infant Microbiome: Implications for Infant Health and Neurocognitive Development. *Nursing Research*, *65*(1), 76.

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A. M., Härkönen, T., Ryhänen, S. J., Franzosa, E. A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E. S., Knip, M., & Xavier, R. J. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatments on strain-level diversity and stability. *Science Translational Medicine*, *8*(343), 343ra81.

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *PLOS ONE*, *7*(3), e33865.

Zinöcker, M. K., & Lindseth, I. A. (2018). The Western Diet–Microbiome-Host Interaction and Its Role in Metabolic Disease. *Nutrients*, *10*(3).

Zymobiomics (2022). Zymobiomics Gut Microbiome Standard. <https://www.zymoresearch.com/products/zymobiomics-gut-microbiome-standard> [Retrieved 04.04.2022]

Appendices

Appendix 1. Purification protocol for human feces material using the Chemagic Magnetic Separation Module I

For research use only. Not for use in diagnostic procedures.

Purification Protocol for Human Feces Material Using the chemagic Magnetic Separation Module I

Protocol name: **chemagic DNA Stool MSMI prefilling VD210526.che**

Positioning Tips and Plates on the Tracking System

Can be done manually or by an integrated robotic system

Position 1:	Rack with Disposable Tips
Position 2:	low-well-plate (MICROTITER SYSTEM) prefilled with 75 µl **Magnetic Beads**
Position 3:	deep-well-plate (riplate SW) containing 800 µl Lysate **Binding Buffer 2** (added automatically) ! ***See "Processing Steps in Detail".***
Position 4:	deep-well-plate (riplate SW) [**Wash Buffer 3** added automatically]
Position 5:	deep-well-plate (riplate SW) [**Wash Buffer 4** added automatically]
Position 6:	deep-well-plate (riplate SW) [**Wash Buffer 5** added automatically]
Position 7:	deep-well-plate (riplate SW) [**Wash Buffer 6** added automatically]
Position 8:	deep-well-plate (riplate SW) prefilled with 250 µl **Elution Buffer 7**

Version 210526

Any further questions?

chemagen Technology technical support: +49 (0) 2401 805-501 | support.chemagen@perkinelmer.com

PerkinElmer

Processing Steps in Detail

Before You Start

- Dissolve lyophilized **Proteinase K** in H_2O (volume is given on the label).

! ***See below "Storage Conditions".***

Preparing the Lysate

1. Place up to 200 mg stool material in an appropriate tube.
2. Add 1.6 ml **Lysis Buffer 1** and mix vigorously by vortexing.
   These suspensions can be stored for 4 weeks frozen.
3. Add 30 µl **Protease**, mix and incubate for **10 minutes** at 70 °C using a thermo mixer.
4. Incubate another **5 minutes** at 95 °C.
5. Centrifuge for **5 minutes** at high speed (>13000 rpm).
6. Transfer 800 µl of the supernatant into a well of the Sample Plate.
7. Continue with "**Protocols Steps**" of the Isolation Protocol.

Version 210526

Any further questions?
chemagen Technology technical support: +49 (0) 2401 805-501 | support.chemagen@perkinelmer.com

PerkinElmer

For research use only. Not for use in diagnostic procedures.

Protocol Steps

1. Select the protocol "**chemagic DNA Stool MSMI prefilling VD210526.che**" and press the [**Insert IDs**] button. Follow the instructions as given in the chemagic QA software. If the enhanced functions are deactivated continue without pressing the [**Insert IDs**] button.
2. Prefill the **Elution Buffer 7** and the thoroughly resuspended **Magnetic Beads** according to the sample positions.

! ***For indication of volumes and sample positions see "Positioning Tips and Plates on the Tracking System".***

3. Place the plates on the tracking system according the instructions given by the chemagic QA software.
4. Place the Sample Plate in **position 3** on the tracking system.
5. Check all plates for accurate orientation and fitting.
6. Close the door and start the process by pressing the [**Start**] button.

Any further questions?
chemagen Technology technical support: +49 (0) 2401 805-501 | support.chemagen@perkinelmer.com

PerkinElmer

Additional Information

Safety Information

Always wear a laboratory coat, disposable gloves and protective goggles. For detailed information please consult the appropriate material safety data sheet (MSDS).

Storage Conditions

All kit components can be stored at room temperature, except the reconstituted **Proteinase**.

The reconstituted **Proteinase K** is stable at 4 °C for 4 weeks. For long term storage we recommend to store the reconstituted **Proteinase K** in aliquots at -20 °C. Do not freeze the **Proteinase K** aliquots after thawing.

Lysis Buffer 1 may form a precipitate upon storage. If necessary, warm to 55 °C to redissolve.

Binding Buffer 2, **Wash Buffer 3**, **Wash Buffer 4** and **Wash Buffer 5** contain ethanol. Longer storage of the buffers without lids should be avoided. If ethanol evaporates the optimal yield cannot be guaranteed.

Prefilled plates can be stored at room temperature if the plates are sealed with a foil to avoid ethanol evaporation and buffer contamination.

General Remarks

Deep-well-plates (riplate SW) and the low-well-plates (MICROTITER SYSTEM) are delivered from chemagen optionally. You can use your own plates also, but the instrument protocol has to be adapted to the specific plates.

For prefilling plates we recommend to use a dispensing system (e.g. Multidrop from Thermolabsystems, µFill from Biotek or PlateMate from Matrix) or a multichannel pipette.

The **Elution Buffer 7** included in this kit is 10 mM Tris-HCl pH 8.0. TE buffer pH 8.0 can also be used without any protocol adjustments. Water pH 8.0 may also be used, but the yield could be slightly decreased.

The **Magnetic Bead** suspension should be mixed vigorously before dispensing, otherwise the suspension is not homogenous and the DNA yield could be low.

Any further questions?
chemagen Technology technical support: +49 (0) 2401 805-501 | support.chemagen@perkinelmer.com

PerkinElmer

UV Measurements/Real Time PCR

In some cases you may find some traces of **Magnetic Beads** left in the eluate. Such particles will not interfere with standard PCR and most downstream applications but may increase the background in UV measurements or could influence real time PCR.
In such a case we recommend to perform an additional separation step using an appropriate **chemagic** magnetic stand in order to separate traces of particles.

Version 210608

Any further questions?
chemagen Technology technical support: +49 (0) 2401 805-501 | support.chemagen@perkinelmer.com

PerkinElmer

Appendix 2. Purification of PCR products (V4) with DynaMag™-96 magnetic stand.

This purification protocol is performed in strip tubes.

Vortex and spin the PCR products. Take 5µL of the sample for gel electrophoresis. Keep the rest of the PCR product in refrigerator while preparing the gel.

Purification:

1. Fill the PCR product with PCR grade water to the volume of 50µL
2. Vortex the AMPure XP magnetic beads. Using a multichannel pipette, add 20µL of beads to each well. Gently pipette mix up and down 10 times. Seal the strip and shake with vertical shaker for 5 min. Briefly spin the tubes.
3. Place the strip tubes onto the magnetic stand. Wait for 5 min.
4. Transfer the supernatant (ca. 70µL) into a new strip tube.
5. Vortex the AMPure XP magnetic beads. Using a multichannel pipette, add 50µL to each well. Gently mix up and down 10 times. Seal the strip and shake with vertical shaker for 5 min. Briefly spin the tubes.
6. Place the strip tubes onto the magnetic stand. Wait for 5 min.

Wash:

7. Discard 100µL of supernatant.
8. Add 200µL of 80% ethanol. Do not mix.
9. Move the strip from one position to another on the magnetic stand, so that the magnetic beads travel from one side of the tube onto another. Repeat three times.
10. Discard 200µL of supernatant.
11. Add 200µL of 80% ethanol. Do not mix.
12. Move the strip from one position to another on the magnetic stand, so that the magnetic beads travel from one side of the tube onto another. Repeat three times.
13. Discard all of the supernatant.

Drying:

14. Let the samples dry unsealed on the magnetic stand for 5–10 min, or until dry.

Elution:

15. Transfer the strip to a regular rack and add 20µL of PCR grade water. Gently pipette up and down 20 times.

16. Incubate the samples on the regular rack for 2 min.

17. Transfer the strip onto the magnetic stand and wait for 2 min.

18. Pipette 18µL of the supertatant into a new strip.

Appendix 3. V4 sequencing index sequences

Appendix table 3. V4 primer index sequences

i5 forward	Sequence	i7 reverse	Sequence
SA501	ATCGTACG	SA701	CGAGAGTT
SA502	ACTATCTG	SA702	GACATAGT
SA503	TAGCGAGT	SA703	ACGCTACT
SA504	CTGCGTGT	SA704	ACTCACTG
SA505	TCATCGAG	SA705	TGAGTACG
SA506	CGTGAGTG	SA706	CTGCGTAG
SA507	GGATATCT	SA707	TAGTCTCC
SA508	GACACCGT	SA708	CGAGCGAC
		SA709	ACTACGAC
		SA710	GTCTGCTA
		SA711	GTCTATGA
		SA712	TATAGCGA

Appendix 4. Testing with Quant-iT 1X dsDNA High Sensitivity Assay kit (ThermoFisher, USA)

DNA concentration measurement was tested in 96-format with Quant-iT HS assay kit (ThermoFisher, USA). PerkinElmer microplate OptiPlate and VICTOR Nivo microplate reader (PerkinElmer, Finland) with user interface program 4.0.7 was used in the test.

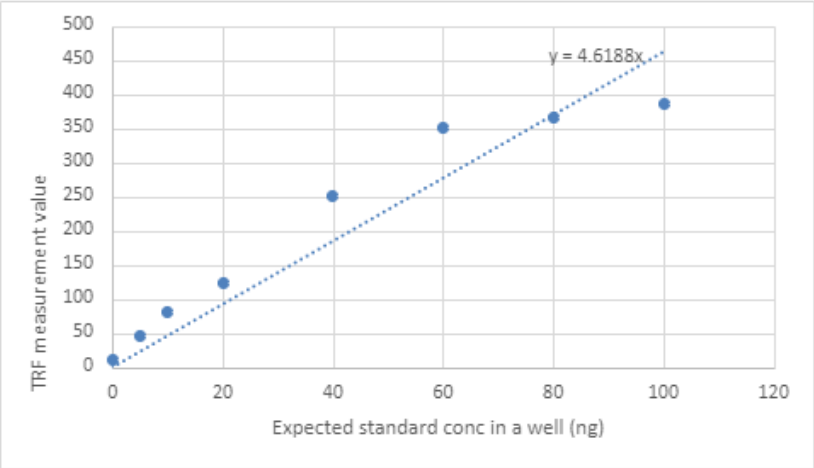
Quant-iT kit includes 10 standards. With the help of those standards, a standard curve can be drawn and the concentration of the samples can be calculated. According to the Quant-iT manual, 10µL of the standards and 2µL of the (unknown) DNA sample should be pipetted into the wells. We tested with Quant-iT and Qubit HS kit (ThermoFisher, USA) standards, whether pipetting 10µL or 2µL of the standards would affect the accuracy of the measurement.

The measurement was performed with following parameters (table 4A):

Appendix table 4A. TRF-EndPoint measurement parameters

PLATE TYPE	PerkinElmer OptiPlate
PLATE FORMAT	96 wells (8X12)
PLATE PARAMETERS	
A1 row-coordinate(µm)	11240
A1 column-coordinate(µm)	14380
Well spacing (µm)	9000
Well diameter(µm)	7130
Well volume(µl)	350
1 OPERATION	TRF-EndPoint
MEASUREMENT UNIT	COUNTS
MEASUREMENT TYPE	Single label
EXCITATION FILTER	480/30nm
EMISSION FILTER	530/30nm
DICHROIC MIRROR	D400
DELAY TIME (µs)	400
EMISSION TIME (µs)	400
MEASUREMENT DIRECTION	Top measurement
MEASUREMENT TIME (ms)	500
Z-FOCUS (mm)	8.5
EXCITATION SPOT SIZE (mm)	4
EMISSION SPOT SIZE (mm)	4
FLASH ENERGY (mJ)	100
MEASUREMENT ORDER	Bi-directional by rows

A standard curve was drawn and the equation to calculate the concentrations was obtained (appendix figure 4B).



Appendix figure 4B. Standard curve of the Quant-iT measurement. Equation y=4,6188x

The concentrations were calculated using the sample TRF-EndPoint values and the equation.

Appendix table 4C. The concentrations of Quant-iT and Qubit standards measured with Quant-iT kit. r_1 and r_2 indicate the replicate measurements 1 and 2.

Sample	Volume pipetted (µL)	Expected conc (ng/µL)	Measured cocn (ng/µL)
Quant-iT standard_r1	10	0	0.173205162
Quant-iT standard_r1	10	0.5	0.736121936
Quant-iT standard_r1	10	1	1.797003551
Quant-iT standard_r1	10	2	2.944487746
Quant-iT standard_r1	10	4	5.326058717
Quant-iT standard_r1	10	6	8.01073872
Quant-iT standard_r1	10	8	7.274616784
Quant-iT standard_r1	10	10	6.776651944
Quant-iT standard_r2	10	0	0.238157097
Quant-iT standard_r2	10	0.5	1.212436131
Quant-iT standard_r2	10	1	1.645449034
Quant-iT standard_r2	10	2	2.403221616
Quant-iT standard_r2	10	4	5.477613233
Quant-iT standard_r2	10	6	7.166363558
Quant-iT standard_r2	10	8	8.55200485
Quant-iT standard_r2	10	10	9.872694206
Quant-iT standard_r1	2	0	0.866025808
Quant-iT standard_r1	2	0.5	2.489824197
Quant-iT standard_r1	2	1	4.00536936
Quant-iT standard_r1	2	2	5.629167749
Quant-iT standard_r1	2	4	8.660258076
Quant-iT standard_r1	2	6	11.90785485
Quant-iT standard_r1	2	8	12.66562744
Quant-iT standard_r1	2	10	20.56811293
Quant-iT standard_r2	2	0	1.299038711
Quant-iT standard_r2	2	0.5	2.165064519
Quant-iT standard_r2	2	1	5.196154845
Quant-iT standard_r2	2	2	5.953927427
Quant-iT standard_r2	2	4	8.660258076
Quant-iT standard_r2	2	6	14.50593228
Quant-iT standard_r2	2	8	15.48021131
Quant-iT standard_r2	2	10	16.12973067
Qubit standard 1_r1	10	0	0.151554516
Qubit standard 1_r2	10	0	0.129903871
Qubit standard 1_r1	2	0	1.407291937
Qubit standard 1_r2	2	0	0.649519356
Qubit standard 2_r1	10	10	10.21910453
Qubit standard 2_r2	10	10	10.17580324
Qubit standard 2_r1	2	10	18.83606131
Qubit standard 2_r2	2	10	22.40841777

The Quant-iT standards were measured with Qubit dsDNA HS assay kit to confirm the concentrations of the Quant-iT standards. 10µL of the Quant-iT standard was used in the measurement (appendix table 4D).

Appendix table 4D. Quant-iT standards measured with Qubit dsDNA HS assay kit. The standards were aliquoted to 8 well strip tubes in order to facilitate pipetting with multichannel pipette. Thus, concentrations from the aliquot and the stock were both measured.

Expected standard conc (ng/µL)	Conc from aliquot (ng/µL)	Conc from stock (ng/µL)
0	0	0
0.5	0.503	0.657
1	1.16	1.3
2	1.88	2.3
4	4.05	5.31
6	6.13	7.65
8	8.82	8.92
10	10.8	12.7

Last, concentration was measured) from faeces, OmniGENE fluid, Chemagic Lysis Buffer 1 (EC) and ZymoBIOMICS Gut Standard (appendix table 4E). These samples were obtained from test extractions prior to this thesis (the tests were performed by this thesis' author).

Appendix table 4E. Feacal, OmniGENE and Chemagic Lysis Buffer 1 concentrations measured with Quant-iT and Qubit. 2µL of the sample was pipetted.

Sample	Quant-iT (ng/µL)	Qubit (ng/µL)	Quant-iT/Qubit	Average Quant/ Qubit	Quant-iT*0.5
EC	1.651177053	0.107	15.43156125	1.730971916	0.825588527
Fecal sample 1	52.24795962	27.2	1.920880868		26.12397981
Fecal sample 2	29.24942209	15.9	1.839586295		14.62471104
Fecal sample 3	27.12648016	16.9	1.605117169		13.56324008
Fecal sample 4	41.98707364	26.7	1.572549575		20.99353682
Fecal sample 5	39.98207293	20.8	1.922215045		19.99103647
Fecal sample 6	20.99353682	8.03	2.614388147		10.49676841
Fecal sample 7	50.71472378	26.9	1.885305717		25.35736189
Fecal sample 8	14.55089696	21.2	0.686363064		7.275448479
Fecal sample 9	26.08767955	21.6	1.207762942		13.04383977
Fecal sample 10	39.7031617	27.8	1.428171284		19.85158085
Fecal sample 11	17.14927142	9.92	1.728757199		8.574635708
Fecal sample 12	38.87168188	24.8	1.567406527		19.43584094
OmniGENE	1.662959652	0	1.662959652		0.831479826
EC	1.974764587	0.065	30.38099365		0.987382294
Gut Standard	10.80923774	4.17	2.592143343		5.40461887

Based on the results, Quant-iT measured concentrations roughly 1,5-2 times higher than Qubit. It appeared that pipetting 10µL of the sample would yield results closer to Qubit measurements (table 4C). Due to the discrepancy of the yields from Quan-iT and Qubit, the DNA concentration measurement with Quant-iT was thought to me unreliable. Thus, in this thesis, the DNA yields were measured with Qubit dsDNA HS assay kit (ThermoFisher).

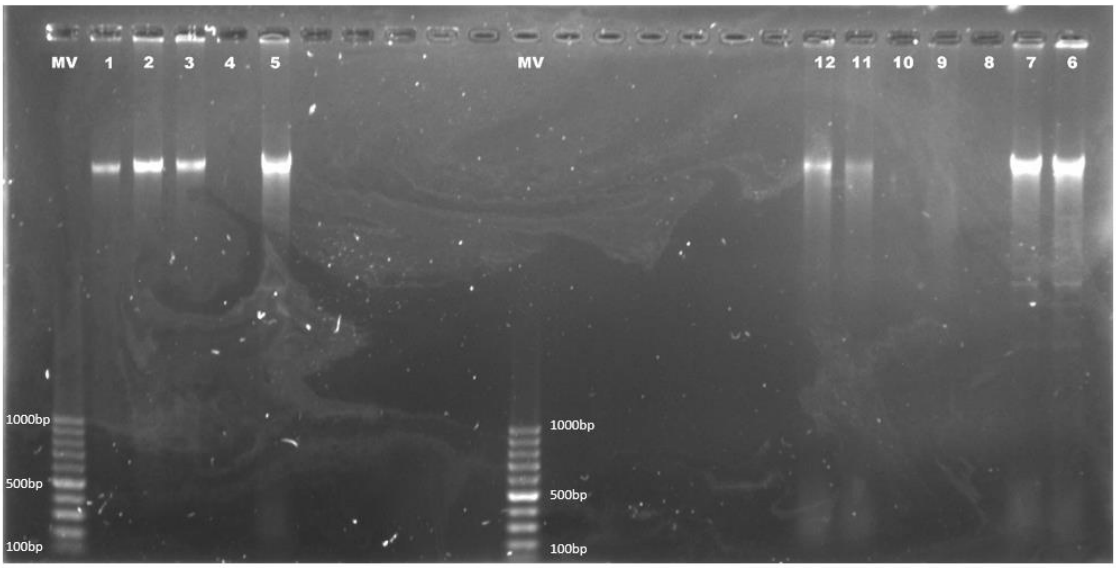
Appendix 5. Testing the Chemagic MSMI extraction kit with DNA Stool 200 H96 kit (PerkinElmer, Finland) prior to the "main" thesis extraction.

Test 1. At first, DNA extraction was tested with manufacturer's original protocol with and without proteinase K (pre-treatment groups Cprot and C, see table 1). Refer to appendix 1 for manufacturer's protocol.

Appendix table 5A. Concentrations of extraction test 1 with Chemagic MSMI system.

DNA ID	Sample	Volume pipetted (µL)	ProtK incubation (+/-)	Conc (ng/µL)	Absolute conc (ng)	Nanodrop A260/280
1	Gut Standard	75	+	4.56	912	2.1
2	Gut Standard	100	+	8.98	1796	2.02
3	Gut Standard	100	-	6.71	1342	1.93
4	Lysis buffer 1	250	-	0		
5	Adult faeces	250	+	18	3600	1.8
6	Adult faeces	250	-	19.8	3960	1.62
7	Adult faeces	250	+	20.2	4040	1.85
8	OMNIgene fluid	250	+	0		
9	RNA/DNA Shield fluid	250	+	low		
10	Lysis buffer 1	250	+	low		
11	Infant faeces	250	+	7.47	1494	1.93
12	Infant faeces	250	+	0.798	159.6	

After DNA extraction, the DNA integrity was assessed with 1% TAE gel.



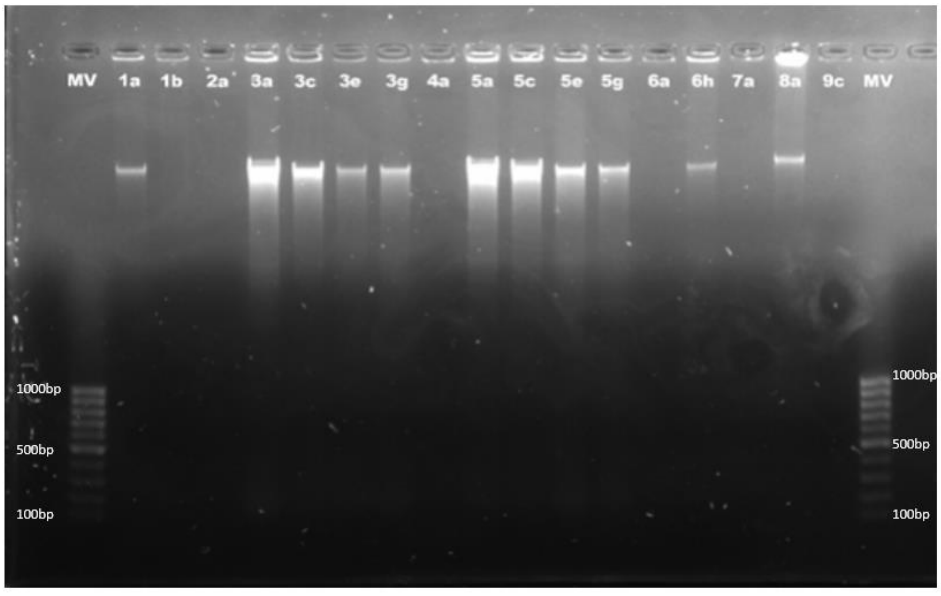
Appendix figure 5B. Gel picture of fecal, Gut Standard, OmniGENE fluid DNA/RNA Shield fluid and Lysis Buffer 1 samples after genomic DNA isolation (test 1). MV=molecular meight/ladder, 1-3=Gut Standard, 4=Lysis Buffer 1, 5-7=Adult feces, 8=OmniGENE fluid, 9=DNA/RNA Shield fluid, 10=Lysis Buffer 1, 11-12=Infant feces. Gel 1% TAE. Ladder GeneRuler 100bp. Gel run 110V, 1h35min.

Test 2. The second test consisted of incorporating bead-beating with Qiangen's PowerBead Pro plates and TissueLyser (25Hz, 2x5 min) with and without proteinase K incubation. These groups were similar to pre-treatment groups CM and CMprot. It was also tested whether adding 50 µL of OMNIgene liquefaction reagent (OMNIgene, USA) would affect the DNA concentration majorly. Liquefaction reagent is sometimes added to the samples to render them pipettable. Fecal samples stored in -80ºC for long periods of time can become dry or otherwise hard to pipette (own experience).

Appendix table 5C. Concentrations of extraction test 2 with Chemagic MSMI system.

DNA ID	Sample	Volume pipetted (µL)	ProtK	Liquefaction (50µL)	DNA conc (ng/µL)	Absolute conc (ng)	Nanodrop A260/280
1a	Gut Standard	75	-	-	4.29	858	1.81
1b	RNA/DNA Shield fluid	250	-	-	0.057	11.4	
2a	Lysis Buffer 1	250	-	-	0.088	17.6	
3a	Adult faeces	250	+	-	29.7	5940	1.69
3c	Adult faeces	200	+	+	17.1	3420	1.73
3e	Infant faeces	250	+	-	7.08	1416	1.97
3g	Infant faeces	200	+	+	9.47	1894	1.9
4a	Lysis Buffer 1	250	+	-	0.104	20.8	
5a	Adult faeces	250	-	-	29.2	5840	1.62
5c	Adult faeces	200	-	+	21.2	4240	1.66
5e	Infant faeces	250	-	-	12.2	2440	1.73
5g	Infant faeces	200	-	+	9.39	1878	1.85
6a	OMNIgene fluid	250	-	-	too low		
6h	Gut Standard	75	+	-	2.44	488	
7a	Lysis Buffer 1	250	-	-	too low		
8a	Gut Standard	75	+	-	3.86	772	1.77
9a	Lysis Buffer 1	250	-	-	too low		

After DNA extraction, the DNA integrity was assessed with 1% TAE gel.



Appendix figure 5D. Gel picture of fecal, Gut Standard, OmniGENE fluid DNA/RNA Shield fluid and Lysis Buffer 1 samples after genomic DNA isolation (test 2). MV=molecular weight/ladder, 1a= Gut Standard, 1b= DNA/RNA Shield, 2a= Lysis Buffer 1, 3a&c=Adult feces, 3e&g=Infant feces, 4a=Lysis Buffer 1, 5a&c=Adult feces, 5e&g=Infant feces, 6a=OMNIgene, 6h=Gut Standard, 7a= Lysis Buffer 1, 8a= Gut Standard and 9a=Lysis Buffer 1. Gel 1% TAE. Ladder GeneRuler 100bp. Gel run 110V, 1h15min.

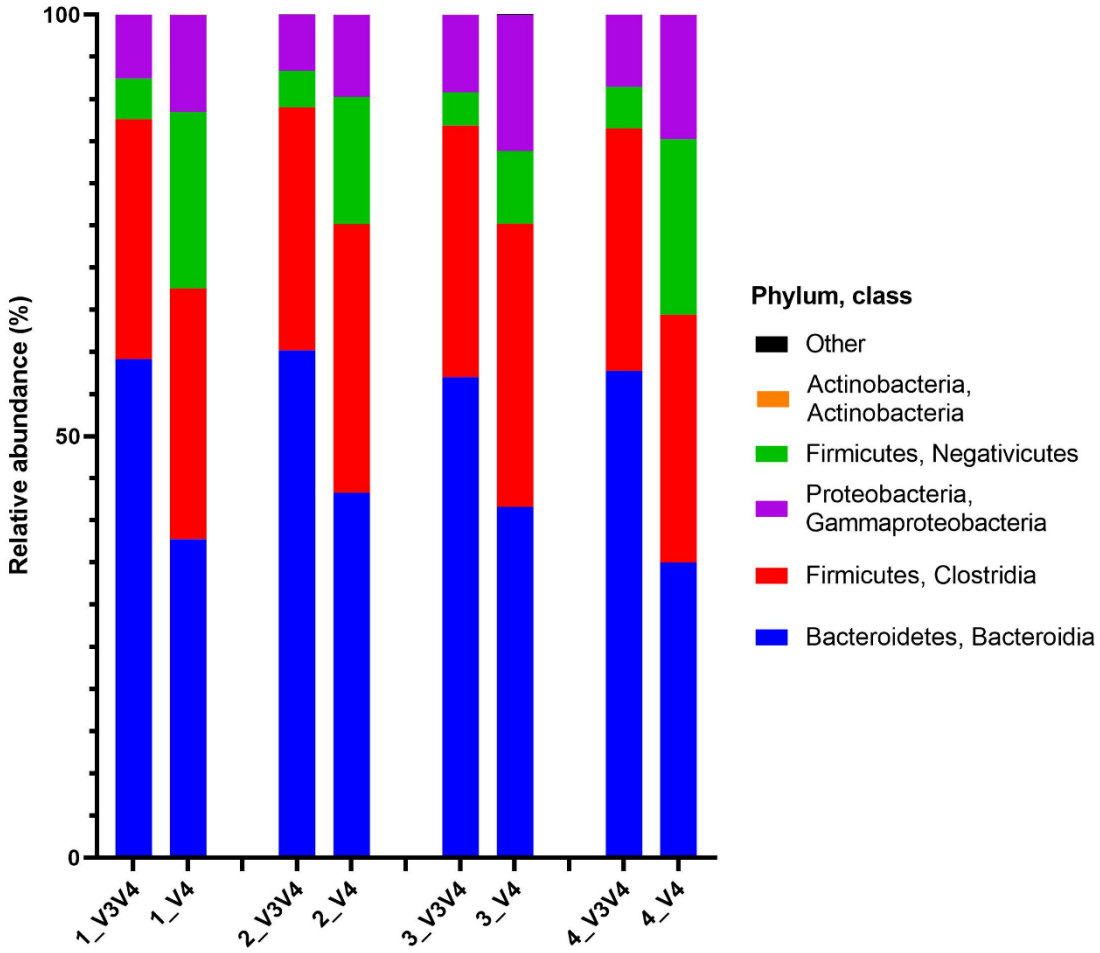
It was concluded, that adding 50µL of liquefaction reagent did not impact the DNA concentration majorly. Additionally, it was decided that the homogenization frequency was decreased from 25 Hz to 15Hz due to Lysis Buffer 1 foaming and potentially causing cross-contamination when removing the plastic seal from the bead plate. Furthermore, the fecal sample volume was decreased from 250µL to 200µL.

Appendix table 6. The genera detected in the ZymoBIOMICS Zymo Gut Standards across pre-treatment groups with V3V4 and V4 sequencing. The manufacturer's (Zymo) expected abundances are shown on the left. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. (n=2)

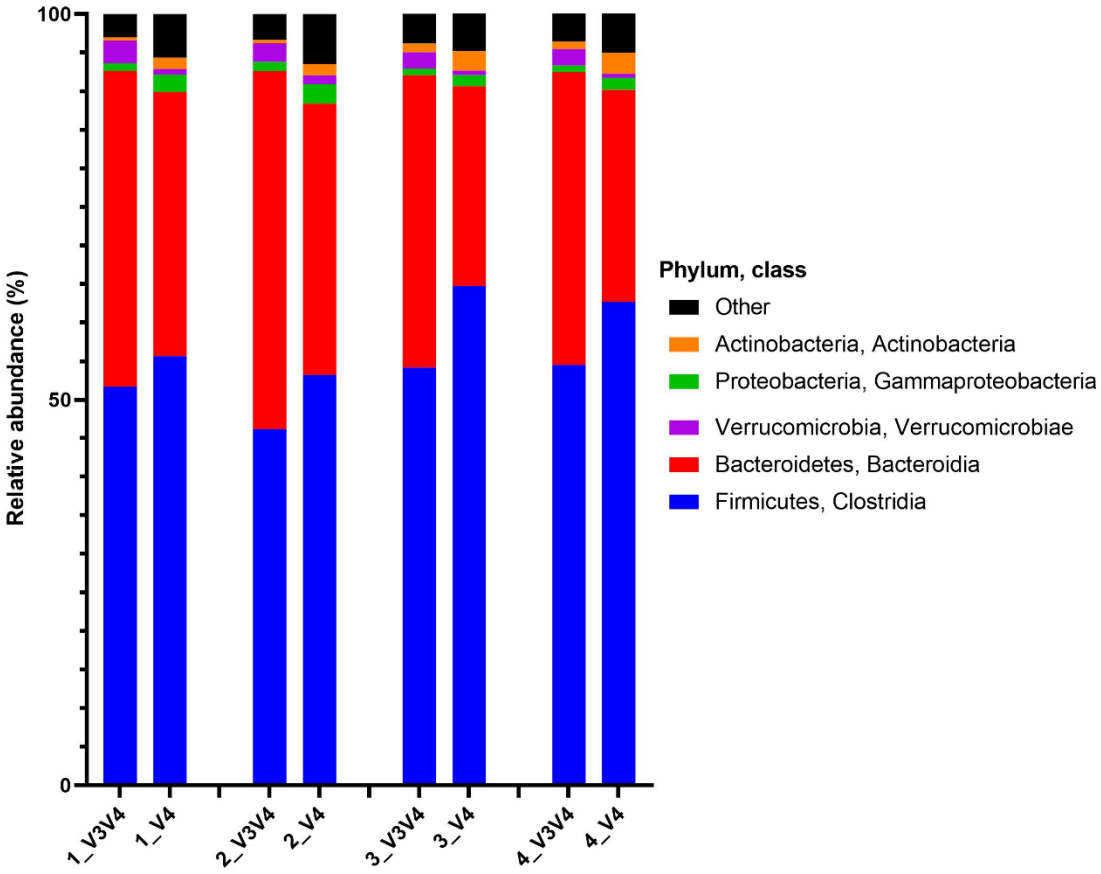
Order, genus	Zymo	1_V3V4	1_V4	2_V3V4	2_V4	3_V3V4	3_V4	4_V3V4	4_V4
Veillonellaceae, Veillonella	15.87	19.75993097	41.31167876	20.65043321	42.44170335	19.75849079	38.96519501	20.64331426	39.45722267
Prevotellaceae, Prevotella	4.89	6.501031728	32.302625	6.729586923	29.90237922	6.671073347	27.00039104	6.70233915	22.02721789
Ruminococcaceae, Faecalibacterium	17.63	18.71411017	4.647996767	19.10222372	5.465893867	19.24702303	6.942127463	18.53823967	7.761238764
Enterobacteriaceae, Escherichia	12.12	10.28658502	8.582638955	10.32994942	9.275131382	10.97233637	10.06607427	10.78423195	10.27439558
Bacteroidaceae, Bacteroides	9.94	14.88163601	2.256249774	16.11150054	2.80428281	16.05050916	3.599488961	15.38719157	4.630133905
Lachnospiraceae, Roseburia	9.89	11.01797913	4.728648826	8.578357965	3.789286119	8.734430845	5.446419199	9.700121091	6.384996668
Fusobacteriaceae, Fusobacterium	7.49	11.54630418	3.490248094	10.56693343	3.219462703	10.51077185	4.115238073	10.17623548	5.032380141
Peptostreptococcaceae, Clostridioides	2.62	3.713567862	0.705382563	4.164832284	0.882161639	4.482929904	1.414627632	4.203977841	1.602087669
Akkermansiaceae, Akkermansia	0.97	2.141708061	0.095274206	2.332389648	0.253764115	1.944251169	0.221049248	2.018154192	0.264094925
Bifidobacteriaceae, Bifidobacterium	8.78	0.678427835	1.004117826	0.675995455	0.913308438	0.783913731	1.11498238	0.905800385	1.261048601
Lactobacillaceae, Lactobacillus	9.63	0.619081795	0.405876952	0.642136272	0.433331435	0.744319622	0.570284986	0.807440143	0.67573822
Methanobacteriaceae, Methanobrevibacter	0.066	0.007251604	0.118100475	0.006891891	0.081133892	0.005762647	0.067730002	0.007582495	0.052488631
Christensenellaceae, Christensenellaceae R-7 group	0	0	0.039143445	0.00028141	0.028487725	0.000267751	0.039641591	0	0.04861711
Lachnospiraceae, Lachnospiraceae NC2004 group	0	0.045328513	0	0.030707175	0	0.0286524	0.000523344	0.04154952	0
Lachnospiraceae, Ambiguous_taxa	0	0	0.020412877	0	0.033035325	0.000234783	0.039793712	0.000247865	0.039090982
Rikenellaceae, Alistipes	0	0	0.028352448	0	0.025143362	0.000234783	0.025328873	0.003298066	0.038007136
Ruminococcaceae, Subdoligranulum	0	0.005596369	0.014316767	0.006610481	0.016323414	0.006466994	0.019975599	0.00856376	0.02589472
Lachnospiraceae, Agathobacter	0	0.034469605	0	0.028941531	0.001475024	0.018861475	0.001046687	0.023895691	0.000262477
Lachnospiraceae, [Ruminococcus] torques group	0	0	0.011370462	0	0.017745591	0	0.029095481	0	0.027381341
Ruminococcaceae, Ruminococcus 1	0	0.012209908	0.013293286	0.007758173	0.010819362	0.007974594	0.014361637	0.006682152	0.013700852
Clostridiales vadinBB60 group, Ambiguous_taxa	0	0	0.022240821	0	0.017547568	0.000267751	0.022148209	0.000495729	0.022104653
Clostridiaceae 1, Clostridium sensu stricto 1	0.0002	0.000549279	0.001719324	0.001103585	0.047321576	0.000737316	0.004678794	0.002468453	0.012732398
Enterobacteriaceae, Enterobacter	0	0	0.006596212	0	0.042417411	0	0.006613699	0	0.013755154
Prevotellaceae, Alloprevotella	0	0	0.009550276	0	0.037671789	0	0.004274302	0	0.014940816
Ruminococcaceae, Ruminococcaceae UCG-002	0	0.000831317	0.00794733	0.000844229	0.012384983	0.001005066	0.017469415	0.001649033	0.017509037
Lachnospiraceae, Lachnospiraceae UCG-008	0	0.013634896	0.001734842	0.011620719	0.003466889	0.009782911	0.00316403	0.009307354	0.004615979
Enterobacteriaceae, Salmonella	0	0.00586361	0.003454166	0.007366496	0.008355955	0.006166276	0.007985643	0.004198409	0.011621401
Marinifilaceae, Odoribacter	0	0	0.011855047	0	0.012171861	0	0.014520107	0	0.013734791
Barnesiellaceae, Barnesiella	0	0	0.009300225	0	0.012635853	0	0.013735091	0	0.014750759
Lachnospiraceae, [Eubacterium] xylanophilum group	0	0.005871009	0.006658281	0.005799333	0.002987797	0.005260113	0.005828683	0.006262249	0.004912395

Appendix 7. Most abundant phyla in infant, adult and senior fecal samples

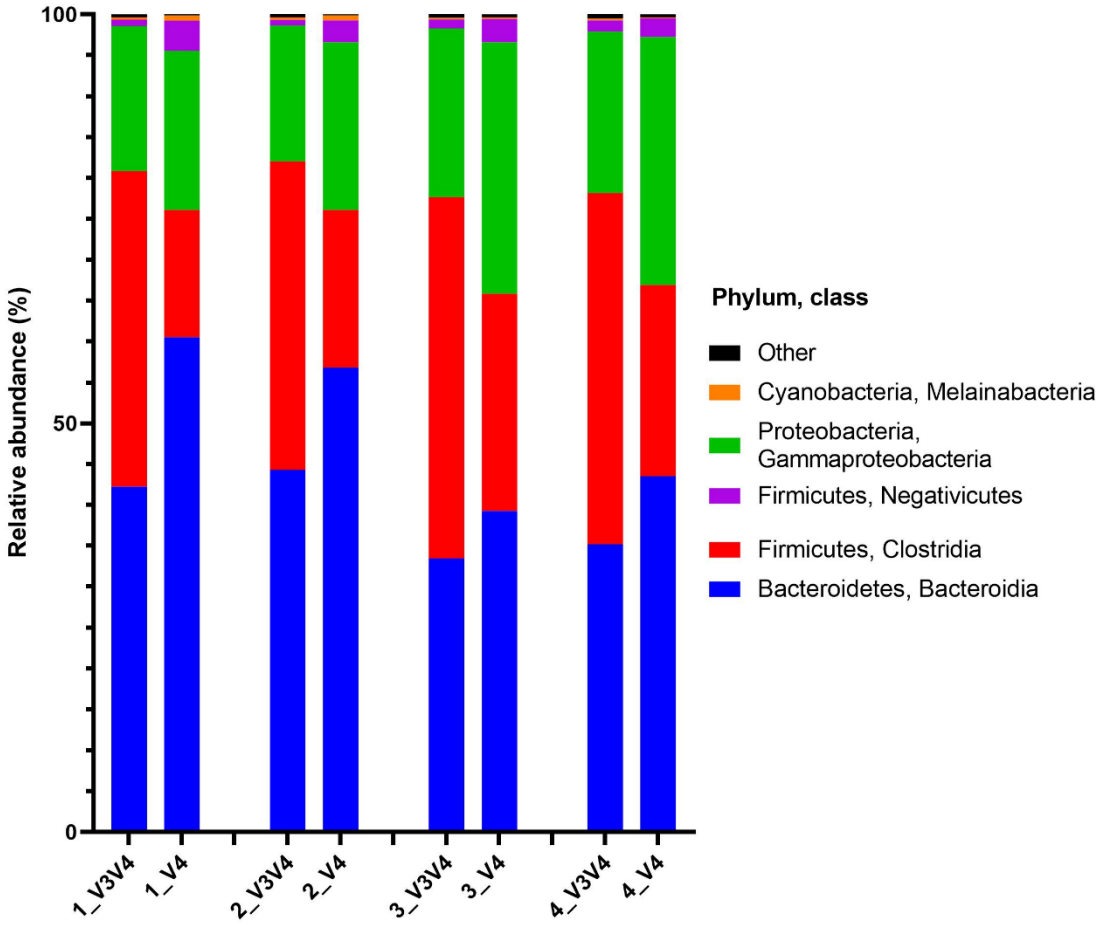
Most abundant phyla are shown in figures 7A-7C.



Appendix figure 7A. Phyla of the infant fecal samples across different pre-treatment groups with V3V4 and V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot.. (n=2)



Appendix figure 7B. Phyla of the adult fecal samples across different pre-treatment groups with V3V4 and V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot. (n=2)



Appendix figure 7C. Phyla of the senior fecal samples across different pre-treatment groups with V3V4 and V4 sequencing. Groups: 1=Cprot, 2=C, 3=CM, 4=CMprot.. (n=2)