FACULTY OF SCIENCE AND TECHNOLOGY

BACHELOR'S THESIS

Study programme / specialisation: Biological Chemistry	The spring semester, 2023 Open
Author: Victoria Vedvik	
Supervisor at UiS: Mark van der Giezen Co-supervisor: Luz Aurora Martinez Contreras External supervisor(s):	
Thesis title: Inflammatory bowel disease and assessment of the intestinal microbiota.	
Credits (ECTS): 20 stp.	
Keywords: Inflammatory bowel disease Gut microbiome 16S PCR amplification Illumina sequencing	Pages: 47 + appendix: 16 Stavanger, June 2023

Abstract

Inflammatory bowel disease (IBD) is a widespread disease, especially in the Western world. IBD is normally classified under two conditions; ulcerative colitis (UC) and Crohn´s disease (CD). The cause of IBD is not yet known, but there are many scientific studies that point in the direction of dysbiosis in the intestines of the diseased patient. Treatment and medications can slow down IBD, but there is no cure for the disease. Affected people suffer from fatigue, food avoidance, and reduced quality of life, among others. It is necessary to assess the microbial environment of the gut and work towards finding the reason, and further, the right treatment.

This thesis is part of a collaboration between the University of Stavanger and Stavanger University Hospital. The project aims to analyze stool samples, from patients diseased with IBD. The methods used are DNA extraction, 16S PCR amplification, gel electrophoresis, and Illumina sequencing. With the use of the 16S rRNA gene, the amplification of the conserved region of bacterial DNA could be done. This is key to assessing microbial diversity and understanding the disease. Illumina sequencing is a short-read sequencing, and a lot of information is lost in the sequencing process. Hopefully, in the future, with the use of long-read sequencing, such as PacBio, the information lost could be revealed, and used for further study of IBD.

The DNA extraction yielded good-quality DNA in 18 (67%) of 27 samples. 7 (26%) of 27 samples did not have good-quality DNA according to the absorbance ratios, A260/A280 and A260/A230. The sequencing results showed a majority of the most common bacteria in the human gut, the phyla Firmicutes and Bacteroidetes. 5 (16%) of 32 samples consisted almost entirely of these two groups, which goes against the theory that diseased gut contains a larger number of the phyla Proteobacteria and Actinobacteria. One sample stood out, containing almost only the phyla Proteobacteria, which is proved to be an indication of disease. Also, the bacteria genus that was most abundant was *Bacteroides*, followed by *Faecalibacterium* and *Blautia*.

Lastly, this project was a small project containing a small number of samples. The results are not statistically reliable but give an indication of what we might find in a diseased gut. Further work would be to collect a large number of samples and proceed with these methods for assessing the bacterial DNA. Also, the use of long-read sequencing could open the possibility of one day finding the cure for IBD.

Acknowledgements

I want to thank my supervisor, Mark van der Giezen, for help and guidance through this project, and the opportunity to write the thesis with such a great team. A huge thanks to Luz Aurora Martinez Contreras, for being patient with me and helping me at the lab.

Further, I want to thank Mitchellrey Magbanua Toleco and Yohannes Seyoum Demissie, for help and guidance in the lab and analysing my results.

Lasly, many thanks to my fellow student Hanne Gjersdal. Thanks for the support and a good discussion partner in this project. I also want to thank my friends and family for support. It has been a pleasure to write and learn from this team.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Inflammatory bowel disease (IBD) is a widespread disease, and the incidence has increased in the last decade. Western Europe is the most exposed to IBD for reasons not yet known [1]. Evidence suggests the reason why some are affected is related to environmental factors, diet, and genetics. This field is developing, and new evidence and discoveries continue to emerge. Intestinal dysbiosis is closely related to IBD [2]. Different medications and lifestyles are suggested to obtain homeostasis in the gut, to prevent this dysbiosis [3]. Through methods like DNA extraction, PCR, and sequencing we can determine some of the gut microbiota, such that treatment and medication can be specialized for each individual. Still, these methods can be improved, such that the cause of IBD someday can be discovered.

This thesis project is part of a collaboration between the University of Stavanger and Stavanger University Hospital. Stavanger University Hospital has been involved in several clinical trials to better understand inflammatory bowel disease, treatment, and improving quality of life for the patients. This leads to the aim of the project which is to collect stool samples provided by the Stavanger University Hospital and learn how assess the presence of microbes in the human intestinal microbiome.

1.2 Human gut microbiome

The human gut consists of many different microbes, including bacteria, archaea, fungi, microbial eukaryotes, and viruses. In a healthy gut the microbiome is usually dominated by bacteria of two specific phyla, Gram-negative Bacteroidetes and Gram-positive Firmicutes [3]. Combined, these take up

about 95% of the gut microbiota. However, diseased individuals tend to have larger amounts of other bacterial phyla, such as Proteobacteria, Verrucomicrobia, Actinobacteria, or Fusobacteria. The gut microbiota still varies much from individual to individual, therefore there is no standard microbiome ecology that every healthy gut shares [4].

As a human develops from infancy to adulthood, the gut microbiota evolves into a beneficial relationship. This relationship is important for the host because it contributes to important and accurate anatomical development and the host's immune system [5]. A disruption in this relationship may be a consequence of intestinal dysbiosis of the gut microbiota. A dysbiosis disrupts the intestinal barrier, which in turn relocates bacterial symbionts to the gut mucosa. An imbalance between pro-inflammatory and anti-inflammatory molecules is the result of an inappropriate immune response. This inflammation, caused by the disruption of the equilibrium in the gut, likely favors the pathogenic bacteria which maybe results in a decrease of the beneficial bacteria in the gut [3].

1.2.1 Dysbiosis and impact of external factors on the human gut

The gut microbiota is likely shaped by environmental factors, diet, and genetics. The diet is considered to be the main environmental factor. This is because the bacterial environment in the gut is shaped after the diet. In Western Europe, the consumption of processed food may correlate with the high number of IBD. These diets include high-processed food, sugar, animal protein, and little vegetables, and have little variety of nutrients [6]. The gut microbiome is affected by what we eat in that way it shapes the composition and function of the microbiome, which in turn affects the homeostasis of the intestinal environment [7]. Mediterranean countries have the lowest incidence of inflammatory bowel disease (IBD). The Mediterranean diet is rich in vegetables and fruit and is not as exposed to processed food compared to Western Europe. This variety in nutrients may be the reason why the incidence is so low [1]. Also, non-western countries have low incidence of IBD, but this number are currently increasing because of a more Western like diet. A study showed that diets with high intake of dietary fiber, fruits and vegetables, among others, decreased the risk of Crohn´s disease. This may be the reason for the low incidences of IBD in non-western countries [8].

Further, the human gut microbiota is defined through the first years of living. What microbial stimulation an infant is exposed to after birth is important for the immune system [9, p. 825]. A normal birth transfers the

mother´s microbiota to the child and affects the microbial diversity in a way that is important for further development. Birth through C-section has been shown to have a negative impact on how the gut microbiota to the infant develops. This may be because it is not exposed to the mother‘s microbiota to such a large extent as in a normal birth [10, p. 10]. Diet, illnesses, and obesity in childhood and early adulthood, also affect the gut microbiota, and make the individual more exposed to allergy, asthma, and other autoimmune illnesses later in life [10, p. 30-33].

A healthy gut has a low concentration of oxygen and is dominated by obligate anaerobe bacteria, that cannot live in environments with oxygen. It is suggested that the dysbiosis in IBD could appear from an increased oxygen level in the gut, and thereby a decrease in strict anaerobic bacteria. Consequently, an increase in facultative anaerobes has been seen in the gut of those with IBD. It is this shift from obligate to facultative anaerobes that affect the oxygen levels and then leads to dysbiosis [11].

1.3 Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a chronic disease that affects the gastrointestinal tract. The pathogenesis of IBD is linked to dysbiosis and inflammation that disrupts metabolic processes in the gut, such as amino acids, short-chain fatty acids, lipids, carbohydrates, proteins, and others. IBD is normally classified under two conditions, ulcerative colitis, and Crohn‘s disease. The diseases are complex, and therefore hard to regulate [5][3]. They have some of the same symptoms but are usually located in different areas of the intestine. Both are detected and treated differently.

1.3.1 Ulcerative colitis and Crohn´s disease

Ulcerative colitis (UC) is usually found in the rectum and lower part of the colon and is a non-transmural inflammatory disease [12]. It is a continuous mucosal inflammation and can extend up the colon. Abdominal pain, bloody diarrhea, and urgency to defecate are some of the typical symptoms of UC. Detection of the disease is usually through endoscopy, and confirmation of the biopsy specimens, to see how far the disease has gotten [13]. The symptom of UC is classified as mild, moderate, or severe [5]. It is also ulcerative wounds in the colon, such that the illness is defined by remissions and relapses [14], this means that the illness in periods has flare-ups of symptoms, and in other times it is almost gone.

Crohn´s disease (CD) is, on the other hand, a transmural inflammatory

disease, that penetrates the colon and exists across the entire wall of an organ or blood vessel. It usually can affect the entire intestinal tract from the mouth to the rectum [12]. The disease severity varies from where it is located in the colon and its behavior. It is also classified as mild, moderate, and severe. The symptoms are usually weight loss, abdominal pain, diarrhea, and more. Since this disease is a transmural inflammation, surgery is often required when complications appear [5]. The cause of Crohn´s disease is suggested to be genetic or exposure to bacteria, according to research. The disease can affect in three ways, inflammatory, obstructive, and fistulating types. Inflammation-induced thickening of the intestinal mucosal wall cause hindrances in the gut as a consequence of the inflammatory and obstructive types. When parts of the bowel intrude the neighboring bowel, this is a result of the fistulating type [15].

1.3.2 Treatment

Different treatments and medications can slow down the symptoms of IBD, but to this day, there is no cure for the disease. An inflammatory response to the intestinal microbes is suggested by different evidence to be the reason for inflammatory bowel disease [5]. Also, as mentioned, intestinal dysbiosis is through evidence suggested to correlate to IBD [16]. The unknown reason for why someone is affected, causes one of the challenges with IBD; the lack of knowledge of the disease, and why some are diseased, affect the possibility to find proper medication and treatment that works [17].

Treatment of CD is dependent on where it is located, its behavior, and the severity of the disease. Usually, patients with CD need at least one surgery throughout life [5]. Since there is not any cure for CD, research is focused on increasing the life quality of patients with CD, which means reducing chronic inflammation in the gut. Current treatment option of the disease is alpha-4 integrin inhibitor and other TNF-alpha inhibitors [15].

The treatment of UC depends on the severity of the symptoms and the disease. Usually, 5-Aminosalycilates are the go-to treatment for mild to moderate disease. If the treatment does not work, there are other options. Still, 15% of the patients need a surgical procedure to ease the symptoms once throughout life [13].

The strict control of the diet could also ease the symptoms of the diseases when no medication works [6]. A diet that contains probiotics and prebiotics could promote a balanced intestinal microbiota environment. Probiotics are microorganisms that can give health benefits to the host when given in the right amounts. Prebiotics can give beneficial physiological effects. They are non-digestible substances and can affect the activity of the gut microbiota.

It is important to mention that the improvement of UC and CD by use of pro- and prebiotics is inconsistent, and the research results are unclear, such that is not a recommended medical solution [18].

1.4 Social challenges related to inflammatory bowel disease

The life quality of a person with IBD is significantly affected by the disease. Permanent alteration in diet, lifestyle, and behavior is required to reduce the symptoms [1]. Patients often focus on their diet to reduce the impact of IBD in daily life [19].

Some patients experience food avoidance as a consequence of IBD. In the paper "Food avoidance in outpatients with Inflammatory bowel disease" they concluded that there was a concern in the avoidance of nutrition and mistrust in advising about diet among patients with IBD [20]. Patients also struggle with mental health problems as a result of losing self-control. According to the paper "Identifying and understanding disease burden in patients with inflammatory bowel disease", patients said available help with mental health problems was most important to handle consequences of IBD, beyond the typical symptoms patients experience. Another symptom of IBD that patients have reported to have most impact on daily life is fatigue [21]. Fatigue can be a cause of anxiety and depression in patients with IBD [22]. Anxiety and depression do affect the work and daily life of those affected. A treatment for IBD would likely ease these symptoms, and increase the life quality for those living with this disease. Data from IBSEN study cohorts of health-related Quality of Life (HRQoL) in patients with IBD, showed that women with CD had a reduced HRQoL, which was a consequence of increased activity of the disease, and reduced work. Still, the result showed that the overall HRQoL did not differ from those with UC/CD and the background population [23].

1.5 Bacterial 16S rRNA gene and Illumina sequencing

To study the human gut microbiota usually two methods are used, either by amplicon sequencing using 16S ribosomal RNA sequencing or by metagenome shotgun sequencing [24]. In this thesis, the assessment of the human gut microbiota will be done by amplicon sequencing using 16S rRNA.

16S rRNA is a part of the ribosome. Ribosomes take care of the protein synthesis in the cell, and are a complex of many proteins. The ribosome in prokaryotic cells contains a small subunit called the 30S and a large subunit 50S. The 16S rRNA gene do encode a ribosomal RNA on the 30S subunit [25].

16S rRNA is about 1500 bp long and is a highly conserved region with few mutations, which makes it beneficial for studying bacteria [26]. The region is so conserved because of their important role in translating mRNA into proteins [25]. By using 16S rRNA PCR, we can target a region of the bacterial DNA that is present in all bacteria [27]. 16S PCR detects the highly variable regions V1-V9 in the bacterial RNA depending on what we want to find, see figure 1.1. These variable regions, show in different bacteria, a sequence diversity. The variable regions must be compared with other regions to differentiate between bacteria, this is because one region alone do not contain enough diversity to separate bacteria. For example the regions most suitable to distinguish between bacteria on genus level, were the regions V2 and V3 [28].

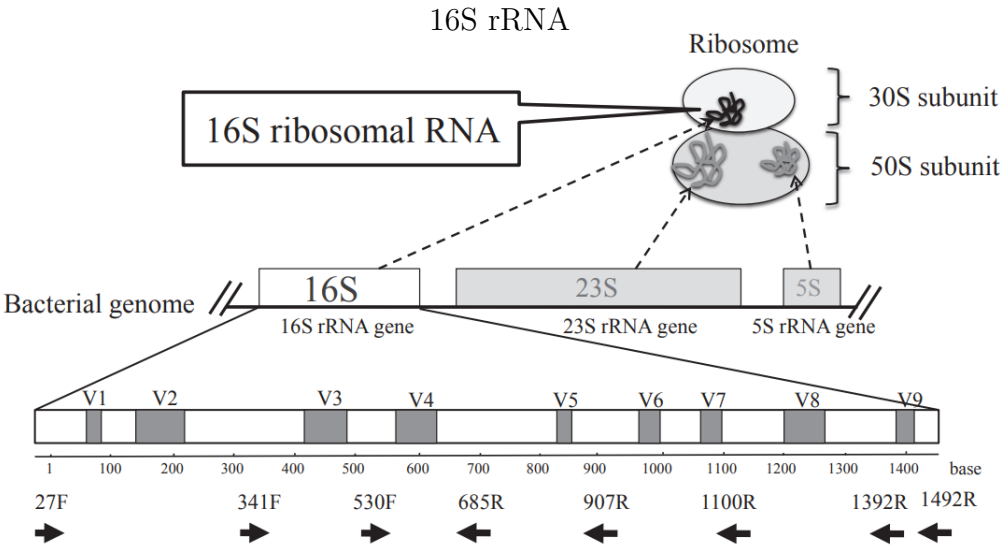


Figure 1.1: The structure of ribosome complex and 16S rRNA gene [25].

The region amplified varies depending on the intended use. It may be used as a marker for bacterial classification or the identification of bacteria in clinical samples. Additionally, the primers also play a role in which sequences are amplified. Primers are chosen based on what the outcome is supposed to be [4]. In this thesis, the primers 27F and 1492R are used. These are usually

the primers chosen to amplify the bacterial ribosomal RNA in the human gut [29].

The regions amplified could then further be sent to sequencing, and through Illumina sequencing, the sequence could be read [4]. Illumina sequencing uses a "sequencing by synthesis" approach. It identifies DNA bases as it puts them together into nucleic acid chains. With each base added a fluorescent light signal is sent, such that the order of the DNA sequence can be determined [30]. Illumina sequencing is a short-read sequencing, and are the most used sequencing method for the last decade. It provides highly accurate sequencing reads, and are affordable. This short-read sequencing has a down side; it reads less than 300 bases, which leads to a loss of very much information. On the other hand, long-read sequencing can produce sequences with the range of 10 kilobases to megabases, such as PacBio sequencing. In short, this type of sequencing uses circular DNA template called a SMRTbell cells that contains double-stranded DNA and DNA polymerase. During the sequencing, the polymerase incorporates fluorescently labeled deoxynucleoside triphosphate into the template. The template are excited and recorded on camera. This is repeated several times, and make it possible to uncover the sequence and identity of each base in the SMRTbell template [31].

1.6 Bioinformatics and analysis

Bioinformatics is a category that contains computational and analytic methods that process biological data, and are very necessary to understand the microbial diversity of the human gut [32]. After sequencing the microbiome is in a format called raw data. This is long sequencing threads with base pairs. With help of programs such as R, or websites as MicrobiomeAnalyst, these threads with base pairs can be interpreted. Through these programs the diversity of bacteria can be visualized with different kinds of plots, making it easier to evaluated the bacteria in each sample.

Further, the data are usually classified into OTU, metadata, and taxonomy data. The taxonomy data contains information about the bacteria, phylum, family, genus, among other. Also, numerical taxonomy are used, and the basic unit of this is OTU, operational taxonomic unit [33]. The metadata describes the information, in other words data about the data [34].

Usually, alpha and beta diversity is discussed when doing a analysis. Alpha diversity point out similarities and differences within one sample [35], while beta diversity compare between the samples [36]. When doing such analysis it is important to make the results as reliable as possible, and look at statistical considerations. Such considerations could be the rarefaction,

normalization and the metrices used to analyse the samples.

Chapter 2

Materials and method

2.1 Approach

This chapter will give an overview of the materials and methods used in this project. The project aims to analyze stool samples from patients with Crohn´s disease and ulcerative colitis. DNA extraction is used to obtain clean and high-quality DNA samples, followed by endpoint PCR amplification and gel electrophoresis to detect the amount and weight of the sample. Sequencing is then performed to provide an overview of the bacterial species in the sample, which can provide an indication of how the patient´s gut microbiota is.

2.2 Materials

2.2.1 Biological materials

Stavanger University Hospital provided samples from patients with early and late stages of Inflammatory Bowel Disease (IBD), which included a mix of Crohn´s Disease and ulcerative colitis. The samples were collected at various time points ranging from V0 to V60, with intervals at V3 and V11, indicating the number of months since diagnosis. The samples were prepared by adding sterile zirconia beads and stored at -80 ^{o}C. In this project, 25 samples were analyzed.

2.2.2 Primers

Primers were used in the master mix for PCR to amplify most of the bacterial 16S rRNA gene [29]. These primers were chosen because they amplify the bacterial ribosomal RNA small subunit (16S). The numbers 27 and 1492 indicates the position wanted to amplify in the rRNA gene.The primers are listed in table 2.1.

Table 2.1: Forward and reverse primer with nucleotide sequence used in this study.

Primers	Nucleotide sequence 5'-to-3'
27F	AGAGTTTGATCCTGGCTCAG
1492R	TACGGYTACCTTGTTACGACTT

2.3 Method

2.3.1 DNA extraction

The DNA extraction was accomplished by the use of the IHMS DNA extraction protocol Q, provided by Costea [37]. DNA extraction is a method to purify DNA, by using physical and chemical methods to separate DNA from cell membranes, proteins and other cell components [38].

Stool samples from Stavanger University Hospital were used to carry out a fecal DNA extraction. Qiagen QIAamp DNA stool kit was used for the extraction. 150 to 200 mg of frozen feces was transferred into a 2 ml tube. The samples were homogenized with 1.0 ml ASL lysis buffer of the kit and 0.3 g of sterile zirconia beads (Ø 0.1 mm) by vortexing for 2 minutes. The buffer was heated at 70 ^{o}C to avoid precipitation. A lid was placed on the samples and wrapped in parafilm to avoid spilling from the tube through vortexing and incubation.

Further, the samples were incubated for 15 minutes at 95 ^{o}C, with a possible variation of 1-2 ^{o}C. The lid and parafilm were removed, and only parafilm was replaced. For the cells to be mechanically lysed, the samples were sonicated for 8 minutes and 15 seconds in a Fastprep^{tm} instrument. To optimize results, the sonication time was split into 2 minutes x 4, with a 1-

minute break between, letting the sample rest every other time and slightly turning it.

The samples were cooled down on the ice for 2 minutes and centrifuged at 16000 x g at 4 ^{o}C for 5 minutes afterward. The supernatant was transferred to a new 2 ml tube and put on ice for later use. The pellet was mixed with 300 µl ASL lysis buffer of the kit. The steps were repeated with a smaller volume (300 µl). Supernatants were pooled into the new 2 ml tube.

260 µl of 10 M ammonium acetate was added to each lysate tube and mixed well before incubating on ice for 5 minutes. The samples were centrifuged at 16000 x g at 4 ^{o}C for 10 minutes. The supernatant was transferred to two 1.5 ml Eppendorf tubes, and an equal amount isopropanol was added. The samples were mixed well and incubated on ice for 30 minutes. Fresh 70% ethanol was made. For 15 minutes, the samples were centrifuged at 16000 x g and 4 ^{o}C. Through aspiration, the supernatant was removed. The nucleic acid pellets were washed with 70% EtOH (0.5 ml). Then it was centrifuged for 3-5 minutes, the alcohol was discarded and the pellet was left to dry. The nucleic acid pellet was dissolved in 100 µl of TE buffer, and the two aliquots were pooled.

For dissolving leftover RNA, 2 µl of DNase-free RNase (10 mg/ml) were added, and the samples were incubated for 15 minutes at 37 ^{o}C. Further, leftover proteins were also dissolved by adding 15 µl proteinase K and 200 µl AL buffer to the supernatant. The samples were vortexed for 15 seconds and incubated at 70 ^{o}C for 10 minutes. 200 µl of ethanol (96-100%) was added to the lysate and mixed by vortexing. The samples were transferred to a QIAamp spin column and centrifuged at 16000 x g for 1 minute at 20 ^{o}C. The flow through was discarded and 500 µl of AW1 buffer (Qiagen) was added, and subsequently centrifuged at 16000 x g for 1 minute at 20 ^{o}C. The flow through was discarded again, and 500 µl of AW2 buffer (Qiagen) was added, then centrifuged at 16000 x g for 1 minute at 20 ^{o}C.

The samples were transferred to clean 1.5 ml Eppendorf tubes and dried by centrifugation at 20 ^{o}C for 1 minute. 150 µl of ATE buffer (Qiagen) was added in two series of 75 µl, rested for 30 seconds, and centrifuged at 16000 x g for 1 minute. Finally, the samples were centrifuged at 16000 x g for 1 minute to elute the DNA. Afterwards, the samples were stored in -20 ^{o}C until further use.

2.3.2 Methods for enhancing DNA sample quality

DNA clean and concentrator

Genomic DNA cleaning and concentrator from Zymo Researcher were used to recover DNA after extraction. The protocol removes any contamination of impurities from the extraction step [39]. Samples with poor quality, were enhanced by implementing the following protocol. The protocol with no changes are listed in Appendix A.

The centrifuge was set to 16 000 g. In a 1.5 ml microcentrifuge tube, 140 µl of ChIP DNA Binding Buffer was added to 70 µl of DNA sample, and mixed thoroughly. The ratio was 2:1 of the DNA Binding Buffer and sample. The mixture was transferred to a Zymo-Spin^{tm} IIC-XLR Column in a Collection Tube. Then, the sample was centrifuged for 30 seconds, and the flow-through discarded. 400 µl of DNA Wash Buffer was added to the column, centrifuged for 1 minute, and repeated. 50 µl of DNA Elution Buffer was added directly to the column matrix and incubated at room temperature for five minutes. The samples were centrifuged for 30 seconds to elute the DNA.

DNA precipitation

Impurities were removed from the DNA by DNA precipitation. The protocol is listed in Appendix A.

24 µl of the sample was added to 40 µl of 10 M ammonium acetate. 101 µl of ice-cold 99% ethanol was added to the DNA solution and mixed by pipetting and gentle vortexing. The final concentration was approximately 2.4 M. The samples were put in the freezer at -20 ^{o}C.

The next day, the centrifuge was set to 4 ^{o}C, and the samples were centrifuged for 30 minutes at 15000 g. The supernatant was removed by pipetting without disturbing the pellet. Furthermore, the DNA pellet was washed by adding 200 µl of 70% ethanol. Then, the samples were centrifuged at 4 ^{o}C for 30 minutes at 15 000 g, closely followed by another 15 minutes at the same temperature and g. The supernatant was removed through pipetting, without disturbing the pellet.

The pellet was air-dried for 5-20 minutes until all the ethanol was gone, and then the lid was closed. The DNA was redissolved in 20 µl of ATE buffer. The mixture was warmed at 37-42 ^{o}C for approximately 10 minutes. The samples were gently vortexed, and the tubes spun.

2.3.3 Quantification of DNA

After the DNA extraction and improvements of the DNA quality, the samples were analyzed with NanoDrop Micro-UV/Vis Spectrophotometer from Thermo scientific. Double-stranded DNA, dsDNA, was used to measure the samples with concentration ng/µl. To analyze the DNA quality, the instrument gave two absorbance ratios, A260/A280 and A260/A230. A260/A280 is a good indicator of protein contamination, and the expected result above 1.822 indicates a pure DNA sample. The absorbance ratio A260/A230 indicates contamination caused by organic compounds, and the targeted result should be 1.8, or preferably 2.0.

2.3.4 PCR-amplification

Polymerase chain reaction, PCR, is a rapid DNA amplification technique. The quantity of bacterial material was determined by endpoint PCR, and thereby measured using gel electrophoresis. This was performed on 16S rRNA from the DNA extraction by DreamTaq DNA polymerase Master Mix from Thermo Fisher Scientific, and a BIO-RAD T100 thermal cycler.

PCR solutions

The master mix was prepared by adding together the solutions in the same order as listed in table 2.2. The concentations of the solutions was multiplied with the number of samples and positive/negative controls. The volume and concentration of each reagent in the Master Mix were selected based on the recommended solutions provided by Thermo Fisher Scientific, which is shown in Appendix A.

Table 2.2: PCR Master mix solution calculated for one sample.

Order	Reagents	Amount (µl), X1
1	Water, nuclease-free	15.3
2	10X DreamTaq Buffer	2
3	dNTP Mix, 10 mM	0.4
4	10 uM Forward primer	0.4
5	10 uM Reverse primer	0.4
6	DreamTaq DNA Polymerase	0.1
	Total volume	18.6

19 µl of the master mix was transferred into each PCR tubes, and 1 µl of each the DNA extracted stool samples was added. 1 µl of the positive

control, ctt1 from Zymo Biomer, and 1 µl of nuclease free water was added as negative control. The samples were spun for few seconds before placed in the thermo cycler.

PCR cycles

The following temperature cycling conditions for BIO-RAD T100 thermal cycler listed in table 2.3, were used to amplify the samples. The PCR product was prepared for agarose gel by combining 10 µl PCR product and 2 µl 6X loading dye. For easier loading of the products onto the gel, 5 µl PCR product and 1 µl 6X loading dye was used.

Table 2.3: Temperature cycling conditions for PCR.

Steps	Temperature	Time	Number of cycles
Initial denaturation	95	1 min	1
Denaturation	95	20 sec	28
Annealing	52	30 sec	
Extension	65	2 min	
Final extension	65	5 min	1

Modifications of the PCR procedure

The PCR was repeated several times, to improve the gel. The samples were diluted to 40 ng/µl, except samples with to low or approximately same concentration as wanted. Due to missing positive control, a new positive control was used, Ctt1 Zymo Biomer. Also a new DNA ladder, Hyperladder 1kb 100 lanes from Meridian was used instead of the DNA ladder from Thermo Scientific. The first two gels did not have the new positive control and Hyperladder from Meridian.

2.3.5 Gel electrophoresis

Gel electrophoresis is a laboratory method that separates DNA based on molecular size [40]. The gel contains small pores which the molecules travels through with the force of an electrical field.

1.5% agarose gel was prepared by adding 1.5 gram per 100 ml. For two gels, the amount used was 300 ml, therefore 4.5 gram agarose was added to 300 ml 1X TAE buffer. To dissolve the agarose, the solution was heated in the microwave for 10-20 seconds, and stirred until all the agarose was dissolved.

10 µl of Gel-red was added to each 100 ml agarose gel and mixed. 300 ml was prepared, before 200 ml was poured into two forms. The gel rested while the PCR samples were prepared. For the first gel, 2 µl of 6X loading dye was added to 10 µl of PCR sample, and the two controls. For the second gel, 1 µl of 6X loading dye was added to 5 µl of PCR sample with 40 ng/µl concentration, and the two controls.

The solidified gel was placed in the gel tank and 12 µl of each of the PCR samples was loaded into the wells. 6 µl of the samples was loaded to wells of the other gel. 3 µl of the DNA ladder was added to the second and last well, followed by the samples. The gel was set to 60 V (BIO-RAD PowerPac Basic machine) and 400 mA for 120 minutes, and later adjusted to 85 V, 400 mA and 90 minutes. When the time were up, BIO-RAD ChemiDoc Touch Imaging System was used to take a pictures of the gels. As mentioned, for easier load onto the gel a smaller amount of samples was added for the next gels. The amount was 6 µl. The pictures of the gel was edited with Fiji Image J and GIMP.

The DNA extracted stool samples were prepared the same way on a gel to verify the quality of the DNA. Since the DNA fragmentation was the goal to assess, the samples was not amplified with PCR, neither were a positive and negative control necessary for these gels. 2 µl of the sample, 8 µl of nuclease free water and 2 µl of loading dye was added, and loaded onto the gels.

2.3.6 Illumina Sequencing and bioinformatics

To assess the bacterial DNA in the samples, the samples were sent to Statens Serum Institutet (SSI) in Denmark for Illumina sequencing. Illumina sequencing is a type of Next-generation sequencing, and is used to determine the order of nucleotides in entire genomes, or in this case targeted regions of DNA [41].

The samples were prepared by adding 50 µl of the DNA extracted samples to new clean 1.5 ml Eppendorf tubes, with double lock. The tubes were labeled with the sample number and month since diagnosis. Then, it was packed into a plastic bag and sent to the sequencing lab in Denmark. This was the required preparation of the samples from SSI, but this way of preparation varies between the companies.

Through bioinformatic tools, an analysis of the sequencing data was performed. From SSI the data received were already labeled with which type of bacteria that are found in the different samples. Three excel files with OTUs, taxonomy and metadata was used as datasets. The data were interpreted

and visualized with the use of the website MicrobiomeAnalyst. Analysis taken into considerations were a bar plot of the different bacteria phyla, alpha diversity plot, beta diversity plot, linear discriminant analysis effect size (LEfSe), and Single-factor analysis.

Chapter 3

Results and discussion

In this project, 25 stool samples were collected from patients with ulcerative colitis and Crohn´s disease. Two stool samples without IBD were used as negative controls. Stavanger University Hospital provided the samples.

The main goal of this thesis is to learn how to assess the microbes in the intestinal of a person with ulcerative colitis or Crohn´s disease. This was achieved through DNA extraction, 16S PCR amplification, gel electrophoresis and Illumina sequencing. Also, different methods that enhance DNA quality were tested on few samples for learning purposes. In this part, the results will be shown and discussion of the results will be done.

3.1 DNA extraction

DNA extraction was performed with the IHMS DNA extraction protocol Q on 27 samples. After the extraction, the concentration and absorbance were measured with NanoDrop Micro-UV/Vis Spectrophotometer from Thermo Scientific. DNA extraction is an important step for the collection of high-quality DNA for further assessment. The quality and quantity of DNA are necessary to measure to obtain trustworthy results and improve the sequencing [42].

The absorbance ratios, A260/A280 and A260/A230 measure the DNA quality and are an indication of how good quality the DNA is. The optimal absorbance ratio for A260/A280 is above 1.822, and for A260/A230 it is above 1.8, preferably 2.0. An absorbance ratio within the range indicates good-quality DNA. A low absorbance ratio, on the other hand, indicates contamination of the sample. This contamination could be proteins, RNA, or alcohol, which could be a consequence of mistakes during the DNA extraction process. Lack of accuracy when pipetting, wrong temperature, minutes,

strength on the machines, or precipitation in buffers and materials could contribute to a poor yield of DNA. The concentration and absorbance ratio of each sample was different. 18 samples (67%) of 27 did have absorbance ratios, A260/A280 and A260/A230, within the recommended range. 7 samples (26%) of 27 did have a low absorbance ratio, see table 3.1.

Table 3.1: Results from DNA extraction. Measured with NanoDrop Micro-UV/Vis Spectrophotometer. NC stands for negative control.

Patient number/visit	Concentration of DNA (ng/µl)	A260/280	A260/230
289 v60	242.6	1.78	0.98
294 v60	48.7	1.82	1.80
295 v60	94.8	1.93	1.85
299 v60	101.9	1.83	1.00
303 v60	214.3	1.85	2.02
304 v60	351.1	1.73	0.77
307 v60	150.9	1.77	1.34
311 v60	74.8	1.87	2.26
315 v60	173.2	1.91	2.29
316 v60	115.0	1.84	2.04
317 v60	137.7	1.79	1.46
321 v60	183.1	1.86	1.84
323 v60	75.1	1.84	1.83
450 v11	25.5	1.82	1.72
455 v11	151.3	1.88	1.84
465 v3	125.5	1.89	2.20
466 v3	123.1	1.74	0.70
468 v3	131.5	1.85	1.81
469 v3	125.3	1.90	2.14
470 v3	162.7	1.77	1.67
471 v0	644.9	1.82	1.82
473 v0	165.7	1.85	1.81
474 v3	170.6	1.92	2.31
475 v0	113.7	1.77	1.61
476 v0	77.4	1.81	0.43
NC 1	154.0	1.53	0.66
NC 2	111.1	1.53	0.74

Further, some samples may contain more DNA than others, and would naturally yield a higher quantity of DNA, than a sample with low DNA con-

centration. High-fiber food cannot be digested by humans [43], and therefore the risk of extracting DNA from food rather than bacterial DNA, is unlikely but can be the cause of poor DNA concentration and quality of some samples. To manage the highest possible concentration of DNA, a higher amount of stool sample could be used in the extraction, to increase the possibility to secure a good amount of DNA.

The absorbance ratio does assess the purity of DNA and is monitored to avoid waste of samples and high costs when sent to sequencing. This is to avoid poor results after the sequencing. Usually, the absorbance ratio A260/A230 is questionable because salt may have an effect on the ratio when a saline elution buffer is used. The concentration of this ratio under 2.0 may indicate contamination of salts because salt also absorbs at the wavelength 230 nm [44]. What concentration of DNA and absorbance ratio needed for sequencing varies among companies that perform such services. This uncertainty may be the reason why some companies do not consider this ratio. Usually, the most important requirements are the concentration of DNA and an absorbance ratio of A260/A280 within the range. In this project, the ratio A260/A230 was not considered when choosing samples for sequencing.

DNA extraction method

There are many different protocols and kits to use when doing a DNA extraction for metagenomic studies. Preferably, the extraction should have effective protocols and affordable kits but still yield the best quality DNA. The concentration of fecal in the samples, use of beads, and vigorous shaking step do influence the quality and quantity of extracted DNA, according to [45]. They tested four different kits for DNA extraction and concluded that the use of beads in the lysing step of DNA extraction and vigorous shaking provided good-quality DNA. About 18 samples (67%) of 27 did have good quality DNA after the extraction, the protocol used did contain beads and shaking. Furthermore, they concluded that the DNA quantity also improves when 10 to 50 mg fecal is used for the DNA extraction. In this thesis, the amount of fecal used was 150 to 200 mg which may have affected the DNA extraction of some samples to poor yield.

In [37], they show that DNA extraction has the largest effects on the microbial composition they have examined. The results of extraction protocols can vary because of the impact of many different variables. In this thesis project, the results of the DNA extraction varied very much in both quality and quantity, and there was no pattern in the results, see table 3.1. In the paper, they concluded that protocol Q with the use of the IHMS Qi-

agen kit yielded the best quality DNA and quantity with the prerequisite of influencing. This is a good benchmark for further and new studies and methods.

Further, in [46], they compared six different DNA extraction protocols. They concluded that bead size does have an impact on the result of extraction. Protocol Q is a standardized protocol that is used in this thesis. However, a novel protocol MP matched protocol Q, and even used less time and was more affordable. They recommended the protocol MP for large-scale human gut metagenomic studies. Given the last consideration, the use of protocol Q was useful in this project, because of the small number of samples analyzed. The paper [47], also concluded that the standardized IHMS protocol Q performed the best extraction of DNA with both bacterial and fungal microbiome research.

3.1.1 DNA quality check

The next step to ensure the quality of the DNA is by doing gel electrophoresis of the newly extracted DNA. In the DNA extraction step, the quantity and quality were assessed through measurements with a NanoDrop Micro-UV/Vis Spectrophotometer, and this gave a marker of how clean the DNA was. Additionally, agarose gel was used. The samples were loaded onto the gel, and the rate at which the molecules migrated through the gel gave an indication of how fragmented the DNA was. DNA is usually a huge molecule; therefore, it would not be able to travel through the gel.

When samples are going to be sequenced, it is important to check the fragmentation of the DNA. For PCR amplification to be successful the DNA need to have high-quality molecular weight to have a long enough sequence for amplification. What was expected to see on the gel was a clear band at the top of the DNA ladder. This would imply DNA that is not fragmented and is in good condition, in other words, high-quality DNA. A fine smear that starts at the top and fades downwards indicates degradation of the DNA. A DNA ladder is added to show where the smear should start, to visualize the size of the DNA.

Even though the bands are weak, sample NC 2 on gel C, and 469 v3 on gel B, have a weak band at the top, with a fine smear which indicates a little fragmented DNA. Sample 289 v60 on gel A and NC 1 on gel C have a spot at the bottom of the smear. This may be because of degraded DNA. Further, the other samples were examined in the same way and gave a result that 2 samples (7%) of 27 showed little fragmented DNA, 20 (74%) of 27 were fragmented to a certain degree, and 5 (18.5%) of 27 did not have a visible band at all. Samples that did not have a band at all, may indicate that the

concentration of DNA was very low, see figure 3.1.

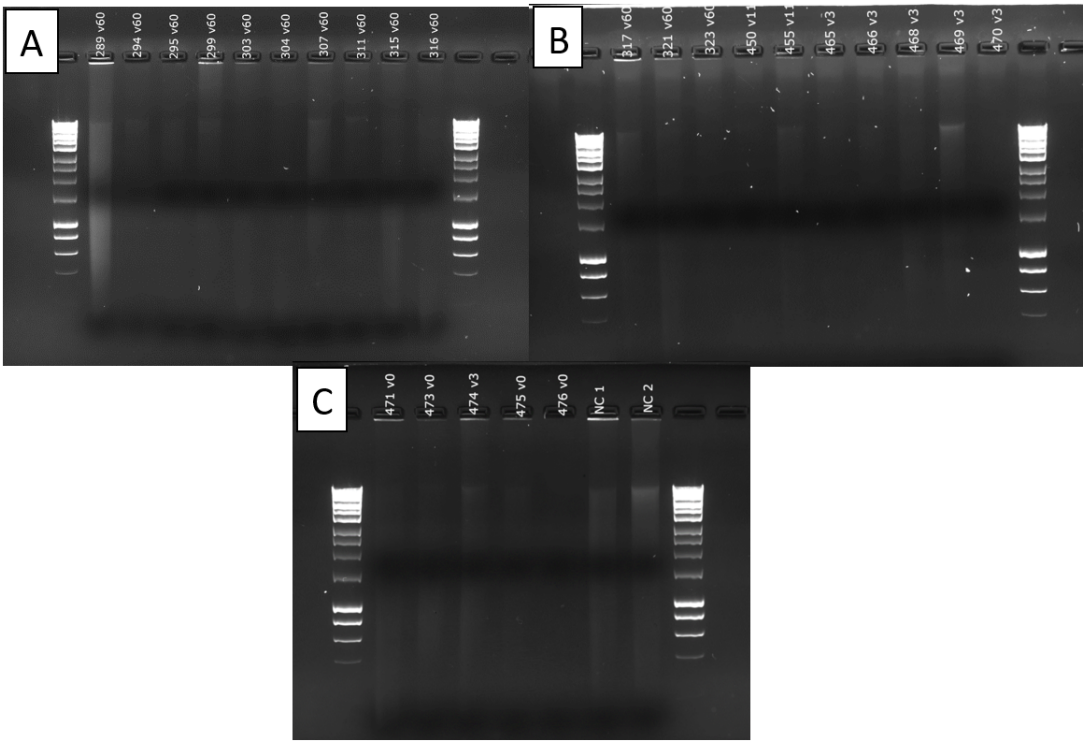


Figure 3.1: A DNA quality check on agarose gel was performed on all the DNA extracted samples. A DNA ladder is added to show the size of the DNA. The gel pictures were taken with BIO-RAD ChemiDoc Touch Imaging System, and edited with Fiji Image J and GIMP.

DNA quality checks were done right after each DNA extraction, but these gels were made to make a more organized overview of the quality of DNA. Therefore, 2 µl of the samples were mixed with 8 µl nuclease-free water to have enough concentration of the sample to do the gels, after a amount of 50 µl was sent to sequencing. This explains weak and no bands on the gels.

The fragmentation could be a consequence of pipetting errors, or too rough processing of the samples. For further work, the techniques could be improved so the DNA in the samples would be preserved better, and practice of the techniques could be done such that pipetting errors did not occur.

3.2 DNA purification

Methods that improve DNA quality after DNA extraction are an effective and time-saving process. Often in large projects, this method could be performed on the samples with low absorbance ratios, instead of doing the DNA extraction protocol once more.

In this project, two such methods were tested on five samples. The methods tested were DNA clean and concentrator protocol from Zymo Researcher and DNA precipitation with ammonium acetate. Both protocols are listed in Appendix A. These two methods were tested on the samples 289 v60, 299 v60, 450 v11, and the two negative controls.

The reason not every sample was tested, is because only ten samples of 27 were going to be sent to sequence. If there was a lack of samples with good quality, DNA clean and precipitation would be a good alternative to improve the samples that had poor DNA quality. In this case, with a project this small, DNA quality improvements on all the samples were not necessary.

3.2.1 DNA clean and concentrator

After the DNA clean protocol from Zymo Researcher was performed on the samples, they were measured with the NanoDrop Micro-UV/Vis Spectrophotometer from Thermo Scientific. DNA clean does remove, among others, enzymes, primers, and nucleotides. Making it possible to clean small and large fragments of the DNA [48].

The concentration of the samples was reduced by approximately 40% for 289 v60, 299 v60, and NC 1. On the other hand, the samples 450 v11 and NC 2 got an increased amount of DNA after the DNA clean, see table 3.2.

Table 3.2: Results before and after DNA clean of five samples. Measured with NanoDrop Micro-UV/Vis Spectrophotometer from Thermo scientific. NC 1 and 2 stands for negative control.

	Before DNA clean	After DNA clean
Samples	Concentration of DNA (ng/µl)	Concentration of DNA (ng/µl)
289 v60	242.6	118.1
299 v60	101.9	59.8
450 v11	25.5	29.0
NC 1	154.0	107.0
NC 2	111.1	128.9

Furthermore, the absorbance ratios of 289 v60, 299 v60, and NC 1 did

not improve, which may indicate further contamination of the samples. 450 v11 and NC 2 did improve. 450 v11 improved to the recommended level, but NC 2 did not improve enough, which could be a consequence of too much contamination that the DNA cleanup did not manage to clean enough DNA. Table 3.3 shows a comparison of DNA quality before and after the use of the DNA clean protocol.

Table 3.3: Comparison of the DNA absorbance ratio of five samples, before and after DNA clean.

	Before DNA clean		After DNA clean	
Samples	A260/A280	A260/A230	A260/A280	A260/A230
289 v60	1.78	0.98	1.71	0.70
299 v60	1.83	1.00	1.76	0.84
450 v11	1.82	1.72	1.85	1.92
NC 1	1.53	0.66	1.55	0.63
NC 2	1.53	0.74	1.77	1.37

3.2.2 DNA precipitation

The second protocol to improve the DNA quality was DNA precipitation with the use of ammonium acetate, and the samples were measured with NanoDrop Micro-UV/Vis Spectrophotometer from Thermo Scientific. Through DNA precipitation, salts are added to the solution with DNA and 70% ethanol. Ethanol neutralizes the DNA structure and makes the DNA precipitate out of the solution by centrifugation. This leads to an increased amount of DNA in the solution [49] [50].

The concentration of the five samples did improve considerably after this protocol, see table 3.4. This may correlate to the contamination of salts in the DNA.

Table 3.4: Results before and after DNA precipitation of five samples. Measured with NanoDrop Micro-UV/Vis Spectrophotometer from Thermo scientific. NC 1 and 2 stands for negative control.

	Before DNA precipitation	After DNA precipitation
Samples	Concentration of DNA (ng/µl)	Concentration of DNA (ng/µl)
289 v60	242.6	314.4
299 v60	101.9	112.4
450 v11	25.5	58.7
NC 1	154.0	341.9
NC 2	111.1	242.9

The absorbance ratios of 289 v60, NC 1, and NC 2 did improve with this protocol, but not to the level expected. It still indicates a lot of contamination, because the ratios are so below the recommended range, 1.822 to 2.0. Sample 299 v60 did improve the absorbance ratio A260/A230, which indicates that the contamination of organic compounds may be decreased. The last sample 450 v11 did not improve at all. The results from the DNA precipitation and a comparison with the DNA quality before and after this protocol is shown in table 3.5.

Table 3.5: Comparison of the DNA absorbance ratio of five samples, before and after DNA precipitation. DNA prec. stands for DNA precipitation.

	Before DNA prec.		After DNA prec.	
Samples	A260/A280	A260/A230	A260/A280	A260/A230
289 v60	1.78	0.98	1.79	1.05
299 v60	1.83	1.00	1.78	1.04
450 v11	1.82	1.72	1.77	1.19
NC 1	1.53	0.66	1.62	0.88
NC 2	1.53	0.74	1.78	1.41

Comparison of the DNA purification methods

From the results, the DNA precipitation method yielded a better concentration of DNA, with 5 samples (100%) out of 5 with increased concentration after the protocol. For the DNA clean, only 1 sample (20%) out of 5 got an increased amount of DNA. Further, both protocols yielded 3 (60%) of 5 samples with improved absorbance ratios. Only the DNA clean protocol

had samples where the improved ratios were within the recommended range, none of the samples for DNA precipitation improved to this level.

Both RNA and DNA do absorb at the same wavelength, and a spectrophotometer cannot distinguish these two. Therefore, even though the concentration and absorbance ratios increased after the protocols, it could still mean that there is contamination of RNA. Some DNA may also be taken out of the sample with the pipetting of the supernatant.

These kinds of methods are the last chance to improve DNA quality. In a larger project, a new extraction of the samples would be the next step if the quality did not get better, given that every sample counts.

3.3 PCR amplification and verification on agarose gel

The goal of this project is to assess the bacteria in a diseased gut of IBD. The detection of bacteria can be done through PCR amplification of the 16S rRNA gene and verification on an agarose gel. The aim of PCR amplification and agarose gel is to verify that the gene did amplify and there is bacterial DNA in each sample. This is important for the interpretation of the sequencing result of the bacterial DNA.

Verification of the samples was performed on 1.5% agarose gel after PCR amplification. Hyperladder 1 kb 100 lanes from Meridian were used as a DNA ladder. A DNA ladder contains bands with different sizes, so it is easier to see the length of the PCR product. The region amplified was about 1500 base pairs (bp) long. The primers were 27F and 1492R [29]. For verification of bacterial DNA in the samples, it is expected to see bands at the size of 1500 bp. The DNA ladder and positive control, which contains bacterial DNA, are used to compare the bands.

16 samples (62%) of 26 showed a band at the right size. 10 samples (38%) of 26 did not have a band at all. This gives a reason to believe that the samples did not amplify with the PCR amplification maybe as a consequence of degraded nucleases in the DNA. It could also be because there was no bacterial DNA in the samples, but this is very unlikely. No or faint bands can be a consequence of many causes, especially those related to PCR. The cycling length of the PCR is affected by the concentration of the samples. In this project, 28 cycles were used, and are in the recommended range that is between 20-35 cycles. It is crucial for amplification that the primers can bind to the sample. If the annealing time is either too short or the temperature too high, the primers would not be able to bind to the

sample, and then cause no/faint bands. Also, factors such as impurities in the primers, low concentration of enzyme, and more could be the reason why some of the samples did not amplify during the PCR amplification [51]. Figure 3.2 shows the 1.5% agarose gel of the 25 DNA extracted samples.

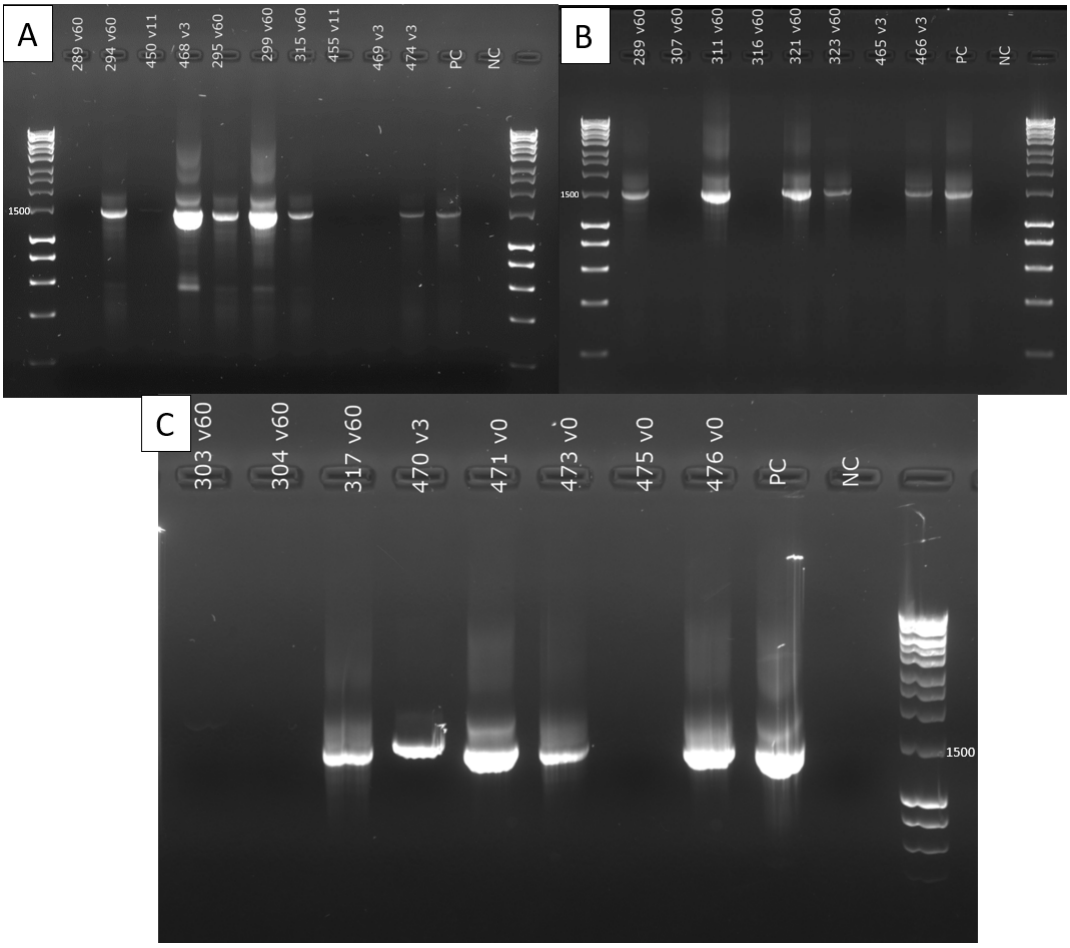


Figure 3.2: Gel electrophoresis of the 16S PCR amplification of the 25 DNA extracted samples from Stavanger University Hospital. The gel pictures were taken with BIO-RAD ChemiDoc Touch Imaging System, and edited with Fiji Image J and GIMP. The expected band size, 1500, is marked on the DNA ladder. PC stands for positive control, with bacterial DNA that gives a band at 1500, and NC stands for negative control, with nuclease free water. NC is expected to give no bands. A) The samples are diluted to 40 ng/µl. B) The samples are diluted to 40 ng/µl. C) The samples are not diluted.

Also, figure 3.2 A shows multiple bands, also called, nonspecific bands, on some of the samples. This could be a consequence of cycling times and

temperatures. Excessive cycling, extension time, annealing time and temperature, and thermal cycler ramping speed can increase the chance for nonspecific bands [51]. There could also be contamination of DNA other than bacteria, this cause a band at other sizes than expected. This could be from food or other microorganisms. In the process of DNA extraction, other cell components and contamination is removed as much as possible from the sample. The bacterial DNA has a region of the size 1500, so when the sample does have smears or marks similar to bands, this is likely because the sample did still have parts of other cell components and contamination after DNA extraction. This is also a reason why cleaning the DNA and avoid pipetting errors is necessary. Smeared bands, as seen in figure 3.2 C, show low or high molecular weight, and do have the same causes as for nonspecific bands.

The two negative controls are missing bands. This indicates that they did not amplify with the PCR step. The reason could be contamination of the samples or too degraded DNA. The DNA quality and quantity of these two controls were very low, and not within the recommended range for absorbance ratio, which underlines the possibility for contamination. Figure 3.3 shows only the two negative controls which did not have IBD.

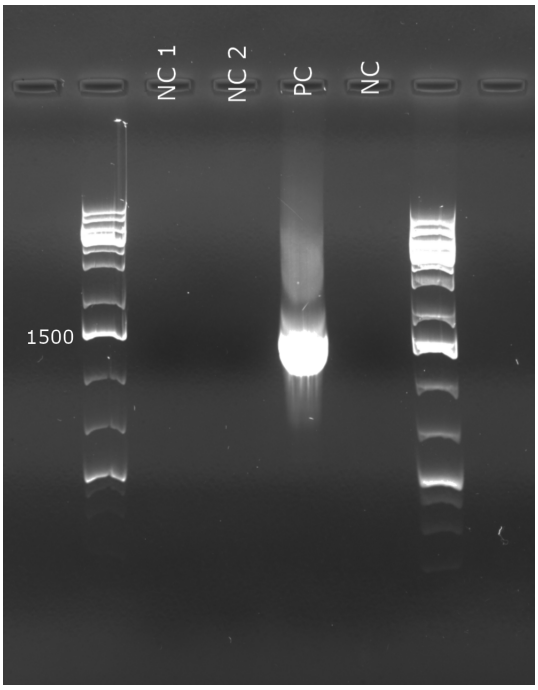


Figure 3.3: Gel electrophoresis of the 16S PCR product of the two samples that did not have IBD. The samples are missing bands on the 1500 mark. The gel pictures were taken with BIO-RAD ChemiDoc Touch Imaging System, and edited with Fiji Image J and GIMP.

3.4 Bacterial composition analysis of Crohn´s disease and ulcerative colitis using 16S rRNA

Due to problems with sequencing the 27 samples extracted in this project, the samples analyzed in this section are old sequencing results from 32 older samples. The approach of the results from sequencing would be done the same way, and the results would most likely look something like the ones presented here.

To understand the amount and differences of the microbes in the gut, bioinformatic tools are used to visualize the data from sequencing. The website MicrobiomeAnalyst was used to visualize the data set in this project. Since not any raw data were received, the use of sequence alignment and modification, such as the use of BLAST and FASTA, was not necessary.

From SSI the data received were already processed. The data were labeled with which type of bacteria that were found in the different samples. Alpha and beta diversity was used to interpret the data. Since only 16S PCR was performed on the samples, the data only shows the different bacterial phyla, not the microbial eukaryotes. The data contained 9 CD positive samples and 23 UC positive samples.

Next-generation sequencing has made it possible to look at microbial diversity on a whole other level [4]. Illumina sequencing is, as mentioned, a short-read sequencing method, that produces reads that are about 300 bases. These short reads lead to a loss of information about the samples. Short-read sequencing is still widely used today because it is a cheap, effective, and accurate method. It also has a lot of analysis tools available [52]. Long-read sequencing can produce up to 10 kilobases to megabases and gives the advantage that much more information can be obtained from this type of sequencing. This ability to produce many bases has a huge effect on sequencing technology. Hopefully, long-read sequencing can lead to resolving regions of the human genome that are lost in short-read sequencing. In the end, this may result in discovering the mechanisms of some diseases, and in this project, the cause of IBD [31].

Each sample is sequenced a different amount of times, which gives a different amount of bacteria that show in the sample. The more the sample is sequenced, the higher the possibility for another bacteria to be detected. When a large number of samples are tested and compared, it is important to take into account the rarefaction. The rarefaction narrows the differences between the samples, and makes it possible to get a reliable comparison in the alpha diversity [35].

3.4.1 Taxonomy

A taxonomy bar chart makes it possible to visualize the groups of bacteria found in all the samples. The bacteria are divided into the phylogenetic levels class, order, family, genus, and species. A bar chart is necessary for interpreting and understanding the bacterial diversity of the human gut. Through such visualization, a picture of the gut is painted. The human gut contains many different bacteria, but the most abundant are the Firmicutes and the Bacteroidetes. Also, bacteria such as Actinobacteria and Proteobacteria are found, but to a much smaller extent. From the sequencing, it is expected to find the Firmicutes and the Bacteroidetes in the samples. The number of other bacteria is also of interest to look at because of what role they may play in IBD.

In 31 (97%) of 32 samples from the sequencing, Firmicutes and Bac-

teroidetes were found. The Actinobacteria was more abundant than the Proteobacteria in almost all the samples. One notable exception from this is sample 442 v3. This sample consists almost entirely of the phyla Proteobacteria. Many Proteobacteria-related diseases contain a degree of inflammation, also IBD. Several studies have shown that Proteobacteria may be a microbial identification of disease [53], see figure 3.4 to the left. The comparison of UC and CD in figure 3.4 to the right, is made to give an indication of the differences of bacteria in the samples. Since there are very few samples to compare, it shows almost no differences. An exception is sample 442 v3, which had 99% Proteobacteria. In a larger study, it would be necessary to compare these two diseases to see what type of bacteria is different based on the type of IBD, whether it is ulcerative colitis or Crohn´s disease. As mentioned earlier, the treatment is different for those with UC and CD, and the reason could be the bacterial diversity.

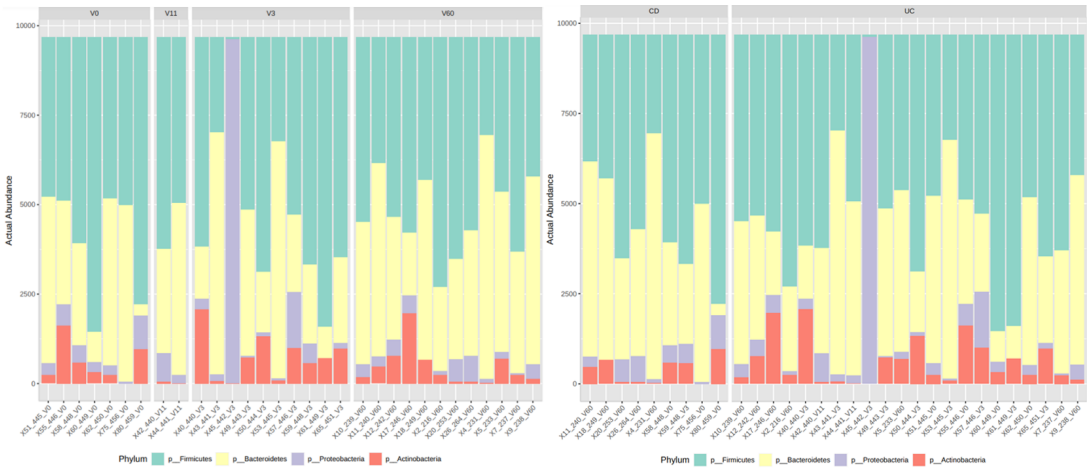


Figure 3.4: A taxonomy bar chart of the distribution of bacteria phyla found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show actual abundance of each bacteria ranging from 0 to 10 000. The x-axis show the samples.

Patients with IBD tend to have larger amounts of Actinobacteria and Proteobacteria in the intestines [4]. Therefore, it is expected to see some amount of these types of bacteria in the taxonomy bar chart from the samples. However, due to the unavailability of a control group, no comparison with healthy individuals was made. A standardized healthy gut is not yet provided, it is not possible to conclude that this amount of bacteria is enough to correlate to the disease. 5 (16%) of 32 samples deviate from this tendency with almost no Proteobacteria and Actinobacteria, even though they have

been diagnosed.

The bacteria genus contains several species within a family. The genus varies a lot between each sample. The genus *Bacteroides* was most abundant, followed by *Faecalibacterium* and *Blautia*, see figure 3.5. The *Bacteroides* is a part of the Bacteroidetes phylum. This bacteria has been shown to mostly have a beneficial function in the human gut but can act as a pathogen if it escapes the intestinal environment. Then, it can destroy tissue and resist the host immune system [54]. *Faecalibacterium* contains one known specie, *F. prausnitzii*, which is the most abundant anaerobic bacteria in the human intestines. Studies have shown that in IBD there is a reduction of this bacteria, especially in Crohn´s disease [55]. Lastly, the genus *Blautia* is known for its probiotic effect in the gut and even relieves inflammatory diseases. The composition of *Blautia* is affected by diet and disease, among others [56]. Each of these bacteria is necessary for the human gut, but a slight disruption in the composition may lead to disease. Therefore, research on the role of bacterial composition in the human gut is important to understand the reason for dysbiosis.

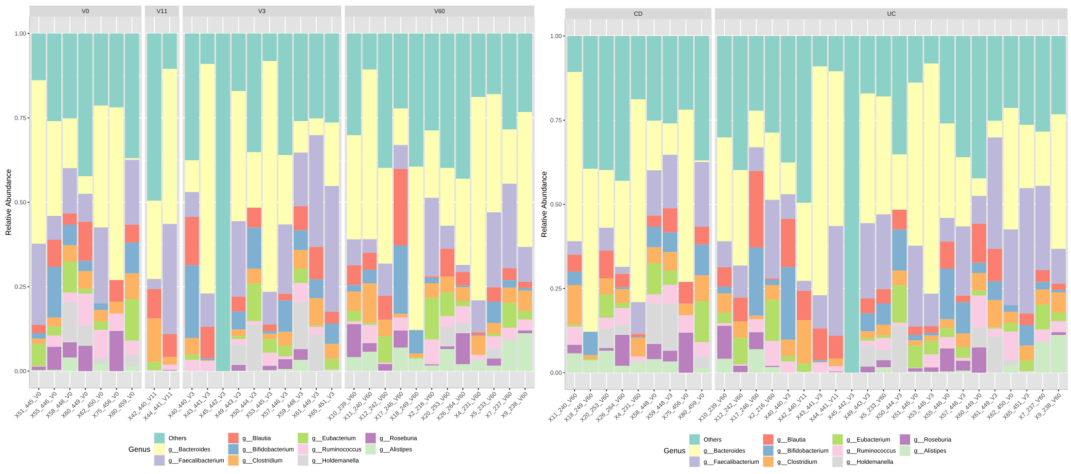


Figure 3.5: A taxonomy bar chart of the distribution of bacteria genus found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show relative abundance of each bacteria ranging from 0.00 to 1.00. The x-axis show the samples.

The results of figure 3.4 and 3.5 give an indication of what bacteria phyla and genera there are in the human gut. A conclusion that the bacteria mention in this section actually correlates with IBD cannot be made due to

the small number of samples. The bacteria class, order, top ten family, and species are listed in Appendix D.

3.4.2 Alpha diversity

Alpha diversity measures the diversity within a sample or a particular area. The features used in the alpha diversity are usually richness and evenness, and the number of species in that area. The richness contains the number of taxonomic groups, and evenness refers to the distribution of the amount of the groups [35]. The statistical method used for the alpha diversity of all samples was T-test/ANOVA. In this project, the observed richness and Shannon richness and evenness were used to visualize the samples in a box plot. Alpha diversity, measured with the feature observed richness, shows the diversity in figure 3.6 ranging from below 10 observed species to above 40 observed species in all of the 32 samples. The Shannon index was ranging from below 1 to above 3. From the box plot of alpha diversity, the diversity of bacteria varies a lot in each sample at the different months since diagnosis. This is a result of both the observed richness and the Shannon index. Alpha diversity is the first step of checking the difference of microbial bacteria in each sample and for further analysis [35].

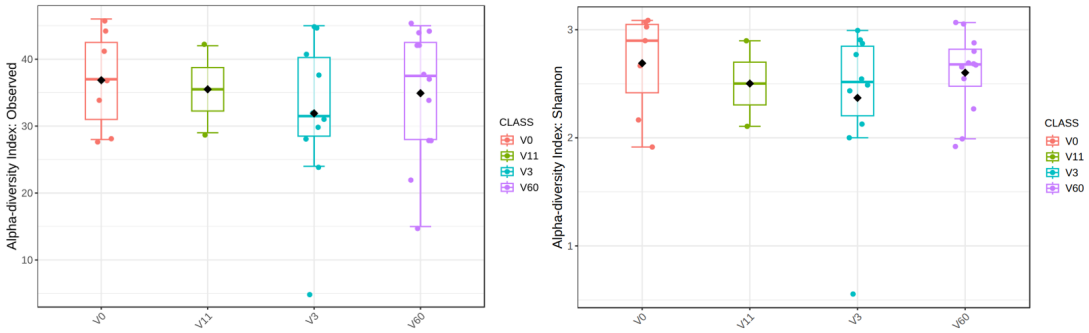


Figure 3.6: Alpha diversity of all the samples distributed in months since diagnosis in a box plot. The figure to left shows the diversity with the observed richness (labeled), and the right figure show Shannon richness and evenness (labeled).

A comparison of the alpha diversity of all samples divided into UC and CD was performed. The box plot showed that within the samples with UC and CD, there was a variety of diversity of bacteria, this was also the case for a comparison between the two diseases, see figure 3.7.

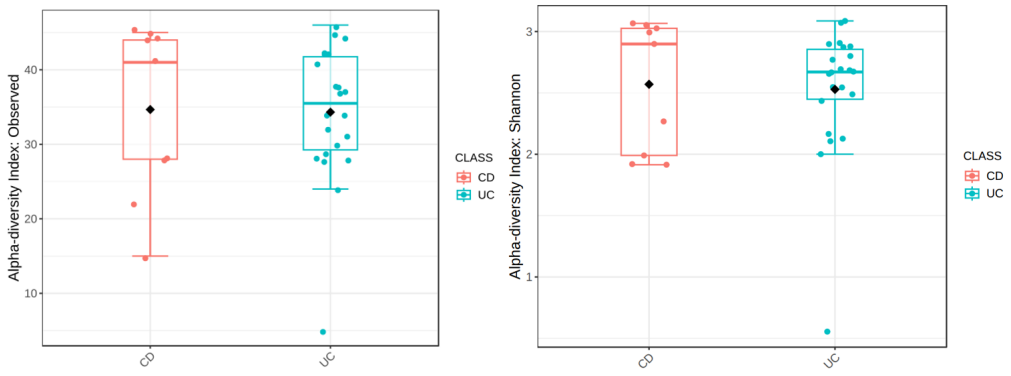


Figure 3.7: Alpha diversity of all samples divided into if the patient have UC or CD in a box plot. The figure to left shows the diversity with the observed richness (labeled), and the right figure show Shannon richness and evenness (labeled).

3.4.3 Beta diversity

Beta diversity tells how similar or different two communities/samples are by quantification [36]. Bray-Curtis index and Jaccard index are both used in the assessment of beta diversity [57]. The dataset was measured with the Bray-Curtis index and the Jaccard index. The Bray-Curtis dissimilarity values abundant species more than the rare species, it is also responsive to the difference in the amount between species [58].

The results of beta diversity divided into groups of months since diagnosis, show little dissimilarity in the samples. The dots present one sample, and many are centered around 0.0. Some individual samples show more dissimilarity, the further apart they are ($p > 0.05$, PERMANOVA). The axis percentage explains the variation in the samples. The statistical method used was PERMANOVA (permutational analysis of variance) and the results were plotted in PCoA (principal coordinate analysis) plot, see figure 3.8.

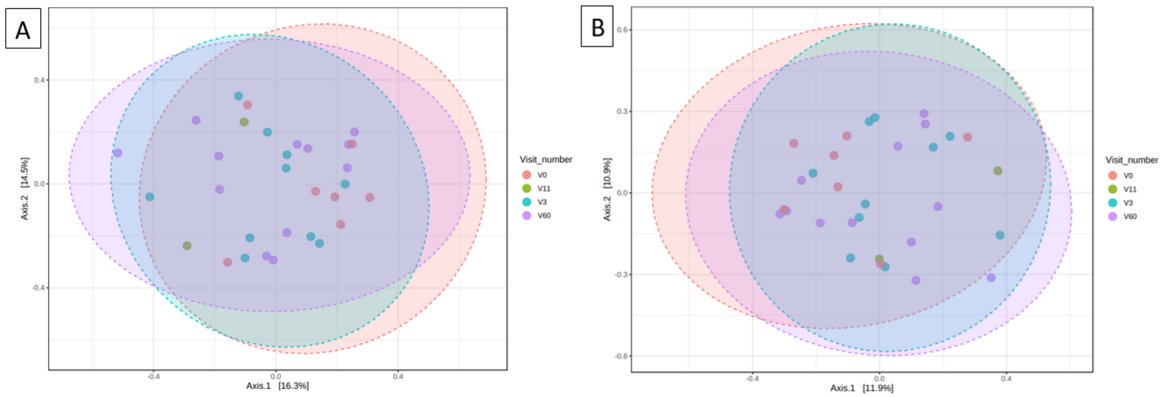


Figure 3.8: Beta diversity of all the samples, where one dot represent one sample and are divided into months since diagnosis. A) The Bray-Curtis index is used. Axis 1 explains 16.3% of the variation, while axis 2 explains 14.5% of the variation. B) The Jaccard index. Axis 1 explains 11.9% of the variation, while axis 2 explains 10.9% of the variation.

For UC and CD PCoA plot of Bray-Curtis also showed little dissimilarity, most of the samples were also centered around 0.0. The Jaccard index plot did somewhat show more dissimilarity, where the samples were more scattered from -0.4 to 0.4 ($p > 0.05$, PERMANOVA). See figure 3.9.

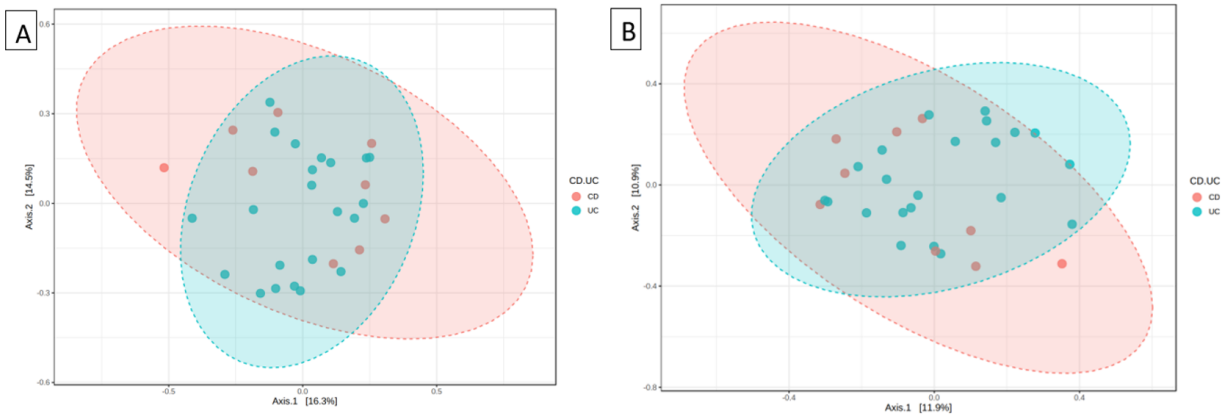


Figure 3.9: Beta diversity of all the samples, where one dot represent one sample and are divided into if the patient have UC or CD. A) The Bray-Curtis index is used. Axis 1 explains 16.3% of the variation, while axis 2 explains 14.5% of the variation. B) The Jaccard index. Axis 1 explains 11.9% of the variation, while axis 2 explains 10.9% of the variation.

3.4.4 Comparison and classification

Usually, comparison and classification of significant features are performed on datasets with a large number of samples. In this case, LefSe (Linear Discriminant Analysis Effect Size) was used, which compares classes and finds the features that may be the reason for differences among the classes [59]. Due to the low number of samples, there were no significant features that could produce such a plot. From the taxonomy plot, the bacteria phyla and genus in the samples were very alike and therefore give reason to believe there are no significant features in the samples. Single-factor analysis was also tested. The single-factor analysis is used to test one factor of a number of samples [60]. In this case, a comparison at the genus level. There were no significant features between the UC and CD groups. Between the groups divided into months since diagnosis, two bacteria were different, one was unclassified and the other one was *Alistipes*.

3.5 Gut microbiome and IBD

Analysis of the gut microbiota of a diseased gut has been accomplished in this project. The bacteria found in the samples after sequencing shows the bacteria composition of a small proportion of people diagnosed with IBD. The bacteria in the gut, which are obtained at the early stages of life, give a beneficial relationship with the host. This relationship is important for the normal development of the body and the immune system [5]. Many studies have shown that a disruption of the composition of bacteria in the gut can lead to disease.

In the paper [61], they mention reduced microbiota diversity and a higher proportion of Gram-negative bacteria than Gram-positive bacteria are more often found in patients with IBD. Even though many studies have looked at the microbes in the diseased gut, they mention that a defined interaction between the host and microbes has not been made. Metagenomic studies aim to understand the bacterial functions in IBD. The paper also looked at studies where the abundance of *F. prausnitzii* is low, is usually found in active UC and CD, and may play a role in the outbreak of IBD. Also, the phyla Bacteroidetes were shown to be more abundant than Firmicutes in both UC and CD.

Further, in [62] they also found that a decrease in bacterial diversity is consistent with microbiota change in IBD. They also mention that a reduction of the bacteria *F. prausnitzii* is associated with IBD and an increased amount of Enterobacteria and the Bacteroidetes. *F. prausnitzii* is a part of the phyla

Firmicutes, which is the most abundant bacteria in the human gut. From the sequencing results in this project, the alpha diversity showed a diversity in bacteria in the samples, but from the findings in the paper, the alpha diversity showed little bacterial diversity. The goal of this study was to look at how the role of gut microbiota in treatment response.

Further, the interest in human gut microbiota has increased, but the identification of the bacterial composition in a diseased gut has proven to be a very hard task, according to [63]. They suggest the use of fecal microbiota transplantation to look at metabolic processes in a diseased gut. They also concluded that bacteria that are usually found in small amounts in a healthy gut are dominant in inflammatory diseases. This is dysbiosis of the diversity of the gut environment. Also, in [64] they mention that dysbiosis which disrupts the balance between pathogenic and commensal bacteria, has been related to IBD. Here they look at characteristics of the microbiome found in IBD and suggest that the use of the microbiome can be useful as a treatment for IBD.

In these papers, microbial diversity is mentioned. In IBD the microbial diversity is reduced and has been linked to the disease. The bacteria phyla Enterobacteria contain *E. coli*, which is usually of low abundance in the gut, but in IBD its amount is increased [61]. An increased amount of *E. coli* may lead to pathogenesis in the disease [65]. The bacteria *F. prausnitzii* is also mentioned to be of low abundance in IBD. This bacteria produces among other anti-inflammatory molecules, that can protect against inflammatory reactions [55]. By evaluating the abundance and each bacteria in the gut, a more directed treatment can be made. The problem is that every human has an individual composition of bacteria, and one treatment is not easy to make. The importance of metagenomic study show in both the understanding of IBD, but also the cure. Assessing the gut microbiota is a huge task, but with many who research this field, sharing the results, will lead to a better understanding.

3.6 Further work

The next natural step for further work in this project would be to collect more samples. An extended trial with 1000 - 10 000 samples, spanned between different gender, culture and countries would make statistically reliable results. Healthy and diseased people should be included in the trial. The composition of bacteria in the gut should be measured under the different stages of disease to understand how the bacteria correlate to IBD. When enough data has been collected, a comparison between the gut microbiota

of one with and without IBD could be made. Hopefully, an extended trial would make it possible to paint a picture of what a healthy gut microbiota would look like, and more importantly what a diseased gut look like. A more general standard of human gut microbiota could be analyzed and set for future references.

A mapping of the gut microbiota could be the start of a more specified research for treatment that could target specific microorganisms that are abundant in the diseased gut. Questions like why and what bacteria flourishing suddenly in the gut, could hopefully be answered. Through long-read sequencing, such as PacBio, this may be seen in the future. Changing Illumina sequencing with PacBio, could be a better step in the sequencing part, because long-read sequencing have access to produce reads of 10 kilobases to megabases, hopefully revealing new information.

In the paper [3], the authors discuss various metabolic processes that may be associated with dysbiosis in IBD. It is uncertain if these metabolic changes are the cause, or consequence of the inflammatory response observed in the gut of IBD patients. They reviewed findings correlated with these metabolic processes. They suggested that investigating certain metabolites, that is found to correlate with microbial community structure, could be important because they may be associated with the disease activity. Detection of such metabolites could serve as a specific target for observing the microbial function in IBD. Lastly, they concluded that there is a huge gap between identification of the metabolites and their direct role in the disease, such there is reason for further work on this field.

Limitations to this project was the amount of time, number of samples and the uncertainty of cause of IBD. With more time, more sample could have been collected. There is no standard for healthy gut, and with the amount of samples being under 100, an extreme caution in conclusion of the results is needed [4].

The composition of human gut microbiota varies in different parts of the world, different cultures and ethnicity. Evidence prove that ethnicity has an impact on gut microbiota and the different compositions [66]. Further, a negative control that is proven to not have UC or CD, could be used to compare the microbial eukaryotes in the gut. In this project, the lack of a negative control from the sequencing, challenges a proper evaluation of the bacterial environment. Because the extracted stool samples are compared with bacteria in gut microbiota of a general standard, it becomes a source of error. An optimal comparison would be between a negative control and positive control from the same country and culture, and preferably same area of the country.

Chapter 4

Conclusion

The goal of this thesis was to learn how to assess the presence of bacterial microbes in the human gut when a person has inflammatory bowel disease. The incidence of inflammatory bowel disease has increased throughout the years. A disease normally classified under ulcerative colitis and Crohn´s disease. Hence, there is much research in this field trying to discover the reason for dysbiosis in the gut.

Methods such as DNA extraction, 16S PCR amplification, gel electrophoresis, and Illumina sequencing are used for the assessment of the microbes in the gut and are critical to achieving good quality and quantity DNA, which are further used to achieve reliable sequencing data. Through bioinformatic tools, we interpreted the data and visualized it to get a better overview of the gut. In this thesis, the DNA extraction yielded good-quality DNA in 18 (67%) of 27 samples. 7 (26%) of 27 samples did not have good-quality DNA according to the absorbance ratios. Further, 16 samples (62%) of 26 did amplify with the 16S PCR amplification, while 10 samples (38%) of 26 did not. Analysis of the sequencing data showed that Firmicutes and Bacteroidetes were most abundant in the gut of someone with UC and CD. Only one sample did diverge from this by containing almost only the Proteobacteria phyla. The most abundant bacteria genus were *Bacteroides*, followed by *Faecalibacterium* and *Blautia*. The dataset did only contain 32 samples, and therefore we cannot conclude that these results are a standard of microbial diversity in a diseased gut, but still show what may be expected to find with a larger number of samples.

Further work would be to extend the number of samples, with this small amount of samples the results cannot be statistically reliable. Improvements in the methods for DNA extraction and clean-up could improve the sample quality, and the use of long-read sequencing could make it possible to discover parts of the genome that are lost in short-read sequencing.

References

[1] L. Roncoroni, R. Gori, L. Elli, *et al.*, "Nutrition in patients with inflammatory bowel diseases: A narrative review," *Nutrients*, vol. 14, no. 4, 2022, ISSN: 2072-6643. DOI: 10.3390/nu14040751.

[2] X.-X. Yan and D. Wu, "Intestinal microecology-based treatment for inflammatory bowel disease: Progress and prospects," *World Journal of Clinical Cases*, vol. 11, no. 1, pp. 47–56, 2023, ISSN: 2307-8960. DOI: 10.12998/wjcc.v11.i1.47.

[3] K. Upadhyay, D. Desai, T. Ashavaid, and A. Dherai, "Microbiome and metabolome in inflammatory bowel disease," *Journal of Gastroenterology and Hepatology (Australia)*, vol. 38, no. 1, pp. 34–43, 2023, ISSN: 0815-9319. DOI: 10.1111/jgh.16043.

[4] C. Allaband, D. McDonald, Y. Vázquez-Baeza, *et al.*, "Microbiome 101: Studying, analyzing, and interpreting gut microbiome data for clinicians," *Clinical Gastroenterology and Hepatology*, vol. 17, no. 2, pp. 218–230, 2019, ISSN: 1542-3565. DOI: 10.1016/j.cgh.2018.09.017.

[5] M. Boirivant and A. Cossu, "Inflammatory bowel disease," *Oral Diseases*, vol. 18, no. 1, pp. 1–15, 2012, ISSN: 1601-0825. DOI: 10.1111/j.1601-0825.2011.01811.x.

[6] S. Lautenschlager, M. Barry, G. Rogler, *et al.*, "Lifestyle factors associated with inflammatory bowel disease: Data from the swiss IBD cohort study," *BMC Gastroenterology*, vol. 23, no. 71, 2023, ISSN: 1471-230X. DOI: 10.1186/s12876-023-02692-9.

[7] P. Zhang, "Influence of foods and nutrition on the gut microbiome and implications for intestinal health," *International Journal of Molecular Sciences*, vol. 23, no. 17, p. 9588, 2022, ISSN: 1422-0067. DOI: 10.3390/ijms23179588. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9455721/ (visited on 04/27/2023).

[8] M. I. El Mouzan, H. S. Winter, A. A. Assiri, *et al.*, "Microbiota profile in new-onset pediatric crohn's disease: Data from a non-western population," *Gut Pathogens*, vol. 10, no. 1, p. 49, 2018, ISSN: 1757-4749. DOI: 10.1186/s13099-018-0276-3. [Online]. Available: https://doi.org/10.1186/s13099-018-0276-3 (visited on 05/30/2023).

[9] M. T. Madigan, K. S. Bender, D. H. Buckley, W. M. Sattley, D. A. Stahl, and T. D. Brock, *Brock biology of microorganisms*, Sixteenth edition, global edition. Harlow: Pearson, 2021, 1124 pp., ISBN: 978-1-292-40479-0.

[10] C. Milani, S. Duranti, F. Bottacini, *et al.*, "The first microbial colonizers of the human gut: Composition, activities, and health implications of the infant gut microbiota," *Microbiology and Molecular Biology Reviews : MMBR*, vol. 81, no. 4, e00036–17, 2017, ISSN: 1092-2172. DOI: 10.1128/MMBR.00036-17. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5706746/ (visited on 04/25/2023).

[11] L. Rigottier-Gois, "Dysbiosis in inflammatory bowel diseases: The oxygen hypothesis," *ISME Journal*, vol. 7, no. 7, pp. 1256–1261, 2013, ISSN: 1751-7370. DOI: 10.1038/ismej.2013.80.

[12] D. Baumgart and W. Sandborn, "Inflammatory bowel disease: Clinical aspects and established and evolving therapies," *Lancet*, vol. 369, no. 9573, pp. 1641–1657, 2007, ISSN: 0140-6736. DOI: 10.1016/S0140-6736(07)60751-X.

[13] J. D. Feuerstein and A. S. Cheifetz, "Ulcerative colitis: Epidemiology, diagnosis, and management," *Mayo Clinic Proceedings*, vol. 89, no. 11, pp. 1553–1563, 2014, ISSN: 0025-6196. DOI: 10.1016/j.mayocp.2014.07.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0025619614006260 (visited on 04/26/2023).

[14] E. F. Stange, S. P. L. Travis, S. Vermeire, *et al.*, "European evidence-based consensus on the diagnosis and management of ulcerative colitis: Definitions and diagnosis," *Journal of Crohn's & Colitis*, vol. 2, no. 1, pp. 1–23, 2008, ISSN: 1873-9946. DOI: 10.1016/j.crohns.2007.11.001.

[15] S. Bandzar, S. Gupta, and M. O. Platt, "Crohn's disease: A review of treatment options and current research," *Cellular Immunology*, vol. 286, no. 1, pp. 45–52, 2013, ISSN: 0008-8749. DOI: 10.1016/j.cellimm.2013.11.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0008874913001871 (visited on 02/01/2023).

[16] A. Lavelle and H. Sokol, "Beyond metagenomics, metatranscriptomics illuminates microbiome functionality in IBD," *Nature Reviews Gastroenterology & Hepatology*, vol. 15, no. 4, pp. 193–194, 2018, Number: 4 Publisher: Nature Publishing Group, ISSN: 1759-5053. DOI: 10.1038/nrgastro.2018.15. [Online]. Available: https://www.nature.com/articles/nrgastro.2018.15 (visited on 04/26/2023).

[17] V. A. Kristensen, R. Opheim, G. Perminow, *et al.*, "Inflammatory bowel disease in south-eastern norway III (IBSEN III): A new population-based inception cohort study from south-eastern norway," *Scandinavian Journal of Gastroenterology*, vol. 56, no. 8, pp. 899–905, 2021, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00365521.2021.1922746, ISSN: 0036-5521. DOI: 10.1080/00365521.2021.1922746. [Online]. Available: https://doi.org/10.1080/00365521.2021.1922746 (visited on 04/19/2023).

[18] N. Shafiee, Z. Manaf, N. Mokhtar, and R. Ali, "Anti-inflammatory diet and inflammatory bowel disease: What clinicians and patients should know?" *Intestinal Research*, vol. 19, no. 2, pp. 171–185, 2021, ISSN: 1598-9100. DOI: 10.5217/ir.2020.00035.

[19] Y. Fan and O. Pedersen, "Gut microbiota in human metabolic health and disease," *Nature Reviews. Microbiology*, vol. 19, no. 1, pp. 55–71, 2021, ISSN: 1740-1534. DOI: 10.1038/s41579-020-0433-9.

[20] A. Marsh, J. Kinneally, T. Robertson, A. Lord, A. Young, and G. Radford –Smith, "Food avoidance in outpatients with inflammatory bowel disease – who, what and why," *Clinical Nutrition ESPEN*, vol. 31, pp. 10–16, 2019, ISSN: 2405-4577. DOI: 10.1016/j.clnesp.2019.03.018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240545771830620X (visited on 04/27/2023).

[21] K. R. Christensen, M. A. Ainsworth, M. Skougaard, *et al.*, "Identifying and understanding disease burden in patients with inflammatory bowel disease," *BMJ Open Gastroenterology*, vol. 9, no. 1, 2022, Publisher: BMJ Specialist Journals Section: Inflammatory bowel disease, ISSN: 2054-4774. DOI: 10.1136/bmjgast-2022-000994. [Online]. Available: https://bmjopengastro.bmj.com/content/9/1/e000994 (visited on 04/27/2023).

[22] T. Stroie, C. Preda, C. Meianu, *et al.*, "Fatigue is associated with anxiety and lower health-related quality of life in patients with inflammatory bowel disease in remission," *Medicina*, vol. 59, no. 3, p. 532, 2023, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1648-9144. DOI: 10.3390/medicina59030532. [Online].

Available: https://www.mdpi.com/1648-9144/59/3/532 (visited on 04/27/2023).

[23] G. Huppertz-Hauss, M. Lie Høivik, L.-P. Jelsness-Jørgensen, *et al.*, "Health-related quality of life in patients with inflammatory bowel disease 20 years after diagnosis: Results from the IBSEN study," *Inflammatory Bowel Diseases*, vol. 22, no. 7, pp. 1679–1687, 2016, ISSN: 1078-0998. DOI: 10.1097/MIB.0000000000000806.

[24] R. Ranjan, A. Rani, A. Metwally, H. S. McGee, and D. L. Perkins, "Analysis of the microbiome: Advantages of whole genome shotgun versus 16s amplicon sequencing," *Biochemical and biophysical research communications*, vol. 469, no. 4, pp. 967–977, 2016, ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2015.12.083. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4830092/ (visited on 04/26/2023).

[25] K. Fukuda, M. Ogawa, H. Taniguchi, and M. Saito, "Molecular approaches to studying microbial communities: Targeting the 16s ribosomal RNA gene," *Journal of UOEH*, vol. 38, no. 3, pp. 223–232, 2016, ISSN: 0387-821X. DOI: 10.7888/juoeh.38.223.

[26] J. E. Clarridge, "Impact of 16s rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases," *Clinical Microbiology Reviews*, vol. 17, no. 4, pp. 840–862, 2004, Publisher: American Society for Microbiology. DOI: 10.1128/CMR.17.4.840-862.2004. [Online]. Available: https://journals.asm.org/doi/full/10.1128/CMR.17.4.840-862.2004 (visited on 04/20/2023).

[27] J. Bador, B. Nicolas, A. Chapuis, *et al.*, "16s rRNA PCR on clinical specimens: Impact on diagnosis and therapeutic management," *Médecine et Maladies Infectieuses*, vol. 50, no. 1, pp. 63–73, 2020, ISSN: 0399-077X. DOI: 10.1016/j.medmal.2019.09.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0399077X18309909 (visited on 02/09/2023).

[28] S. Chakravorty, D. Helb, M. Burday, N. Connell, and D. Alland, "A detailed analysis of 16s ribosomal RNA gene segments for the diagnosis of pathogenic bacteria.," *Journal of microbiological methods*, vol. 69, no. 2, pp. 330–339, 2007, ISSN: 0167-7012. DOI: 10.1016/j.mimet.2007.02.005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2562909/ (visited on 05/28/2023).

[29] H. Heuer, M. Krsek, P. Baker, K. Smalla, and E. M. Wellington, "Analysis of actinomycete communities by specific amplification of genes encoding 16s rRNA and gel-electrophoretic separation in denaturing gradients," *Applied and Environmental Microbiology*, vol. 63, no. 8, pp. 3233–3241, 1997, ISSN: 0099-2240. DOI: 10.1128/aem.63.8.3233-3241.1997.

[30] A. Bräutigam and U. Gowik, "What can next generation sequencing do for you? next generation sequencing as a valuable tool in plant research," *Plant Biology*, vol. 12, no. 6, pp. 831–841, 2010, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1438-8677.2010.00373.x, ISSN: 1438-8677. DOI: 10.1111/j.1438-8677.2010.00373.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1438-8677.2010.00373.x (visited on 04/26/2023).

[31] G. A. Logsdon, M. R. Vollger, and E. E. Eichler, "Long-read human genome sequencing and its applications," *Nature Reviews Genetics*, vol. 21, no. 10, pp. 597–614, 2020, Number: 10 Publisher: Nature Publishing Group, ISSN: 1471-0064. DOI: 10.1038/s41576-020-0236-x. [Online]. Available: https://www.nature.com/articles/s41576-020-0236-x (visited on 05/28/2023).

[32] A. Bayat, "Bioinformatics," *BMJ : British Medical Journal*, vol. 324, no. 7344, pp. 1018–1022, 2002, ISSN: 0959-8138. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1122955/ (visited on 02/20/2023).

[33] CD-genomics. "Operational taxonomic unit (OTU) and OTU clustering - CD genomics," CD genomics. (2023), [Online]. Available: https://www.cd-genomics.com/microbioseq/operational-taxonomic-unit-otu-and-otu-clustering.html (visited on 05/30/2023).

[34] A. Gjersdal and T. H. Nätt, *Metadata*, in *Store norske leksikon*, 2023. [Online]. Available: https://snl.no/metadata (visited on 05/30/2023).

[35] A. D. Willis, "Rarefaction, alpha diversity, and statistics," *Frontiers in Microbiology*, vol. 10, 2019, ISSN: 1664-302X. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02407 (visited on 05/25/2023).

[36] X. Su, "Elucidating the beta-diversity of the microbiome: From global alignment to local alignment," *mSystems*, vol. 6, no. 4, 2021, ISSN: 2379-5077. DOI: 10.1128/mSystems.00363-21.

[37] P. I. Costea, G. Zeller, S. Sunagawa, *et al.*, "Towards standards for human fecal sample processing in metagenomic studies," *Nature Biotechnology*, vol. 35, no. 11, pp. 1069–1076, 2017, Number: 11 Publisher: Nature Publishing Group, ISSN: 1546-1696. DOI: 10.1038/nbt.3960. [Online]. Available: https://www.nature.com/articles/nbt.3960 (visited on 04/26/2023).

[38] N. Gupta, "DNA extraction and polymerase chain reaction," *Journal of Cytology*, vol. 36, no. 2, pp. 116–117, 2019, ISSN: 0970-9371. DOI: 10.4103/JOC.JOC_110_18. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6425773/ (visited on 03/06/2023).

[39] Zymo-Researcher. "Genomic DNA clean & concentrator-10," Zymo Research International. (2023), [Online]. Available: https://zymoresearch.eu/products/genomic-dna-clean-concentrator-10 (visited on 05/28/2023).

[40] E. Jacobsen, *Elektroforese*, in *Store norske leksikon*, 2022. [Online]. Available: https://snl.no/elektroforese (visited on 06/09/2023).

[41] Illumina. "Next-generation sequencing (NGS) — explore the technology," Illumina. (2023), [Online]. Available: https://www.illumina.com/science/technology/next-generation-sequencing.html (visited on 05/28/2023).

[42] E. Kia, B. MacKenzie, D. Middleton, *et al.*, "Integrity of the human faecal microbiota following long-term sample storage," *PLoS ONE*, vol. 11, no. 10, 2016, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0163666.

[43] L. Aabakken and B. A. Waaler, *Fordøyelsessystemet*, in *Store medisinske leksikon*, Store medisinske leksikon, 2020. [Online]. Available: https://sml.snl.no/ford%C3%B8yelsessystemet (visited on 05/24/2023).

[44] G. Lucena-Aguilar, A. M. Sánchez-López, C. Barberán-Aceituno, J. A. Carrillo-Ávila, J. A. López-Guerrero, and R. Aguilar-Quesada, "DNA source selection for downstream applications based on DNA quality indicators analysis," *Biopreservation and Biobanking*, vol. 14, no. 4, pp. 264–270, 2016, ISSN: 1947-5535. DOI: 10.1089/bio.2015.0064. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4991598/ (visited on 05/22/2023).

[45] M. Ariefdjohan, D. Savaiano, and C. Nakatsu, "Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens," *Nutrition Journal*, vol. 9, no. 1, 2010, ISSN: 1475-2891. DOI: 10.1186/1475-2891-9-23.

[46] F. Yang, J. Sun, H. Luo, *et al.*, "Assessment of fecal DNA extraction protocols for metagenomic studies," *GigaScience*, vol. 9, no. 7, 2020, ISSN: 2047-217X. DOI: 10.1093/gigascience/giaa071.

[47] K. Fiedorová, M. Radvanský, E. Němcová, *et al.*, "The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery," *Frontiers in Physiology*, vol. 10, 2019, ISSN: 1664-042X. DOI: 10.3389/fmicb.2019.00821.

[48] Qiagen. "DNA and PCR cleanup — DNA sample concentration — QIAGEN," Qiagen. (2023), [Online]. Available: https://www.qiagen.com/us/product-categories/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up (visited on 05/26/2023).

[49] M. Green and J. Sambrook, "Precipitation of DNA with ethanol," *Cold Spring Harbor Protocols*, vol. 2016, no. 12, pp. 1116–1120, 2016, ISSN: 1940-3402. DOI: 10.1101/pdb.prot093377.

[50] Bitesizebio. "DNA clean-up: 5 easy tips," Bitesizebio. Section: DNA / RNA Manipulation and Analysis. (2021), [Online]. Available: https://bitesizebio.com/142/5-ways-to-clean-up-a-dna-sample/ (visited on 05/26/2023).

[51] Bio-Rad. "PCR troubleshooting — bio-rad," Bio-Rad. (2023), [Online]. Available: https://www.bio-rad.com/en-no/applications-technologies/pcr-troubleshooting?ID=LUSO3HC4S (visited on 05/26/2023).

[52] S. Amarasinghe, S. Su, X. Dong, L. Zappia, M. Ritchie, and Q. Gouil, "Opportunities and challenges in long-read sequencing data analysis," *Genome Biology*, vol. 21, no. 1, 2020, ISSN: 1474-7596. DOI: 10.1186/s13059-020-1935-5.

[53] G. Rizzatti, L. R. Lopetuso, G. Gibiino, C. Binda, and A. Gasbarrini, "Proteobacteria: A common factor in human diseases," *BioMed Research International*, vol. 2017, 2017, ISSN: 2314-6133. DOI: 10.1155/2017/9351507. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5688358/ (visited on 05/30/2023).

[54] H. M. Wexler, "Bacteroides: The good, the bad, and the nitty-gritty," *Clinical Microbiology Reviews*, vol. 20, no. 4, pp. 593–621, 2007, ISSN: 0893-8512. DOI: 10.1128/CMR.00008-07. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2176045/ (visited on 06/13/2023).

[55] Y. Cao, J. Shen, and Z. Ran, "Association between faecalibacterium prausnitzii reduction and inflammatory bowel disease: A meta-analysis and systematic review of the literature," *Gastroenterology Research and Practice*, vol. 2014, 2014, ISSN: 1687-6121. DOI: 10.1155/2014/872725.

[56] X. Liu, B. Mao, J. Gu, *et al.*, "Blautia—a new functional genus with potential probiotic properties?" *Gut Microbes*, vol. 13, no. 1, 2021, ISSN: 1949-0976. DOI: 10.1080/19490976.2021.1875796. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7872077/ (visited on 06/13/2023).

[57] Z. Yang, F. Xu, H. Li, and Y. He, "Beyond samples: A metric revealing more connections of gut microbiota between individuals," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 3930–3937, 2021, ISSN: 2001-0370. DOI: 10.1016/j.csbj.2021.07.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2001037021002993 (visited on 05/30/2023).

[58] C. Ricotta and J. Podani, "On some properties of the bray-curtis dissimilarity and their ecological meaning," *Ecological Complexity*, vol. 31, pp. 201–205, 2017, ISSN: 1476-945X. DOI: 10.1016/j.ecocom.2017.07.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1476945X17300582 (visited on 05/30/2023).

[59] N. Segata, J. Izard, L. Waldron, *et al.*, "Metagenomic biomarker discovery and explanation," *Genome Biology*, vol. 12, no. 6, R60, 2011, ISSN: 1465-6906. DOI: 10.1186/gb-2011-12-6-r60. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3218848/ (visited on 05/30/2023).

[60] D. Selvamuthu and D. Das, "Single-factor experimental design," [Abstract], in *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*, D. Selvamuthu and D. Das, Eds., Singapore: Springer, 2018, pp. 223–264, ISBN: 9789811317361. DOI: 10.1007/978-981-13-1736-1_7. [Online]. Available: https://doi.org/10.1007/978-981-13-1736-1_7 (visited on 06/13/2023).

[61] G. Loh and M. Blaut, "Role of commensal gut bacteria in inflammatory bowel diseases," *Gut Microbes*, vol. 3, no. 6, pp. 544–555, 2012, ISSN: 1949-0976. DOI: 10.4161/gmic.22156. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3495792/ (visited on 06/13/2023).

[62] C. Mah, T. Jayawardana, G. Leong, *et al.*, "Assessing the relationship between the gut microbiota and inflammatory bowel disease therapeutics: A systematic review," *Pathogens*, vol. 12, no. 2, p. 262, 2023, ISSN: 2076-0817. DOI: 10.3390/pathogens12020262. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9965214/ (visited on 06/13/2023).

[63] A. R. Watson, J. Füssel, I. Veseli, *et al.*, "Metabolic independence drives gut microbial colonization and resilience in health and disease," *Genome Biology*, vol. 24, no. 1, p. 78, 2023, ISSN: 1474-760X. DOI: 10.1186/s13059-023-02924-x. [Online]. Available: https://doi.org/10.1186/s13059-023-02924-x (visited on 06/13/2023).

[64] Y.-H. Shen, H. Zhu, L. Zhou, *et al.*, "In inflammatory bowel disease and extraintestinal manifestations: What role does microbiome play?" *Engineered Regeneration*, vol. 4, no. 4, pp. 337–348, 2023, ISSN: 2666-1381. DOI: 10.1016/j.engreg.2023.04.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S266613812300035X (visited on 06/13/2023).

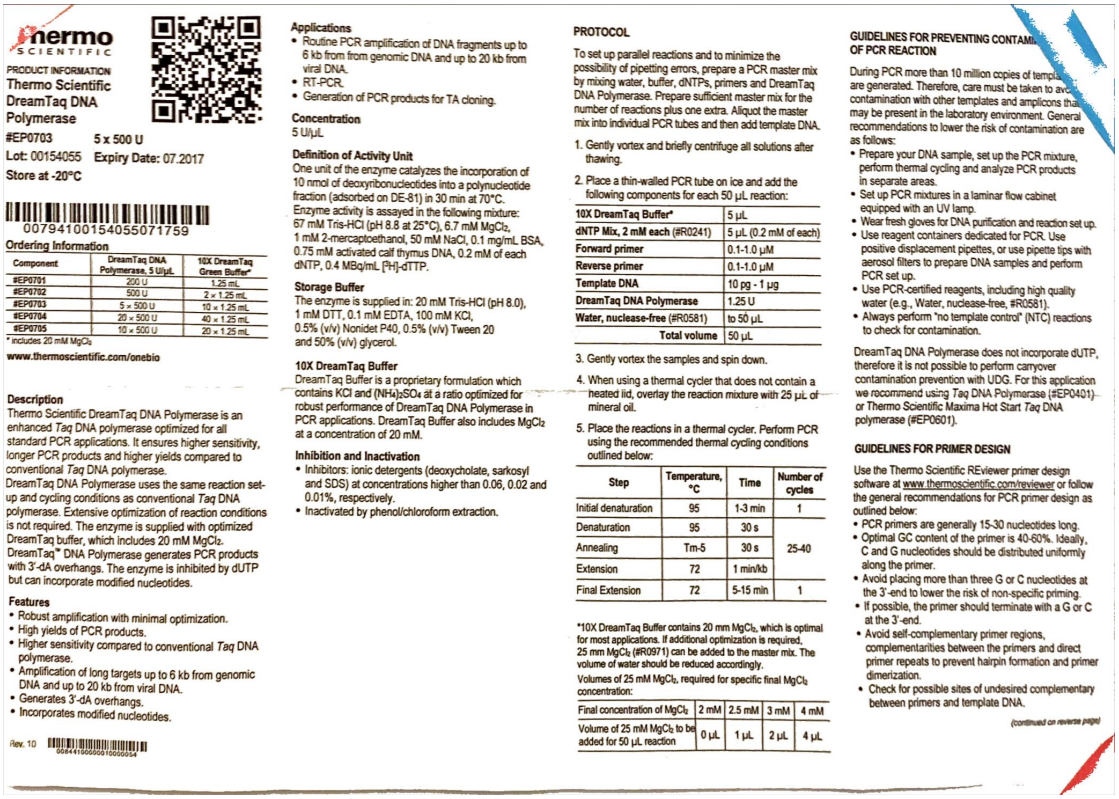
[65] C. Palmela, C. Chevarin, Z. Xu, *et al.*, "Adherent-invasive escherichia coli in inflammatory bowel disease," *Gut*, vol. 67, no. 3, pp. 574–587, 2018, Publisher: BMJ Publishing Group Section: Recent advances in clinical practice, ISSN: 0017-5749, 1468-3288. DOI: 10.1136/gutjnl-2017-314903. [Online]. Available: https://gut.bmj.com/content/67/3/574 (visited on 06/14/2023).

[66] J. Dwiyanto, M. Hussain, D. Reidpath, *et al.*, "Ethnicity influences the gut microbiota of individuals sharing a geographical location: A cross-sectional study from a middle-income country," *Scientific Reports*, vol. 11, no. 1, 2021, ISSN: 2045-2322. DOI: 10.1038/s41598-021-82311-3.

Appendix A

Appendix A

A.1 Protocol from Thermo Fisher Scientific: PCR Master Mix



PRODUCT INFORMATION

Thermo Scientific DreamTaq DNA Polymerase

#EP0703 **5 x 500 U**

Lot: 00154055 **Expiry Date:** 07.2017

Store at -20°C

0079410015405507175 9

Ordering Information

Component	DreamTaq DNA Polymerase, 5 U/µL	10X DreamTaq Green Buffer*
#EP0701	200 U	1.25 mL
#EP0702	500 U	2 × 1.25 mL
#EP0703	5 × 500 U	10 × 1.25 mL
#EP0704	20 × 500 U	40 × 1.25 mL
#EP0705	10 × 500 U	20 × 1.25 mL

* includes 20 mM $MgCl_2$

www.thermoscientific.com/onebio

Description

Thermo Scientific DreamTaq DNA Polymerase is an enhanced *Taq* DNA polymerase optimized for all standard PCR applications. It ensures higher sensitivity, longer PCR products and higher yields compared to conventional *Taq* DNA polymerase.

DreamTaq DNA Polymerase uses the same reaction set-up and cycling conditions as conventional *Taq* DNA polymerase. Extensive optimization of reaction conditions is not required. The enzyme is supplied with optimized DreamTaq buffer, which includes 20 mM $MgCl_2$.

DreamTaq™ DNA Polymerase generates PCR products with 3'-dA overhangs. The enzyme is inhibited by dUTP but can incorporate modified nucleotides.

Features

- Robust amplification with minimal optimization.
- High yields of PCR products.
- Higher sensitivity compared to conventional *Taq* DNA polymerase.
- Amplification of long targets up to 6 kb from genomic DNA and up to 20 kb from viral DNA.
- Generates 3'-dA overhangs.
- Incorporates modified nucleotides.

Rev. 10

Applications

- Routine PCR amplification of DNA fragments up to 6 kb from from genomic DNA and up to 20 kb from viral DNA.
- RT-PCR.
- Generation of PCR products for TA cloning.

Concentration

5 U/µL

Definition of Activity Unit

One unit of the enzyme catalyzes the incorporation of 10 nmol of deoxyribonucleotides into a polynucleotide fraction (adsorbed on DE-81) in 30 min at 70°C. Enzyme activity is assayed in the following mixture: 67 mM Tris-HCl (pH 8.8 at 25°C), 6.7 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 50 mM NaCl, 0.1 mg/mL BSA, 0.75 mM activated calf thymus DNA, 0.2 mM of each dNTP, 0.4 MBq/mL [^{3}H]-dTTP.

Storage Buffer

The enzyme is supplied in: 20 mM Tris-HCl (pH 8.0), 1 mM DTT, 0.1 mM EDTA, 100 mM KCl, 0.5% (v/v) Nonidet P40, 0.5% (v/v) Tween 20 and 50% (v/v) glycerol.

10X DreamTaq Buffer

DreamTaq Buffer is a proprietary formulation which contains KCl and $(NH_4)_2SO_4$ at a ratio optimized for robust performance of DreamTaq DNA Polymerase in PCR applications. DreamTaq Buffer also includes $MgCl_2$ at a concentration of 20 mM.

Inhibition and Inactivation

- Inhibitors: ionic detergents (deoxycholate, sarkosyl and SDS) at concentrations higher than 0.06, 0.02 and 0.01%, respectively.
- Inactivated by phenol/chloroform extraction.

PROTOCOL

To set up parallel reactions and to minimize the possibility of pipetting errors, prepare a PCR master mix by mixing water, buffer, dNTPs, primers and DreamTaq DNA Polymerase. Prepare sufficient master mix for the number of reactions plus one extra. Aliquot the master mix into individual PCR tubes and then add template DNA.

1. Gently vortex and briefly centrifuge all solutions after thawing.
2. Place a thin-walled PCR tube on ice and add the following components for each 50 µL reaction:

10X DreamTaq Buffer*	5 µL
dNTP Mix, 2 mM each (#R0241)	5 µL (0.2 mM of each)
Forward primer	0.1-1.0 µM
Reverse primer	0.1-1.0 µM
Template DNA	10 pg - 1 µg
DreamTaq DNA Polymerase	1.25 U
Water, nuclease-free (#R0581)	to 50 µL
Total volume	50 µL

3. Gently vortex the samples and spin down.
4. When using a thermal cycler that does not contain a heated lid, overlay the reaction mixture with 25 µL of mineral oil.
5. Place the reactions in a thermal cycler. Perform PCR using the recommended thermal cycling conditions outlined below:

Step	Temperature, °C	Time	Number of cycles
Initial denaturation	95	1-3 min	1
Denaturation	95	30 s	25-40
Annealing	Tm-5	30 s	
Extension	72	1 min/kb	
Final Extension	72	5-15 min	1

*10X DreamTaq Buffer contains 20 mm $MgCl_2$, which is optimal for most applications. If additional optimization is required, 25 mm $MgCl_2$ (#R0971) can be added to the master mix. The volume of water should be reduced accordingly.

Volumes of 25 mM $MgCl_2$, required for specific final $MgCl_2$ concentration:

Final concentration of $MgCl_2$	2 mM	2.5 mM	3 mM	4 mM
Volume of 25 mM $MgCl_2$ to be added for 50 µL reaction	0 µL	1 µL	2 µL	4 µL

GUIDELINES FOR PREVENTING CONTAMI... OF PCR REACTION

During PCR more than 10 million copies of templa... are generated. Therefore, care must be taken to avo... contamination with other templates and amplicons tha... may be present in the laboratory environment. General recommendations to lower the risk of contamination are as follows:

- Prepare your DNA sample, set up the PCR mixture, perform thermal cycling and analyze PCR products in separate areas.
- Set up PCR mixtures in a laminar flow cabinet equipped with an UV lamp.
- Wear fresh gloves for DNA purification and reaction set up.
- Use reagent containers dedicated for PCR. Use positive displacement pipettes, or use pipette tips with aerosol filters to prepare DNA samples and perform PCR set up.
- Use PCR-certified reagents, including high quality water (e.g., Water, nuclease-free, #R0581).
- Always perform "no template control" (NTC) reactions to check for contamination.

DreamTaq DNA Polymerase does not incorporate dUTP, therefore it is not possible to perform carryover contamination prevention with UDG. For this application we recommend using *Taq* DNA Polymerase (#EP0401) or Thermo Scientific Maxima Hot Start *Taq* DNA polymerase (#EP0601).

GUIDELINES FOR PRIMER DESIGN

Use the Thermo Scientific REviewer primer design software at www.thermoscientific.com/reviewer or follow the general recommendations for PCR primer design as outlined below:

- PCR primers are generally 15-30 nucleotides long.
- Optimal GC content of the primer is 40-60%. Ideally, C and G nucleotides should be distributed uniformly along the primer.
- Avoid placing more than three G or C nucleotides at the 3'-end to lower the risk of non-specific priming.
- If possible, the primer should terminate with a G or C at the 3'-end.
- Avoid self-complementary primer regions, complementarities between the primers and direct primer repeats to prevent hairpin formation and primer dimerization.
- Check for possible sites of undesired complementary between primers and template DNA.

(continued on reverse page)

A.2 Protocol from Qiagen QIAmp DNA stool kit: IHMS DNA extraction protocol Q

IHMS DNA extraction protocol Q

Fecal DNA extraction with the use of Qiagen QIAamp DNA stool kit

1. Homogenize the 150 to 200mg frozen feces with 1.0mL ASL lysis buffer of the kit by vortexing for 2min in a 2mL tube containing 0.3g of sterile zirconia beads Ø 0,1mm zirconia (BioSpec, Cat. No. 11079101z). [if buffer shows precipitate, heat at 70°C before use]
2. Incubate for 15min at 95°C.
3. Cells are mechanically lysed by running the Fastprep™ Instrument for 8min15sec (series of beating 1 min and resting 5 min are preferable).
4. Samples are allowed to cool down on ice for 2min.
5. Samples are centrifuged at 16000 x g, 4°C, for 5min.
6. Supernatant is transferred to a new 2mL tube.
7. The pellet is mixed with 300µL ASL lysis buffer of the kit, and steps 2-5 are repeated.
8. Supernatants are pooled in the new 2mL tube.

9. Add 260µl of 10M ammonium acetate to each lysate tube, mix well, and incubate on ice for 5 min.
10. Centrifuge at 16000 g, 4°C, for 10min.
11. Transfer the supernatant to two 1.5mL Eppendorf tubes, add one volume of isopropanol, mix well, and incubate on ice for 30 min.
12. Centrifuge at 16000 g, 4°C, 15min, remove the supernatant using aspiration, wash nucleic acids pellet with 70 % EtOH (0,5mL) and dry the pellet under vacuum for 3min.
13. Dissolve the nucleic acid pellet in 100µL of TE (Tris-EDTA) buffer and pool the two aliquots.
14. Add 2µL of DNase-free RNase (10mg/mL) and incubate at 37°C, 15 min.
15. Add 15µL proteinase K and 200µL AL buffer to the supernatant, vortex for 15sec and incubate at 70°C for 10 min.
16. Add 200µL of ethanol (96-100%) to the lysate, and mix by vortexing.
17. Transfer to a QIAamp spin column and centrifuge at 16000 g for 1min, at Room Temperature (RT).
18. Discard flow through, add 500µL buffer AW1 (Qiagen) and centrifuge at 16000 g for 1min, at RT.
19. Discard flow through, add 500µL buffer AW2 (Qiagen) and centrifuge at 16000 g for 1min, at RT
20. Dry the column by centrifugation at RT for 1min.
21. Add 200µL Buffer AE (Qiagen), incubate for 1min at RT
22. Centrifuge for 1min at 16000 g to elute DNA.

Quality control: use 1% agarose gel
Sample concentration: use Nanodrop or Qubit

Nature Biotechnology: doi:10.1038/nbt.3960

A.3 Protocol from Zymo Researcher: DNA clean

Protocol

Buffer Preparation

✓ *Before starting*: Add 24 ml 100% ethanol (26 ml 95% ethanol) to the 6 ml **DNA Wash Buffer** concentrate. Add 96 ml 100% ethanol (104 ml 95% ethanol) to the 24 ml **DNA Wash Buffer** concentrate.

Sample Processing

All centrifugation steps should be performed between 10,000 - 16,000 x g.

1. In a 1.5 ml microcentrifuge tube, add 2-5 volumes of **ChIP DNA Binding Buffer** to each volume of DNA sample[1] (see table below). Mix thoroughly.

Application	DNA Binding Buffer : Sample	Example
Plasmid, genomic DNA (>2 kb)	2 : 1	200 µl : 100 µl
PCR product, DNA fragment	5 : 1	500 µl : 100 µl

2. Transfer mixture to a provided **Zymo-Spin™ IIC-XLR Column**[2] in a **Collection Tube**.

3. Centrifuge for 30 seconds. Discard the flow-through.

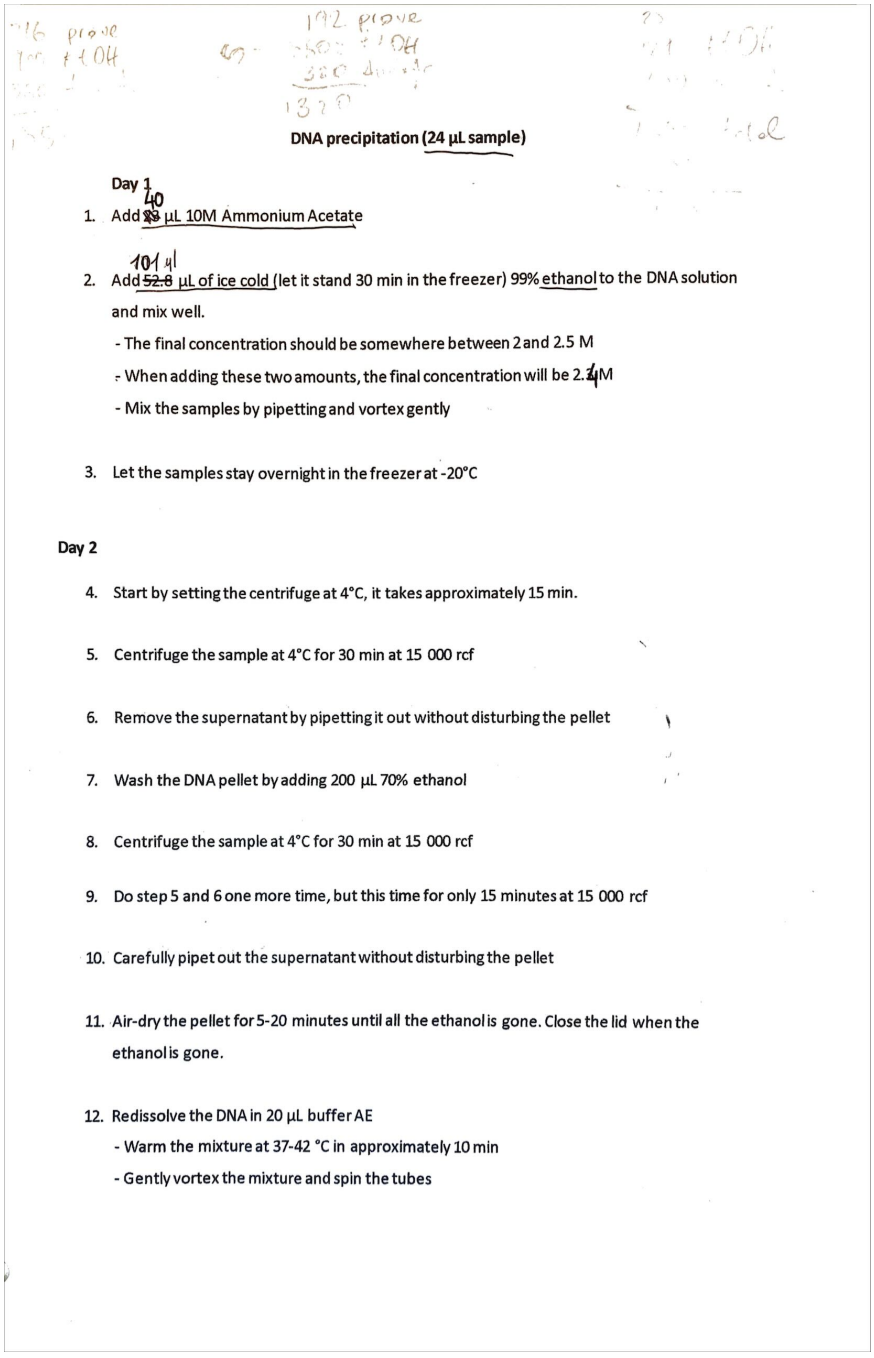
4. Add 400 µl **DNA Wash Buffer** to the column. Centrifuge for 1 minute. Empty the **Collection Tube**. Repeat the wash step.

5. Transfer the column to a 1.5 ml microcentrifuge tube. Add ≥ 50 µl[3] **DNA Elution Buffer**[4] or water[5] directly to the column matrix and incubate at room temperature for five minutes and centrifuge for 30 seconds to elute the DNA.

   Ultra-pure DNA is now ready for use.

[1] It may be necessary to add RNase A to cell lysates prior to performing the procedure to ensure RNA-free DNA will be recovered in Step 5.
[2] The sample capacity of the column is 900 ul. It may be necessary to load and spin a column multiple times if a sample has a volume larger than 900 ul.
[3] To increase concentrations of first elution, use ≥ 35 µl **DNA Elution Buffer**. To increase yield, load the eluate directly on the matrix a second time, incubate for 3 minutes at room temperature, and centrifuge again.
[4] **DNA Elution Buffer**: 10mM Tris-HCl, pH 8.5, 0.1mM EDTA
[5] Elution of DNA from the column is dependent on pH and temperature. If water is used, make sure the pH is >6.0. The total yield may be improved by eluting the DNA with 60-70°C **DNA Elution Buffer**.

A.4 Protocol for DNA precipitation: Ammonium acetate



DNA precipitation (24 µL sample)

Day 1

1. Add ~~20~~ 40 µL 10M Ammonium Acetate
2. Add ~~52.8~~ 101 µl µL of ice cold (let it stand 30 min in the freezer) 99% ethanol to the DNA solution and mix well.
   - The final concentration should be somewhere between 2 and 2.5 M
   - When adding these two amounts, the final concentration will be 2.4M
   - Mix the samples by pipetting and vortex gently
3. Let the samples stay overnight in the freezer at -20°C

Day 2

4. Start by setting the centrifuge at 4°C, it takes approximately 15 min.
5. Centrifuge the sample at 4°C for 30 min at 15 000 rcf
6. Remove the supernatant by pipetting it out without disturbing the pellet
7. Wash the DNA pellet by adding 200 µL 70% ethanol
8. Centrifuge the sample at 4°C for 30 min at 15 000 rcf
9. Do step 5 and 6 one more time, but this time for only 15 minutes at 15 000 rcf
10. Carefully pipet out the supernatant without disturbing the pellet
11. Air-dry the pellet for 5-20 minutes until all the ethanol is gone. Close the lid when the ethanol is gone.
12. Redissolve the DNA in 20 µL buffer AE
    - Warm the mixture at 37-42 °C in approximately 10 min
    - Gently vortex the mixture and spin the tubes

Appendix B

Appendix B

Overview of gels in this project. The gels are sorted into four groups, depending on when the samples were extracted. After each extraction, PCR amplification and gel verification was performed on the samples.

B.1 Gel verification of 10 samples

Table B.1: The samples connected to the gels in figure B.1 and B.2. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months
1	289 v60
2	294 v60
3	450 v11
4	468 v3
5	295 v60
6	299 v60
7	315 v60
8	455 v11
9	469 v3
10	474 v3
11	Positive control
12	Negative control

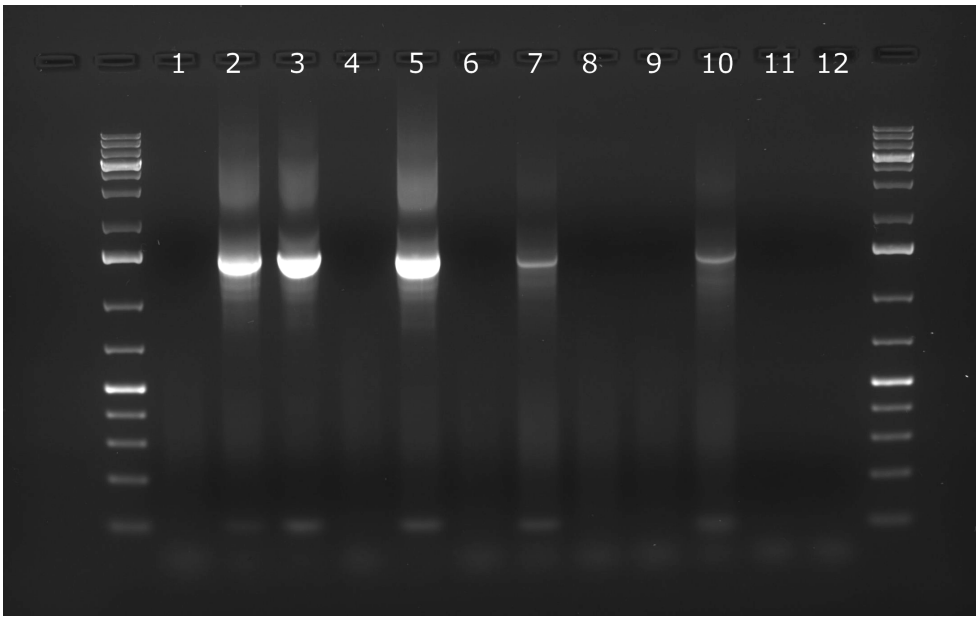


Figure B.1: Samples from 14.02.23. The positive control did not work, well 11.

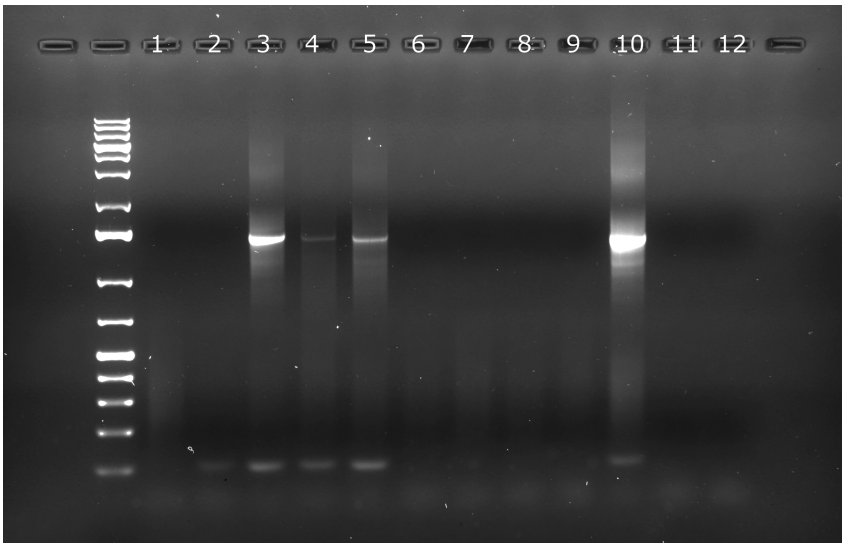


Figure B.2: Diluted amples from 22.02.23, diluted to 40 ng/µl. The positive control did not work, well 11.

B.2 Gel verification of 8 samples

Table B.2: The samples connected to the gel in figure B.3. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months
1	289 v60
2	307 v60
3	311 v60
4	316 v60
5	321 v60
6	323 v60
7	465 v3
8	466 v3
9	Positive control
10	Negative control

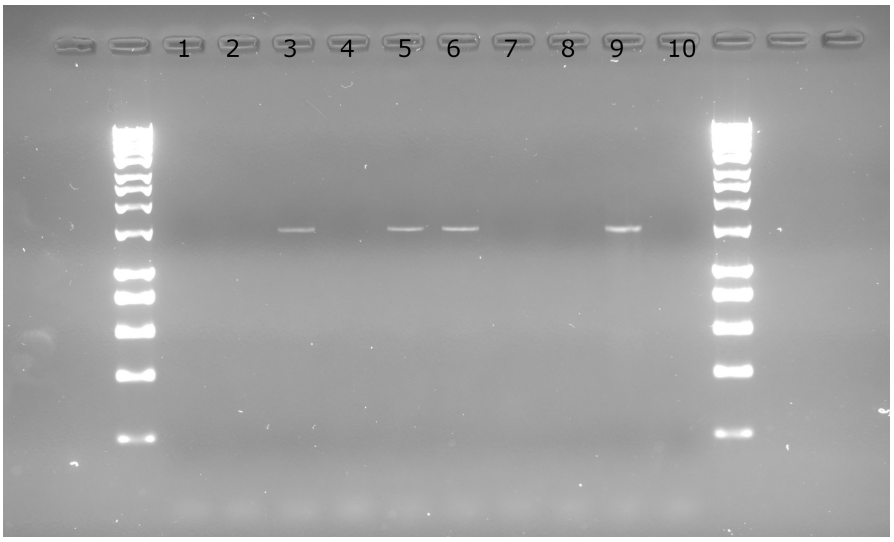


Figure B.3: Samples from 10.03.23.

Table B.3: The samples connected to the gel in figure B.4. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months
1	289 v60
2	307 v60
3	311 v60
4	316 v60
5	321 v60
6	323 v60
7	465 v3
8	466 v3
9	303 v60
10	304 v60
11	317 v60
12	Positive control
13	Negative control

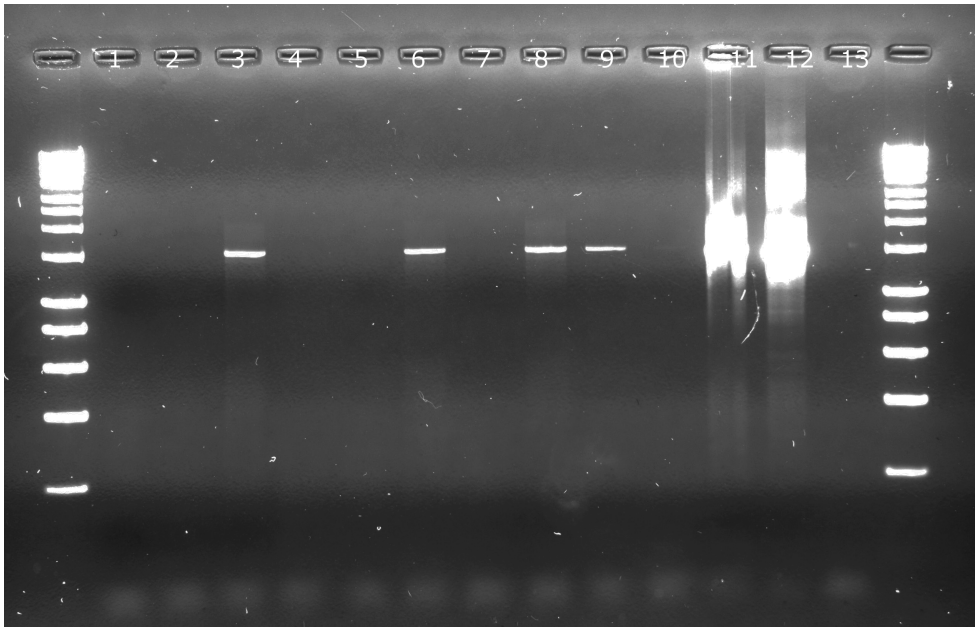


Figure B.4: Samples from 14.03.23.

B.3 Gel verification of 8 samples

Table B.4: The samples connected to the gels in figure B.5 and B.6. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months
1	303 v60
2	304 v60
3	317 v60
4	470 v3
5	471 v0
6	473 v0
7	475 v0
8	476 v0
9	Positive control
10	Negative control

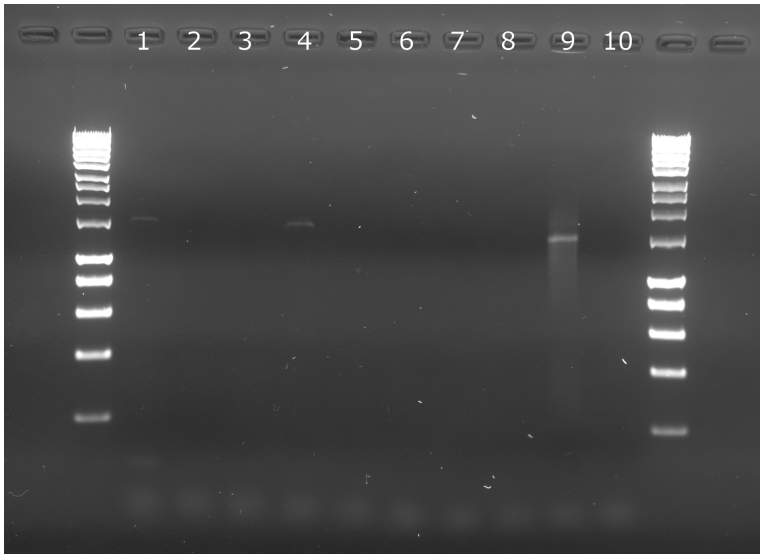


Figure B.5: Samples from 21.03.23. These samples were diluted to 40 ng/µl.

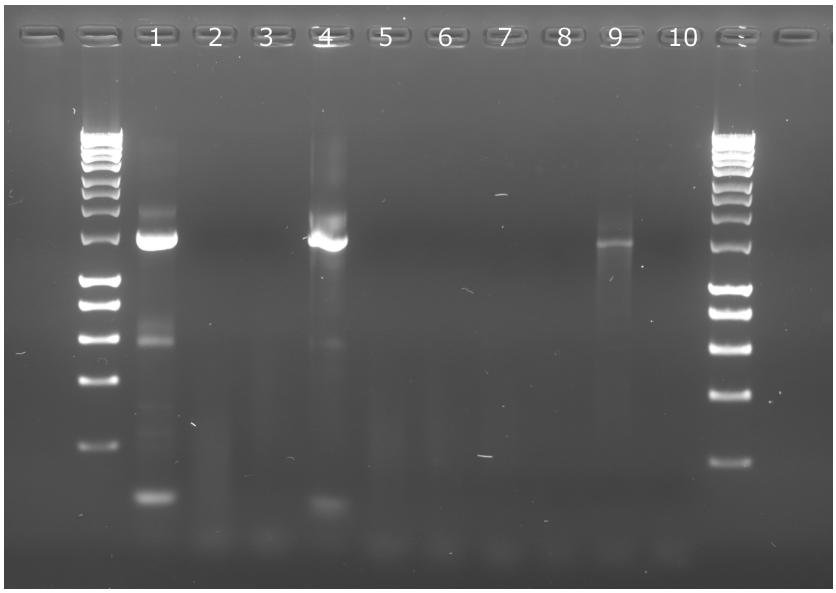


Figure B.6: The same samples as above, but this time it was not diluted.

B.4 Gel verification of negative samples

Table B.5: The samples connected to the gels in figure B.7, B.8, B.9 and B.10. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months
1	246 v60
2	298 v60
3	440 v3
4	470 v3
5	NC 1
6	NC 2
7	Positive control
8	Negative control

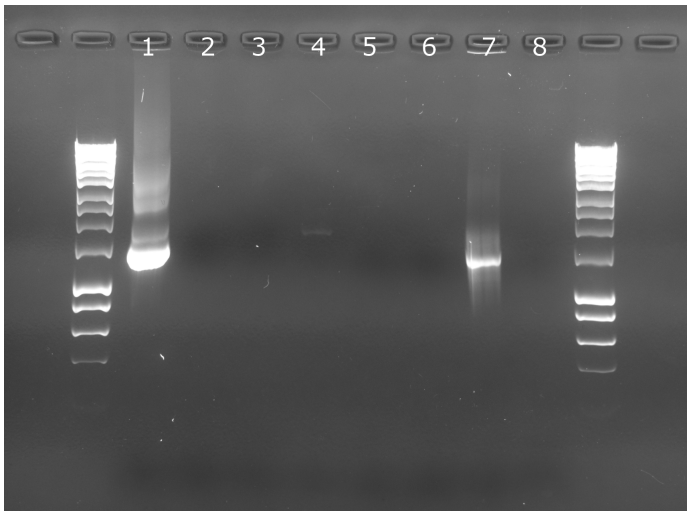


Figure B.7: Diluted samples from 28.03.23. The first four samples in this gel, is not used in this thesis at all. The last two are NC 1 and NC 2.

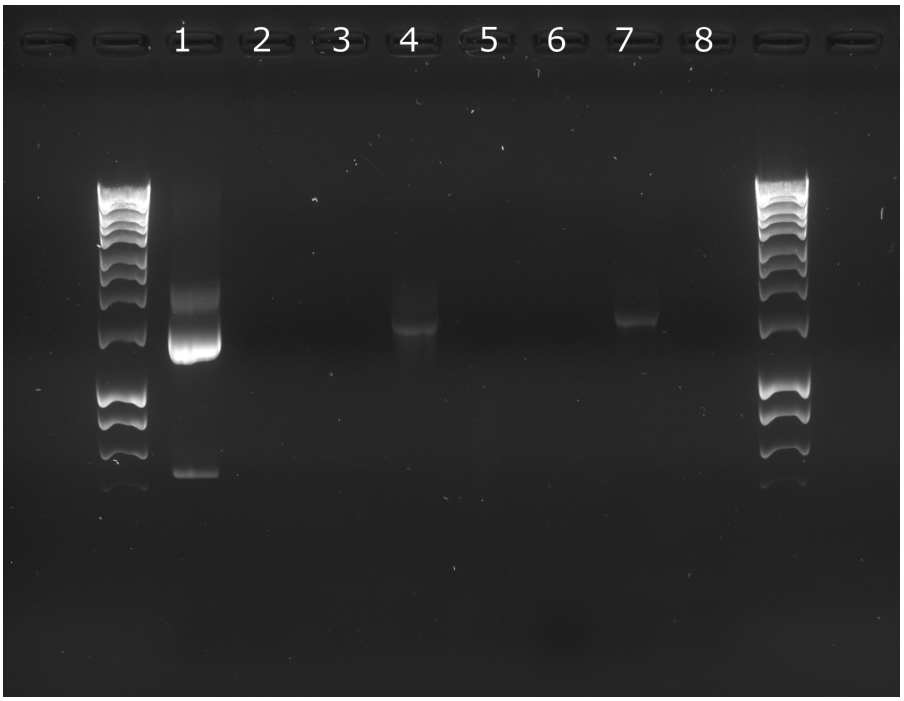


Figure B.8: These are the same samples as above, but not diluted.

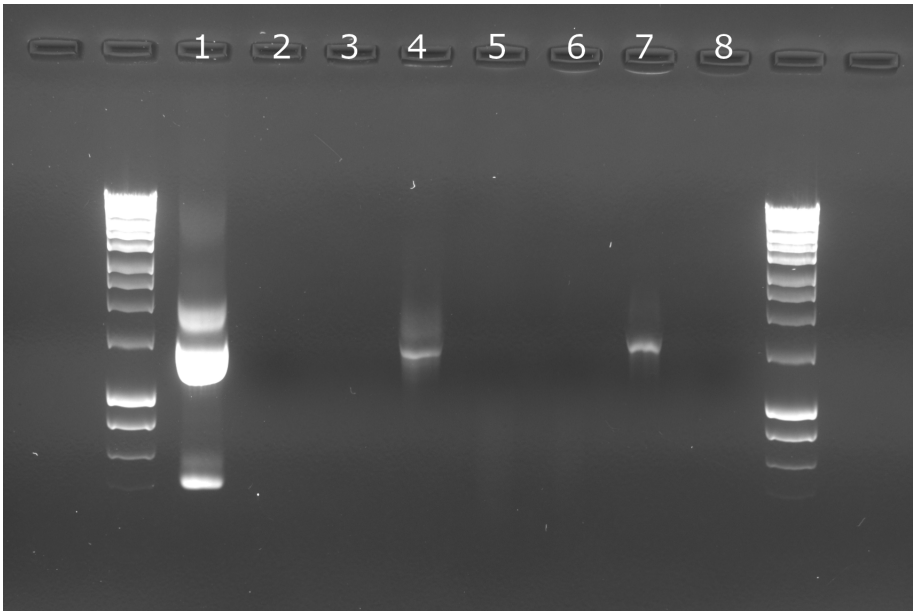


Figure B.9: Samples from 29.03.23.

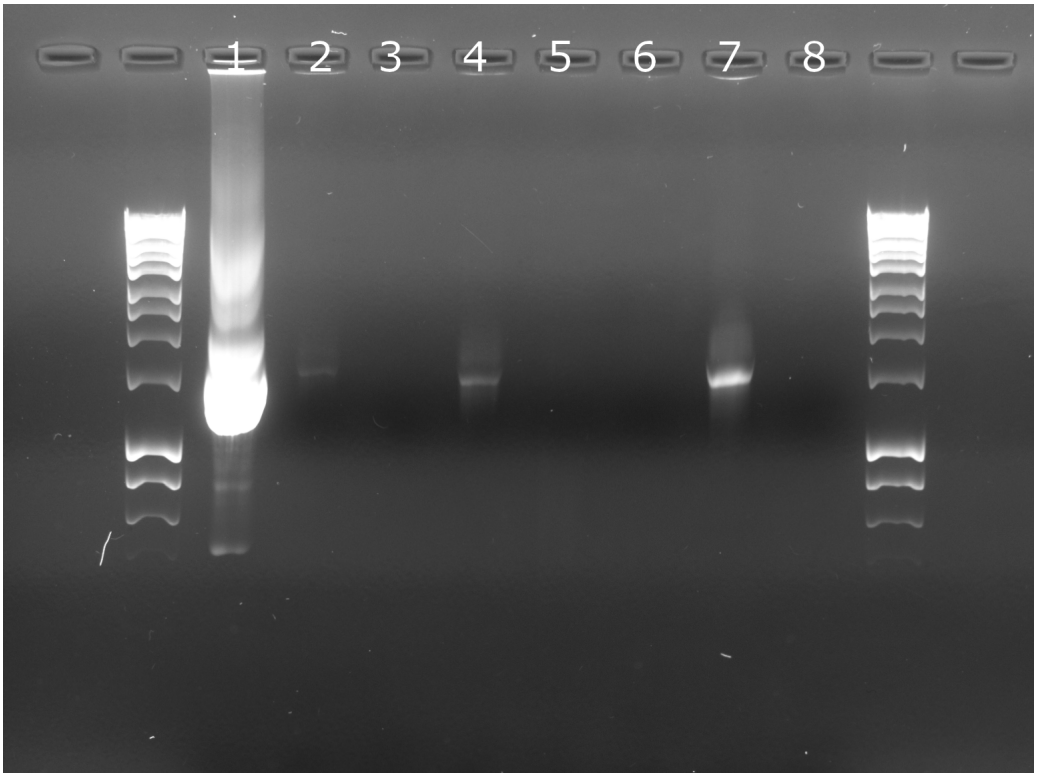


Figure B.10: Samples from 30.03.23.

Appendix C

Appendix C

C.1 Overview of all samples used in the thesis

Table C.1: Overview of the samples used in this project. "No samples" is short for number of samples. Patient number/months is since last visit.

No samples	Patient number/months	No samples	Patient number/months
1	246 v60	16	440 v3
2	289 v60	17	450 v11
3	294 v60	18	455 v11
4	295 v60	19	465 v3
5	298 v60	20	466 v3
6	299 v60	21	468 v3
7	303 v60	22	469 v3
8	304 v60	23	470 v3
9	307 v60	24	471 v0
10	311 v60	25	473 v0
11	315 v60	26	474 v3
12	316 v60	27	475 v0
13	317 v60	28	476 v0
14	321 v60	29	Negative control 1
15	323 v60	30	Negative control 2

Appendix D

Appendix D

Taxonomy bar chart of the bacteria found in 32 samples, divided into the phylogenetic levels: class, order, family, and species.

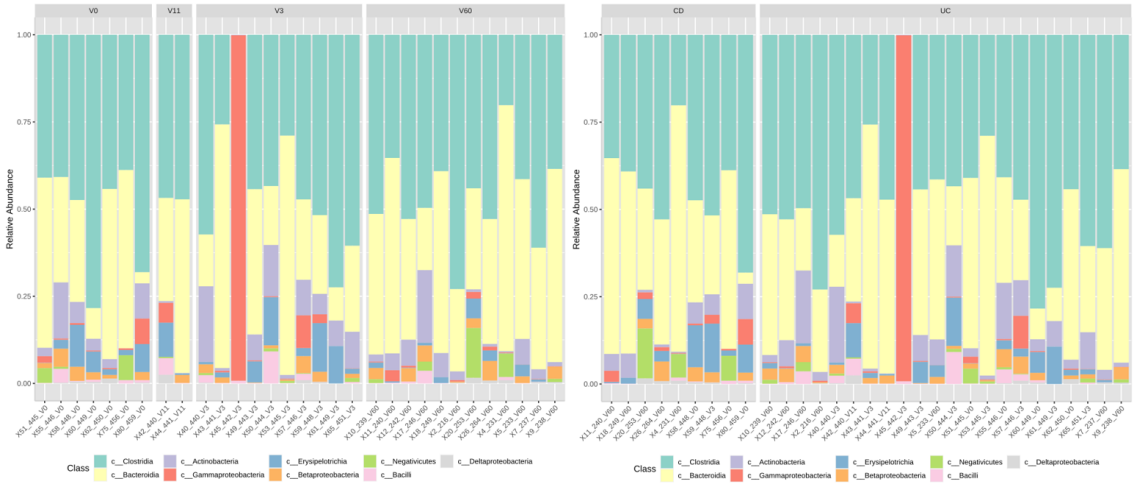


Figure D.1: A taxonomy bar chart of the distribution of bacteria class found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show relative abundance of each bacteria ranging from 0.00 to 1.00. The x-axis show the samples.

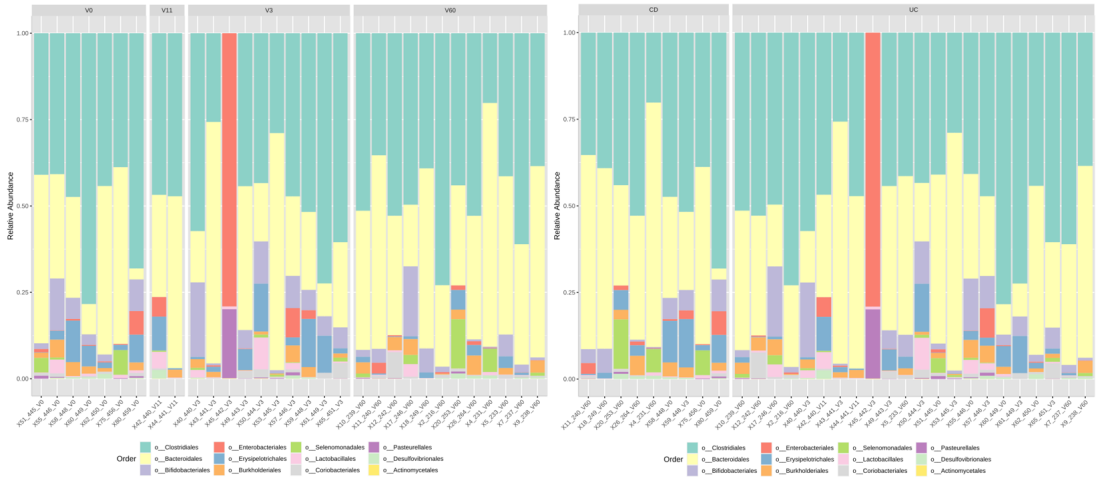


Figure D.2: A taxonomy bar chart of the distribution of bacteria order found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show relative abundance of each bacteria ranging from 0.00 to 1.00. The x-axis show the samples.

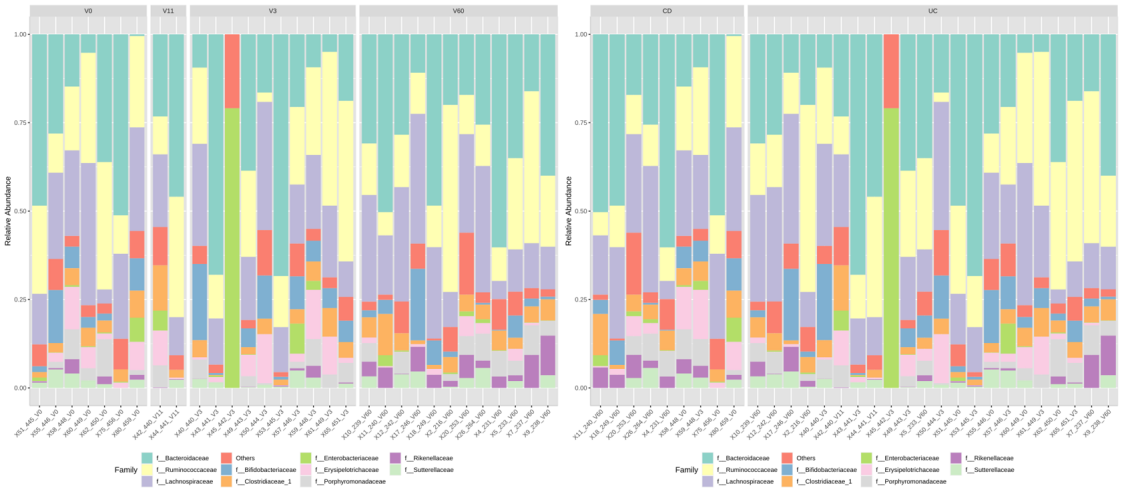


Figure D.3: A taxonomy bar chart of the distribution of bacteria family found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show relative abundance of each bacteria ranging from 0.00 to 1.00. The x-axis show the samples.

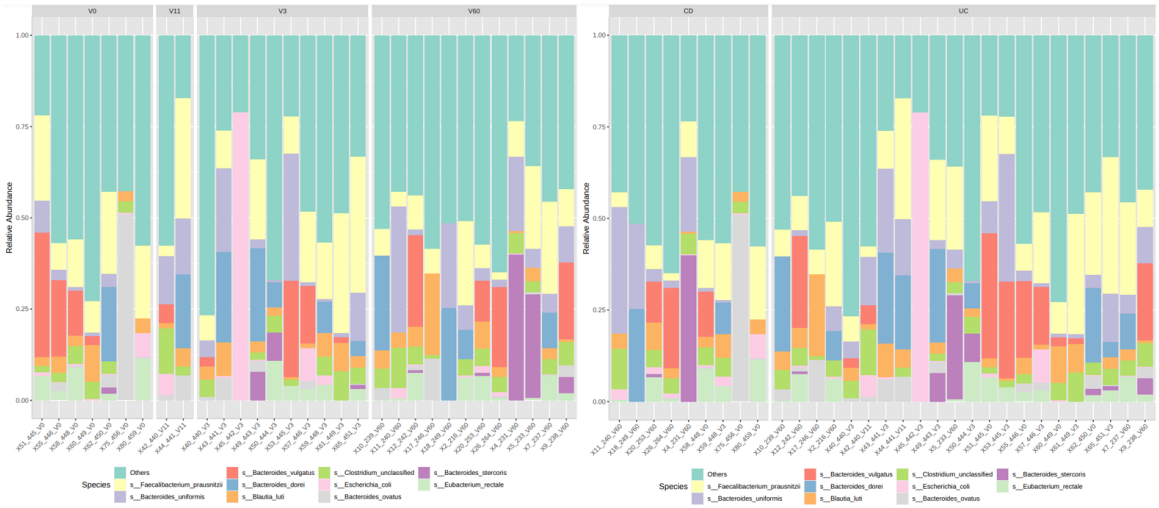


Figure D.4: A taxonomy bar chart of the distribution of bacteria species found in 32 samples, divided into groups based on months since diagnosis (left) and patients with UC and CD (right). The y-axis show relative abundance of each bacteria ranging from 0.00 to 1.00. The x-axis show the samples.