Detection of the Archaeome in the Human Gastrointestinal Tract

Hafsteinn Esjar Baldursson

Verk-og náttúrufræðideild
Háskóli Íslands
2020

Detection of the Archaeome in the Human Gastrointestinal Tract

Hafsteinn Esjar Baldursson

12 Eininga ritgerð sem er hluti af
Baccalaureus Scientiarum gráðu í Lífefna- og Sameindalíffræði

Leiðbeinendur
Viggó Þór Marteinsson
Sigurlaug Skírnisdóttir
Stephen Knobloch

Líf- og umhverfisvísindadeild
Verkfræði- og náttúruvísindasvið
Háskóli Íslands
Reykjavík, júní 2020

Detection of the Archaeome in the Human Gastrointestinal Tract.
12 eininga ritgerð sem er hluti af *Baccalaureus Scientiarum* gráðu í lífefna- og sameindalíffræði

Líf- og umhverfisvísindadeild
Verkfræði- og náttúruvísindasvið
Háskóli Íslands
Askja-Sturlugata 7
101 Reykjavík

Sími: 525 4000

Skráningarupplýsingar:
Hafsteinn Esjar Baldursson, 2020, Detection of the Archaeome in the Human Gastrointestinal Tract, BS ritgerð, Líf- og umhverfisvísindadeild, Háskóli Íslands, 46 bls.

Útdráttur

Mannslíkaminn hefur að geyma billjónir örvera sem eru jafnvel fleiri en líkamsfrumurnar. Einnig býr hann yfir hundrað sinnum stærra bakteríuerfðamengi en erfðamengi mannsfrumnanna sjálfra. Rannsóknir á örveruflóru mannsins, svo sem hefðbundnar greiningar á 16S rRNA genum, hafa einblínt á bakteríuhluta hennar vegna þess að skort hefur aðferðir til þess að greina og meta aðra þætti, eins og arkeuflóruna. Í þessu verkefni verður leitast eftir að bæta þekkt vandamál við greiningu arkea í meltingarfærum.

Í verkefninu voru bornar saman sýnatökuaðferðir og aðferðasett. Með þekktum aðferðum við PCR mögnun tókst að magna hluta 16S rRNA gensins og raðgreining genabútanna gat farið fram með Illumina Miseq raðgreiningartækni. Niðurstöðurnar sýndu lítinn mun á örveruflóru milli sýnanna, sem fengust með ólíkum sýnatökuaðferðum. Niðurstöður tveggja sýna bentu til þess að 0,4% örveruflórunnar tilheyrðu arkeum, nánar tiltekið tegundinni *Methanobrevibacter smithii*, þegar notaður voru almennir vísar. Með stigskiptri PCR uppsetningu greindist einu raðbrigði meira (samtals tvö) en bæði komu þó frá *M. smithii.*

Að lokum var gerð magngreining á dreifkjörnungum í saursýnum með qPCR og SYBR aðferð. Niðurstöður bentu til þess að u.þ.b. 10^7 eintök 16S rRNA bakteríugena væru til staðar í hverju sýni. Einnig að hlutfall arkea af dreifkjörnungum væri 0,5% til 1% í sýnum þar sem arkeur greindust. Ekki greindust arkeur í öllum sýnum sem til skoðunar voru. Helsta vandamálið sem upp kom við verkefnið var að mæla DNA styrk staðla af nægilegri nákvæmni til að gera staðalkúrfur sambærilegar milli mælinga. Auðveldlega má bæta úr því við frekar rannsóknir með ítarlegri mælingum á styrk staðla fyrir qPCR.

Abstract

The human body carries trillions of microbial cells, even outnumbering human cells, and houses about hundred times more bacterial genes then human genes. Research on the human microbiome has focused on the bacterial domain and because of methodological problems, the presence and abundance of archaea has been consistently overlooked in conventional 16S rRNA gene amplicon surveys. In this project several of the known methodological problems to detect archaea in gastrointestinal tract are addressed in order to improve detection methods and quantify archaeal signatures.

Sampling methods and specialized fecal sampling kits were compared. Also,different methodological set-ups using different primer conditions and previously established PCR procedures were tested. Amplification of the 16S rRNA gene was successful and enabled amplicon analysis with the Illumina Miseq next generation sequencing. Little difference in microbial communities was detected between samples acquired with different sampling kits. Data analysis revealed about 0.4% archaeal signatures of *Methanobrevibacter* in two of the stool test samples when using universal primers.With a nested methodological PCR setup, one additional archaeal sequence variant was detected despite originating from the same archaeal species.

Lastly, the archaeal and bacterial signatures were quantified using a qPCR SYBR approach, revealing a bacterial load of 10^7 copies of 16S rRNA gene per total DNA in the stool samples. Of the total stool DNA between 0.5% and 1% were attributed to archaeal signatures. The main issue of the project was in accurately measuring the DNA concentration of standard samples and optimisation of the standard. This could easily be resolved in future studies with more precise quantification allowing for more accurate results.

Table of Contents

Figures

Tables

Abbreviations

DNA: Deoxyribonucleic acid

PCR: Polymerase chain reaction

qPCR: Quantitative/Real-time polymerase chain reaction

NEB: New England Biolabs

CFU: Colony Forming Unit

dsDNA: Double stranded DNA

BR: Broad range

HS: High sensitivity

Bp: Base pairs

ASV: Amplicon Sequence Variant

RSV: Ribosomal Sequence Variant

NTC: No template control

Ct: Cycle threshold

Tm: melting temperature (of nucleic acid strand)

Acknowledgements

I want to thank Matis and especially Viggó Þór Marteinsson, Sigurlaug Skírnisdóttir and Stephen Knobloch for giving me the opportunity and resources to work on this project, as well as Matis's chef, Guðlaug for keeping me and my microbiota well-nourished and healthy during my project. I particularly want to thank Stephen and Sigurlaug for the assistance, guidance, and expert advice they gave me during my time at Matis, even despite a global pandemic. I would also like to thank them for taking the time to read and correct my essay. At last I would like to thank my supervisory teacher at the University, Snædís Huld Björnsdóttir for accepting this project and proofreading my essay.

1 Introduction

1.1 Microorganisms

Microorganisms are both prokaryotes and eukaryotes, such as bacteria, archaea, yeast and fungi and also viruses. The classification includes microorganisms that are both host dependent and those that can reproduce outside of a host (Schulz & Jørgensen, 2001). Classifications of microorganisms has relied on several frameworks through the past century. In recent decades, DNA sequencing has provided a common framework that enables scientist to link different organism together based on the 16S rRNA gene. A phylogenetic tree inferred from rRNA sequences has little resemblance to the taxa defined tree by classical phenotypic criteria. The latter definition relied heavily on morphology, which in retrospect is an almost useless indicator of relationships (Olsen & Woese, 1993). Molecular comparisons via the small subunit rRNA sequences show that life on this planet divides into three primary groupings known as bacteria, archaea, and eukaryotes. The differences that separate them are of a more profound nature than the differences that separate typical kingdoms, such as animals and plants and are therefore classified in domains above the level of kingdoms. The three domains can be linked together in a universal phylogenetic tree defined by Woese and Olsen (Olsen & Woese, 1993; Woese et al., 1990).

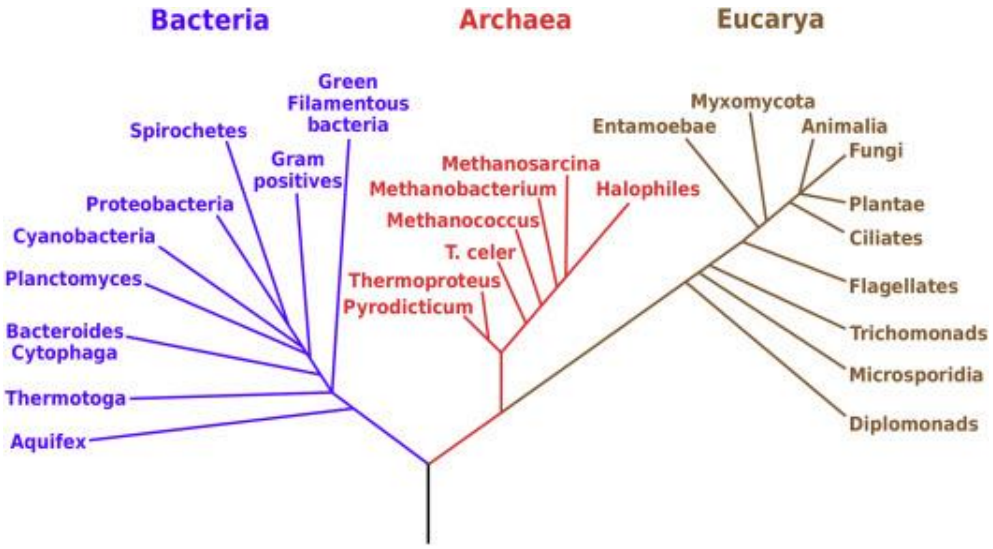

Figure 1: Rooted universal phylogenetic tree showing the three domains based upon 16S rRNA sequences. (Olsen & Woese, 1993).

Most of the earth's prokaryotes reside in three large habitats: open ocean, soil, and the sediment/soil subsurface totalling about $4–6 \times 10^{30}$ cells and represent a large portion of the known genetic diversity of life (Whitman et al., 1998). They are an essential component of the earth's biota (ecological community of microorganisms), producers of important components of the atmosphere and catalyse crucial transformations in the biogeochemical cycles of the biosphere. Likewise they are an essential part of higher life forms inhabiting the earth, each of whom has its own unique microbiome (genetic material of all the microbes), crucial for their viability and that plays important roles in nutrition and health

(Conlon & Bird, 2015; Whitman et al., 1998). In animals, prokaryotes are abundant on the skin, and within the gastrointestinal tract and their numbers are enormous (Drasar & Barrow, 1985; Sender et al., 2016). For comparison, based on the surface area of a human adult, the total number of prokaryotes on the skin is about 3×10^8 cells. But that is far below the number of prokaryotic cells found in the digestive system. The vast majority of the microbes in the gastrointestinal tract live in the colon, which harbors around 3.2×10^{11} prokaryotic cells per millilitre of organ content (Cummings & MacFarlane, 1997; Sender et al., 2016).

1.2 The Human Gut Microbiome

The human body carries trillions of microbial cells, even outnumbering human cells, and houses about hundred times more bacterial genes then human genes (Sender et al. 2016). The majority resides in the gut, contributing to energy harvest from food. They synthesize essential amino acids and vitamins and process components of the diet otherwise indigestible such as fibres, or polysaccharides. Some microbial products can be toxic to the host, especially if the microbiota gets imbalanced and its population changes (Qin et al., 2010). Dietary means, particularly the use of a range of fibers strengthens the hosts immune defence. Fiber consumption may be the best way of maintaining a healthy gut microbiota population that helps to keep harmful bacteria at bay by competing for nutrients and sites of colonization (Conlon & Bird, 2015). Host immune defences strengthened by the microbiota includes a mucus barrier along the intestine that prevent potentially harmful bacteria from causing damage to tissues and migrating through the epithelial barrier (Conlon & Bird, 2015; Salzman et al., 2009).

The importance of microbial communities to human health drives research to uncover the diversity and function of the microbiome. Numerous studies have reported changes in the gut microbiota during a wide range of diseases like obesity, diabetes, and liver diseases as well as cancer and even neurodegenerative diseases (Cani, 2018). However, the observed microbial associations with specific outcomes tend to vary between studies. The variability could result from biological differences between studies but might also be due to artefacts induced by differences in sample collection, storage and other factors (Vogtmann et al., 2017). Individuals may have up to several hundred species of microbes within their gut which are largely shared between healthy adults according to findings from the human microbiome project (Turnbaugh et al., 2007). Some of the most commonly found or recognized genera of gut bacteria in adults are *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, *Clostridium*, *Escherichia*, *Streptococcus* and *Ruminococcus*. Approximately 60% of the bacteria belong to the *Bacteroidetes* or *Firmicutes* phyla (Bäckhed et al., 2005). Methane producing archaea most commonly *Methanobrevibacter smithii* have been detected in about 50% of individuals (Scanlan et al., 2008).

1.3 Detection of the Archaeome in the Human Gastrointestinal Tract

While the era of metagenomics and metatranscriptomics has begun, 16S rRNA gene amplicon sequencing still remains one of the most used methods to explore microbial communities. The majority of microbiome studies is bacteria-centric in spite of the bacteriome only being one part of the big picture (the microbiome) that contains microorganisms such as archeae, viruses, phages, yeast and fungi (Pausan et al., 2019). Due to the fundamentally different biology of *Archaea* it is not until recently that microbiome studies have started to shed light on the human archaeome (Cani, 2018; Eisenstein, 2018; Koskinen et al., 2017; Pausan et al., 2019). The under-representation of archaea in microbiome studies can be explained by a few factors: Primer mismatches of the "universal primers", inadequate detection of archaeal diversity, sampling methods or DNA extraction methods (Raymann et al., 2017). Sometimes the abundance of archaeal DNA is simply too low in the studied samples to be detected or the extraction methods are unsuitable. The incompleteness of the 16S rRNA gene reference databases can lead to archaeas not being classified or even the removal of archaeal signatures from sequencing data. Therefore Pausan et al. (2018) sought to identify the optimal primer pairs and methods for amplicon sequencing and quantification of the archaeomes in human samples. They showed that the choice of the primers influences substantially the detection of archaeal communities in amplicon-based microbiome studies. For improved archaeal detection, it has been recommended to use a nested PCR approach with the primer pair 344f-1041R for the first PCR, followed by a second PCR with the primer pair 519F-806R. Furthermore, for quantifying the number of archaeal 16S rRNA gene copies the use of the SYBR approach based on the primer pair 806aF-958aR is recommended (Pausan et al., 2018).

2 Objective

The overall objective of this project was to address several of the known methodological problems to detect archaea in the gastrointestinal tract. The specific aims were (1) to compare two different fecal sampling kits for 16S rRNA gene based microbiome analysis, (2) to compare universal prokaryotic primers and archaeal specific primers (the latter with a nested PCR approach) for archaeal 16S rRNA gene detection and relative abundance comparison, and (3) to quantify bacterial and archaeal concentrations in fecal samples using qPCR.

3 Materials and Methods

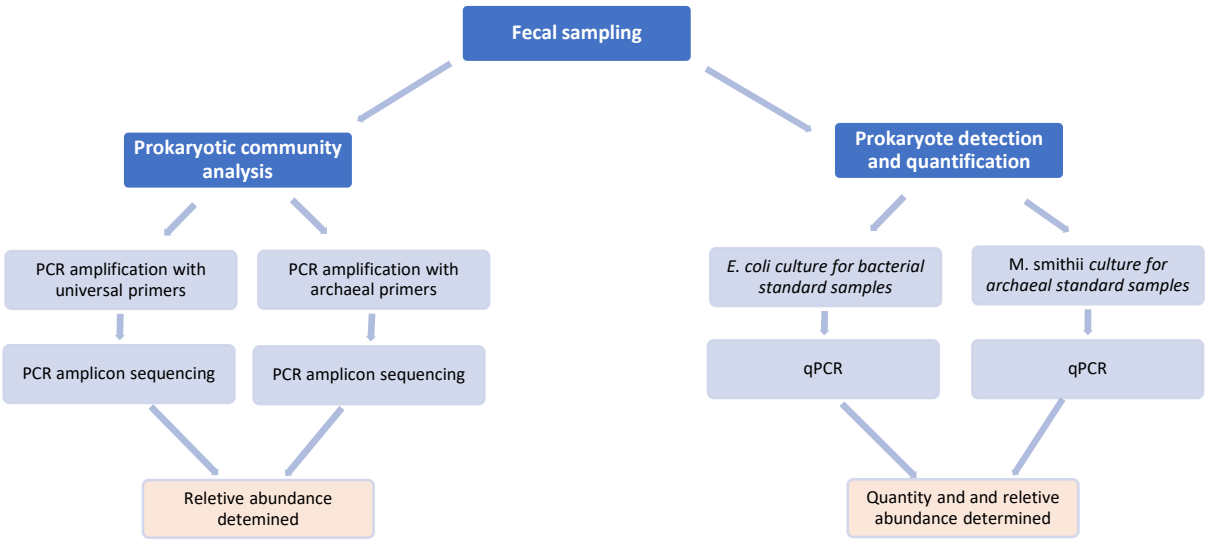

Figure 2: An overview flow chart showing main steps and methods.

3.1 Faecal sampling

The fecal sampling was done using specialized kits to ensures sample stability during storage/transport at ambient temperatures without the need for refrigeration or specialized equipment. Two different kits were chosen for fecal sampling, used according to the manufacturers protocols (see appendix A), on the one hand it was the DNA/RNA shield™ fecal collection tube (ZYMO research) and on the other hand the OMNIgene-GUT (OM-200) (DNA genotek).

3.2 Prokaryotic community analysis

3.2.1 DNA extraction from stool samples

For the isolation of microbial genomic DNA from the stool samples, QIAamp PowerFecal Pro DNA Kit was used according to the protocol from the manufacturer (QIAGEN) (see appendix A). 800 µl of lysis buffer were added to the bead tube beforehand to make it easier to empty the tip by pipetting up and down, instead of after pipetting the fecal matter.

3.2.2 Fluorometric quantification and Nano Drop purity assessment of stool sample DNA

3.2.2.1 Fluorometric quantification

Sample and standard DNA was quantified using Qubit® fluorometer 1,0 (Invitrogen) using the Quant-iT® fluorescence technology. Measurements were made by following the

manufacturers Quant-iT assays abbreviated protocol (Invitrogen) seen in appendix A with both dsDNA BR (Broad range) assay kit (Invitrogen) and dsDNA HS (High sensitivity) assay kit (Invitrogen) depending on sample concentration.

3.2.2.2 Purity assesment

The purity and concentrations of DNA extracts was assessed using a Thermo Scientific NanoDrop™ 1000 Spectrophotometer. (NanoDrop 1000 Spectrophotometer, V3.7 User's Manual, Thermo Scientific). For DNA concentration estimations, it is not as accurately as the fluorometer method since it cannot differentiate between DNA, RNA, or proteins.

3.2.3 PCR amplification using universal primers for prokaryotes

The 16S rRNA gene was amplified from the extracted stool DNA using the universal primers shown in table 1 and the Q5® High-Fidelity DNA Polymerase (NEB). The universal primer pair F515-R806 should produce 477 bp amplicons with indexes, containing 291 bp from the V4 region of the 16S rRNA gene (Walters et al., 2016). All samples had concentration of 2 ng/µl for a total of 10 ng in each reaction. Q5 was used due to its high fidelity (~280 times higher than Taq) and is a thermostable DNA polymerase with 3′→ 5′ exonuclease activity, fused to a processivity-enhancing Sso7d domain to support robust DNA amplification ((Pezza & Sun, n.d.; Potapov & Ong, 2017) The reaction was performed with the components and thermal profiling shown in table 2. A positive control for the reaction was included by adding DNA from *Carnobacterium divergens;* RF36, ISCAR 7473 with a known DNA template instead of fecal sample DNA. A negative control or no template control was made by replacing DNA with water so there would be no amplification. The making of reaction mixes was conducted under aseptic conditions.

Table 1: Primer sequences used in PCR amplification of the 16S rRNA gene (universal primers EMP-short, F515- R806).

Primer name	Short name	Index + primer Sequence (5'-> 3')
16S-bact-515f (Parada et al., 2016)	515F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG+ GTGYCAGCMGCCGCGGTAA
16S-bact-806r (Apprill et al., 2015)	806R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG+ GGACTACNVGGGTWTCTAAT

Table 2: PCR components and reaction.

	PCR mix
Information	**1x 25 µl reaction**
5x Q5 Enhancer	5
5x Q5 Buffer	5
EMP 515F primer, 10 µM	1.25
EMP 806R primer, 10 µM	1.25
dNTP, 40 mM (NEB)	0,1
Q5 polymerase 5U/µl	0.3
DNA 2ng/µl	5
water fill up	7.1
Total	**25 µl**

Table 3: Thermal profile for PCR amplification.

	Temperature	**Time**	
Initial Denaturation	98°C	30 sec	
Denaturation	98°C	10 sec	
Annealing	50°C	30 sec	*30 cycles
Extension	72°C	30 sec	
Final Extension	72°C	2 min	
Hold	4°C	Forever	

3.2.4 Agarose gel electrophoresis

Gel electrophoresis was used to verify the presence of PCR products of the correct size and quality. A 1% agarose gel was prepared in 1x TAE buffer with 5% SYBR® safe (Invitrogen). By first adding 5 µl of 6x loading dye purple (NEB) to 5 µl of PCR product, a volume of 10 µl could be loaded on to the gel. A quick-Load® 100 bp ladder (NEB) was used as a size reference. The gel was run in a TAE buffer at a voltage of 90V-120V for 30 – 50 minutes depending on gel size. Visualizations were made with Gel Doc® XR system (Bio Rad).

3.2.5 Nested PCR amplification using archaea specific primers

Nested PCRs were performed by two consecutive reactions where 5 µl from PCR 1 were added to a second reaction mix with more specific primers with indexes as previously recommended by Pausan, et al. (2018). The first PCR was intended to select the archaeal community of interest and the second to further amplify the archaeal product. The process was optimized to give a satisfactory outcome of 16S rRNA gene amplification by making time changes to the denaturing step, doing a temperature gradient for the annealing step, improve the extension step duration and finally by optimizing the number of cycles. The best results were confirmed using gel electrophoresis. The 16S rRNA gene was amplified from the extracted stool DNA using the two pairs of primers shown in table 4. The Q5® High-Fidelity DNA Polymerase (NEB) was used (see tables 5 and 6). A positive control for the reaction was made by adding 1 µl of DNA from a mock control sample (Matis) (50 ng/µl)

made from 20 strains of which 2 were archaeal strains (*Thermococcus barophilus* and *Pyrococcus abyssi*). All other samples in this reaction have a concentration of 10 ng/µl for a total of 50 ng in each reaction. A negative control was made both for PCR 1 and PCR 2 by adding water instead of DNA so there should be no amplification. The making of reaction mixes was conducted under sterile conditions.

Table 4: Primer sequences used for the nested PCR amplification from archaea (Pausan et al., 2018).

PCR	**Primer name** (Klindworth et al. 2012)	**Short name**	**Index + primer Sequence (5'-> 3')**
Archaea PCR1	S-D-Arch-0344-a-S-20	344F	ACGGGGYGCAGCAGGCGCGA (no index)
	S-D-Arch-1041-a-A-18	1041R	GGCCATGCACCWCCTCTC (no index)
Archaea PCR2	S-D-Arch-0519-a-S-15	519F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG+ CAGCMGCCGCGGTAA
	S-D-Arch-0786-a-A-20	806R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG+ GGACTACVSGGGTATCTAAT

Table 5: Nested PCR components and reaction.

PCR 1 (PCR 2)	**PCR mix**
Information	1x 25 µl reaction
5x Q5 Enhancer	5
5x Q5 Buffer	5
Arch 344F primer, 10 µM (Arch 519F primer, 10 µM)	1.25
Arch 1041R primer, 10 µM (Arch 806R primer, 10 µM)	1.25
dNTP, 40 mM (NEB)	0,13
Q5 polymerase 5U/µl	0.25
DNA 2ng/µl	5
water fill up	7.3
Total	**25 µl**

Table 6: Thermal profile used for PCR.

	Temperature	Time	
Initial Denaturation	98°C	30 sec	
Denaturation	98°C	10 sec	*25 cycles (*30cycles PCR2)
Annealing	56°C	30 sec	
Extension	72°C	30 sec	
Final Extension	72°C	2 min	
Hold	4°C	Forever	

3.2.6 PCR amplicon sequencing and statistical analyses

after amplification of the V4 region of 16S rRNA gene from archaea and bacteria with PCR next generation sequencing on Miseq (16S Metagenomic Sequencing by Illumina) was performed at Matís. Workflow was carried out according to the Illumina provided protocol "16S Metagenomic Sequencing Library Preparation" by Matis staff.
The obtained MiSeq sequence data was processed using the open source package DADA2 (Divisive Amplicon Denoising Algorithm; (Callahan et al., 2016). The DADA2 turns paired-end fastq files into merged, denoised, chimera-free, and inferred sample sequences called ribosomal sequence variants (RSVs). The final product is a higher-resolution analogue of the traditional OTU table, which records the number of times each exact amplicon sequence variant was observed in each sample. The taxonomic affiliations were determined using SILVA v128 as the reference database (Quast et al., 2013). The resulting RSV table includes the number of reads and observed ribosomal sequence variants, each row corresponds to non-chimeric inferred sample sequence with a separate taxonomic classification.

3.3 Prokaryote detection and quantification

3.3.1 Cultures for qPCR standards

An *E. coli* culture was prepared and grown overnight and then frozen for later use in qPCR standard preparation. First, a fresh *E. coli* culture was made by adding a few colonies to 30 ml of LB medium in an Erlenmeyer flask and grown overnight with shaking at 37°C. Then, serial dilutions were made in an MRD medium (9 ml MRD: 1 ml culture). 1 ml samples of each dilution were frozen in 1,5 ml Eppendorf tubes for later use. 100 µl of diluted cultures, from 10^-3 down to 10^-7 in replicates of two, were spread on Petri dishes containing LB-medium, making it 10^-4 down to 10^-8. The diluted cultures were then grown over night at 37°C and the density calculated, CFU/ml.

DNA extraction from cell cultures

The DNA was extracted from an *E. coli* culture with MasterPure™ complete DNA and RNA purification kit according. The manufacturers protocol (Epicenter) was followed for cell samples but with the following changes: Part D, step 8. 50µl of QiAamp power fecal pro kit solution 6 (TE buffer) was used to resuspend the pellets and 200µl for the undiluted *E. coli* culture sample (to dissolve the pellet sufficiently). The pellet did not resuspend at first and therefore the solution was heated to ca 60°C for one minute for the pellet to dissolve.

DNA was previously extracted from the archaea *M. smithii* culture by Matis staff following the same procedure.

3.3.2 DNA quantification and quality assessment of standard samples

The same methods followed as described under section 2.2.2 Fluorometric quantification and Nano Drop purity assessment of stool sample DNA.

3.3.3 qPCR

Absolute quantification method of bacteria and archaea was used with SYBR® Green dsDNA-binding dye (BioRad) from SsoAdvanced™ Universal SYBR Green supermix (BioRad) that contains an Sso7d fusion polymerase, a dsDNA-binding protein that increases speed and processivity and increases tolerance to PCR inhibitors, which provides increased stability for the polymerase-template complex, enhancing performance. The manufacturers iQ™ SYBR® Green supermix Instruction Manual was followed. The reactions were performed according to the QuantStudio™ Design and Analysis Software user guide (Thermo Fisher) on the QuantStudio™ 3 and 5 Real-Time PCR system. The primers shown in table 7. were used as suggested by Pausan et al. (2018) with the components and method as shown in table 8 and 9. The following changes were made to the procedure: An extension step of 72°C was not added as recommended in the protocol from BioRad and QuantStudio. Instead the annealing step was set to 60°C for the bacteria qPCR and 53°C for the archaea qPCR as instructed by Pausan et al. (2018). Negative controls (no template controls) were included during qPCR amplification. For better comparison between samples, three additional stool test samples, from Matís (named SS, Xav and SK), were added to the qPCR runs. The samples were acquired with the same methods as the ES samples. (see material and methods, 3.1 sampling and 3.2.1 DNA extraction, page 19).

Table 7 Primer sequences used for qPCR for bacteria and archaea (Pausan et al., 2018).

PCR	Primer name	Short name	primer sequence (5' -> 3')
Bacteria qPCR	S-D-Bact-0337-a-S-20	qPCR_338bF	ACTCCTACGGGAGGCAGCAG
	S-*-Univ-0517-a-A-15	qPCR_517uR	GWATTACCGCGGCKGCTG
Archaea PCR	S-D-Arch-0787-a-S-20	qPCR_80a6F	ATTAGATACCCSBGTAGTCC
	S-D-Arch-0958-a-A-19	qPCR_806aR	YCCGGCGTTGMTCCAATT

Table 8: Quantitative PCR components and reaction mix for both archaea- or bacteria depending on primer pair.

Components	**concentration**	**µl**
SsoAdvancedTM Universal SYBR Green supermix (BioRad)	1x	10
Arch-qPCR_806aF/ Bact-qPCR_338bF	300nM (0,3 µM)	0,6
Arch-qPCR_968aR/ Bact-qPCR_517uR	300nM (0,3 µM)	0,6
Water		7,8
DNA (different dilutions of DNA)		1
Total Volume		**20**

Table 9: Thermal profile used for quantitative PCR.

Target and primers		**qPCR conditions**					
target gene	**Primer pair**	**hold stage**	**Initial denaturation**	**Denaturation**	**Annealing**	**No. of cycles**	**Melting curve**
16S rRNA Bacteria	338bF-517uR	2', 50°C	3', 98°C	15", 98°C	60", 60°C	35	60-95°C
16S rRNA Archaea	806aF-958aR	2', 50°C	3', 98°C	15", 98°C	60", 53°C	35	60-95°C

3.3.4 Standard curve and data analysis

To establish a standard curve, five tenfold dilution of standard sample (*E. coli, M. smithii)* were performed. The amount analysed by the qPCR and the resulting Ct values were then plotted against the log of the starting quantity giving a straight line when primers are at optimal capacity and sample mix and loading is done correctly. The standard curve is then used to derive the quantity of the test samples.
The optimal R^2 value is >0.99 where 1 indicates a perfect linear relationship of Ct values. Amplification efficiency (E) of 2.02 means that for each cycle the amount of product doubles and primers are 100% efficient. This efficiency is calculated from the slope(s) of the standard curve according to the following formulas: E = 10(-1/slope)-1. For 100% efficiency, the slope of the standard curve is -3.32. A good reaction should have an efficiency between 90% and 110%, which corresponds to a slope between -3.58 and -3.10 (*Important Parameters of Quantitative PCR (QPCR) Analysis*, n.d.). The cycle threshold (Ct) was determined using the single threshold in the QuantStudio™ Design and Analysis Software. Detection limits were defined based on the average Ct values of non-template controls (duplicates) and the corresponding standard curves of the positive controls.
After being able to establish a standard curve the bacterial and archaeal communities were quantified in three successful qPCR tests each. Using the 338bF-517uR universal primers (universal tests 1, 2 and 3). The tests had a primer efficiency between 102% -109%. The archaeal communities were quantified using the 806aF-958aR primers (arch. Tests 1, 2 and 3) and the tests had a primer efficiency between 100% -104%. All successful qPCR tests had linear correlation between dilution point with in the standard curve ($R^2 > 0.99$).

4 Results

4.1 Prokaryotic community analysis with universal and archaea-specific primers

4.1.1 Microbial genomic DNA extracted from stool samples

For the DNA extraction 250 µl of fecal matter, collected with the two sampling kits, were used. Different viscosity of matter from the sampling kits resulted in different amounts being transferred to the DNA extraction kit depending on whether it came from OM-200 (has less lysis buffer) or the DNA Shield kit. The samples from each kit were called ES1 and ES2 respectively. The final amount was weighted and calculated as shown in table 10.

Table 10: Weight of samples for DNA extraction.

Specimen taken from sampling kits	ES1 (OM-200) (g)	ES2 (Shield) (g)
Empty DNA extraction kit tube	3,04	3,05
Lysis buffer added	3,90	3,91
Sample added (total weight)	4,06	4,18
Weight of sample	**0,16**	**0,27**

After DNA extraction with the QIAamp PowerFecal Pro Kit (Qiagen), the DNA was quantified using a Qubit fluorometer, showing a concentration of 75 ng/µl in sample ES1 and 10.5 ng/µl in sample ES2.

4.1.2 PCR amplification of the partial 16S rRNA gene using universal prokaryotic primers

With the universal bacterial primers 515F-806R, 291 bp of the V4 region of the 16S rRNA gene were amplified using a PCR, resulting in a 477 bp amplicon with indexes. The quality and success of the amplification was then verified using gel electrophoresis, confirming that an amplicon of the appropriate size was acquired (see fig. 3).

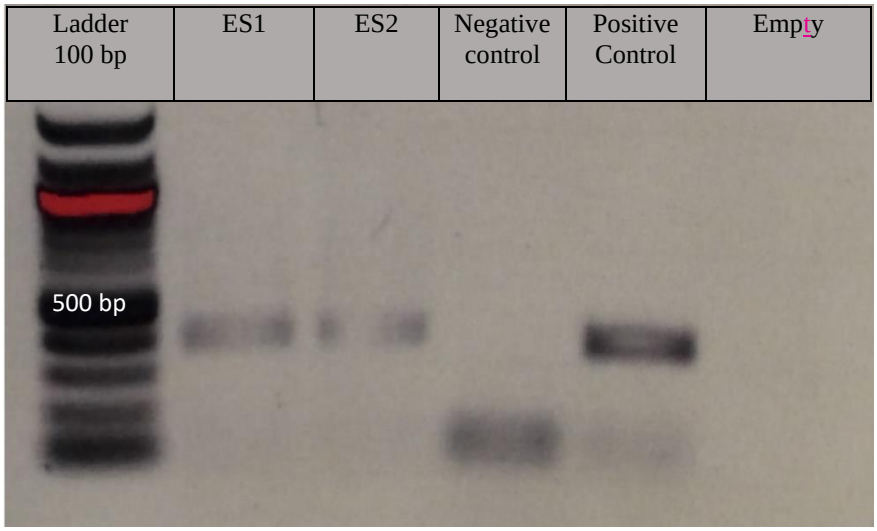

Figure 3: Amplification of prokaryotic 16S rRNA genes in samples ES1 and ES2. The samples were loaded on to the gel in the order portrayed. DNA from both samples was amplified, and of the correct size (~477), as well as the Positive control (C. divergens; RF36). There was no amplification in the negative control apart from what is thought to be primer dimers.

4.1.3 Amplification of partial 16S rRNA gene in archaea and optimization of PCR amplification methods for Miseq amplicon sequencing

For the amplification of archaeal rRNA genes a more specific nested PCR approach has to be taken where the first PCR is intended to select the archaeal community of interest. The first PCR was carried out by using the primer pair 344F-1041R and should result in an amplicon size of 735 bp. The quality and success of the amplification was then verified using gel electrophoresis. The positive control in this case did not work as such since it only contained bacterial DNA (*C. divergens*; RF36). As expected, there was no amplification for the positive control, as the gel electrophoresis revealed (Fig. 4).

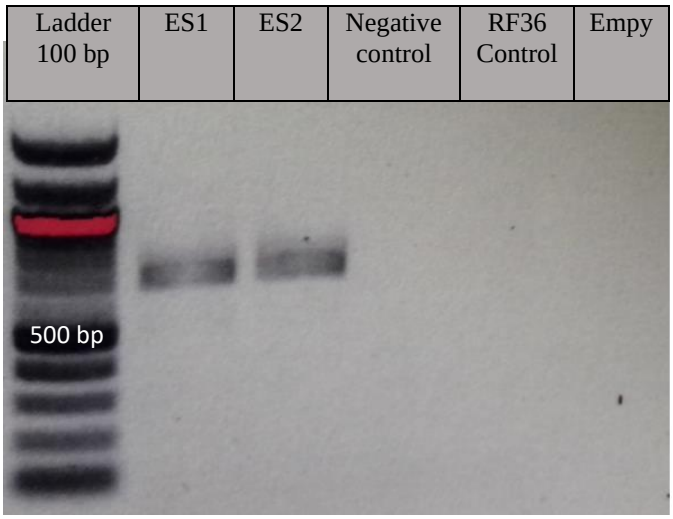

Figure 4:First PCR, Amplification of archaeal 16S rRNA genes in samples ES1 and ES2. The samples were loaded on to the gel in the order portrayed. DNA from both samples was amplified and resulted in clear bands of the correct size (~735). No amplification in negative controls. Positive control missing from the PCR.

The second PCR is meant to further amplify the archaeal product with more specificity (nested approach). The PCR was carried out by using the primer pair 519F-806R and should result in an amplicon size of 389 bp with indexes. The quality of the amplification was then verified as before with electrophoresis, showing successful and unspecific amplification. Bands at ~735 bp were probably remains from the first PCR (5 µl of the first PCR used in the nested PCR), and the bigger band (~1150 bp) nonspecific binding of the primers (fig. 5).

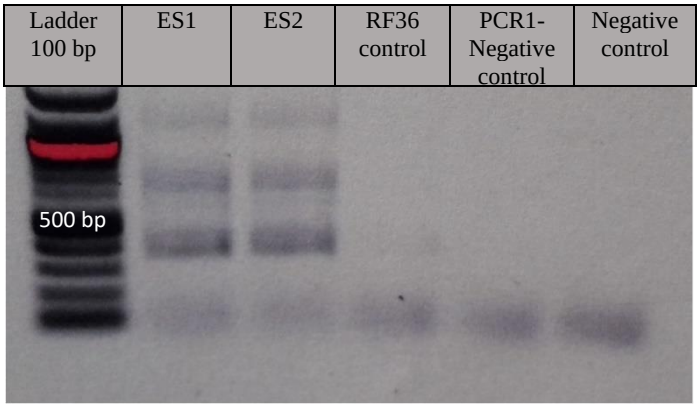

Figure 5: PCR2, Amplification of archaeal 16S rRNA genes in samples ES1 and ES2. The samples were loaded on to the gel in the order portrayed. *DNA from both samples was amplified with clear bands and of the correct size (~389)but also containing unspecific amplification. No amplification in negative controls. The left negative control was 5 µl of the negative control product from PCR1, instead of DNA, and the right was deionized water instead of DNA.*

The procedure was then performed again with changes in denaturation time, from two minutes to 30 seconds, to possibly yield less unspecific amplicons. For the positive control, a Mock sample from Matis was used, made from 20 strains and there of 2 archaeal strains (*Thermococcus* and *Pyrococcus*).

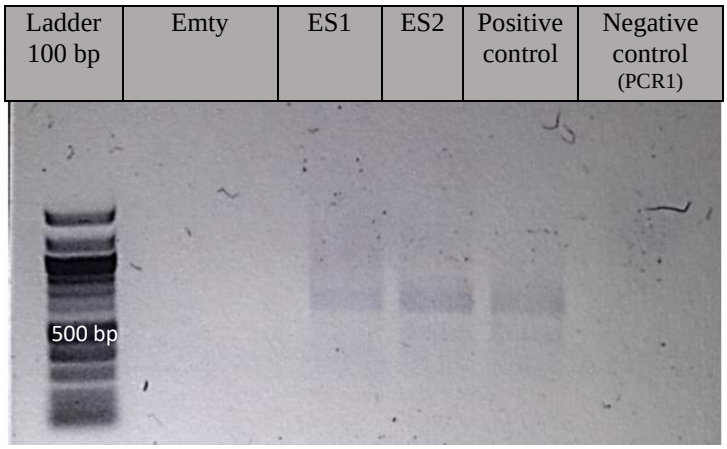

Figure 6: PCR1, Amplification of archaeal 16S rRNA genes in samples ES1 and ES2. The samples were loaded on to the gel in the order portrayed. DNA from both samples and positive control was amplified and resulted in clear bands of the correct size (~735). There was no amplification in the negative control.

The second PCR was done with an annealing temperature gradient for acquiring more specificity of the primers at higher temperatures. The temperature was raised in increments of 2-3°C from 58°C to 63°C. There was unspecific amplification identical to the previous

test (see fig. 5) and was the same between test samples and positive control. Annealing temperature did not seem to have any effects, besides from more primer dimers forming at 58°C, showing clearer bands at a size around 100 bp then at higher annealing temperatures as seen in figure 7. Negative control (-ctrl) from PCR1 showed amplification at 58°C at a size of c.a. 500bp which could indicate unspecific priming in that sample since other negative controls have no amplification. (fig. 7)

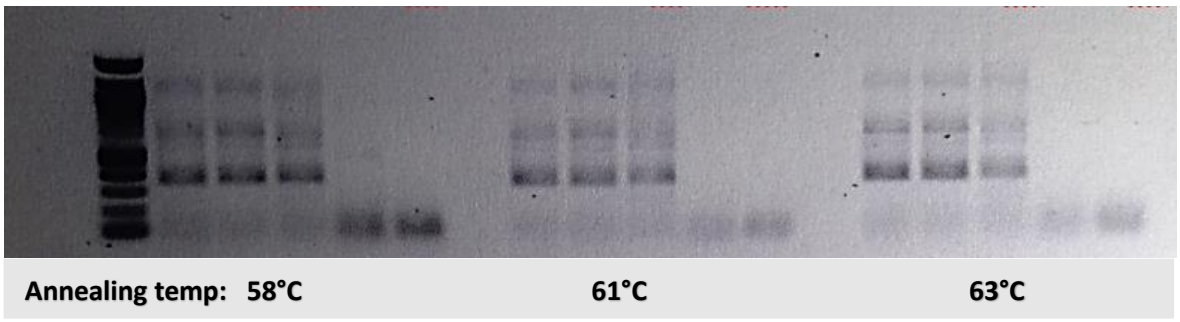

Figure 7: PCR2, Amplification of archaeal 16S rRNA genes in samples ES1 and ES2 at different annealing temperatures. The samples were loaded on to the gel in the following order for each annealing temp: ES1; ES2; positive control; negative control (PCR1); negative control. A 100 bp ladder was also added in well nr 1. at 58°C. Successful amplification with clear bands regardless of annealing temp.

4.1.4 Prokaryotic community analysis and fecal sampling kit assessment with Miseq amplicon sequencing

Next generation sequencing was performed by Matis on an Illumina Miseq after amplification of the V4 region of the 16S rRNA gene in prokaryotes with both archaea specific primers, using nested PCR approach, and universal primer, using a single-step PCR. The data was analysed based on the DADA2 algorithm for arranging and correcting Illumina-sequenced amplicon errors. DADA2 infers sample sequences exactly and resolves differences of as little as 1 nucleotide (Callahan et al., 2016). After correcting the sequenced amplicon errors the final product is an amplicon sequence variant (ASV) table**,** a higher-resolution analogue of the traditional operational taxonomic unit table (OUT table), which records the number of times each exact amplicon sequence variant was observed in each sample. Taxonomy can then be assigned to the output sequences. The ASV table includes the number of reads and observed ribosomal sequence variants (ASVs) obtained for all samples covering the *Archaea* and *Bacteria* domains. This revealed that 0.4% of the ASVs from both samples (ES1 and ES2) using the universal primer set were of archaeal origin and exclusively assigned to *Methanobrevibacter smithii.* This corresponded to between 160-170 reads out of total 39000-42000 reads, for the samples ES1 and ES2 respectively, when using the 515F/806R primer pair in a single PCR.

In total, 205 bacterial ASVs were observed in sample ES1 and 184 ASVs in sample ES2, of that was only one archaeal ASV in both samples 12 ASVs were detected in the NTC (table 11.) It is important to keep in mind the 16S rRNA gene copy number variability between species of archaea and bacteria. According to the ribosomal RNA database by Schmidt laboratory (Stoddard et al. 2015) the average copy number for bacteria in the ES samples are 5.3 16S copies but only 2 copies are present in the genome of *M. smithii*.

Table 11: Amplicon sequencing when using the universal 515F/806R primer pair. Total ASVs and total reads.

Samples	NTC	ES1	ES2
Total ASVs	12	205	184
Total reads	170	42025	39816

With the nested approach using the 344F-1041R/519F-806R primers, further detection of archaea was made possible. The 389 bp amplicons from all three annealing temperatures were sequenced. The number of reads ranged from 59431-77455 in sample ES1 and 68137-77704 in sample ES2 with average total reads, for both archaea and bacteria, of 72209 between both samples. Despite allowing more reads only one RSV was detected, coming from *M. smithii* as with the universal primers. The number of observed bacterial RSVs were 37 in total.

Based on this amplicon sequencing data it is not clear that the archaeal primers are more suitable for detecting archaeal diversity since their use only resulted in two archaeal ASVs detected, both from *M. smithii*, compared to one with the universal primers.

4.2 Detection and quantification of bacterial and archaeal communities in fecal samples

For absolute quantification, samples for the standard curve were prepared from an *E. coli* culture against which the test samples were then compared. The total density of the culture was estimated as 2.1*10^9 CFU/ml by counting colonies on an agar plates as demonstrated in table 12.

Table 12: Counts of CFU on agar plates. Total density of E. coli *culture was estimated at 2.1*10^9 CFU/ml.*

Culture dilutions	**Petri dish A**	**Petri dish B**
10^-7	215 colonies	218 colonies
10^-8	21 colonies	21 colonies

After DNA extraction from an undiluted culture, the DNA was dissolved in 200 µl of buffer and then quantified using Qubit with dsDNA BR assay kit. This resulted in 59.4 ng/µl, which gives approximately 10^7 CFU for every 50 ng standard sample and 2*10^5 CFU/ng. Considering there are 7 16S rRNA gene copies in *E. coli,* (Stoddard et al., 2015) there are 1.4*10^6 copies/ng of DNA. Average 16S copy number in bacteria is 4.2 (Větrovský & Baldrian, 2013) but approximately 5.3 in the ES samples as mentioned before (section 4.1.4) (Stoddard et al. 2015).

4.2.1 Adjustment of qPCR conditions

The fecal samples were measured with Qubit as well as three additional samples from Matis, called SS, Xav and SK, (see appendix B, table 15) and then diluted to 10 ng/µl. An aliquot of the *E. coli* standard sample was diluted serially 5 times with a tenfold dilution each (50 ng/µl to 0,005 ng/µl) and used to make a standard curve. The reaction was prepared as stated in materials and methods page 24-25.

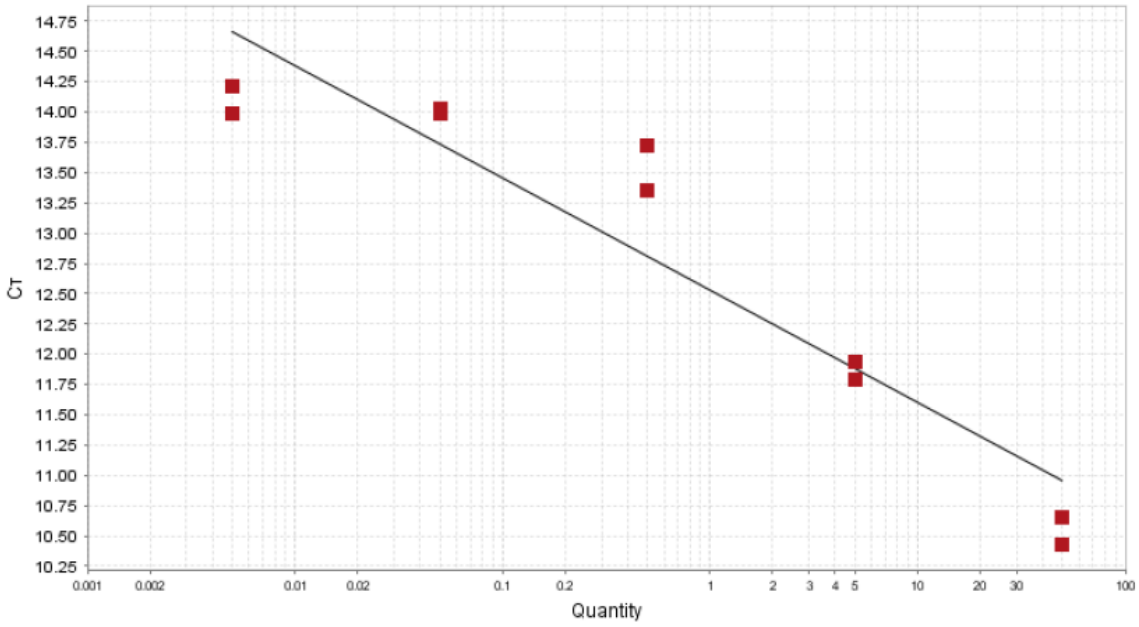

Figure 8: The Standard curve from the E. coli standard samples. Samples were diluted 5 times with a tenfold dilution (50 ng/µl to 0,005 ng/µl). Making of standard curve was unsuccessful, with a slope of -0.927 and therefore inapplicable for quantification of test samples.

The standard curve could not be fixed when doing a second qPCR test where the standards were made with greater care and accuracy when pipetting and using 3 µl instead of 1 µl of sample DNA. The second qPCR had very similar results and the slope of the standard curve was -0.716. In both runs the "no template control" (NTC) showed signs of amplification with a cycle threshold (Ct) on pair with the lowest standard concentration, indicating potential formation of primer dimers or the presence of contamination. To test for primer dimer formation the qPCR product was run on an agarose gel (see figure 9).

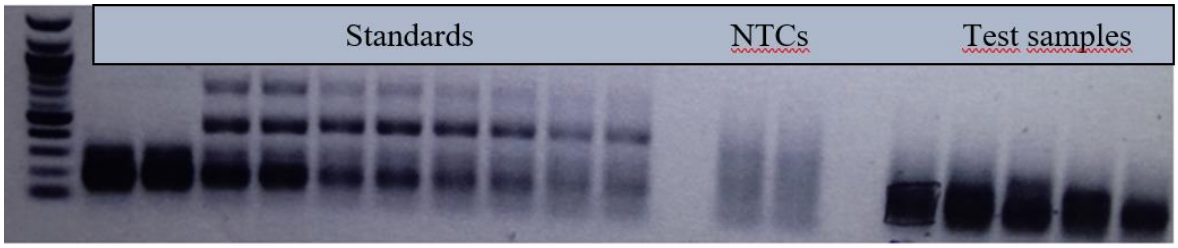

Figure 9: qPCR products run on an agarose gel. Samples were loaded in the following order; 100 kb Ladder, 2x 50 ng standard, 2x 5 ng standard, 2x 0.5 ng standard, 2x0,05 ng standard, 2x 0,005 ng standard, empty well, 2x NTC, empty well, ES1, ES2, SS, Xav, SK.

No clear bands in NTC but a lot of smear, often caused by high cycle numbers or high annealing temperature. High number of cycles also increases possibility of unspecific binding as seen in standard samples in figure 9 above.

In order to eliminate smearing and double bands, PCR was done with the same qPCR reaction mix and qPCR primers but 30 cycles, instead of 40 cycles and with an annealing temperature gradient, raising the temperature in increments of 2-3°C from 55°C to 65°C. (figure 10).

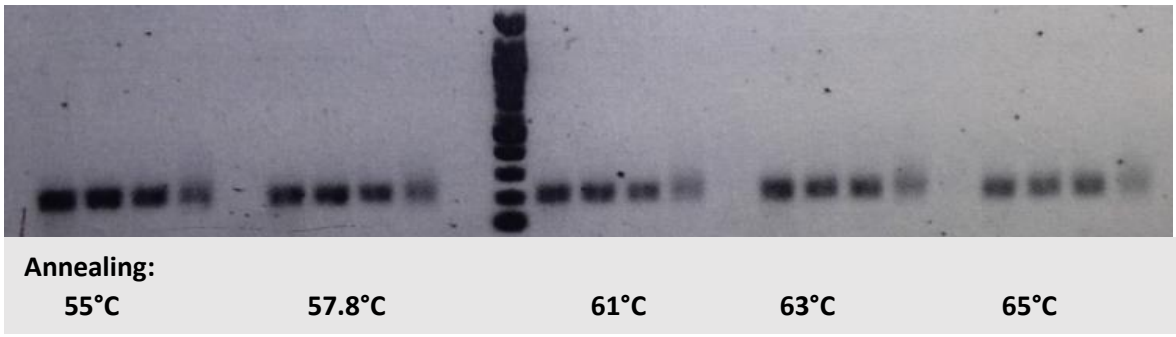

Figure 10: PCR results with qPCR reaction mix, run on an agarose gel with electrophoresis. Samples were loaded in the following order; 50 ng standard, 5 ng standard, 0.5 ng standard, ES1, NTC and ran with different annealing temperatures as portrayed. *Number of PCR cycles set to 30 cycles instead of 40. Annealing temperature gradient, raising the temperature in increments of 2-3°C from 55°C to 65°C.*

The annealing temperature had little effect on the outcome, yet no double bands appeared on the gel. The results indicate that a lower cycle number prevents the formation of over amplification and smearing. Because fluorescence is measured and detected, above the background noise, in the earlier cycles effects this late in the PCR process should not have a big impact on the standard curve.
There are many variables that can affect the amplification in a PCR and therefore not certain the cycle number change, by itself, resolves the issue at hand. For verification further analysis would be needed.

By lowering the number of cycles, an optimal standard curve was acquired (figure 11) to then evaluate the concentration of bacteria in the stool samples. *E. coli* standard sample and the test samples were quantified with Qubit beforehand as seen in table 12. DNA concentration of the undiluted *E. coli* was measured at 68.2 ng/µl.

The undiluted standard sample ($6.8*10^1$ ng of DNA) was consistently not following the standard curve because of delayed amplification resulting in a lower cycle threshold. It can be an indicator of inhibitors in the sample or a concentration above the detection limit. The sample was therefore omitted from the standard curve that consequently was located out of range for for the test samples, making the quantification potentially less accurate.
To better substantiate the possibility of inhibitors, purity of the *E.coli* DNA was assessed with a nanodrop 1000 spectrophotometer, indicating organic contamination with a 260/230 ratio of 1.25. The highest concentration sample was therefore not used for constructing a standard curve for bacterial quantification but was replaced with another tenfold dilution (10^{-3}) for it to consist of 5 serial 1:10 dilutions.

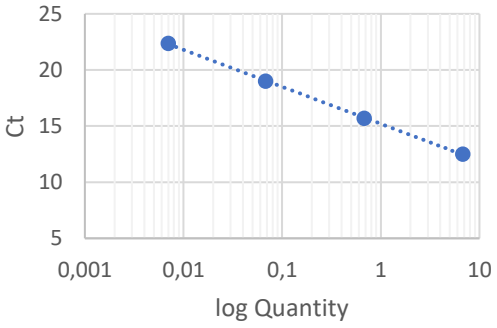

Figure 11: Standard curve established with the E. coli standard samples. Samples were diluted 5 times with a tenfold dilution (68 ng/µl to 0,007 ng/µl) The highest dilution point did not follow the standard curve and was omitted. The slope of the standard curve is -3.275, R^2 equals 1 and the efficiency is 102%.

4.2.2 Quantification of bacterial communities in fecal samples

Bacterial communities were quantified in three qPCR tests (tests 1, 2 and 3) using the 338bF-517uR primers. The tests had a primer efficiency between 102% -109%. The results are plotted in Fig. 12 and an example of numerical results are shown for test 1 in table 13.

Table 13: Quantity according to qPCR result from test 1. Test samples were quantified with dsDNA Br assay kit and Qubit before qPCR run.

Samples	ES1	ES2	SS	Xav	SK
Concentration-qubit (ng/µl)	3.5	3.9	4.2	4.2	3.2
qPCR Input (ng)	10.5	11.8	12.6	12.5	9.7
Results for 3 µl of DNA	14.7	15.3	13.6	18.2	13.8
Results for ng/µl of DNA	**4.9**	**5.1**	**4.5**	**6.1**	**4.6**
Results for CFU/µl of DNA	**$9.8*10^5$**	**$1*10^6$**	**$9*10^5$**	**$1.2*10^6$**	**$9.2*10^5$**
Output: Input ratio	**140%**	**130%**	**108%**	**145%**	**142%**

When the ratio between sample input, measured with Qubit, is compared to the qPCR results it appears that the samples contain over 100% bacterial signature. This overestimation can partially be explained by 16S copy number variations between bacterial species and the standard containing above average number copies. When the variation had been considered by using the average copy number of 5.3 in bacteria detected in the ES samples (see section 4.1.4.), the samples from test 1 contained a 95,4% bacterial signature on average. Another factor is the Qubit quantification of both the standard sample and the fecal test samples (fig. 12) presumably not accurately depicting the DNA concentration.

Quantification of the bacterial load of all stool samples contained around 10^7 copies/total DNA (on average 6ng samples). Obtained by applying the copy number of *E. coli* ($1.4*10^6$ copies/ng) standard samples to the standard curve.

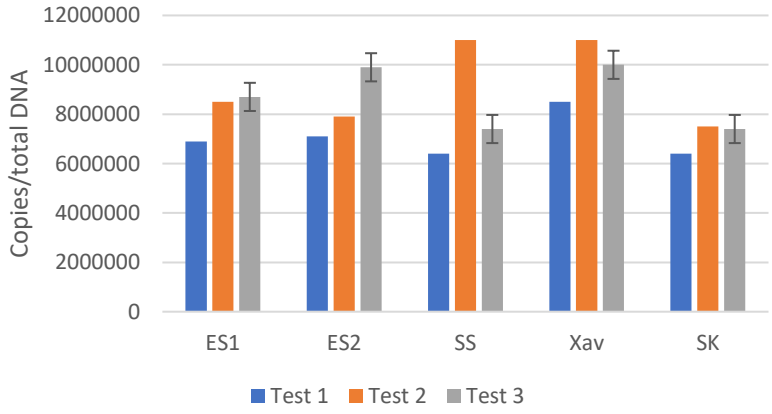

Figure 12: Values of total copies of 16S rRNA gene for bacteria per total sample DNA. Test 3 is an average of 3 replicates per sample, showing standard error. Tests 1 and 2 had only one measurement for each sample.

4.2.3 Quantification of archaeal communities in fecal samples

The standard curve for the archaeal qPCR was established with five serial tenfold dilution of a 0.73 ng/µl of *M.smithii* DNA. The undiluted standard sample ($7*10^{-1}$ ng) did not amplify, indicating inhibitors in the sample and was therefore omitted from the standard curve. Purity of the *M. smithii* sample was assessed with a nanodrop 1000 spectriophotometer, indicating a great amount of organic contaminants with a 260/230 ratio of 0.07 (table 16). However, these results are very unreliable considering that nanodrop can only measure accurately down to 2 ng/µl, which can exaggerate the results. Another tenfold dilution was added ($0.7*10^{-6}$) instead so that the standard curve would consist of five serial 1:10 dilutions. Since the standard samples contains the same and only specie found in the stool samples, with a genome containing two 16S rRNA gene copies, there is no need for adjusting to copy number variations and the quantification becomes more accurate. (see section 4.1.4.)

Archaeal communities were quantified in three qPCR tests (tests 1,2 and 3) using the 806aF-958aR primers. The tests had a primer efficiency between 100% -104%. No amplification was detected in an *E. coli* sample when using archaeal primers, an indication that the primers only amplify archaeal 16S rRNA genes which is a key factor. The ratio between bacterial and archaeal 16S rRNA genes was determined using the average for each sample. For sample ES1 the ratio was 200:1.1 (0.55% archaeal signatures) and 125:1 (0.8% archaeal signatures) for sample ES2. Samples SS, Xav and SK had concentration below the detection limit. The results are plotted in Fig. 13.

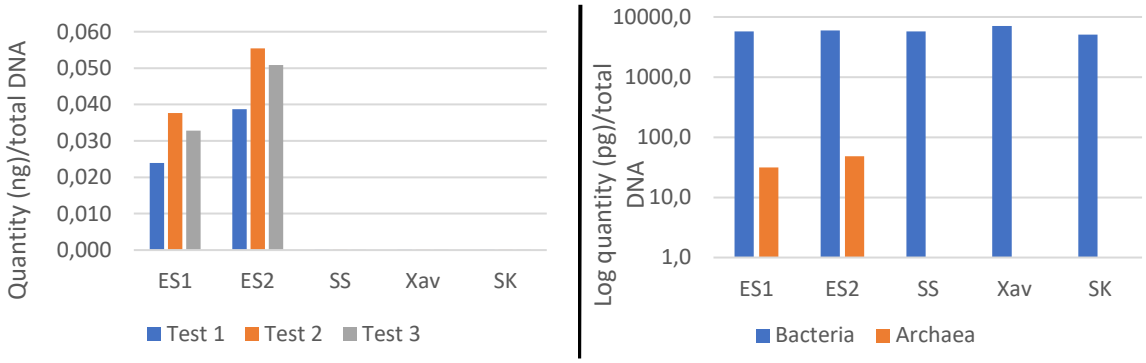

Figure 13: A. Shows ng of archaeal DNA of total DNA in samples. Tests 2 and 3 are an average of 3 replicates per sample. B. Logarithmic values of pg of 16S rRNA archaeal gene (orange, primer pair 806aF-958aR) and bacteria (blue, primer pair 338bF-517uR)

5 Discussion

5.1.1 Discrepancy in fluorometric quantification

Throughout this project, standard samples and test samples were repeatedly quantified with the Qubit fluorometer. It turned out that the concentration values varied between measurements adding an uncertainty to methods based on concentrations acquired with Qubit, for example constructing a standard curve. Overestimation of standard sample concentrations gives a proportional overestimation of the test samples. Stool test samples did not all contain the same concentration because of error with the Qubit early on when diluting the samples. Figure 14. shows how the concentration fluctuated between Qubit measurements.

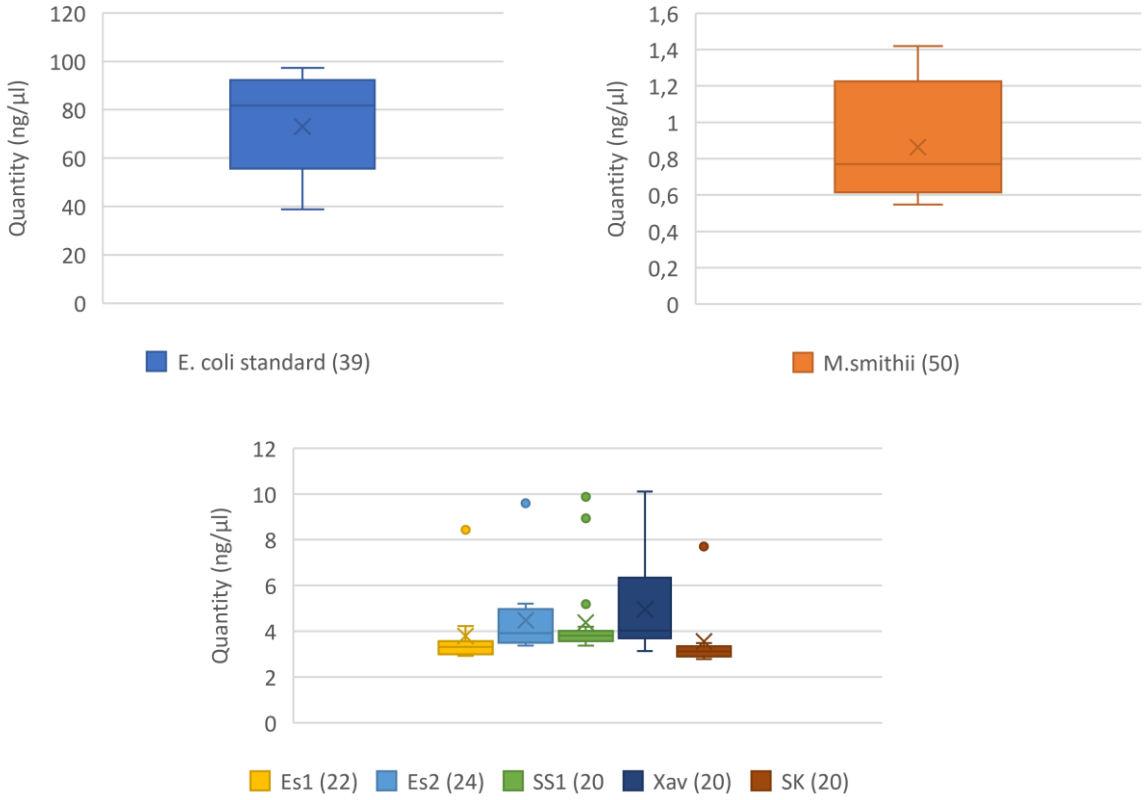

Figure 14: Boxplot showing discrepancy between Qubit measurements. Brackets indicate number of measurements

Qubti has been shown to give values that are not proportional to the dilution ratio, in contrast with NanoDrop and qPCR, values can be significantly lower and that DNA structural transition can occur at a critical salt concentration that is presumed to contribute to the stabilization of dsDNA; concentrations below a specific concentration of this factor induces a defined structural change in dsDNA, resulting in a second stable structure of dsDNA with a low affinity for fluorescent dye (Nakayama et al. 2016).

5.1.2 Prokaryotic community analysis and fecal sampling kit assessment

Miseq amplicon sequencing was successful following PCR amplification using the universal bacterial primers 515F-806R and for the amplification of archaeal rRNA genes with a nested PCR approach using the archaeal primer pairs 344F-1041R and 519F-806R as recommended by Pausan et al. (2018). The data was then analysed based on the DADA2 algorithm for arranging and correcting Illumina-sequenced amplicon errors.

After correcting the sequenced amplicon errors, the final product was an ASV table, obtained for samples ES1 and ES2, covering the *Archaea* and *Bacteria* domains. This revealed that 0.4% of the ASVs from both samples using the universal primer set were of archaeal origin and exclusively assigned to *Methanobrevibacter smithii.* When comparing the PCR methods using the sequencing results, it is not clear which method is more suitable for detecting archaeal diversity in the stool samples. With the nested approach, despite allowing more reads, only two ASVs were detected in the stool samples, both coming from *M. smithii.* With the universal primers one archaeal ASV was detected, also coming from *M. smithii.* Observed bacterial RSVs were 37 in total with the archaeal primers and 226 with the universal primers. Notably, stool samples have been found to contain the overall lowest diversity of archaea with only 3-5 identified archaeal RSV's while skin and oral samples contain higher diversity, with 5 to 49 RSV's as discovered using the nested approach (Pausan et al. 2018). The samples ES1, ES2 and SS therefore seem to have in common containing only one archaea species and in low abundance. Previous studies have shown that archaeal signatures are detected in about 50% of individuals, similar to the archaeal detection in this project where 2 out of 4 individuals had archaeal signatures in there microbiome (Conlon & Bird, 2015). No obvious difference was detected between sampling methods. Based on user experience the DNA/RNA shield (ZYMOgen) had the upper hand, according to the subject donating samples ES1 and ES2. The DNA shield kit is both better suited for sampling and creates a less viscous sample, allowing for easier pipetting.

5.1.3 Quantification of bacterial and archaeal communities in fecal samples

By observing the cell count and the rRNA copy number per *E. coli* cell, the total copy numbers were estimated at 1.4*10^6 copies/ng of *E. coli* DNA. Average 16S rRNA gene copy number in bacterial genomes is 4.2 (Větrovský & Baldrian, 2013) but approximately 5.3 in the ES samples as revealed by taking the average copy numbers for the most abundant bacterial species (according to the tax table acquired by amplicon sequencing, section 4.1.4) and the rrnDB (Stoddard et al. 2015). (Exact calculation not shown)

The goal at first was to establish an optimal standard curve and with a few trial runs and by lowering the number of cycles in the qPCR, unspecific primer binding was eliminated and the standard curve was improved and by the fourth trial the test samples could be quantified. In the process the annealing temperature was tested and seemed to have little effect on the outcome, yet no double bands appeared on the gel. The results indicate that a lower cycle number prevents the formation of over amplification and smearing but shorter denaturing step only increases smear. Because fluorescence is measured and detected during the earlier cycles, effects on the fluorescence late in the PCR process should not have a big impact on the quality of the standard curve. There are many variables that can affect the amplification

in a PCR and therefore not certain that lowering the cycle number was enough by itself to fix the standard curve for the qPCR tests. For verification further analysis would be needed. Shortening the elongation step could also prevent smear and unspecific binding for future tests.

In total there were three successful bacterial qPCRs and three archaeal qPCRs that were repeated and had acceptable primer efficiencies (100- 110%) and linear correlation between dilution point within the standard curve ($R^2 > 0.99$). Quantification of the bacterial load of all stool samples contained around 10^7 copies/total DNA, when copy variation is not taken into account, and the ES1 and ES2 samples contained between 0.5% and 1% archaeal signatures (Fig. 13 B). The amplicon sequencing gave similar results with 0.4% archaeal signature when using the universal primer pair. There are a few uncertainties in the qPCR results, like copy number variation and fluorometric quantification of the test samples that can influence the relative abundance of archaea and bacteria. Comparing the sequencing and qPCR results indicates that the amplicon sequencing after using universal primers gives a good estimation of archaeal signatures and therefore might not be necessary doing a qPCR for measuring relative abundance if amplicon sequencing is available. More samples and comparison between them would be needed for a reliable conclusion.

The improvements in future quantifications should be making sure the standard sample is correctly quantified. If using the Qubit fluorometer, a sample with a known concentration could be used as a reference. The same goes for all test samples and being sure they all have the same DNA concentration for easier comparison.

6 Conclusion

No clear difference was detected between stool sampling methods. Based on user experience the DNA/RNA shield (ZYMOgen) had the advantage, according to the subject donating samples ES1 and ES2. The DNA shield kit is both better suited for sampling and creates a less viscous sample, allowing for easier pipetting.

Amplicon sequencing and qPCR methods from Pausan et al. (2018) were successfully reproduced and could be used for future quantification. Amplicon sequencing after using universal primers gave a good estimation of archaeal signatures in the microbiome and one archaeal ASV was detected from *M. smithii*. With the archaeal specific primers two ASVs were detected, both from *M. smithii*. It therefore might not be necessary to use qPCR for measuring relative abundance if amplicon sequencing is available.

References

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, *75*(2), 129–137. https://doi.org/10.3354/ame01753

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. In *Science* (Vol. 307, Issue 5717, pp. 1915–1920). American Association for the Advancement of Science. https://doi.org/10.1126/science.1104816

Cani, P. D. (2018). Human gut microbiome: hopes, threats and promises. *Gut*, *67*, 1716–1725. https://doi.org/10.1136/gutjnl-2018-316723

Conlon, M. A., & Bird, A. R. (2015). The impact of diet and lifestyle on gut microbiota and human health. In *Nutrients* (Vol. 7, Issue 1, pp. 17–44). MDPI AG. https://doi.org/10.3390/nu7010017

Cummings, J. H., & MacFarlane, G. T. (1997). Role of intestinal bacteria in nutrient metabolism. In *Clinical Nutrition* (Vol. 16, Issue 1, pp. 3–11). Churchill Livingstone. https://doi.org/10.1016/S0261-5614(97)80252-X

Drasar, B. S., & Barrow, P. A. (1985). Intestinal microbiology. *Intestinal Microbiology.*, *No. 10*.

Eisenstein, M. (2018). Microbiology: Making the best of PCR bias. *Nature Methods*, *15*(5), 317–320. https://doi.org/10.1038/nmeth.4683

Important Parameters of Quantitative PCR (qPCR) Analysis. (n.d.).

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), e1–e1. https://doi.org/10.1093/nar/gks808

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., & Moissl-Eichinger, C. (2017). First insights into the diverse human archaeome: Specific detection of Archaea in the gastrointestinal tract, lung, and nose and on skin. *MBio*, *8*(6), 1–17. https://doi.org/10.1128/mBio.00824-17

Olsen, G. J., & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *The FASEB*

Journal, *7*(1), 113–123. https://doi.org/10.1096/fasebj.7.1.8422957

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pausan, M. R., Csorba, C., Singer, G., Till, H., Schöpf, V., Santigli, E., Klug, B., Högenauer, C., Blohs, M., & Moissl-Eichinger, C. (2018). Measuring the archaeome: detection and quantification of archaea signatures in the human body. *BioRxiv*, 334748. https://doi.org/10.1101/334748

Pausan, M. R., Csorba, C., Singer, G., Till, H., Schöpf, V., Santigli, E., Klug, B., Högenauer, C., Blohs, M., & Moissl-Eichinger, C. (2019). Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02796

Pezza, J. A., & Sun, L. (n.d.). *Polymerase Fidelity: What is it, and what does it mean for your PCR? | NEB*. Retrieved May 10, 2020, from https://international.neb.com/tools-and-resources/feature-articles/polymerase-fidelity-what-is-it-and-what-does-it-mean-for-your-pcr

Potapov, V., & Ong, J. L. (2017). Examining sources of error in PCR by single-molecule sequencing (PLoS ONE (2017) 12:1 (e0169774) DOI: 10.1371/journal.pone.0169774). In *PLoS ONE* (Vol. 12, Issue 7). Public Library of Science. https://doi.org/10.1371/journal.pone.0181128

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Salzman, N. H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjöberg, J., Amir, E., Teggatz, P., Barman, M., Hayward, M., Eastwood, D., Stoel, M., Zhou, Y., Sodergren, E., Weinstock, G. M., Bevins, C. L., Williams, C. B., & Bos, N. A. (2009). Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunology*, *11*. https://doi.org/10.1038/ni.1825

Scanlan, P. D., Shanahan, F., & Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiology*, *8*(1), 79. https://doi.org/10.1186/1471-2180-8-79

Schulz, H. N., & Jørgensen, B. B. (2001). Big Bacteria. *Annual Review of Microbiology*, *55*(1), 105–137. https://doi.org/10.1146/annurev.micro.55.1.105

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *Revised Estimates for the Number of Human and Bacteria Cells in the Body*, 036103. https://doi.org/10.1101/036103

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., & Schmidt, T. M. (2015). rrnDB:

Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, *43*(D1), D593–D598. https://doi.org/10.1093/nar/gku1201

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. In *Nature* (Vol. 449, Issue 7164, pp. 804–810). Nature Publishing Group. https://doi.org/10.1038/nature06244

Větrovský, T., & Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, *8*(2), e57923. https://doi.org/10.1371/journal.pone.0057923

Vogtmann, E., Chen, J., Amir, A., Shi, J., Abnet, C. C., Nelson, H., Knight, R., Chia, N., & Sinha, R. (2017). Comparison of Collection Methods for Fecal Samples in Microbiome Studies. *American Journal of Epidemiology*, *185*(2), 115–123. https://doi.org/10.1093/aje/kww177

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2016). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *MSystems*, *1*(1). https://doi.org/10.1128/msystems.00009-15

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 95, Issue 12, pp. 6578–6583). National Academy of Sciences. https://doi.org/10.1073/pnas.95.12.6578

Woese, C. R., Kandlert, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya (Euryarchaeota/Crenarchaeota/kingdom/evolution). In *Proc. Nati. Acad. Sci. USA* (Vol. 87).

Appendix A

1.1 Protocols

1.1.1 Fecal sampling

Copy of protocol for DNA/RNA Shield™ Fecal Collection Tube:

DNA/RNA Shield™ Fecal Collection Tube instructions

1. Prepare and collect fecal specimen using preferred fecal specimen collection set/kit. Note: Method of collecting the fecal sample must prevent feces from falling into toilet water to avoid sample contamination.
2. Unscrew the collection tube cap and use the spoon to scoop **one spoonful** of feces (approximately 1 gram or 1 mL in volume) from a sample.
3. Place the sample in the collection tube.
4. Tighten the cap and shake to mix the contents thoroughly (invert 10 times) to create a suspension.
   Note: Some fecal material may be difficult to re-suspend. As long as the material is suspended, the sample is stabilized. foaming/frothing during shaking is normal.
5. Dispose of unused fecal material and thoroughly wash hands according to your institution's guidelines.

Copy of protocol for OMNIgene-GUT fecal collection kit (OM-200):

OMNIgene-GUT fecal collection kit instructions

1. Important preparation:
   - Empty bladder before beginning the collection.
   - Collect fecal sample free of urine or toilet water.
   - Toilet paper or tissues may be required.
2. While holding the yellow tube top, unscrew only the purple cap from the kit and set aside for later use.
   Important:
   Do not remove the yellow tube top.
   Do not spill the stabilizing liquid in the tube.
3. Use the spatula to collect the small amount of fecal sample.
4. Transfer the fecal sample into the yellow tube top. Repeat until the sample fills the yellow tube top.
   Important: Do not push sample into the tube.
5. Scrape horizontally across the tube top to level the sample and remove any excess. Wipe exterior of tube and top with toilet paper a-or tissue as needed.
6. Pick up the purple cap with the solid end facing down and screw onto the yellow tube top until tightly closed.

7. Shake the sealed tube as hard and fast as possible in a back and forth motion for a minimum of 30 tseconds.
8. The fecal sample will be mixed with the stabilizing liquid in the tube; not all particles will dissolve.
   Important: Continue shaking if large particles remain.
9. Place spatula in original packaging or wrap in toilet paper and discard in garbage.

1.1.2 DNA Extraction from stool

Copy of protocol for QIAamp PowerFecal Pro DNA Kit:

Important notes before starting

- Ensure that the PowerBead Pro Tubes rotate freely in the centrifuge without rubbing.
- If Solution CD3 has precipitated, heat at 60°C until precipitate dissolves.
- Perform all centrifugation steps at room temperature (15–25°C).

Procedure

1. Spin the PowerBead Pro Tube briefly to ensure that the beads have settled at the bottom. Add up to 250 mg of stool and 800 µl of Solution CD1. Vortex briefly to mix.
2. Secure the PowerBead Pro Tube horizontally on a Vortex Adapter for 1.5– 2 ml tubes (cat. no. 13000-V1-24). Vortex at maximum speed for 10 min.
   Note: If using the Vortex Adapter for more than 12 preps simultaneously, increase the vortexing time by 5–10 min.
3. Centrifuge the PowerBead Pro Tube at 15,000 x g for 1 min.

4. Transfer the supernatant to a clean 2 ml Microcentrifuge Tube (provided).

Note: Expect 500–600 µl. The supernatant may still contain some stool particles.

5. Add 200 µl of Solution CD2 and vortex for 5 s.

6. Centrifuge at 15,000 x g for 1 min at room temperature. Avoiding the pellet, transfer up to 700 µl of supernatant to a clean 2 ml Microcentrifuge Tube (provided).

Note: Expect 500–600 µl.

7. Add 600 µl of Solution CD3 and vortex for 5 s.

8. Load 650 µl of the lysate onto an MB Spin Column and centrifuge at 15,000 x g for 1 min.

9. Discard the flow-through and repeat step 8 to ensure that all of the lysate has passed through the MB Spin Column.

10. Carefully place the MB Spin Column into a clean 2 ml Collection Tube (provided). Avoid splashing any flow-through onto the MB Spin Column.

11. Add 500 µl of Solution EA to the MB Spin Column. Centrifuge at 15,000 x g for 1 min.

12. Discard the flow-through and place the MB Spin Column back into the same 2 ml Collection Tube.

13. Add 500 µl of Solution C5 to the MB Spin Column. Centrifuge at 15,000 x g for 1 min.

14. Discard the flow-through and place the MB Spin Column into a new 2 ml Collection Tube (provided).

15. Centrifuge at up to 16,000 x g for 2 min. Carefully place the MB Spin Column into a new 1.5 ml Elution Tube (provided).

16. Add 50–100 µl of Solution C6 to the center of the white filter membrane.

17. Centrifuge at 15,000 x g for 1 min. Discard the MB Spin Column. The DNA is now ready for downstream applications.

Note: We recommend storing the DNA frozen (–30 to –15°C or –90 to –65°C) as Solution C6 does not contain EDTA. To concentrate DNA, refer to the Troubleshooting Guide.

1.1.3 DNA extraction from *E. coli* cells

Copy of protocol for MasterPure™ complete DNA and RNA purification kit:

Cell Samples (e.g., mammalian cell culture, buccal cells, E. coli)

1. Dilute 1 µl of Proteinase K into 300 µl of Tissue and Cell Lysis Solution for each sample.
2. Pellet cells by centrifugation (0.5-1 x 106 mammalian cells; 0.1-0.5 ml of an overnight culture of E. coli) and discard the supernatant, leaving approximately 25 µl of liquid.
3. Vortex for 10 seconds to resuspend the cell pellet.
4. Add 300 µl of Tissue and Cell Lysis Solution containing the Proteinase K and mix thoroughly.
5. Incubate at 65°C for 15 minutes; vortex every 5 minutes.
6. Cool the samples to 37°C and add 1 µl of 5 µg/µl RNase A to the sample; mix thoroughly.
7. Incubate at 37°C for 30 minutes.
8. Place the samples on ice for 3-5 minutes and then proceed with total DNA precipitation in Part D.

D. Precipitation of Total DNA (for all biological samples)

1. Add 150 µl of MPC Protein Precipitation Reagent to 300 µl of lysed sample and vortex vigorously for 10 seconds.
2. Pellet the debris by centrifugation at 4°C for 10 minutes at ≥10,000 x g in a microcentrifuge. If the resultant pellet is clear, small, or loose, add an additional 25 µl of MPC Protein Precipitation Reagent, mix, and pellet the debris again.
3. Transfer the supernatant to a clean microcentrifuge tube and discard the pellet.
4. Add 500 µl of isopropanol to the recovered supernatant. Invert the tube 30-40 times.

5. Pellet the DNA by centrifugation at 4°C for 10 minutes in a microcentrifuge.
6. Carefully pour off the isopropanol without dislodging the DNA pellet.
7. Rinse twice with 70% ethanol, being careful to not dislodge the pellet. Centrifuge briefly if the pellet is dislodged. Remove all of the residual ethanol with a pipet.
8. Resuspend the DNA in 35 µl of TE Buffer

1.1.4 DNA quantification

Copy of protocol for Quant-iT ™ Assays abbreviated protocol:

IMPORTANT: Ensure all assay reagents are at room temperature before you begin.

1. Set up your tubes: you'll need two tubes for the standards (three for the protein assay) and one tube for each of your samples.
2. Make the Quant-iT™ Working Solution by diluting the Quant-iT™ reagent 1:200 in Quant-iT™ buffer. 200 uL of Working Solution are required for each sample and standard.
3. Prepare Assay Tubes according to the table below.

Table 14: Assay tube preperation.

	Standard Assay tubes	User Sample Assay tubes
Volume of Working Solution (from step 2) to add	190 µL	180-199 µL
Volume of Standard (from kit) to add	10 µL	---
Volume of user sample to add	---	1-20 µL
Total volume in each assay tube	200 µL	200 µL

4. Vortex all tubes for 2–3 seconds.
5. Incubate the tubes for 2 minutes at room temperature (15 minutes for the Quant-iT™ protein assay).
6. Read tubes in Qubit™ fluorometer.
7. Multiply by the dilution factor (see Manual) to determine concentration of your original sample. Alternatively, choose Calculate sample concentration to have the Qubit™ fluorometer perform this multiplication for you.
8. * Use only thin-wall, clear 0.5 mL PCR tubes. Acceptable tubes include
9. Qubit™ assay tubes (500, Invitrogen Cat. no. Q32856) or Axygen PCR05-C tubes (VWR, part number 10011-830).

Appendix B

1.1 DNA quantification with Qubit and assessment using Nanodrop

Table 15 : Quantification of stool sample DNA

Samples	ES1 (75 ng/µl)	ES2 (10.5 ng/µl)	SS1	Xav	SK
Concentration (ng/µl)	283	30.6	248	585	232
Dilution to 10 ng/µl	3.5 : 96.5	32.7 : 67.3	4 :96	1.7 : 98.3	4.3 : 95.4

Table 16: Quantification of stool sample DNA and E. coli *culture*

Samples	ES1	ES2	SS1	Xav	SK	*E. coli* culture
Concentration according to previous dilutions (ng/µl)	10	10	10	10	10	59.4
Concentration revaluated (ng/µl)	3.77	3.93	3.99	2.83	2.78	55.6

Table 17: Sample purity and quantification results from nanodrop 1000 UV-Vis spectrophotometer

Samples	*E.coli* (9,25ng/µl)	*M.smithii*	ES1	ES2	SS	Xav	SK
nanodrop 260/280	2.13	2.02	1.53	1.74	1.58	1.35	1.18
nanodrop 260/230	1.25	0.07	0.29	0.11	0.34	0.55	0.54
nanodrop concentration (ng/µl)	58.8	41.6	4.9	5.5	5.4	6	5.1