Simo Soila

Creating a DNA Metagenomics Library

Studies: Parkinson Disease Stool Samples

Helsinki Metropolia University of Applied Sciences

Bachelor of Engineering

Biotechnology and Food Engineering

Creating a DNA Metagenomics Library

26 January 2016

Helsinki Metropolia University of Applied Sciences

Author Title Number of Pages Date	Simo Soila Creating a DNA Metagenomics Library 47 pages + 9 appendices 26 January 2016
Degree	Bachelor of Engineering
Degree Programme	Biotechnology and Food Engineering
Specialisation option	Biomedicine
Instructors	Juha E. A. Knuuttila, Senior Lecturer Petri Auvinen, Laboratory Director Pedro Pereira, Researcher Matias Rantanen, Research Technician Harri Kangas, Research Technician

Extracting DNA from samples and creating a NGS library for metagenomics samples is a modern way of studying the genetic diversity within organisms and environments. Although the theoretical and practical details of these methods are well known, there is no single, standardized way to perform such a study.

The purpose of this thesis was to create an NGS DNA metagenomics library for a Parkinson's disease study. This thesis explains the theory behind the process and contemplates on the possible improvements that could be made to the laboratory protocols.

The thesis project started by extracting stool DNA, after which the NGS library was created and analysed. Concentration values gained from DNA extractions, PCR and Illumina sequencing were documented and reviewed in the results section of this thesis.

The DNA extraction concentration values were distributed normally when looked at statistically. From 138 samples, only one failed to produce any results due to a small amount of DNA during the extractions. The Illumina reads obtained were suitable for further analysis, although the quality of the reads could have been better.

By studying the results, it can be concluded that the creation of the DNA metagenomics library was a success although some of the samples in the pool were present in greater proportions than others. The standard deviation for all reads was ≈ ± 3 611 000 reads, and the average was ≈ 5 055 000 reads.

Keywords	Parkinson's disease, metagenomics, DNA library, Illumina, sequencing, stool, DNA extraction, DNA

Tekijä Otsikko Sivumäärä Aika	Simo Soila Metagenomisen DNA kirjaston luominen 47 sivua + 9 liitettä 26.1.2016
Tutkinto	Opinnäytetyö
Koulutusohjelma	Bio- ja elintarviketekniikka
Suuntautumisvaihtoehto	Biolääketekniikka
Ohjaajat	Juha E. A. Knuuttila, lehtori Petri Auvinen, laboratoriojohtaja Pedro Pereira, tutkija Matias Rantanen, tutkimusteknikko Harri Kangas, tutkimusteknikko

DNA:n eristäminen ja NGS metagenomiikkakirjaston luominen on moderni tapa tutkia geneettisen materiaalin monimuotoisuutta ympäristönäytteissä sekä kudosnäytteissä. Vaikka näiden menetelmien teoria ja käytäntö ovat yleisesti ottaen selvitettyjä, ei kokonaisprosessille ole yhtä ja oikeata tapaa suorittaa.

Opinnäytetyön tavoite oli luoda metagenomisia DNA kirjastoja osana Parkinsonin taudin tutkimusta. Tekstissä käsitellään myös menetelmiin pohjautuvaa teoriaa ja pohditaan mahdollisia parannuksia, joita niihin voisi soveltaa.

Työ aloitettiin eristämällä ja puhdistamalla ulostenäytteiden DNA, jonka jälkeen niistä muokattiin DNA kirjasto, joka analysoitiin Illuminan sekvensaattorilla. DNA eristyksistä sekä PCR ja Illumina ajoista saadut arvot kirjattiin ylös ja niitä tarkastellaan tämän opinnäytetyön tuloksissa.

Näytteiden eristyksien pitoisuudet olivat normaalijakautuneita. 138 näytteestä ainoastaan yksi ei tuottanut tulosta vähäisen DNA määränsä takia. Illuminan sekvensointi onnistui, vaikkakin sekvensoinnin laatu olisi voinut olla parempi.

Tuloksista voitiin päätellä, että metagenomisen DNA kirjaston luominen onnistui, vaikkakin jotkin näytteistä tuottivat enemmän sekvenssejä kuin toiset. Illumina ajon keskihajonta kaikille näytteille oli ≈ ± 3 611 000 sekvenssilukua ja keskiarvo ≈ 5 055 000 sekvenssilukua.

Avainsanat	Parkinsonin tauti, metagenominen, DNA kirjasto, Illumina, sekvensointi, uloste, DNA:n eristäminen, DNA

Contents

Abbreviations

IBS	Irritable bowel syndrome
CNS	Central nervous system
PD	Parkinson’s disease
SDS	Sodiumdodecylsulfate
T4	Refers to enterobacteria phage T4
ANOVA	Analysis of variance
NGS	Next generation sequencing
AATI	Advanced analytical technologies Inc.
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
dNTP	Deoxynucleotide

I would like to thank the staff members of the DNA Sequencing and Genomics lab for the opportunity to have my thesis done as a part of their study and for the help they provided me during my work there. I would especially like to thank researchers Pedro Pereira and Velma Aho, research technicians Harri Kangas and Matias Rantanen and laboratory engineer Lars Paulin for guiding me in practical laboratory work throughout the study. I would also like to thank the instructors senior lecturer Juha Knuutila and laboratory director Petri Auvinen for providing me with insights on the subject and for being the backbone for my thesis support.

1 Introduction

This thesis is based on previous findings which suggest that there is an association between certain gut bacteria and the Parkinson's disease. The Thesis is part of a metagenomics study conducted by the DNA sequencing and genomics laboratory at Institute of Biotechnology of Helsinki University.

The purpose of this thesis was to create a metagenomics DNA library from PD stool samples for the Parkinson's disease study, and to describe in theory and practice the methods used when creating a metagenomics library. In addition, some of the working methods and their effects on the DNA products have been reviewed through scientific literature and personal observations. Metagenomics studies the genetic material found in an environment or in other genetically complex systems.

The protocols and other methods used during the laboratory work were either provided by the laboratory staff or came with the kits and they can be found in the appendix.

2 Parkinson's disease

Parkinson's disease is a slowly proceeding movement disorder typically diagnosed between ages 50 to 70. The difficulties in movement are caused by the loss of dopaminergic neurons in the substantia nigra. The effects of the disease are better described in Figure 1. Even though scientists have managed to connect Parkinson's disease with genetic changes, environmental factors and heredity, the major cause for losing the dopaminergic neurons remains unknown [1, 2]. This could be one of the reasons why non-motor symptoms of the Parkinson's disease are increasingly more studied.

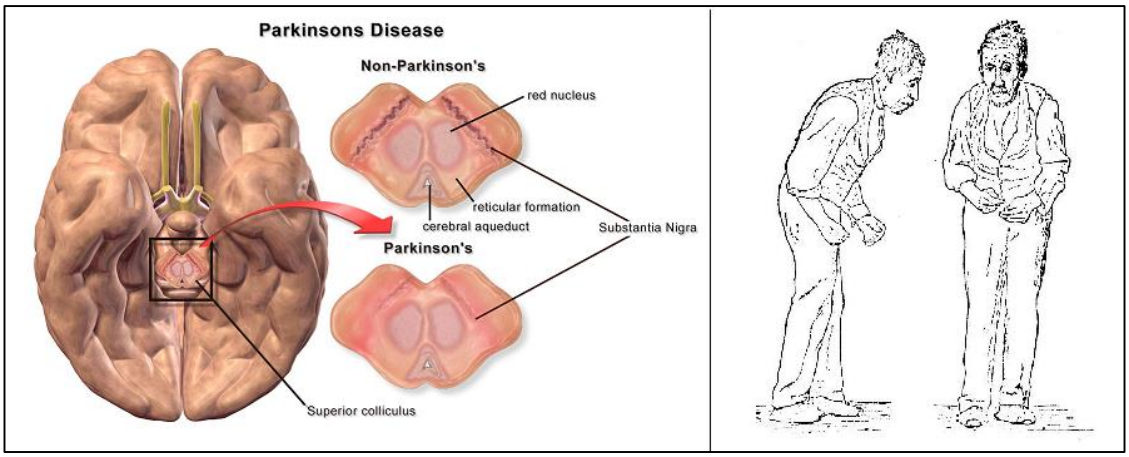


Figure 1. Figure on the left shows the differences in brain structure between a healthy person and a person who suffers from Parkinson's disease [3]. Figure on the right is a sketch of a person who has Parkinson's. It depicts the poor posture, slowed movements and tremor contained by pinching the fingers together [4].

2.1 Connection with gut microbiota

Only recently has the correlation between gut microbiota and the central nervous system been studied. These studies have found that gut microbiota could play a major role in how certain disorders such as IBS, depression, anxiety and chronic pain form, all of which have relations to the CNS. In nature, there are even some examples where small parasites have taken control over the entire nervous system of a simple organism [5].

Figure 2. Parasitic fungi, ophiocordyceps unilateralis, infested dead ant. The fungi controls the ant and forces it to move into higher grounds. There it makes the ant to strike its jaws deep into a plant and remain there until dead [6].

2.1.1 Previous research

Many observations have concluded that gastrointestinal dysfunction is common in Parkinson's disease. As a non-motor feature, this connection has not been studied as much as the primary symptoms of Parkinson's, even though the interaction between the gut and the central nervous system has been suggested to exist before [7].

Bulding on these previous studies, researchers [8] have found a connection between the gut microbiota and Parkinson's disease. In the study, gut microbiota of 72 PD- and 72 control patients were compared through high-throughput sequencing of 16S rRNA genes for phylogenetic marker analysis. The results suggest that the patients who have Parkinson's, tend to lack at least one family of bacteria that is commonly present in control subjects, and that some of the motor phenotypes could be related to the abundances of specific bacterial species within the gut [8].

3 Researching the gut microbiota from stool samples

The gut microbiota varies with the hosts genetic background, living environment and lifestyle. These micro-organisms work in a symbiotic way within the intestinal track of the host, which is beneficial for the host and the microorganism [9]. They are a part of the so called normal flora which includes all of the symbiotic micro-organisms in the human body.

Human stool is ideal for metagenomics study since it consists of as [9] describes, "nearly 200 prevalent bacterial species and approximately 1000 uncommon species". It is also a good way to study human diseases since every person, sick or healthy, produces it, and since it contains so much information on how the human body is functioning.

3.1 Metagenomics

Metagenomics, also known as environmental genomics, studies the genetic material found in an environment. Rather than researching a single microbe, it involves studying

the whole community of microbes found for example in water, soil or stool. The method is to extract the whole community DNA from an environmental sample, purify it and create a DNA library from it. The libraries are then used to study the variations in microbial communities or to study the genomic structures of individual species [10].

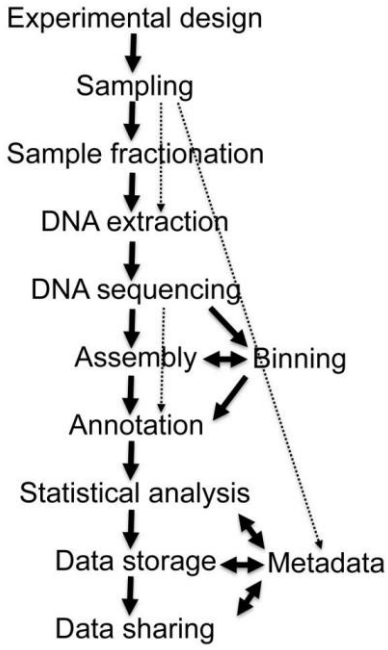


Figure 3. A protocol for executing a metagenomics study [11].

As [9] describes, gut microbiota can be studied in different ways such as descriptive metagenomics and functional metagenomics. The first concentrates on studying the community's structure, microbial variation and amount. The latter concentrates on studying the interactions between host/microbe and microbe/microbe functions.

Many microbes cannot be cultured under standard laboratory conditions. Even if they could be cultured, studying the interactions between all the micro-organisms would be impossible due to the technical constraints involved and the time consumed during the process. This is why the metagenomics approach works better when studying complex relations between microbes.

The metagenomics approach has been successful in generating large numbers of metagenomics sequence datasets that help us to understand the functions and relations of gut microbiota better [9].

3.2 Processes

Collecting a stool sample is fairly easy due to the abundance and composition of the sample material. However, the sample handling protocols vary depending on the study. For example a medical stool sample might require a certain diet before sampling and has to be delivered to medical staff within days of the collection [12], whereas with metagenomics samples the diet is normal, stool is homogenized in a preserving buffer and can be stored in a freezer for days or months before actually being studied. [13]

Sampling, storing and processing has been found to have an effect on the results of DNA sequencing [14, 15]. For example, in a study conducted by [14], a stool sample that had been left in room temperature for 2 weeks before freezing it at -80 °C had lost nearly all high-molecular weight fragments. This type of fragmenting also occurred while unfreezing the samples from -20 °C in 1 h prior to freezing them again in -80 °C. The best conditions for high-molecular weight fragments were when the sample was frozen immediately after sampling in -20 °C and then transferred to -80 °C, or when the samples were kept in room temperature after sampling for 3 h before freezing at -80 °C [14]. Although storage conditions prior to DNA isolation might affect the DNA fragment size, they do not affect the variation of bacterial species significantly unless the sample has been left in room temperature for a time period of 2 weeks or more prior to freezing. The thawing of the samples at the beginning of the DNA extractions has an impact on the taxonomic composition of the samples at the genus and species level [14].

Table1. [14] Changes in the composition of the bacterial taxa in the studied stool samples. The values on the table represent the percentage of each major taxon out of the total number of sequences.

Taxonomic comparison for 3 main bacterial taxa between frozen and unfrozen samples

Taxon	F*	UF1h*	UF3h*	*p* value F vs UF1h	*p* value F vs UF3h
Bacteroides;uncultured bacterium	19	13	9	0.044	9.68e-05
Prevotellaceae;uncultured;human gut metagenome	7	6	3	0.6804	0.0222
Bifidobacterium;uncultured bacterium	2	4	8	0.2257	0.0007

Statistical analysis was performed using Poisson regression model; p value < 0.05 is considered significant; n = 4 subjects; * Values are mean proportion of sequences (%).

F = frozen; UF1h = unfrozen during 1 h; UF3h = unfrozen during 3 h; Taxonomy is indicated at the genus level and if not possible at the family level.

Table 1 shows that the samples that had been unfrozen before the extractions for a longer time, had differences between their major bacterial taxa.

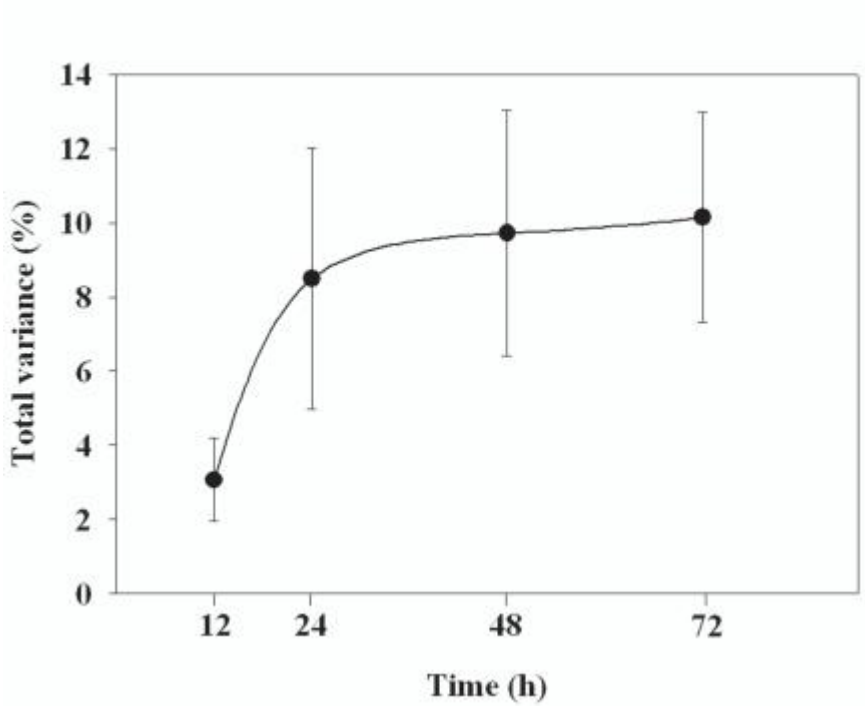


Figure 4. "The % change in bacterial community composition in the stool samples compared to samples frozen immediately. Each point is the mean of samples from four individuals with the bars representing the standard error about the mean." [15]

Figure 4 shows that if the samples are not frozen immediately after the sampling, their bacterial community composition will change gradually, the more time it stays in a room temperature.

3.2.1 DNA extraction

DNA extraction is a crucial step when studying the gut microbiota since it has a direct effect on the outcome of the downstream analysis. There are many ready to go protocols for extracting the DNA. Choosing the right protocol is essential since choosing the wrong method could have a great impact on the purity and amount of the DNA gained. A higher yield and a better quality of DNA will give more accurate results [16, 17]. Many of the protocols, if not all, follow the same principle: break the cell wall and liberate the DNA, remove PCR inhibitors and proteins and elute the DNA.

Stool sample cell lysis, as described by [16], can be done either mechanically, chemically or enzymatically. Mechanical bead lysing is one of the most efficient ways of cell lysing since it not only breaks the cell wall but homogenizes the sample even further. The lysing and homogenization effect of the beads give the reagents in the next step an ideal environment for removing PCR inhibitors from the sample. Although bead lysis provides an efficient way of exposing the DNA, it is considered to be too destructive for chromosomal DNA studies [16].

Stool samples contain many inhibitors that consist mostly of bile salts and complex polysaccharides [18, 19]. Bile salts are conjugated bile acids and form approximately 0.7 % of the total bile secreted from the liver. If not removed properly from the sample, they can inhibit the functioning of DNA polymerase or other reagents, depending on cofactors, in the PCR reaction. [18] Complex polysaccharides, at least in plant DNA preparations, can contaminate and inhibit both the restriction enzyme treatments and the PCR [19]. The removal of these inhibitors is done with bile salt binding chemicals. Shown by [20], bile salts can be inhibited by binding them to molecules such as cholestyramine, meciadanol, sucralfate or aluminium hydroxide and magnesium hydroxide [20]. Since the binding removes most of the inhibitors from the samples, the remaining inhibitors can be removed through a simple washing step. For example, in [21] a silica membrane type of DNA binding and washing step reduced the amount of PCR inhibitors from 12.5 % to 1.1 %.

Commonly, the removal of inhibitors is done by binding the DNA molecule into another molecule and washing its surroundings or by pelleting the inhibitors via centrifugation step and removing the supernatant.

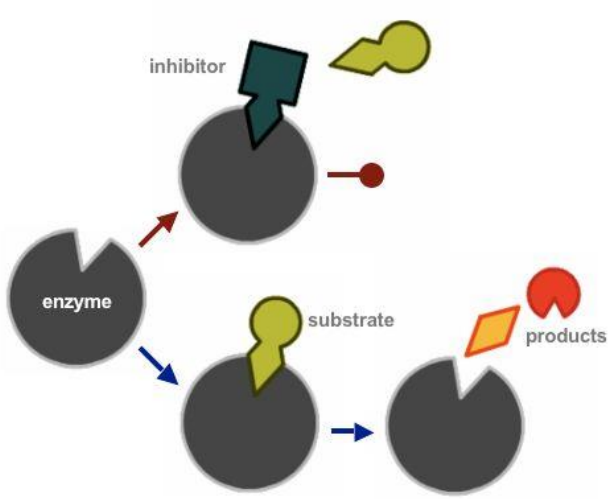


Figure 5. Inhibitor inhibiting the functioning of an enzyme at the top. Enzyme helping in the production of a molecule at the bottom [22].

Stool samples contain many proteins that need to be removed since excess amounts of protein can inhibit the PCR or damage the DNA. One of the most commonly used proteinases is proteinase K. As described by the manual [23], it cleaves peptide bonds at the carboxylic sides of aliphatic, aromatic or hydrophobic amino acids. In [24] it was observed that proteinase K together with dodecylsulfate would provide complete protection from ribonuclease to the tested polysomal RNA. It was also noted that in addition to dodecylsulfate, urea stimulates the activity of proteinase K as well. In [24] the sample that had been treated with only proteinase k, liberated only 7 % of the total amount of aromatic aminoacids, whereas, with dodecylsulfate it was as high as 93 %. Many of the manufacturers take this factor into account and for example [23] manual says “The activity of the enzyme is stimulated by 0.2 - 1 % SDS or by 1 - 4 M urea” [23, 24].

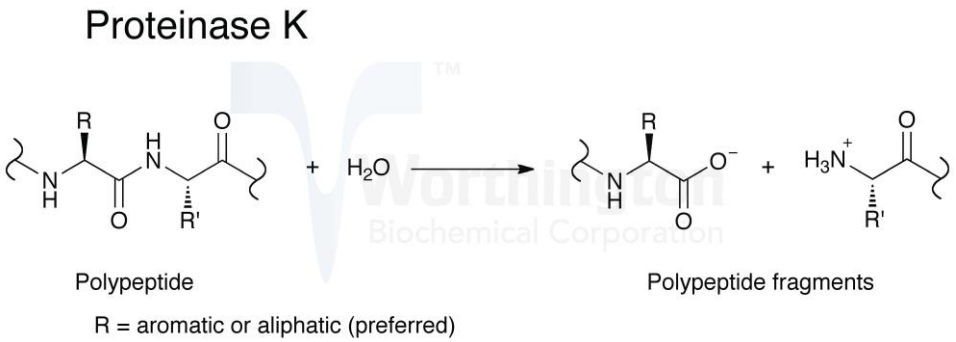


Figure 6. An example of the chemical reaction for the activity of proteinase K on polypeptides [25].

The washing and elution steps are the final steps when processing the DNA. Washing step purifies the DNA from most of the remaining contaminant proteins and PCR inhibitors. The washing step can be performed, for example, with the help of a silicate filter. Silicate binds to the DNA in the presence of strong salts. This is possible due to the negative charge of both, DNA and silica. The positively charged ionised salt molecules set between these negative charges and form a hydrogen bond. This bond will not break as long as the amount of salt stays high and pH stays stable. This allows for the DNA to be washed with salts and ethanol, removing impurities even further from the sample [26]. However, a study conducted by [16] notes that the efficiency of the DNA binding with silica based elution can be as little as 21 % of the initial amount of DNA.

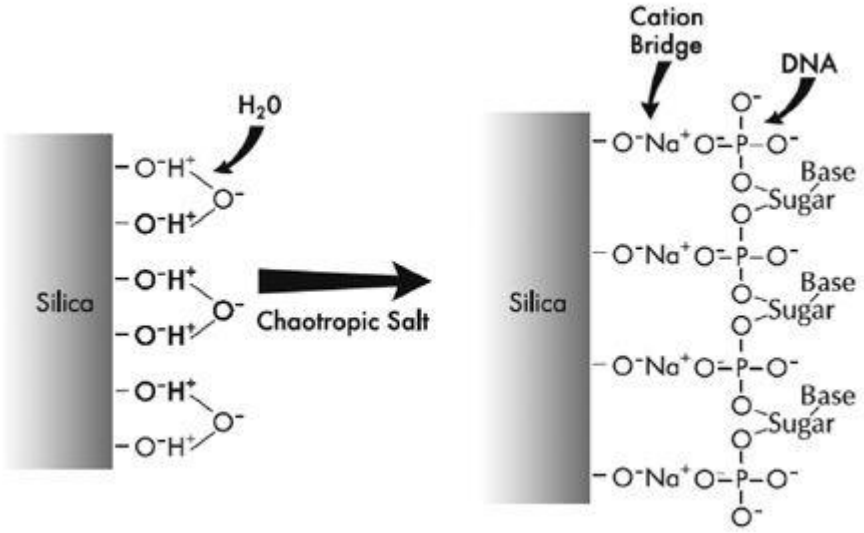


Figure 7. Principle behind the silicate membrane binding of the DNA. In the presence of chaotropic salt, the DNA molecule binds to the hydrolysed Na^+ with hydrogen bonds [27].

The DNA, bound on the silicate filter, is eluted when the concentration of the binding salt is diluted enough. The dilutant is called elution buffer and is a substance designed for storing the samples and the elution of DNA. The DNA is eluted through the filter with a centrifuge.

3.2.2 DNA sequencing library

A DNA sequencing library consists of fragments of DNA that represent the genetic diversity of the environment or a single organism. The creation of these libraries can vary since there are different protocols for cDNA libraries and DNA libraries. The principle behind the workflow is: fragment the DNA, size select the appropriate fragments, repair the ends of the fragments, attach adapters and amplify the fragments with PCR [28].

Fragmenting the DNA can be done enzymatically, chemically or mechanically. Mechanical fragmentation can be done acoustically by concentrating highly dense soundwaves at the DNA, to break the structure vertically. The breaking is done by cavitation bubbles formed by the soundwaves. Cavitation bubbles are formed when soundwaves separate and form a gap in between the water molecules. The gap absorbs energy until it implodes and shears anything around it. The efficiency of the fragmentation depends on the purity of the DNA, the fragmentation on / off time, the concentration of the sample, temperature and purity of the water, the intensity of the soundwaves etc. [28, 29]. The cavitation process is depicted in Figure 8. The length of the resulting DNA fragments is determined by the wanted library size and the limitations of the sequencing equipment [28].

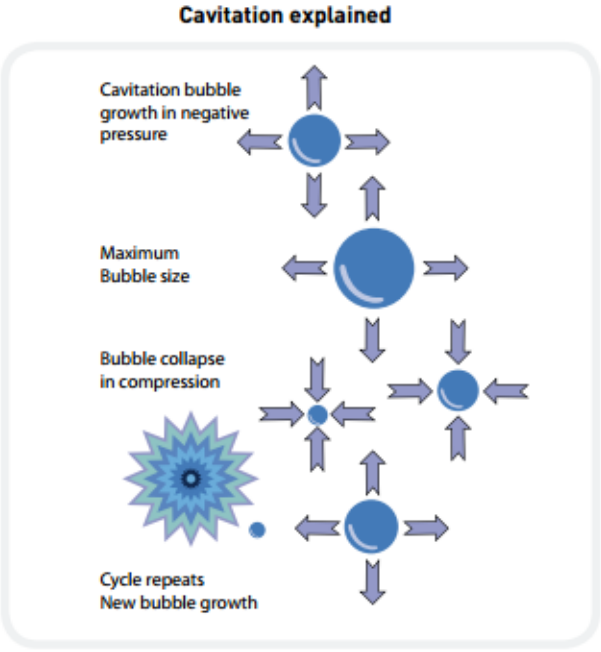


Figure 8. Sonication forms cavitation bubbles that implode and damage the surrounding DNA and other material [29].

The fragmentation of the dsDNA creates fragments that can be partially double stranded (ds) and partially single stranded (ss). This type of broken DNA is not compatible with the adapter ligation step and will cause an insufficient library dataset for the extracted samples. The ends of the fragments are repaired by using several DNA building reagents: T4 polynucleotide kinase, T4 DNA polymerase, Taq DNA polymerase, BSA, ATP and dNTP. In this process, ATP and dNTP work as building blocks for the DNA synthesis [Appendix 4, Norppa-library protocol].

T4 DNA polymerase catalyses 5’ → 3’ nucleotide synthesis on a DNA template. It also has 3’ → 5’ exonuclease activity which removes the hanging nucleotides from the 3’ end. These qualities produce blunt dsDNA strands. BSA works to stabilize the exonuclease activity of the T4 DNA polymerase. At the same time it also inactivates the contaminating nucleases and proteases and prevents the DNA from binding on the sides of the testing tube, keeping the reagents separated from the wall [30-32]. At the same time as the T4 DNA polymerase repairs or cleaves the strands at 5’ or 3’ end, the T4 polynucleotide kinase catalyses a transfer reaction of phosphate from ATP to the 5’ end of the DNA strand. T4 polynucleotide kinase also works as 3’ phosphatase, preventing the addition of more nucleobases at the 3’ end of the fragments. The phosphorylation modification allows for the DNA to be ligated later with another nucleotide sequence [33]. Since the

T4 DNA polymerase works as 3’ → 5’ exonuclease for ssDNA and dsDNA, it needs to be inactivated. Both of the T4 enzymes are inactivated by placing the samples in 75 °C for 10 min. During the inactivation, it is probable that the DNA fragments will sustain some damage done by the T4 DNA polymerase at the 3’ ends. If the strands get damaged, they are repaired in 5’ → 3’ direction by the Taq DNA polymerase which stays active on high temperatures. The correct function of the Taq DNA polymerase in this reaction however is to add an A-overhang onto the 3’ end of the DNA fragment. The end repair product is a double stranded, blunt ended, 5’ phosphorylated and 3’ A-tailed DNA fragment [Appendix 4, Norppa-library protocol], [30, 32-36].

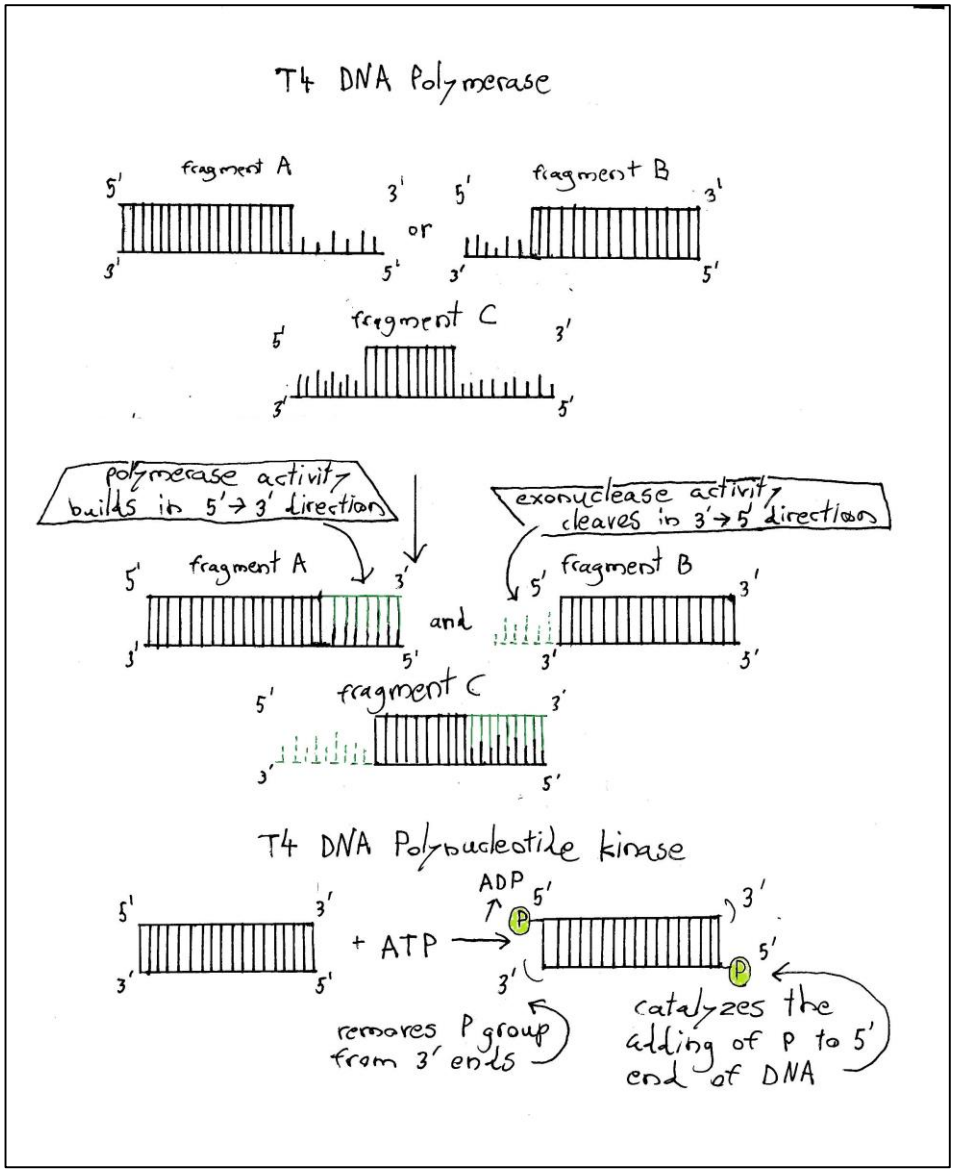


Figure 9. End repair process explained on the molecular level. The effect of two enzymes on the DNA fragments in the presence of ATP, dNTP, BSA and the reaction buffers.

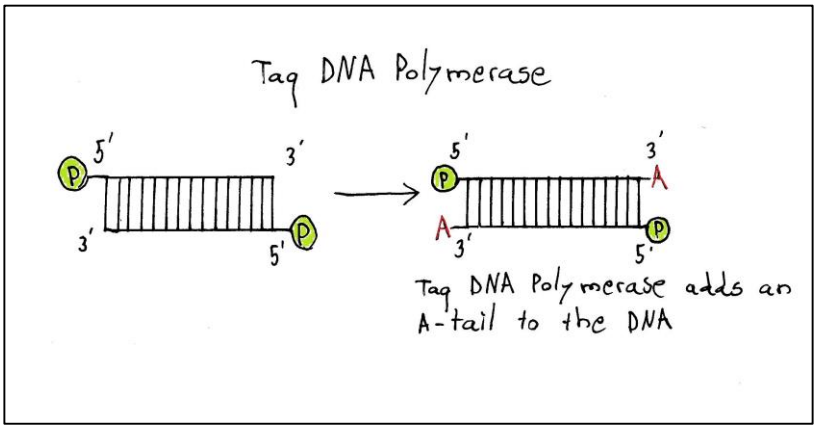


Figure 10. The function of Taq DNA Polymerase enzyme on the DNA fragment in the presence of ATP, dNTP, BSA and the reaction buffers during end repair process.

For the sequencer to be able to differentiate the fragmented sequences from each other, the fragments need to be labelled. The first step is to ligate an adapter to the phosphorylated and A-tailed fragments. The adapter consists of dsDNA with a blunt end and a sticky end. The blunt end will be ligated with phosphorylated and A-tailed dsDNA, and the sticky part will be ligated with label. T4 DNA ligase catalyses this reaction and unites the blunt dsDNA with the blunt part of the adapter. In addition, it also ligates missing nucleobases onto the DNA strand if provided with additional dNTPs. After the adapter ligation, the DNA is ready for the PCR and the adding of indexes [37] [Appendix 4, Norppa-library protocol].

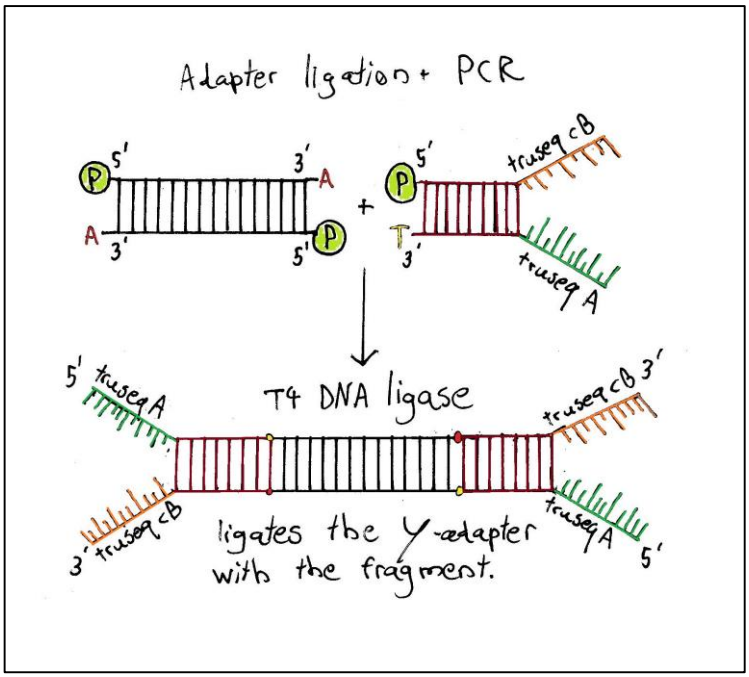


Figure 11. An example of the Adapter ligation reaction with T4 DNA ligase. At the top, the Y. adapter is on the right and the product from the end repair is on the left.

The DNA index complex consists of a primer, n amount of bp indexes and a universal complementary sequence used by the Illumina sequencer. During the PCR, the complementary primer binding site of the DNA index complex will attach to the complementary adapter primer binding site B. The DNA polymerase then copies the strand once. After the first copy, another primer, consisting of the primer and a universal sequence, attaches itself to the adapter primer binding site cA. The DNA strand is now attached with both primers and will copy itself for additional rounds of PCR cycles. The final product consists of the primers, index bp and universal sequences and the insert [38] [Appendix 4, Norppa-library protocol].

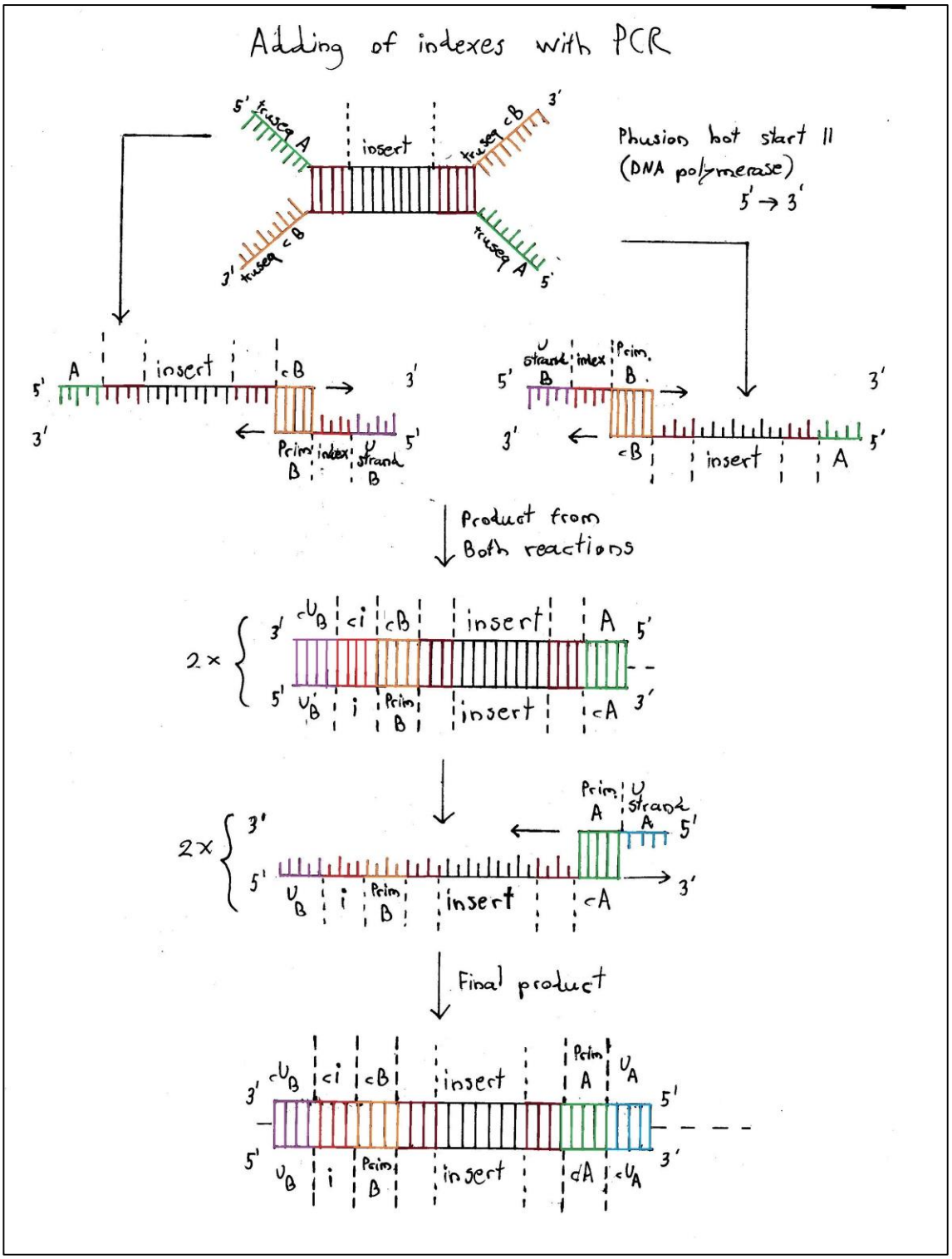


Figure 12. Binding of the index primers into a Y- adapter ligated DNA insert. For Illumina reads there are also two individual universal sequences ligated at the end of both primers next to the indexes with names U- strand A, U_A and U-strand B, U_B.

3.2.3 Illumina sequencing

The final library will be analysed with a sequencing machine. For Illumina machines, the protocol is such: the library is clustered, amplified and analysed in three or four sessions of DNA synthesis. The sample attaches to the bottom of the flow cell which has complementary universal sequences attached on it. First the forward sequence is read, then the index sequence and then the reverse sequence. If there are two indexes, there will be an additional read between the second and the third read.

In clustering, DNA strands are copied isothermally through bridge amplification. The individually marked DNA fragments have two universal sequences attached that are complementary to the probes attached on the Illumina flow cells. The DNA fragment bends on the flow cell and forms a bridge with the other complementary probe. The DNA strand is then copied while attached to both probes. When denatured, it leaves us with two copies of the DNA strands that are then amplified in the same method prior to the reads **[39].**

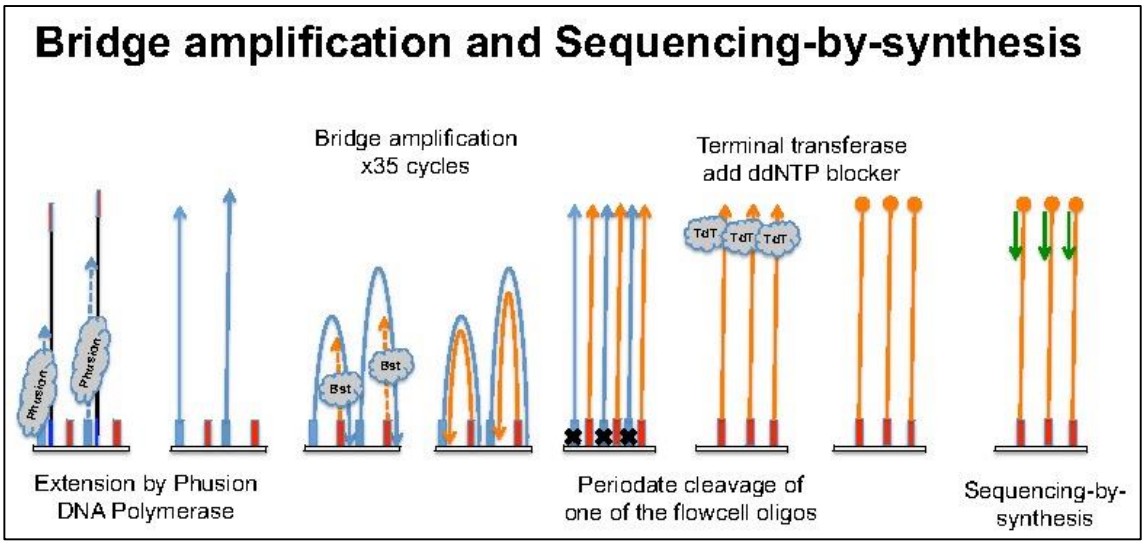


Figure 13. The bridge amplification done by the Illumina machine. The probes are attached to the bottom. DNA strands attach and build up on these probes [40].

Reverse strands are washed, leaving the forward strands on the flow cells. Sequencing primer is then added, followed by the adding of fluorescent marked nucleotides. The clusters are bombarded with a light source that reacts every time a new complementary

nucleotide is added to the DNA strand. Fluorescent marked nucleotides have a terminated 3’ end which means that they lack the 3’ OH group needed for the adding of the next base. The next nucleotide can be added after the OH group has been attached to the terminated nucleotide. Unattached nucleotides are washed from the flowcell before the adding of OH group. For this reason fluorescent marked terminated nucleotides get added one by one to the DNA strand. When bound to the strand, they emit an individual light signal that is registered on the sequencer thus giving the order of the complimentary nucleotides on the DNA strand. The registered base is determined by the average of all strands in the cluster. The more strands there are amplified within the cluster, the more accurate the signal becomes. After the insert has been read, another sequencing primer is attached to the index primer and the fluorescent nucleotides are analysed. After this, the DNA strand is bent again to form a bridge with the reverse probe. The ssDNA strand is then copied, denatured and the forward strand washed away. Reverse strand is then read with the same principle as the forward strand [39].

3.3 Contamination

When dealing with samples that are going to be processed with PCR, it is important to understand all the sources of possible contaminants. As the research on intensive care unit bacterial communities [41] shows, even the places deemed to be most clean may contain a complex bacterial community. In addition, other studies prove that microbial activity can occur even in altitudes as high as the upper troposphere 15 km [42]. So if microbes can rise to altitudes as high as this, they surely have no trouble reaching the 3 m altitudes represented by a common laboratory. As [43] shows, contaminating microbes can also be found within reagents. This is especially difficult when dealing with samples that have a low amount of DNA [41-43].

The spreading of bacteria through surfaces and room air is noted in today’s laboratories by using aseptic techniques. All the handling of the reagents and samples can be done inside a laminar which filters the air very clean. In addition, the laminar can be cleaned with 80 % ethanol after every task. Combined with the use of proper laboratory equipment such as laboratory jacket, rubber gloves, sterile pipet heads and tubes, many of the contaminants can be countered.

3.4 Batch effects

During a metagenomics study, a large number of samples are processed. Most of the time these samples arrive to laboratories in batches. Because of tight schedules within the laboratories, these batches are commonly processed separately. For assessing the working methods, the order in which samples are handled, has to be randomised. Randomization is done to minimize the so called batch effects.

Batch effects are caused by unknown technical variables present in the study. Such a variable can have an effect on the results and distort them. The distorting variable can be from simple source, such as damaged laboratory equipment, or as complex as a variable created by the combined effect of ten other variables [44]. These combined variables are common in biological studies where multiple reagents and conditions are present.

Batch effects have been studied for a long time through various standardised mathematical methods such as the ANOVA tests. Studying the results enable us to specify the distorting variables and to narrow down the possible contaminants. These types of studies are also crucial when optimizing laboratory methods.

4 DNA extractions of PD stool samples

Gut microbiota DNA extractions were done to 138 samples and X lambda DNA controls. The extractions were done in 2 batches, first with 119 samples and the second with 19 samples. 11 extractions were repeated due to low amounts of DNA. The first batch was done 17.3 - 2.4 and the second batch 29.4 - 4.5. The average amount of DNA extractions done daily was 10 samples, starting with 2 and 5 samples and finishing up with 15 samples a day. The stool samples were collected from Parkinson's disease patients, their spouses and control patients in Finland. Samples were stored in PSP kit's buffer after sampling, frozen and shipped to the laboratory in dry ice. They were then moved into a -80 °C freezer to await for the extractions [13].

Most of the stool samples were processed accordingly to the same protocol during the stool extractions. Only one of them, P_41, failed to give a sufficient amount of DNA for

making a library. The protocol for the extractions was carried out accordingly to the protocol depicted in the appendix [Appendix 5, PSP spin stool DNA plus kit protocol].

4.1 Extraction order and thawing of the samples

The extraction order for the samples was determined by randomizing the previous listing order of the samples with an excel function. The correct samples could be picked with the help of the box coordinates sheet provided by the sampling laboratory. Samples were then moved into a laminar in a room temperature and thawed. The thawing usually took approximately 1 h.

Most of the times the samples would thaw equally. Some samples, however, would melt more rapidly than others. This type of unequal melting would happen each time the samples were thawed. Though as [14] and [15] suggests, there is no significant additional damage done to the DNA fragments or variation of bacterial species, unless the samples would be kept melted for extra hours or days.

4.2 Sample homogenization and prelysis

After thawing, each sample was transferred into a 2.0 ml safe lock tube. This would be done in 2 portions. First portion would take 1.0 ml and the second 0.4 ml of sample. The samples were then moved into a thermomixer to maximize the amount of bacterial DNA. Meanwhile the laminar was cleaned and the stool moved back into the freezer. After the thermomixer, 5 zirconia beads were added manually to each sample tube and the tubes were vortexed, two at a time, with a table vortex for 2 minutes. After this, the solids were pelleted with a centrifuge.

The order of the sampling would be determined by the order in which the samples were taken from the freezer. Since some of the samples would have solid stool in them, it was better to take 1.0 ml first to ensure that there was enough of the sample for the next steps. Adding the zirconia beads manually was sometimes difficult since they would drop from the tubes in clusters. It would also take some time to get used to mixing the tubes for 2 minutes with a table vortex.

4.3 Removal of PCR inhibitors and the second sample clean-up

The supernatants were moved into Inviadsorb tubes which bind the inhibitors in faeces efficiently [13]. After mixing and incubating at room temperature, the solids were pelleted to the bottom. After the centrifugation, supernatant was transferred into a 1.5 ml Eppendorf and the centrifugation from the previous step repeated. Meanwhile, 25 µl of the proteinase K was pipetted into a new set of 1.5 ml Eppendorf tubes.

When transferring the supernatant from previous samples, it is important to not suspend the pellet by acting too hastily. It is also good to set the tubes in the centrifuge so that the pellet will appear on a certain side. When the stool is pelleted for the first time, it sometimes creates a sticky layer on top of the supernatant. This makes it more difficult to get the clear supernatant below it. This sticky substance often appears if there is too much solid particles in the samples. In some cases, the sticky substance would not pellet even after recentrifugation.

Another factor that affects the difficulty of transferring the supernatant is the colour of the samples. The Inviadsorb tube is full of black powder. When the sample, that is dark-brown, is added to the tube, the border between supernatant and the pellet becomes unclear. The suspension of the pellet with the supernatant can be avoided by placing the sample between a source of light or by pipetting the supernatant slowly.

4.4 Proteinase K digestion

The previous supernatant was transferred to the tubes containing the proteinase K. They were then thermomixed for 10 minutes.

Some of the samples had less supernatant during this step than required. This is why it’s important to have enough of liquid sample in the beginning of the extractions. Since the samples were no longer inside safe locked tubes, they would often burst open or leak while being moved from the thermomixer to the laminar. The leaks were absorbed with paper and the bursting avoided by letting the samples stand for a minute or two before lifting them from the thermomixer.

4.5 Binding of the DNA

After the digestion, the binding buffer was added and the mixture vortexed for 15 seconds. The mixture was then transferred onto the spin filter in two batches, 700 µl batch and 300 µl batch. Both of the batches were incubated at room temperature for 2 minutes before centrifuging.

The samples need to be vortexed separately to avoid any leakage. Even then it was important to keep the tube caps tightly closed with your hands since the samples were mixed with runny isopropanol. Although vortexing would be faster, it often causes a mess that needs to be cleaned afterwards. The mixture needs to be moved onto the spin filter in two batches since it can only contain 700 µl of sample by each centrifugation.

4.6 Washing steps and ethanol removal

When all the sample was bound to the filter membrane, the filter was first washed with washing buffer 1 and centrifuged. Receiver tube was changed and filter filled with washing buffer 2 and centrifuged. Receiver tube was emptied and re-used in ethanol removal. Ethanol was removed by centrifuging the sample at full speed for 4 minutes. Meanwhile the elution buffer was retrieved from the incubator.

Both of the washing buffers are runny due to alcohol. This leads to dripping when pipetting. Ethanol can dissolve the markings on the test tubes, thus you need to be careful when pipetting. Sometimes all the sample would not centrifuge properly on the filter. This might be due to solids clogging the filter. Eventually the liquid will get centrifuged but the DNA will not be as pure as it should.

4.7 DNA elution

Preheated elution buffer is added on the spin filter and incubated for 5 minutes. Meanwhile, an ice bath is prepared. After incubation, the samples are centrifuged into 1.5 ml receiver tube. After centrifugation, they are transferred to better tubes that are more ideal for longer times of freezing. The tubes are then put on ice and placed near the nanodrop machine. Samples are then measured with nanodrop and the results printed out. After

this, the laminar is and the used equipment are cleaned and the equipments moved back to their original places.

5 DNA libraries of PD stool samples

A protocol constructed by the DNA laboratory staff [Appendix 4 , Norppa-library protocol] was used, The process was done for the previously extracted 137 DNA samples. Although most of the washing steps were done with a pipetting machine, Magnatrix, some were purified manually with the same reagents and protocol [Appendix 3, Clean up protocol].

5.1 Cutting the DNA with Bioruptor

The sample DNA was added into Bioruptor tubes so that the concentration would be 1 ug [DNA] / 100 µl [H_2O]. The fragmented DNA was washed with Magnatrix or manually accordingly to the washing protocol [Appendix 3, Clean up protocol].

Magnatrix machinery is isolated by a plastic cover but it contains no air filtration system. The possibility for a contamination is, however, faint since the DNA is being purified in the process.

5.2 End-repair, adapter ligation, PCR and size separation

All the reagents that require thawing were thawed and placed on an ice bath. While the DNA was being purified, the master mix for end repair was prepared. The sample plate was moved on the ice bath and the required enzymes, T4 polynucleotide kinase, T4 DNA polymerase and DreamTaq polymerase were added to the cooled master-mix. Master-mix was homogenized by pipetting and a portion of the mix was added to each of the wells containing the samples. The 96 well plate was moved into a PCR machine and an end repair program was performed accordingly to [Appendix 4, Norppa-library protocol]. When the program had finished, the samples were purified with Magnatrix or manually [Appendix 3, Clean up protocol].

A master-mix containing the T4 DNA ligase buffer and the Y- adapter was prepared for the adapter ligation step and put on an ice bath. The samples were then placed on the ice bath and the enzyme T4 DNA ligase was added to the master-mix. Master-mix was then added to every well and the plate containing the samples was moved into a PCR machine. An adapter ligation program was used [Appendix 4, Norppa-library protocol]. After adapter ligation, another DNA purification step was performed with Magnatrix or manually [Appendix 3, Clean up protocol].

The primer indexes were pipetted individually to the bottom of the wells and a master-mix containing MQ water, 5x phusion HF buffer, dNTP and PCR_Truseq_A was prepared and placed on an ice bath. The Magnatrix plate was placed on the ice bath and the enzyme Phusion HotStart II was added to the master-mix. The samples were transferred to the wells containing the indexes, and the master-mix was added to these wells. The mixture was homogenized and divided so that each well would have 50 µl of sample in them. Samples were then vortexed and centrifuged briefly and moved into a PCR machine. A PCR program with 18 cycles was performed [Appendix 4, Norppa-library protocol]. After the run, the size separation program was performed with Magnatrix or manually [Appendix 2, DNA size separation step protocol].

Doing the washing steps manually has its' problems. The manual aspect is a variable that creates variations between the samples. It would be more difficult to transfer the DNA into small strip tubes than it was to add the DNA into a Magnatrix well plate.

6 Sequencing

Preparing the Illumina sequencer is done accordingly to the protocol. This step reveals if there has been any mistakes during previous experiments.

6.1 Pooling and purifying the samples and concentrating the pool

The index labelled DNA samples were pooled together in an amount of 13.8 ng / sample. The total volume of the pool would be then ≈ 850 µl and the concentration ≈ 2.2 ng / µl. The pool would then be concentrated with a special filter column a few times, reducing

the volume to ≈ 20.0 µl and rising the concentration up to ≈ 67.0 ng / µl. A manual washing step was then performed two times, selecting the appropriate base size for the IlluminaNextSeq500.

6.2 Sequencing the DNA library with IlluminaNextSeq500

A part of the concentrated sample was pipetted and diluted. Part of this diluted sample was then pipetted into a new tube, which would then be optimized for the Illumina machine. After pooling the samples together, the optimized pool was transferred to a fully prepared Illumina reagent plate, the plate inserted into the Illumina machine and the sequencing program started.

7 Results

The result focus on the values measured from the DNA extractions and PCR products since they offer the best view on how the creation of the library succeeded. The Fragment Analyzer, AATI, report shows the quality of the DNA and the Illumina sequencing results the functioning of the DNA library.

7.1 DNA extractions, nanodrop and values

DNA extractions were done in a randomised order in a period of 16 days. The results are shown in table 1.

Table 1. Concentrations, dates and sample names of all the DNA extractions in a chronological order 17.3 - 4.5.2015, excluding the repeated samples. Nanodrop measures the purity and amount of the DNA and Qubit measures the amount of the DNA.

Sample ID	Extraction date	DNA µg / mL Nanodrop	DNA µg / ml Qubit
C_99	17.3.2015	169.71	56.1
CS_119	17.3.2015	50.49	17.7
CS_47	18.3.2015	22.54	5.48
P_51	18.3.2015	8.72	3.51
C_103	19.3.2015	664.73	120

PS_15	19.3.2015	5.49	2.79
CS_73	20.3.2015	70.85	65.7
C_137	20.3.2015	267.71	71.2
P_60	20.3.2015	416.84	120
C_69	20.3.2015	166.78	92.1
P_119	23.3.2015	77.23	37.3
C_95	23.3.2015	330.69	120
P_58	23.3.2015	27.42	9.75
CS_71	23.3.2015	139.51	19.8
P_50	23.3.2015	159.37	92.5
P_16	23.3.2015	43.6	21.7
CS_114	23.3.2015	9.38	3.5
CS_100	23.3.2015	118.12	35.2
C_75	23.3.2015	91.11	45
P_105	23.3.2015	64.8	43.4
C_30	24.3.2015	24.19	12
C_35	24.3.2015	266.32	99.6
C_136	24.3.2015	12.2	3.5
P_71	24.3.2015	21.4	9.58
CS_107	24.3.2015	223.42	74
C_102	24.3.2015	15.6	6.55
C_74	24.3.2015	18.63	7.57
C_116	24.3.2015	16.55	5.96
P_77	24.3.2015	167.03	80.3
CS_68	24.3.2015	227.27	92.1
C_119	25.3.2015	1.31	0.3
C_147	25.3.2015	117.49	36.6
P_63	25.3.2015	121.63	52.7
C_47	25.3.2015	64.31	35.8
P_59	25.3.2015	35.05	14
C_48	25.3.2015	235.65	94.4
P_48	25.3.2015	102.76	51.1
P_114	25.3.2015	105.83	52.5
P_107	25.3.2015	25.24	14.4
P_100	25.3.2015	41.44	9.44
P_31	26.3.2015	170.96	76.4
C_49	26.3.2015	404.21	120
CS_65	26.3.2015	508.35	120
C_23	26.3.2015	25.07	6.83
C_8	26.3.2015	181.07	97.9
C_86	26.3.2015	52	15.7
C_90	26.3.2015	27.33	9.14
C_80	26.3.2015	333.42	120
C_96	26.3.2015	205.95	98.6
P_53	26.3.2015	93.54	30.6
P_61	27.3.2015	145.78	63.7
P_57	27.3.2015	40.8	14.6
P_42	27.3.2015	224.75	74.2
P_69	27.3.2015	113.04	100

C_59	27.3.2015	68.31	32.2
C_68	27.3.2015	190.95	40.9
P_87	27.3.2015	66.32	22
P_66	27.3.2015	24.32	19.1
P_56	27.3.2015	166.85	62
CS_41	27.3.2015	125.29	56.3
P_126	27.3.2015	316.4	120
P_68	27.3.2015	375.72	120
C_146	27.3.2015	199.45	68.6
C_118	27.3.2015	48.42	31
P_65	27.3.2015	44.53	13.1
P_99	30.3.2015	51.57	19.5
P_79	30.3.2015	71.05	33.1
CS_103	30.3.2015	179.22	63.8
C_105	30.3.2015	78.3	47.1
C_89	30.3.2015	79.6	84.1
C_15	30.3.2015	239.94	97.5
C_82	30.3.2015	127.37	43.1
C_142	30.3.2015	6.46	2.7
CS_87	30.3.2015	383.65	120
C_135	30.3.2015	174.9	85.4
P_94	30.3.2015	23.18	8.95
P_120	30.3.2015	92.47	47.4
P_83	30.3.2015	16.73	3.8
P_47	30.3.2015	166.46	61
P_8	30.3.2015	21.89	4.6
C_19	31.3.2015	98.5	78.6
C_88	31.3.2015	26.02	8.9
C_33	31.3.2015	133.87	28.9
C_104	31.3.2015	94.52	34.3
CS_116	31.3.2015	39.26	11
C_20	31.3.2015	36.86	14.7
CS_66	31.3.2015	54.17	23.5
C_65	31.3.2015	309.85	120
P_88	31.3.2015	57.1	24.1
CS_74	31.3.2015	45.08	18.6
P_103	31.3.2015	94.65	35.5
C_107	31.3.2015	152.93	56.3
P_115	31.3.2015	366.96	120
P_43	31.3.2015	54.08	19.9
C_98	31.3.2015	12.9	1.9
C_51	1.4.2015	67.62	16
P_74	1.4.2015	68.69	20.8
C_123	1.4.2015	423.33	120
P_116	1.4.2015	198.26	96.4
C_84_P	1.4.2015	20.23	5.89
C_131_P	1.4.2015	505.06	120
C_121_P	1.4.2015	44.58	14.2
C_60_P	1.4.2015	591.64	120

C_10_P	1.4.2015	524.51	120
P_21_P	1.4.2015	124.82	60.3
P_104_II_P	1.4.2015	61.47	19.9
C_115_P	1.4.2015	142.69	32.7
C_125_II_P	1.4.2015	109.04	33.9
C_152_II_P	1.4.2015	42.83	12
C_109_P	1.4.2015	350.07	120
C_8_P	2.4.2015	78.52	29.4
C_127_P	2.4.2015	60.91	10.9
P_62_P	2.4.2015	70.45	25.4
C_56_P	2.4.2015	63.21	32
C_93_P	2.4.2015	136.68	48.9
P_110_P	2.4.2015	71.23	46.9
C_99_II_P	2.4.2015	9.43	1.9
C_106_P	2.4.2015	136.91	39.1
CS_62	29.4.2015	315.6	120
P_118	29.4.2015	251.9	94.1
P_85	29.4.2015	57.68	100
C_148	29.4.2015	487.7	120
P_52	30.4.2015	350.51	120
C_10	30.4.2015	86.4	120
C_131	30.4.2015	27.64	68.5
C_85	30.4.2015	31.87	56.2
C_87	30.4.2015	344.08	120
C_84	30.4.2015	79.46	110
C_127	30.4.2015	8.72	42.8
P_104	30.4.2015	39.34	22.4
C_111	30.4.2015	140.96	99.1
C_110	4.5.2015	316.72	120
C_93	4.5.2015	139.4	120
C_152	4.5.2015	626.01	120
C_140	4.5.2015	103.99	46.7
P_62	4.5.2015	388.02	120
CS_31	4.5.2015	115.73	120

DNA extractions were done in two batches. This was due to the samples arriving at different times to the laboratory. Figure 14 shows the amount of the samples processed each day and the time gap between the batches.

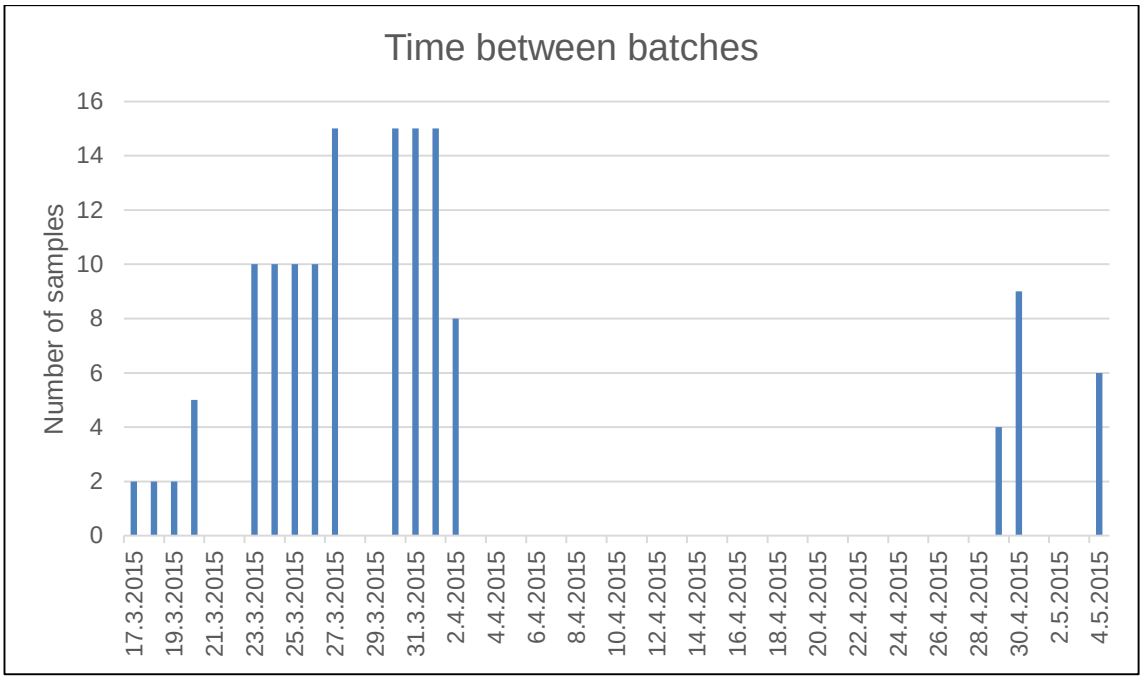


Figure 14. The amount of samples processed daily and the time between the two batches. The work was not continued during the weekends and had to be halted when waiting for more samples to arrive. Repeated samples excluded.

From the raw values presented here, it is possible to evaluate the working process through the methods of standard variation. Since the samples were processed in two separate batches and were not randomised together, they are looked at separately.

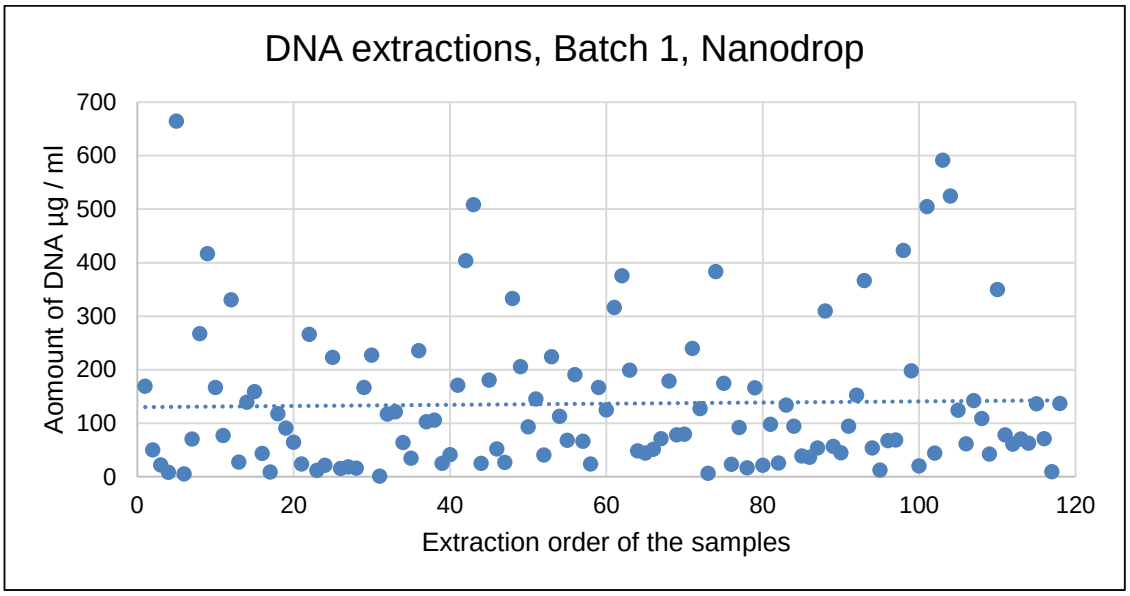


Figure 15. Nanodrop values of the Batch 1 DNA extractions done chronologically. Y-axis shows the amount of DNA measured by the nanodrop and the X-axis shows the extraction order of the Batch 1 stool samples listed in Table 1. Repeated samples excluded.

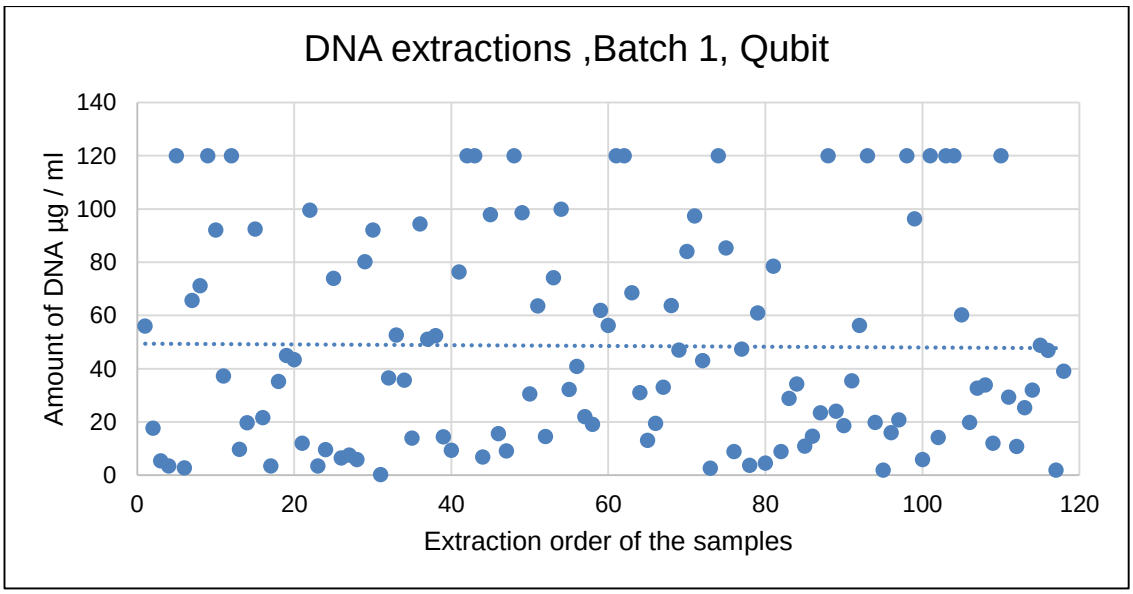


Figure 16. Qubit values of the Batch 1 DNA extractions done chronologically. Samples that were too high in DNA amount, are marked on the chart as 120 µg / ml. The Y-axis shows the amount of DNA measured by the Qubit and the X-axis shows the extraction order of the Batch 1 stool samples listed in Table 1. Repeated samples excluded.

The figures for Batch 1 show us that the experiments obey the standard variation as they should. The slight rise in Nanodrop values and the decrease in DNA amounts is not significant enough to suggest any worrying trends. If the trend lines would be steeper, you could say that the quality of the extracted samples are decreasing as the experiments are executed.

The second Batch was not as big as the first one. From the figures 17 and 18, it can be observed that the amount of DNA is greater in these samples.

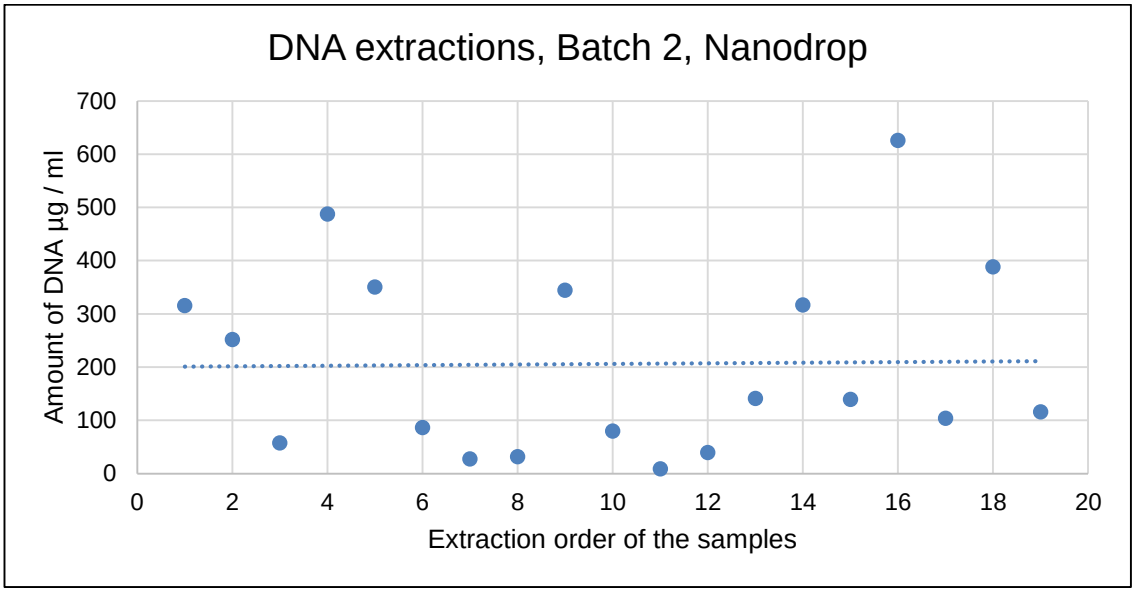


Figure 17. Nanodrop values of the Batch 2 DNA extractions done chronologically. Y-axis shows the amount of DNA measured by the nanodrop and the X-axis shows the extraction order of the stool samples listed in Table 1. Repeated samples excluded.

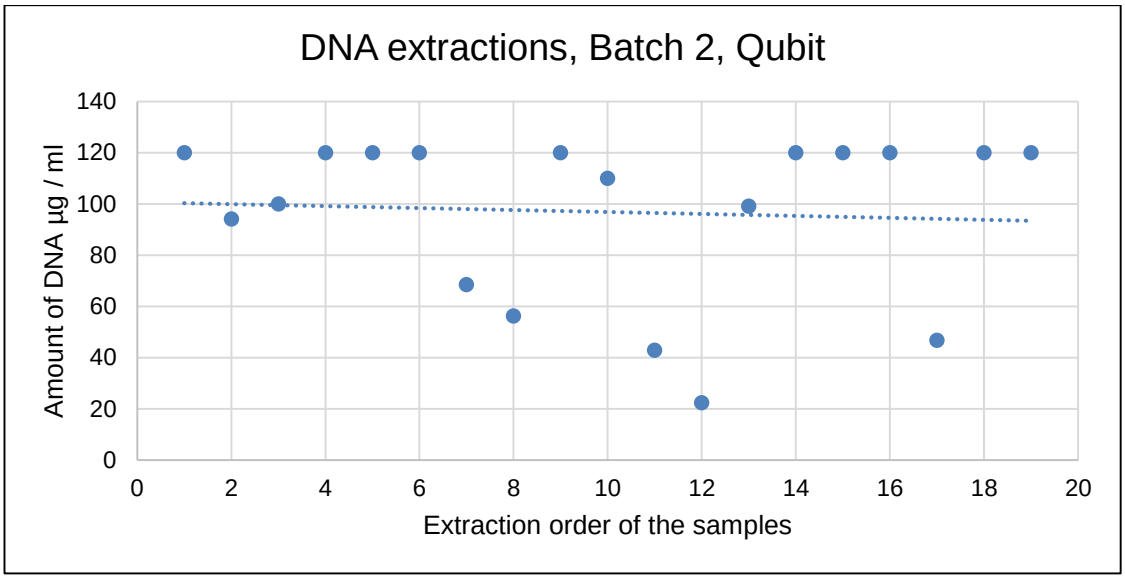


Figure 18. Qubit values of the Batch 2 DNA extractions done chronologically. Samples that were too high in DNA amount, were marked on the chart as 120 µg / ml. The Y-axis shows the amount of DNA measured with the Qubit and the X-axis shows the extraction order of the Batch 2 stool samples listed in Table 1. Repeated samples excluded.

Similarly the extracted samples from Batch 2 show that the samples are standardly variated and that the trend lines either rise or decrease slightly. From this we can conclude that the experiments were carried out in a similar manner and that the samples in Batch 2 consisted of more DNA enriched samples than Batch 1.

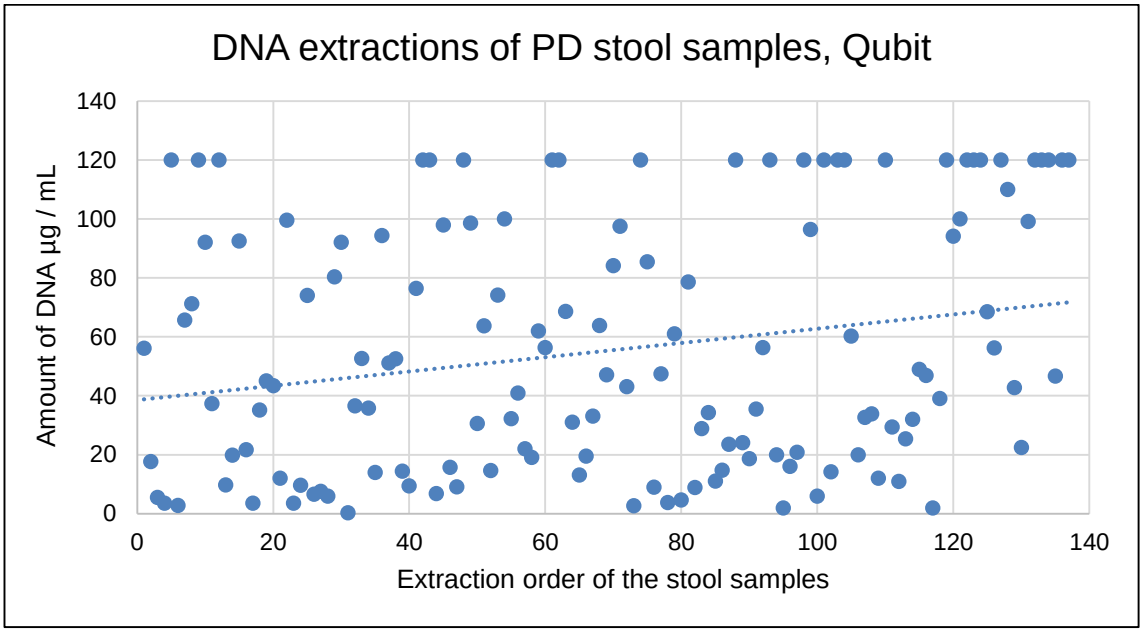


Figure 19. Qubit values of the DNA extractions done chronologically. Samples that were too high in DNA amount, were marked on the chart as 120 µg / ml. The Y-axis shows the amount of DNA measured with the Qubit and the X-axis shows the extraction order of the stool samples listed in Table 1. Repeated samples excluded.

If both of the batches were fitted into the same pictures, it would not be possible to analyse the laboratory work properly since the variances would not be accurate. Figure 19 shows that randomising the samples prior to the tests is essential for analysing the working methods.

7.2 DNA amount after PCR

Table 1 shows the DNA yields of the extractions. Table 2 presents the DNA concentration values after the DNA samples had been processed with fragmentation, end-repair, adapter ligation, several washing steps, the adding of indexes, PCR and size separation.

Table 2. DNA amounts of PCR products of the PD stool samples in a chronological order. The table also shows in which method the washing steps were done during Bioruptor, end repair, adapter ligation and the PCR. The samples were processed from the largest amounts to the smallest amounts of starting DNA. High amounts of DNA are shown with green colours and low amounts are shown with red colours.

Sample Index	PCR Date	DNA Concentration µg / mL	Washing Steps Done
C_103	9.4.2015	4.95	Manually
C_95	9.4.2015	3.90	
P_60	9.4.2015	4.99	
C_80	15.4.2015	9.99	Manually
CS_65	15.4.2015	15.90	
C_49	15.4.2015	12.30	
P_68	15.4.2015	13.70	
CS_87	15.4.2015	11.40	
C_123	15.4.2015	9.67	
C_109_P	15.4.2015	10.50	
C_131_P	15.4.2015	8.30	
C_60_P	15.4.2015	7.55	
C_10_P	15.4.2015	8.17	
C_99	24.4.2015	4.84	Manually
C_137	24.4.2015	7.10	
C_69	24.4.2015	10.60	
CS_73	24.4.2015	10.90	
C_75	24.4.2015	11.80	
CS_100	24.4.2015	13.60	
P_50	24.4.2015	11.40	
CS_68	24.4.2015	3.98	
CS_107	24.4.2015	3.30	
C_35	24.4.2015	4.17	
C_65	24.4.2015	3.94	
P_115	24.4.2015	4.83	
P_119	6.5.2015	3.37	Manually
P_105	6.5.2015	5.92	
P_77	6.5.2015	7.06	
C_147	6.5.2015	6.47	
C_48	6.5.2015	3.87	
P_114	6.5.2015	4.11	
C_47	6.5.2015	3.71	
P_48	6.5.2015	6.45	
P_63	6.5.2015	4.01	
P_53	6.5.2015	3.91	
C_96	6.5.2015	5.17	
P_31	6.5.2015	5.16	
C_8	11.5.2015	1.87	Magnatrix
P_61	11.5.2015	2.58	

P_42	11.5.2015	1.97	
P_69	11.5.2015	3.49	
C_59	11.5.2015	3.14	
P_56	11.5.2015	3.12	
P_87	11.5.2015	2.33	
C_68	11.5.2015	2.06	
P_126	11.5.2015	3.18	
CS_41	11.5.2015	5.92	
C_118	11.5.2015	2.74	
C_146	11.5.2015	3.07	
P_120	18.5.2015	1.38	Magnatrix
C_135	18.5.2015	2.57	
C_89	18.5.2015	1.14	
CS_103	18.5.2015	1.54	
C_105	18.5.2015	1.30	
P_79	18.5.2015	1.31	
P_47	18.5.2015	1.50	
C_15	18.5.2015	1.46	
C_33	18.5.2015	4.98	
C_107	18.5.2015	1.38	
P_103	18.5.2015	1.67	
P_88	21.5.2015	1.70	Magnatrix
C_19	21.5.2015	1.74	
C_104	21.5.2015	2.28	
CS_66	21.5.2015	2.46	
P_74	21.5.2015	1.52	
P_116	21.5.2015	2.92	
C_125_II_P	21.5.2015	1.88	
C_115_P	21.5.2015	1.80	
P_21_P	21.5.2015	1.87	
C_106_P	21.5.2015	1.48	
P_110_P	21.5.2015	2.01	
C_93_P	21.5.2015	1.55	
CS_47	25.5.2015	1.63	Magnatrix
P_58	25.5.2015	2.25	
C_116	25.5.2015	1.65	
C_102	25.5.2015	3.09	
C_74	25.5.2015	1.63	
P_71	25.5.2015	1.90	
P_100	25.5.2015	1.82	
C_86	25.5.2015	1.85	
C_90	25.5.2015	2.23	
P_62_P	25.5.2015	2.41	
C_8_P	25.5.2015	2.29	
C_56_P	25.5.2015	2.57	
C_23	27.5.2015	1.61	Magnatrix
P_57	27.5.2015	3.21	
P_65	27.5.2015	2.25	
P_66	27.5.2015	1.95	

P_94	27.5.2015	1.70	
P_99	27.5.2015	1.84	
CS_74	27.5.2015	2.40	
P_43	27.5.2015	1.48	
C_88	27.5.2015	2.06	
C_20	27.5.2015	2.33	
CS_116	27.5.2015	2.88	
C_51	27.5.2015	2.72	
C_148	1.6.2015	3.96	Magnatrix
P_52	1.6.2015	1.13	
C_127	1.6.2015	1.65	
C_111	1.6.2015	1.14	
C_110	1.6.2015	1.52	
C_93	1.6.2015	2.44	
C_152	1.6.2015	1.31	
P_62	1.6.2015	1.47	
CS_31	1.6.2015	1.10	
C_82	1.6.2015	1.09	
C_152_II_P	3.6.2015	2.23	Magnatrix
C_84_P	3.6.2015	2.15	
C_121_P	3.6.2015	2.30	
P_104_II_P	3.6.2015	2.16	
C_127_P	3.6.2015	2.30	
CS_62	3.6.2015	1.80	
P_118	3.6.2015	2.00	
P_85	3.6.2015	4.92	
C_10	3.6.2015	1.98	
C_131	3.6.2015	1.66	
C_87	3.6.2015	2.73	
C_84	3.6.2015	2.15	
P_8	11.6.2015	2.12	Magnatrix
PS_15	11.6.2015	1.70	
CS_114	11.6.2015	1.78	
C_98	11.6.2015	5.94	
C_119	11.6.2015	3.83	
P_83	11.6.2015	8.41	
P_51	11.6.2015	2.51	
C_142	11.6.2015	1.40	
C_136	11.6.2015	2.38	
C_99_II_P	11.6.2015	2.67	
C_85	11.6.2015	2.23	
P_104	11.6.2015	1.90	
CS_119	15-16.6.2015	0.35	Magnatrix
CS_71	15-16.6.2015	1.05	
P_16	15-16.6.2015	0.67	
C_30	15-16.6.2015	0.18	
P_107	15-16.6.2015	0.40	
P_59	15-16.6.2015	0.34	
C_140	17.6.2015	3.52	Manually

C_109_P	10.7.2015	3.26	

Most of the samples produced an sufficient amount of DNA during this step but some of them required additional PCR cycles. Although the order of execution was from the largest to smallest amount of DNA, it seems that the samples which were processed with manual washing and size separation steps produced greater yields than the samples processed with a Magnatrix program.

7.3 Fragment Analyzer

The sample size was checked with Fragment Analyzer after every completed PCR program. An ideal report has the fragment size centred near 400 base pairs. All of the Fragment Analyzer results can be found in the appendix [Appendix 1, Fragment Analyzer figures]. An ideal report would look like the one depicted in Figure 20.

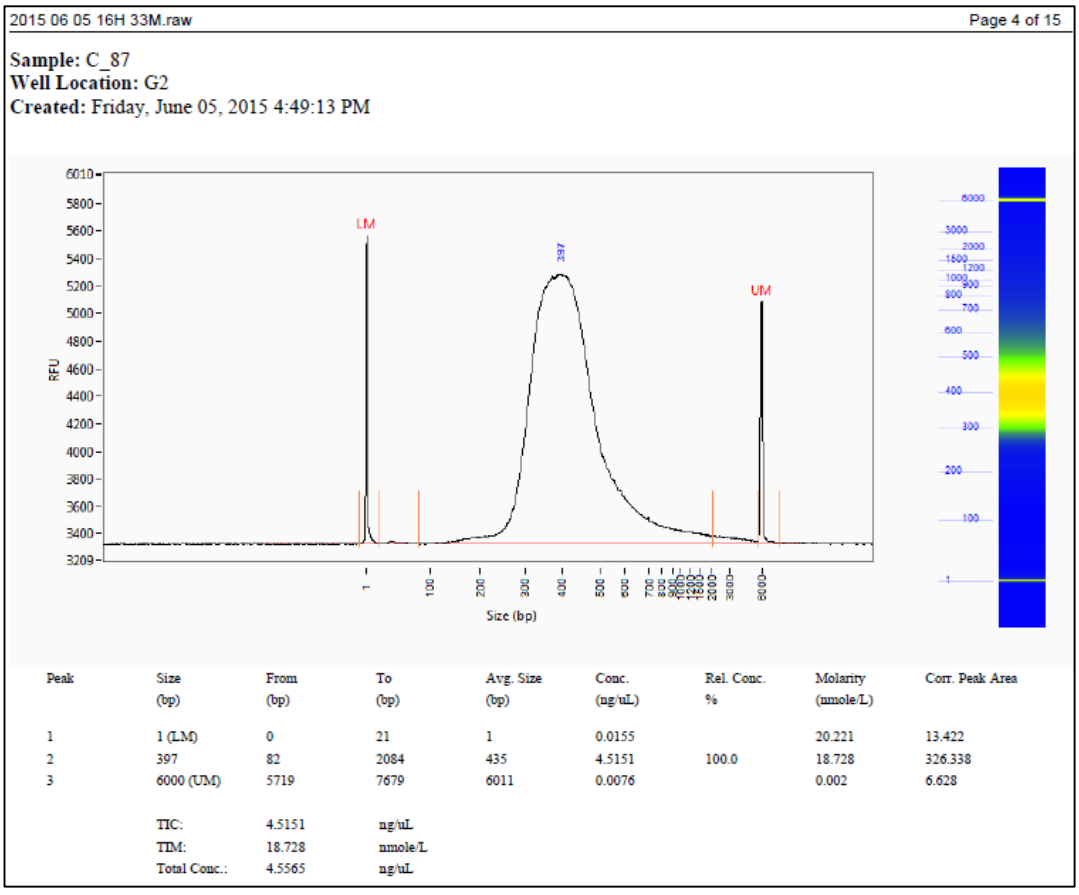
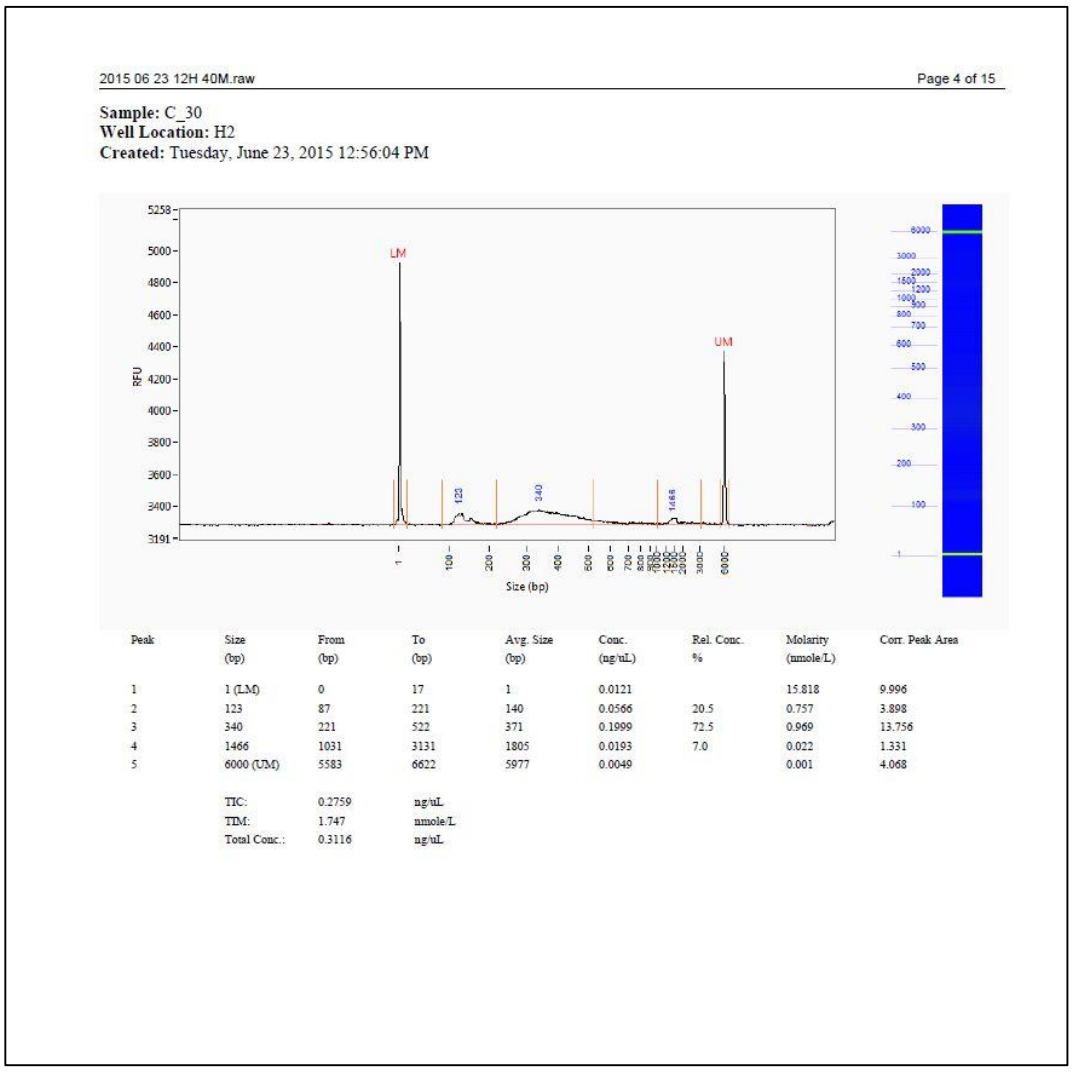


Figure 20. Fragment Analyzer picture of the processed stool DNA sample C_87. The report shows the concentration of the sample and the size and amount of the DNA fragments.

Some of the samples ended up with a lower DNA concentration than others. These samples would have less DNA extracted during DNA extractions and would produce eventually less reads than others. Figure 21 shows one of these samples after the size separation.



Peak	Size (bp)	From (bp)	To (bp)	Avg. Size (bp)	Conc. (ng/uL)	Rel. Conc. %	Molarity (nmole/L)	Corr. Peak Area
1	1 (LM)	0	17	1	0.0121		15.818	9.996
2	123	87	221	140	0.0566	20.5	0.757	3.898
3	340	221	522	371	0.1999	72.5	0.969	13.756
4	1466	1031	3131	1805	0.0193	7.0	0.022	1.331
5	6000 (UM)	5583	6622	5977	0.0049		0.001	4.068

TIC:	0.2759	ng/uL
TIM:	1.747	nmole/L
Total Conc.:	0.3116	ng/uL

Figure 21. Fragment Analyzer picture of the processed stool DNA sample C_30. The report shows the concentration of the sample and the size and amount of the DNA fragments.

Most of the figures produced by Fragment Analyzer were good and within the expected fragment size. Some of the samples had a concentration that was below the average, but even they were all successfully sequenced with the Illumina sequencing machine. The difficult samples were P_16, C_30, P_59, P_107 and CS_119.

7.4 Sequencing quality

Sequencing results can be evaluated by using a sequencing quality control software such as FastQC. By looking at the pictures and comparing them with the ideal control pictures, we can determine how successful the sequencing was. The per base sequence quality pictures and per tile sequence quality pictures were chosen to represent the Illumina results.

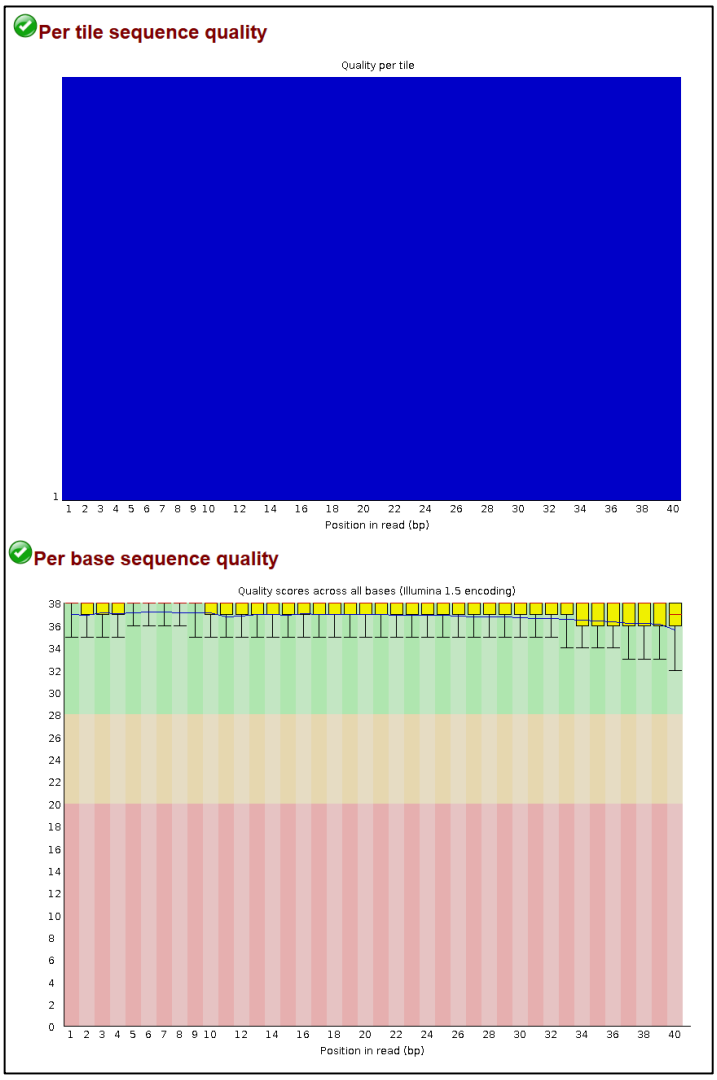


Figure 22. Ideal Fast QC report for a sample sequenced on the Illumina nextseq 500 platform. Per tile sequence quality shows the results in tile colours from blue to red. The redder the tile is, the worse the quality of that base in comparison with the other tiles. Per base sequence quality shows the overview of the range of quality in the same positions as the per tile sequence quality picture. Information on Fast QC reports can be found in [45].

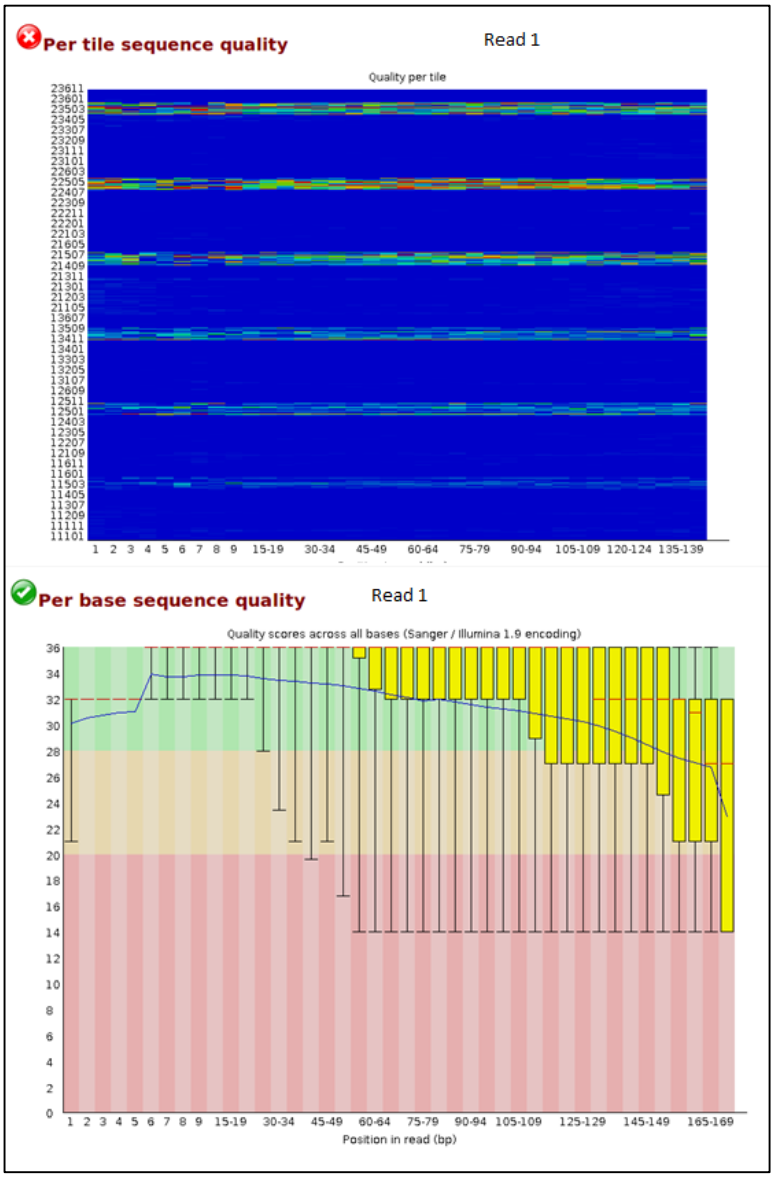


Figure 23. Illumina Fast QC per tile sequence quality and per base sequence quality report for the forward reads of the pooled stool DNA samples.

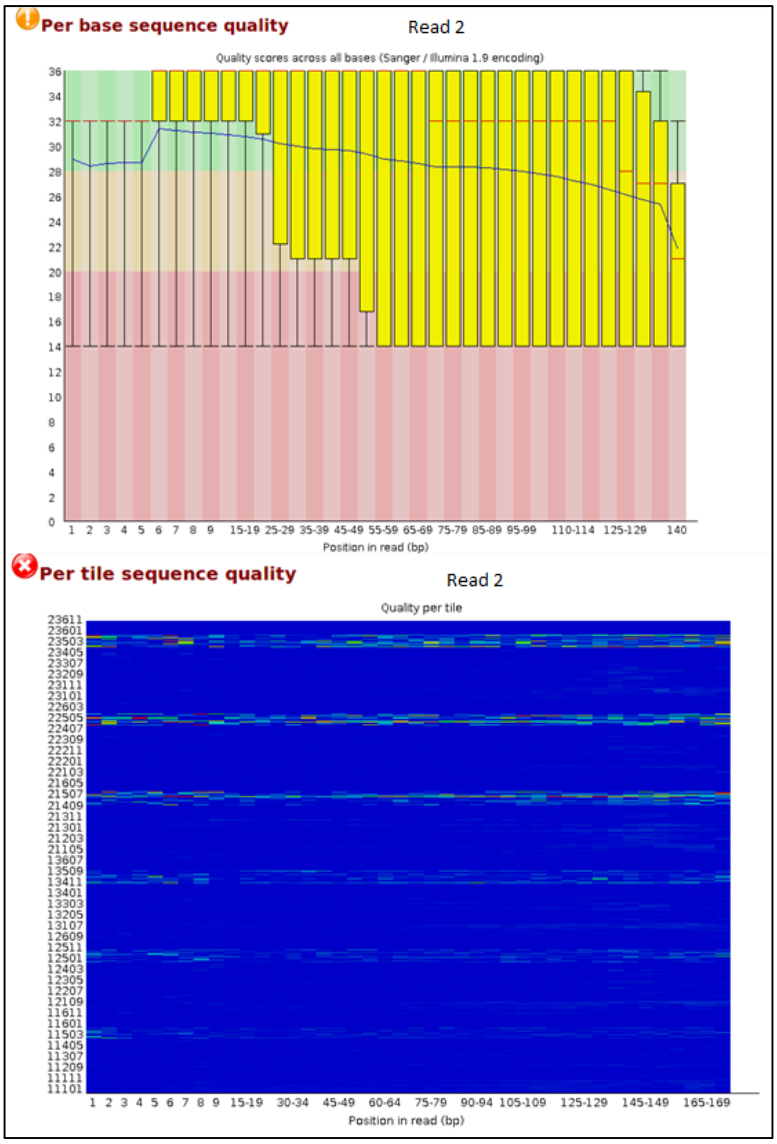


Figure 24. Illumina fast QC per tile sequence quality and per base sequence quality report for the reverse reads of the pooled stool DNA sample.

From these reports we can conclude that the quality of the reads was not as good as it should have been. Especially the tile qualities are far from optimal.

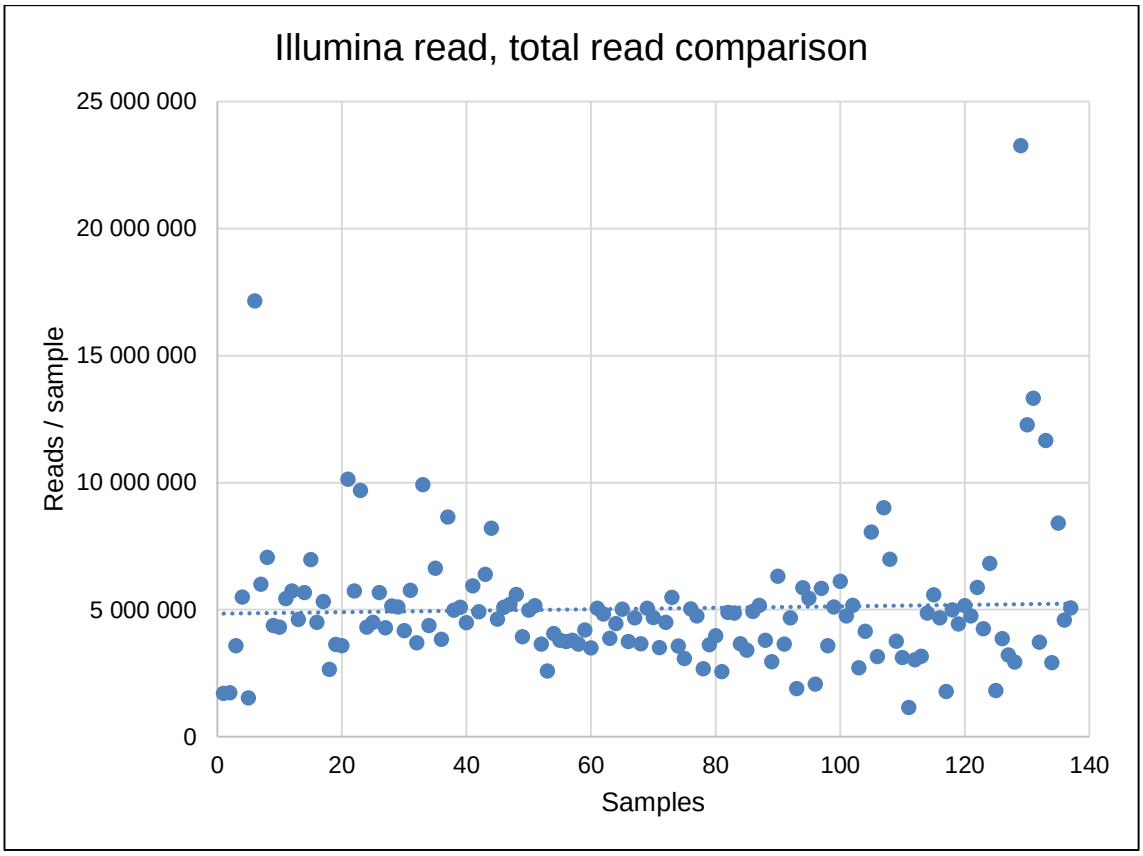


Figure 25. All of the reads produced by the DNA library. One dot represents the amount of reads produced by one sample. The closer to the trend line the dot is, the more evenly it has produced reads in comparison with the other samples. Undetermined fragments were excluded from the figure. They produced 31 457 834 reads which is 4 % from the total.

By looking at Figure 25, we can see that some of the samples produced more reads than the others. This can be due to uneven amounts of DNA in the pool sample or due to reads with poor quality. On the basis of the Fast QC reports for the forward and reverse reads, the latter would seem more probable. This could have been caused by inserting too much of the sample inside the machine.

8 Conclusions

The goal of the thesis was to create a DNA metagenomics library for the 138 PD stool samples and to provide a view on how it could be done. Although DNA could not be extracted from one of the samples, the majority of the samples were successfully sequenced with the Illumina platform. Since the amount of DNA did not correlate with the time parameter in either of the batches, the extractions can be reviewed as trustworthy. The quality of the sequence reads was not optimal. Both these measurements, however, show that the creation of the DNA library was a success, although the sequencing could have succeeded better.

The practical side of the project was a success and there was plenty of useful information in other articles to help in understanding the library preparation process better. Some improvements could be made to the DNA plus kit protocol. At the moment, only the first thermomixer step is done with safe lock tubes. The second thermomixing causes the tubes to pop open randomly. This could lead to cross contamination. After the thesis, Harri Kangas had noticed that the product size from the Bioruptor step seemed to be smaller than it should have. This led to the small fragmented DNA to be washed away in the first washing step, reducing the DNA concentration. This observation was made when the DNA amount was reduced too much during the whole process. The programming for the Bioruptor is being optimized for the next experiments.

References

1. Farlow J, Pankratz ND, Wojcieszek J, Foroud T. Parkinson Disease Overview. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al, editors. GeneReviews(R). Seattle (WA): University of Washington, Seattle; 1993.

2. Parkinsonin tauti [Internet]. Duodecim Terveyskirjasto: © 2015 Kustannus Oy Duodecim; 25.5.2015 [cited 11/11/2015]. Available from: http://www.terveyskirjasto.fi/terveyskirjasto/tk.koti?p_artikkeli=dlk00055.

3. Blausen.com staff. "Blausen gallery 2014". Wikiversity Journal of Medicine 1 (2) [Internet]. 2015 [cited 12/8/2015].

4. Gowers WR. A manual of diseases of the nervous system. London: Churchill; 1886.

5. Mayer EA, Knight R, Mazmanian SK, Cryan JF, Tillisch K. Gut microbes and the brain: paradigm shift in neuroscience. J Neurosci. 2014 Nov 12;34(46):15490-6.

6. Zombie ants have been manipulated by secretions from a killer fungus [Internet]. Bug Bitten: BioMed Central; 2014 [cited 12/11/2015]. Available from: http://blogs.biomedcentral.com/bugbitten/2014/11/05/zombie-ants-have-been-manipulated-by-secretions-from-a-killer-fungus/.

7. Pfeiffer RF. Gastrointestinal dysfunction in Parkinson's disease. Lancet Neurol. 2003 Feb;2(2):107-16.

8. Scheperjans F, Aho V, Pereira PA, Koskinen K, Paulin L, Pekkonen E, et al. Gut microbiota are related to Parkinson's disease and clinical phenotype. Mov Disord. 2015 Mar;30(3):350-8.

9. Mandal RS, Saha S, Das S. Metagenomic Surveys of Gut Microbiota. Genomics, Proteomics & Bioinformatics. 2015 6;13(3):148-58.

10. Sensen CW. Handbook of genome research: genomics, proteomics, metabolomics, bioinformatics, ethical, and legal issues. Weinheim: Wiley-VCH; 2005.

11. Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2012 Feb 9;2(1):3,5783-2-3.

12. Stool Analysis [Internet].: eMedicineHealth; 2012 [cited 23/10/2015]. Available from: http://www.emedicinehealth.com/stool_analysis-health/article_em.htm.

13. PSP Spin Stool DNA/ Plus Kit (REF 1038100x00) [Internet].: STRATEC Molecular GmbH; 2015 [cited 10/8/2015]. Available from: http://www.stratec.com/share/molecular/Manuals/Single/Pathogens/PSPSpinStool_StoolPlusKit.pdf.

14. Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, et al. Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC Microbiol. 2012 Jul 30;12:158,2180-12-158.

15. Roesch LF, Casella G, Simell O, Krischer J, Wasserfall CH, Schatz D, et al. Influence of fecal sample storage on bacterial community diversity. Open Microbiol J. 2009;3:40-6.

16. Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M, Kekkonen RA, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods. 2010 5;81(2):127-34.

17. Smith B, Li N, Andersen AS, Slotved HC, Krogfelt KA. Optimising bacterial DNA extraction from faecal samples: comparison of three methods. Open Microbiol J. 2011;5:14-7.

18. Barrett, K.E. and Barman, S.M. and Boitano, S. and Brooks, H. Ganong's Review of Medical Physiology, 24th Edition. . 2012.

19. Monteiro L, Bonnemaison D, Vekris A, Petry KG, Bonnet J, Vidal R, et al. Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. J Clin Microbiol. 1997 Apr;35(4):995-8.

20. Lipsett P, Gadacz TR. Bile salt binding by maalox, sucralfate, and meciadanol: in vitro and clinical comparisons. J Surg Res. 1989 Nov;47(5):403-6.

21. Boddinghaus B, Wichelhaus TA, Brade V, Bittner T. Removal of PCR inhibitors by silica membranes: evaluating the Amplicor Mycobacterium tuberculosis kit. J Clin Microbiol. 2001 Oct;39(10):3750-2.

22. Image:Inhibitor schematic [Internet]. OpenWetWare: OpenWetWare; 2010 [cited 12/11/2015]. Available from: http://openwetware.org/wiki/Image:Inhibitor_schematic.png.

23. PRODUCT INFORMATION Proteinase K (recombinant), PCR grade [Internet].: Thermo Scientific; 2012 [cited 23/10/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0012880_Proteinase_K_%28recombinant%29_PCR_Grade_eo0491_UG.pdf.

24. Hilz H, Wiegers U, Adamietz P. Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of 'masked' proteins. Eur J Biochem. 1975 Aug 1;56(1):103-8.

25. Proteinase K [Internet].: Wothington Biochemical Corporation; 2015 [cited 12/11/2015]. Available from: http://www.worthington-biochem.com/PROK/.

26. Melzaka KA, Sherwooda CS, Turnerb RFB, Haynesc CA. Driving forces for DNA adsorption to silica perchlorate solutions. . 1996;181(2):635.

27. DNA binding to silica in the precence of chaotrophic salt [Internet].: Arizona State University; 2007 [cited 12/11/2015]. Available from: http://bioenergy.asu.edu/photosyn/courses/bio_343/lab/experiment-ii.html.

28. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. BioTechniques. 2014 Feb 1;56(2):61,4, 66, 68, passim.

29. PROTOCOL DNA Shearing for Bioruptor® Standard & Bioruptor® Plus [Internet].: diagenode; 2012 [cited 10/16/2015]. Available from: https://www.diagenode.com/files/protocols/DNA_Shearing_for_Standard_and_Plus_protocol.pdf.

30. PRODUCT INFORMATION T4 DNA Polymerase [Internet].: Thermo SCIENTIFIC; 2012 [cited 10/16/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0012013_T4_DNAPolymerase_5_UuL_100U_UG.pdf.

31. Khare V, Eckert KA. The proofreading 3'-->5' exonuclease activity of DNA polymerases: a kinetic barrier to translesion DNA synthesis. Mutat Res. 2002 Dec 29;510(1-2):45-54.

32. Bovine Serum Albumin [Internet].: invitrogen corporation; 2008 [cited 10/14/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/bsa_man.pdf.

33. PRODUCT INFORMATION T4 Polynucleotide Kinase [Internet].: Thermo SCIENTIFIC; 2012 [cited 10/14/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0011986_T4_Polynucleotide_Kinase_UG.pdf.

34. [Internet].: OpenWetWare; 2009 [cited 11/5/2015]. Available from: http://openwetware.org/wiki/Addition_of_3%27_A_overhangs_to_PCR_products.

35. PRODUCT INFORMATION dNTP Mix, 25 mM each, [Internet].: Thermo SCIENTIFIC; 2012 [cited 10/14/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0012969_dNTP%20Mix%2025mM_UG.pdf.

36. PRODUCT INFORMATION Thermo Scientific DreamTaq DNA Polymerase [Internet].: Thermo SCIENTIFIC; 2013 [cited 10/14/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0012035_DreamTaqDNAPolymerase_200U_UG.pdf.

37. PRODUCT INFORMATION T4 DNA Ligase, 30 Weiss U/µL [Internet].: Thermo SCIENTIFIC; 2014 [cited 10/16/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0011989_T4_DNA_Ligase_30_Weiss_5000_Weiss_U_UG.pdf.

38. PRODUCT INFORMATION Thermo Scientific Phusion Hot Start II High-Fidelity DNA Polymerase [Internet].: Thermo SCIENTIFIC; 2012 [cited 10/14/2015]. Available from: https://tools.thermofisher.com/content/sfs/manuals/MAN0012401_Phusion_HotStartII_DNAPolymerase_500U_UG.pdf.

39. NextSeq Series Desktop Sequencer [Internet].: © 2015 Illumina, Inc; 2015 [cited 12/2/2015]. Available from: http://www.illumina.com/systems/nextseq-sequencer/technology.html.

40. Introduction to Illumina sequencing [Internet].: Tomato Epigenome Database; 2013 [cited 12/11/2015]. Available from: http://ted.bti.cornell.edu/cgi-bin/epigenome/method-1.cgi.

41. Oberauner L, Zachow C, Lackner S, Hogenauer C, Smolle KH, Berg G. The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. Sci Rep. 2013;3:1413.

42. DeLeon-Rodriguez N, Lathem TL, Rodriguez-R LM, Barazesh JM, Anderson BE, Beyersdorf AJ, et al. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc Natl Acad Sci U S A. 2013 Feb 12;110(7):2575-80.

43. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014 Nov 12;12:87,014-0087-z.

44. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010 Oct;11(10):733-9.

45. FastQC [Internet].: Babraham Bioinformatics [cited 12/7/2015]. Available from: http://www.refworks.com/refworks2/default.aspx?r=references|MainLayout::init.

Appendix 1

Fragment Analyzer figures

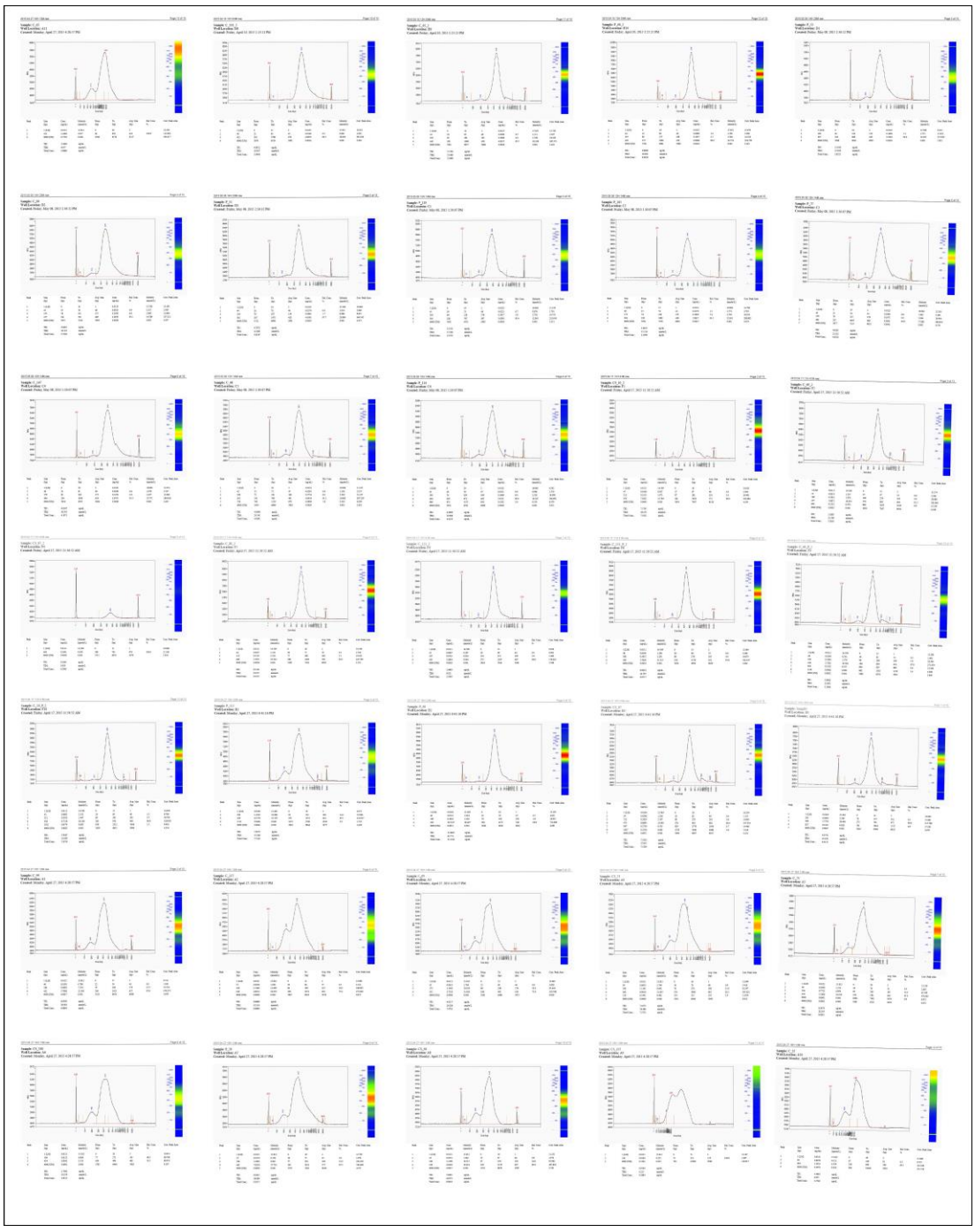


Figure 26. Fragment Analyzer figures for 35 samples compressed into one picture April 27th. The pictures show the amounts and overall distributions of different sized DNA fragments within the samples. 400 bp was the wanted size.

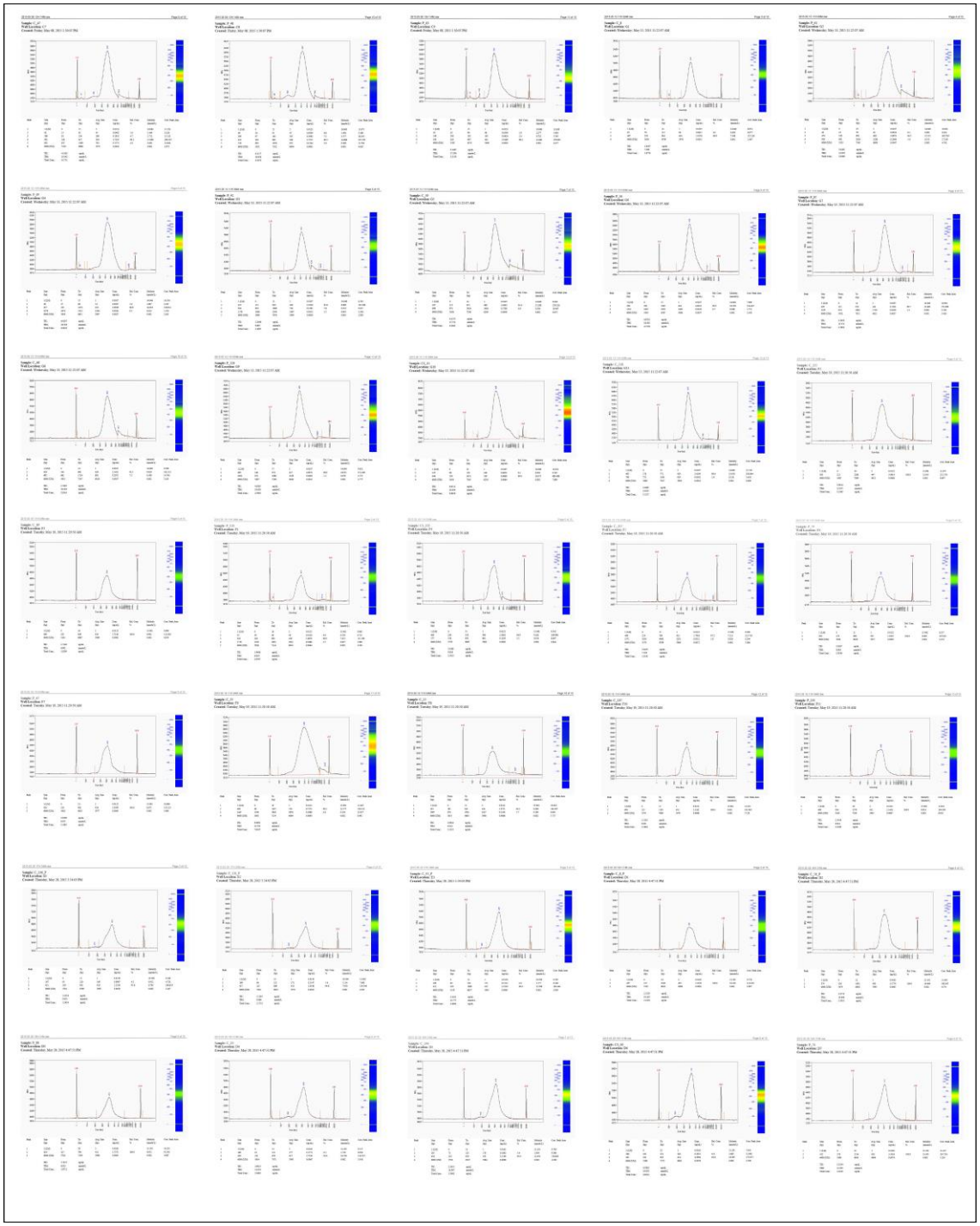


Figure 27. Fragment Analyzer figures for 35 samples compressed into one picture May 8th - May 28th. The pictures show the amounts and overall distributions of different sized DNA fragments within the samples. 400 bp was the wanted size.

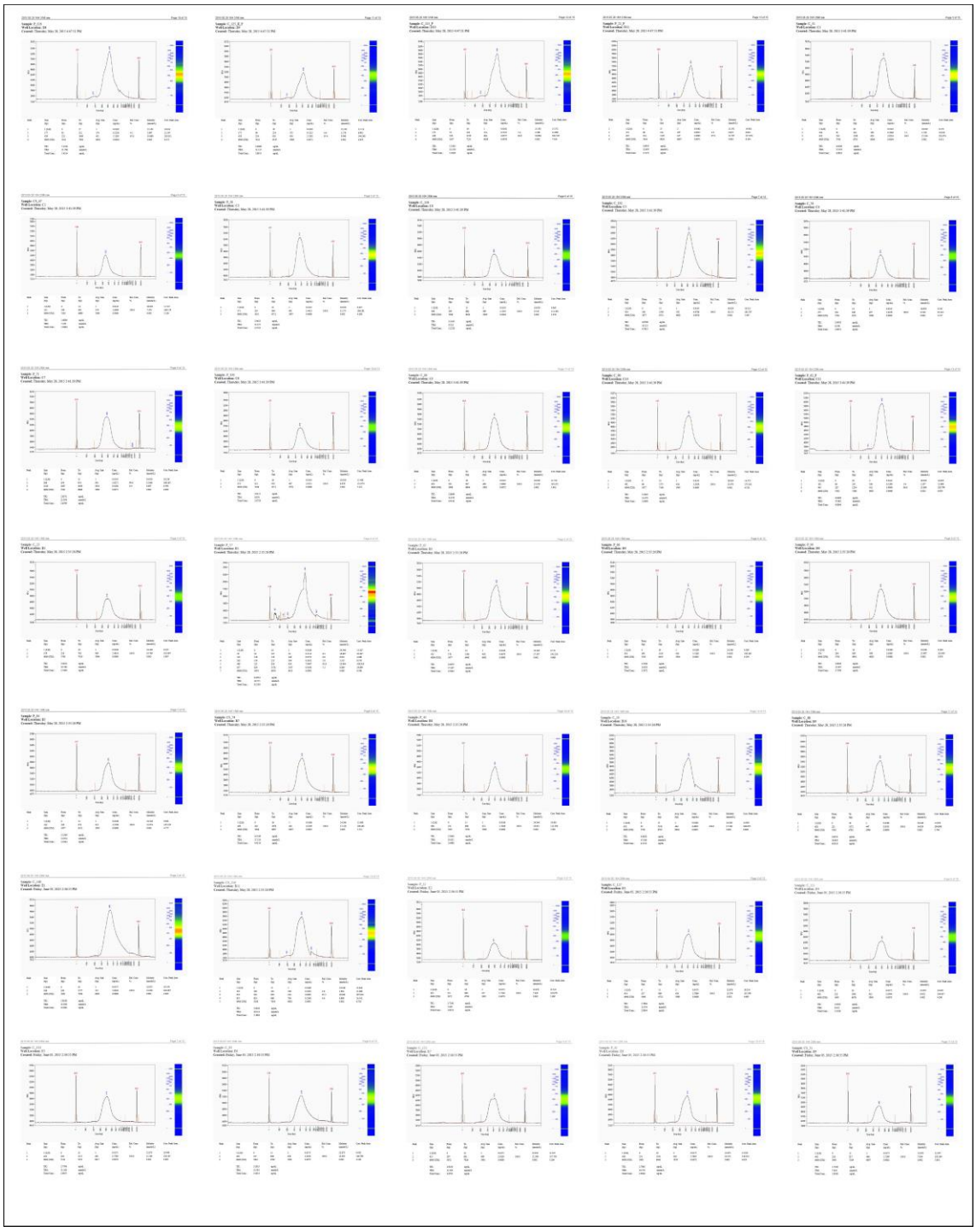


Figure 28. Fragment Analyzer figures for 35 samples compressed into one picture May 28th - June 5th. The pictures show the amounts and overall distributions of different sized DNA fragments within the samples. 400 bp was the wanted size.

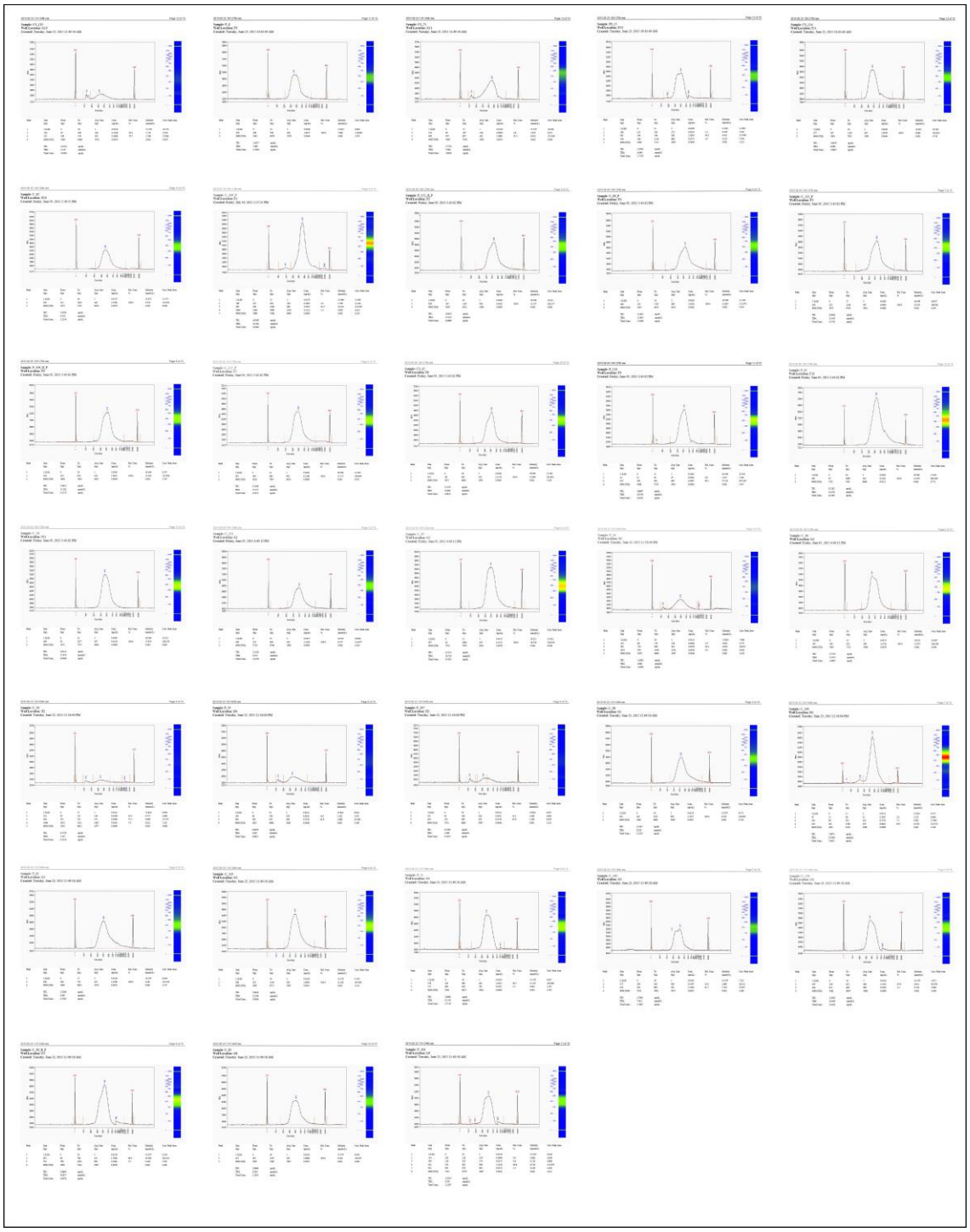


Figure 29. Fragment Analyzer figures for 33 samples compressed into one picture June 23rd. The pictures show the amounts and overall distributions of different sized DNA fragments within the samples. 400 bp was the wanted size

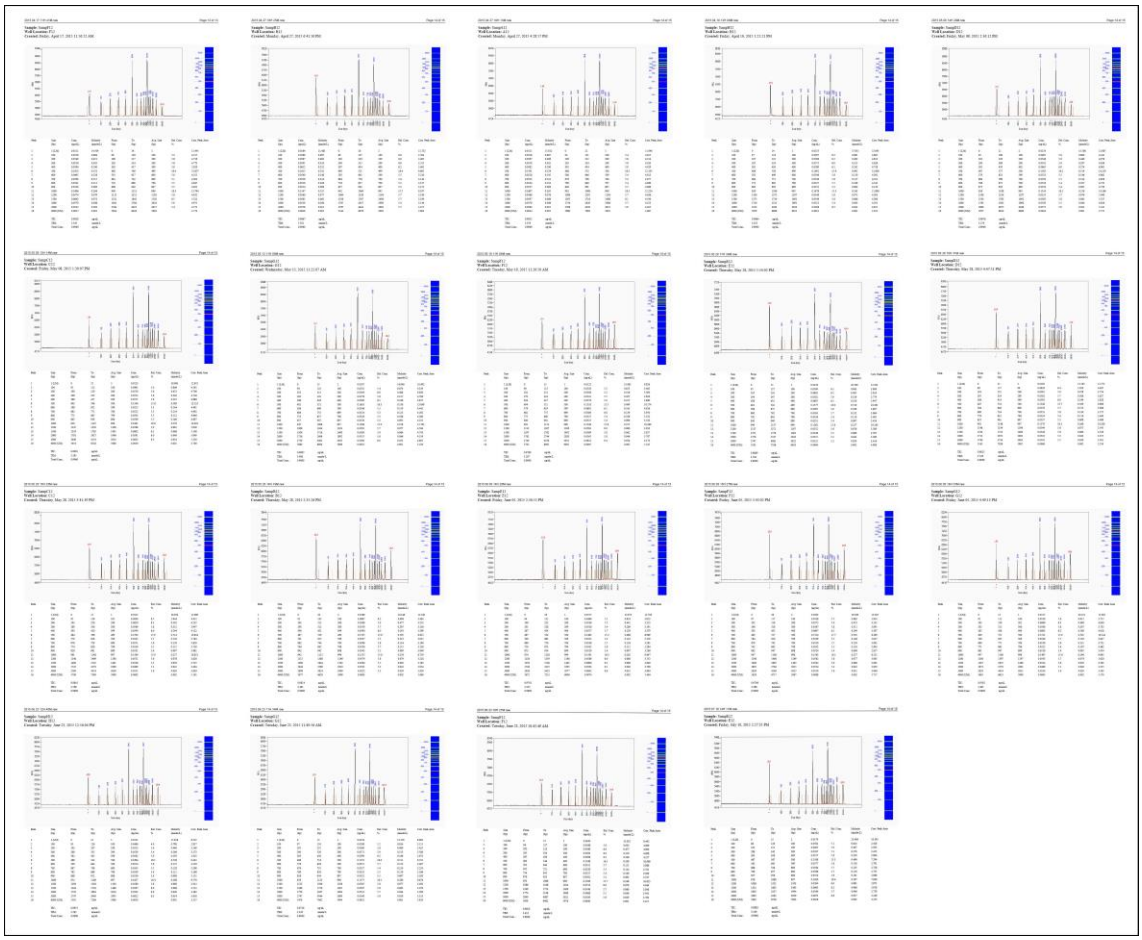


Figure 30. Fragment Analyzer figures for the controls April 27th - June 23rd.

Appendix 2

The size separation step protocol

PEG-size separation 2.0
29.6.2012, Harri

Start off by determing the right PEG concentrations for the binding step. Greater PEG concentration makes the binding of smaller fragments possible. Roughly described: 7 % of PEG binds product bigger than 1000 bp, 9 % bigger than 300-400 bp and 11 % bigger than 150-200 bp. The values are an approximation and the accurate values change between the batches. Check the out coming size with a control sample.

P:\PEG\ -directory contains the recipes for the size separation "kokoerottelureseptit.xlsx" -file. Input the needed PEG % to the field and write down the needed 5 M NaCl, MQ:n and PEG-stock amounts.

1. Take carboxy-beads into a clear Eppendorf tube. (15 µl binds 1 µg of DNA, do not use less than 10 µl without a good reason)
2. Place the Eppondorf on a magnet and **remove the liquid**.
3. Remove the Eppendorf off the magnet and add **5 M NaCl and MQ**. Suspend the beads with a pipet.
4. Add the sample
5. **Add PEG stock 1** (25 % PEG, 1,5 M NaCl). Pipet slowly because PEG is thick and sticks to the tip of the pipet easily when pipetted quickly. Mix well.
6. Incubate for **6** min. Do not incubate past this time limit! You can mix the sample gently during the incubation for more accurate results, though it is not mandatory.

   Meanwhile prepare the 2nd step Eppendorf tube:

   1.1. Take unused beads to a new eppendorf and remove the buffer with a magnet

   1.2. Add 1 M NaCl and suspend the beads

   1.3. Add stock 2 (25 % PEG/1 M NaCl) the needed amount and mix

7. **Separate the beads** from the old Eppendorf tube with a magnet.
8. **Move the supernatant** to step 2. eppendorf tube and mix with a pipet. Throw the step 1 eppendorf away with the beads (These beads have longer DNA than is required. If you want to keep the long products, do the washing steps starting with the step 11.)
9. Incubate for **10 min**
10. Separate the beads with a magnet and remove the supernatant.
11. Add 400 µl of fresh **80 % EtOH** and mix the beads.
12. **Remove the EtOH** with a magnet.
13. **Dry** the beads. 5min RT is enough.
14. **Elution**: add for example. 25 µl **EB**, 0,1x TE or MQ. Suspend the beads well and incubate for 3-5 minutes.
15. Separate the beads with a magnet. Pipet the liquid into a new eppendorf.

(Translated into enlgish by Simo)

Protocol 1. Size separation and washing protocol for the PD stool sample DNA.

Appendix 3

Clean up protocol.

PEG-cleanup 2.0
9.19.2013, Harri

Start by finding out the correct PEG % to be used. The higher the concentration, the shorter DNA is bound to beads.

Roughly: 7 % PEG binds longer than 900-1000 bp products. 9 % over 300-400 bp. And 11 % over ~150bp. Lower than 5,5 % and the precipitation does not happen at all. 5,8 is still safe, it will bind longer than ~1,5kb DNA. Check the correct % before starting!

Calculator is found at P:\PEG\ -folder. Enter your PEG % and follow the recipe.

1. Take MyOne beads to clean eppendorf. (~20 µl for 1µg DNA, dont use <10µl without a good reason)
2. put tube on magnet and **remove supernatant.**
3. Remove tube from magnet and add **5 M NaCl and MQ**
4. Mix well by pipetting up and down
5. Add sample and mix
6. **Add PEG stock 1** (25 % PEG, 1,5 MNaCl). Use lowbind pipette tips and pipette slowly. Stock 1 is viscous. Mix well
7. Incubate **10 min**. It is a good idea to mix 1-2 during the incubation.
8. **Separate** the beads on magnet
9. **Remove supernatant** (Save it if you wish to keep products shorter than the cutoff! This will need another bind according to size-selection recipes.)
10. Add 400 µl or more of **fresh 80 % EtOH** while still on magnet
11. **Remove EtOH** bypipetting. Wash another time and remove again.
12. Repeat the wash step
13. **Dry beads for** 5min RT. Beads should be left still a bit moist to make elution easier (light brown colour is too dry)
14. Remove from magnet
15. Elution: Add EB, 0,1x TE or MQ to elute the DNA. For routine use use half of the starting volume. Otherwise you can elute to any volume as long as the beads are suspended to the buffer.
16. Mix Well and incubate at least 5 min at RT.
17. Separate on magnet and transfer the DNA containing supernatant to a clean tube.

Protocol 2. Washing step protocol for the PD stool sample DNA.

Appendix 4

Norppa PCR protocol for NGS library

Norppa-libraries NextSeq500

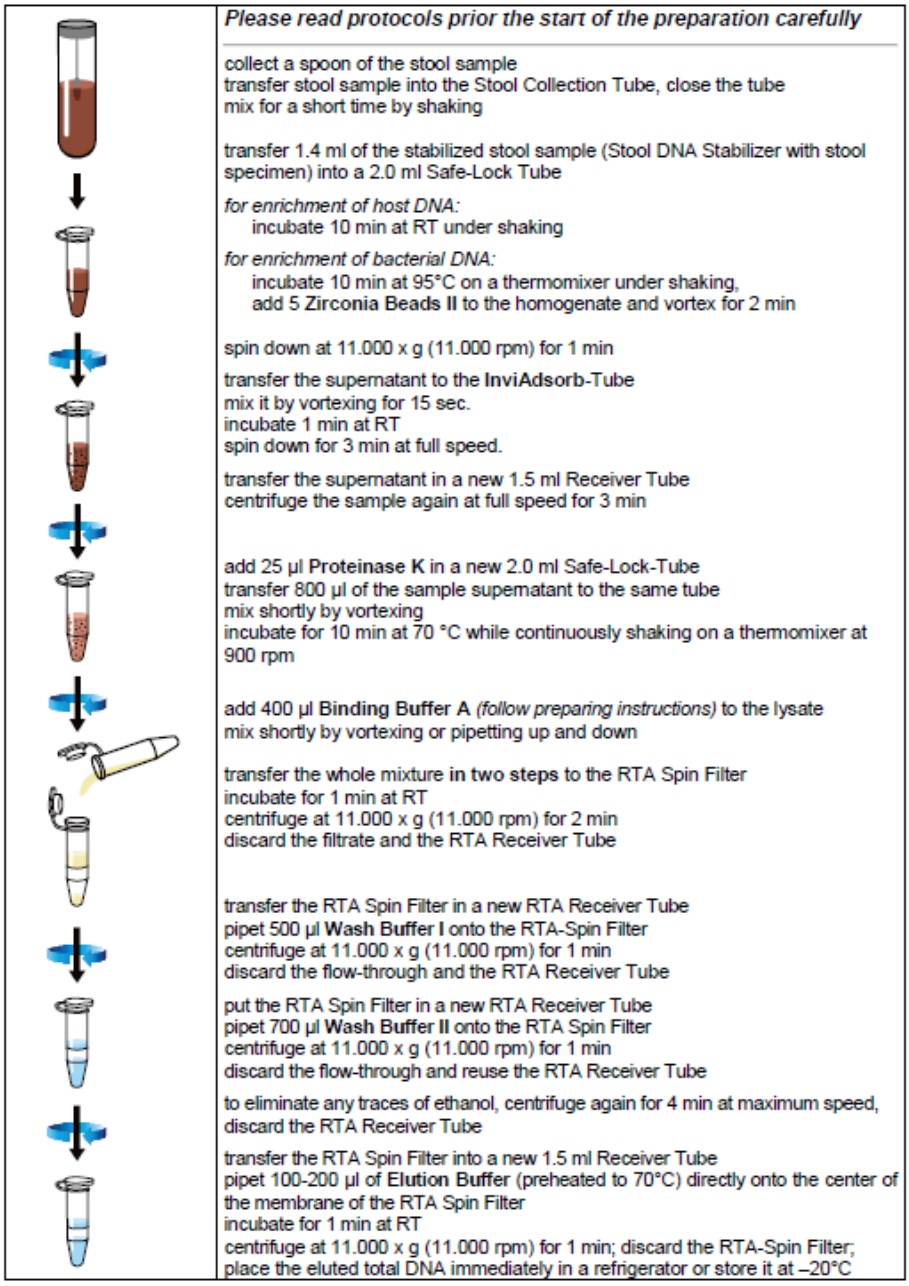
- Samples are found in boxes at the 67 - 20°C -freezer
- Fragmentation with Bioruptor 1 µg (Or 100 µl, if the concentration is small), V = 100 µl;
- Bioruptor-program: 16 cycles, 30s/30s
- Washing step with 9,25 % PEG by magnatrix, elution 25 µl EB
- **End Repair (V = 50 µl)**
  - **MQ** **10,1 µl**
  - **T4 Polynucleotide Kinase Buffer A (10x)** **2,5 µl**
  - **T4 DNA Polymerase Buffer (5x)** **5 µl**
  - **BSA (10 mg/ml)** **0,25 µl**
  - **ATP (10 mM)** **2,5 µl**
  - **dNTP (25 mM)** **0,4 µl**
  - **T4 Polynucleotide Kinase (10 U/µl)** **2,5 µl**
  - **T4 DNA Polymerase (5 U/µl) 1,5 µl**
  - **DreamTaq Polymerase (5 U/µl)** **0,25 µl**
  - **DNA** **25 µl**
- **End Repair –program**
  - **25 °C** **20 min**
  - **72 °C** **20 min**
  - **4 °C** **10 min**
- Washing step with 9,25 % PEG by Magnatrix, elution 32 µl EB
- **Adapter ligation (V = 40 µl)**
  - **DNA** **32 µl**
  - **T4 DNA Ligase Buffer (10x)** **4 µl**
  - **Y-adapteri (TruSeq_A + TruSeq_cB)** **2 µl**
  - **T4 DNA Ligase (30 U/µl)** **2 µl**
- Incubation 60 min, +25 °C
- Washing step with 8,7 % PEG by Magnatrix, elution 20 µl EB
- **PCR (V = 100 µl)**
  - **MQ** **75,2 µl**
  - **5x Phusion HF Buffer** **20 µl**
  - **dNTP (25 mM)** **0,8 µl**
  - **PCR_TruSeq_A (10 pmol)** **1 µl**
  - **Ind_8bp_XXX (10 pmol)** **1 µl**
  - **Phusion HotStart II (2 U/µl) 1 µl**
  - **DNA** **2,5 ng**
- **PCR-program (18 cycles)**
  - **95 °C** **5 min**
  - **98 °C** **20 s**
  - **60 °C** **15 s**
  - **72 °C** **30 s**
  - **go to 2 x 18 times**
  - **72 °C** **5 min**
  - **4 °C** **5 min**
- Size separation with 9 – 8,5 % PEG (targetted size is approximately 350-400 bp) -> run with HS-chip Bioanalyser

Protocol 3. Norppa PCR protocol for the PD stool sample DNA.

Appendix 5

PSP spin stool DNA plus kit protocol

Scheme of the PSP® Spin Stool DNA *Plus* Kit



Please read protocols prior the start of the preparation carefully

collect a spoon of the stool sample
transfer stool sample into the Stool Collection Tube, close the tube
mix for a short time by shaking

transfer 1.4 ml of the stabilized stool sample (Stool DNA Stabilizer with stool specimen) into a 2.0 ml Safe-Lock Tube

for enrichment of host DNA:
incubate 10 min at RT under shaking

for enrichment of bacterial DNA:
incubate 10 min at 95°C on a thermomixer under shaking,
add 5 **Zirconia Beads II** to the homogenate and vortex for 2 min

spin down at 11.000 x g (11.000 rpm) for 1 min

transfer the supernatant to the **InviAdsorb**-Tube
mix it by vortexing for 15 sec.
incubate 1 min at RT
spin down for 3 min at full speed.

transfer the supernatant in a new 1.5 ml Receiver Tube
centrifuge the sample again at full speed for 3 min

add 25 µl **Proteinase K** in a new 2.0 ml Safe-Lock-Tube
transfer 800 µl of the sample supernatant to the same tube
mix shortly by vortexing
incubate for 10 min at 70 °C while continuously shaking on a thermomixer at 900 rpm

add 400 µl **Binding Buffer A** *(follow preparing instructions)* to the lysate
mix shortly by vortexing or pipetting up and down

transfer the whole mixture **in two steps** to the RTA Spin Filter
incubate for 1 min at RT
centrifuge at 11.000 x g (11.000 rpm) for 2 min
discard the filtrate and the RTA Receiver Tube

transfer the RTA Spin Filter in a new RTA Receiver Tube
pipet 500 µl **Wash Buffer I** onto the RTA-Spin Filter
centrifuge at 11.000 x g (11.000 rpm) for 1 min
discard the flow-through and the RTA Receiver Tube

put the RTA Spin Filter in a new RTA Receiver Tube
pipet 700 µl **Wash Buffer II** onto the RTA Spin Filter
centrifuge at 11.000 x g (11.000 rpm) for 1 min
discard the flow-through and reuse the RTA Receiver Tube

to eliminate any traces of ethanol, centrifuge again for 4 min at maximum speed,
discard the RTA Receiver Tube

transfer the RTA Spin Filter into a new 1.5 ml Receiver Tube
pipet 100-200 µl of **Elution Buffer** (preheated to 70°C) directly onto the center of the membrane of the RTA Spin Filter
incubate for 1 min at RT
centrifuge at 11.000 x g (11.000 rpm) for 1 min; discard the RTA-Spin Filter;
place the eluted total DNA immediately in a refrigerator or store it at –20°C

20

PSP® Spin Stool DNA Kit 0515
PSP® Spin Stool DNA Plus Kit 0515

Protocol 4. PSP spin stool DNA plus kit protocol scheme used for the extractions of the PD stool samples