



Norges miljø- og  
biovitenskapelige  
universitet

Master's Thesis 2016 60 ECTS  
Institute of Chemistry, Biotechnology, and Food Sciences

# IS-profiling as a Screening Tool for Intestinal Dysbiosis in Adenoma Development

J. Christopher Noone  
MSc Biotechnology, molecular biology

# TABLE OF CONTENTS

Acknowledgements .....	4
Abstract.....	5
Abbreviations.....	6
1. Introduction.....	7
1.1 Polyps and cancer .....	7
1.1.1 Colorectal cancer .....	7
1.1.2 Polyps.....	7
1.1.3 Adenomas.....	8
1.1.4 Colorectal carcinoma .....	9
1.1.5 The molecular pathogenesis of CRC.....	9
1.1.6 Screening via colonoscopy .....	11
1.2 The human microbiota.....	11
1.2.1 Diet as risk factor for colorectal cancer .....	13
1.2.2 Species Diversity.....	13
1.2.3 Methods for studying species diversity .....	13
1.2.4 IS-profiling .....	14
1.2.5 DNA-extraction.....	16
1.3 CRC-project at Ahus .....	16
1.4 Summary of Project Aims.....	17
2. Materials and Methods .....	18
2.1. Materials .....	18
2.1.1 Bacterial strains:.....	18
2.1.2 Patient and subject stool samples .....	19
2.1.3 Commercial kits, buffers, chemicals, media, and Enzymes .....	19
2.1.4 Oglionucleatides, Reference DNA, primers, cloning vectors, and size markers. ....	22
2.1.5 Instruments, software, and diverse materials. ....	23
2.2. Methods.....	24
2.2.1 Part 1: Technical evaluation of IS-profiling.....	24
2.2.1.1 DNA extraction of bacterial cultures .....	24
2.2.1.2 IS-profiling procedure .....	24
2.2.1.3 Sensitivity and specificity of IS-profiling .....	25
2.2.1.4 Cloning of IS-pro fragments .....	26
2.2.1.5 Sequencing of recombinant plasmids .....	26
2.2.1.6. Reproducibility of IS-pro .....	26
2.2.2. Part 2: Comparison of DNA extraction kits.....	27
2.2.2.1 Fecal samples .....	27
2.2.2.2 DNA extraction with PowerMag Microbiome RNA/DNA Isolation Kit (adapted for manual pipetting) .....	27
2.2.2.3 DNA extraction with QIAamp Stool DNA Mini Kit .....	27
2.2.2.4 DNA extraction with Stratec Molecular's PSP Spin DNA Kit.....	28
2.2.2.5 DNA extraction with Qiagen's QIASymphony SP System .....	28
2.2.2.6 Quantification of bacterial DNA by 16S q PCR .....	28
2.2.2.7 Quantification of human DNA by Quantifiler.....	29
2.2.2.8 Bacterial diversity analysis by IS-profiling .....	29
2.2.3 Part 3: Comparison of bacterial composition and human DNA concentration in fecal samples from polyp patients and controls.....	30
2.2.3.1. Samples .....	30
2.2.3.2. DNA extraction.....	30
2.2.3.3. Quantitation of Bacterial DNA present in stool samples .....	31
2.2.3.4 Quantification of human DNA present in stool samples .....	31

2.2.3.5 IS-pro .....	31
2.2.3.6 Raw data processing .....	31
2.2.3.7 Comparison of the bacterial composition in stool samples.....	31
3. Results.....	33
3.1 Technical evaluation of IS-profiling (IS-pro).....	33
3.1.1 Optimization of IS-pro.....	33
3.1.2 Sensitivity and specificity of IS-profiling .....	33
3.1.3 Cloning of IS-Pro fragments .....	35
3.1.4 Reproducibility of IS-pro .....	36
3.2 Part 2: Comparison of DNA extraction kits.....	39
3.2.1 Quantification of total bacterial DNA by 16S qPCR.....	40
3.2.2 Quantification of human DNA by Quantifiler.....	42
3.2.3 Bacterial diversity analysis by IS-profiling .....	46
3.3 Part 3: Comparison of the bacterial composition and human DNA concentration in fecal samples from polyp patients and controls.....	48
3.3.1 Quantitation of bacterial DNA present in stool samples.....	48
3.3.2 Quantitation of Human DNA present in stool samples .....	49
3.3.3 Comparison of the bacterial composition in stool samples.....	50
3.3.4 Identification of group specific bands.....	54
4. Discussion .....	56
4.1 Part 1: Technical evaluation of IS-profiling .....	56
4.2 Part 2: Comparison of DNA extraction kits.....	58
4.3 Part 3: Comparison of the bacterial composition and human DNA in fecal samples from polyp patients and controls.....	59
References .....	63
Appendices.....	66

---

# ACKNOWLEDGEMENTS

This study was carried out at Akershus University Hospital's Section for Molecular Biology and in the laboratory of EpiGen.

I would like to thank the following people and institutions for their help in the completion of this master's thesis.

Hege Smith Tunsjø, thank you for your inspiration and your positive attitude. Thanks you also for sharing your expertise, knowledge, and experience in the completion this thesis.

Vahid Bemanian, thank you for taking the time to impart your knowledge of molecular biology, and the fruits of your experience during the course of the writing of this thesis.

Trude Haga Flatås, thank you to for your cooperation and encouragement for final push to complete this thesis.

Thanks to Katleen Vranckx of Applied Maths, Ghent, Belgium. Applied Maths' course in BioNumerics and their comprehensive helpdesk advice concerning the use of BioNumerics analysis software helped enormously.

Thanks to Fredrik A. Dahl at Statistikkpoliklinikken, Akershus University Hospital for helping me to sort out the appropriate statistical analyses for my data.

Janne Beate Utåker, thank you for you help with all things administrative at UNMB.

Finally, thanks to Akershus University Hospital for allowing me to complete this master's education during work hours.

---

# ABSTRACT

The collective microbial genome of the human gut, the microbiome, is an essential component to the homeostasis and wellbeing of an individual. Dysbiosis here is implicated in several disorders, making it an important focal point for research. The current thesis evaluates a non-invasive method of analyzing the delicate balance of the normal intestinal flora in the quest for a establishing a screening method by which to facilitate the discovery of indications of colorectal cancer in its earliest stages.

The method, IS-profiling, is a fragment analysis of the bacterial 16S-23S rRNA inter-genic spacer region, using phylum specific fluorescent PCR primers. The species specific variation in IS-fragment length across patients' intestinal flora produces a molecular fingerprint via capillary gel electrophoresis.

The method's necessity for an efficient method of bacterial DNA extraction precipitated an evaluation of DNA extraction methods which led to the choice of the Stratec PSP Stool DNA Extraction Kit. It provided both sufficient amounts of bacterial DNA and the broadest bacterial diversity when compared to the other extraction methods tested here.

Stool samples were collected from two groups: 1. the healthy group, a group of presumed healthy volunteers; and 2. the polyp group, patients who were scheduled for removal of adenomas of the large intestine and/or rectum. The samples were collected 1-2 days before colonoscopy related bowel preparation.

In addition to IS-profiling of the bacterial composition, total bacterial 16S rRNA concentration, and human DNA concentration were also analyzed.

Results of group analysis using the jackknife test revealed an apparent similarity among the samples of the polyp group, placing them correctly in 91.67% of blind attempts. No such success was recorded in the healthy group (39.39% correct). IS-profiling revealed no obvious common peak pattern to either group. It is assumed that the success of the jackknife test is a due to some obscure complex relationship between several peaks inconspicuous to manual interpretation.

Results of 16S bacterial rRNA qPCR showed a healthy group average concentration that was approximately 10 ng/μl lower than that of the polyp group's. Human DNA was present in a concentration over ten times greater in the polyp group compared to the healthy group.

In conclusion, IS-profiling presents itself as a useful, but possibly limited, tool for non-invasive screening for intestinal microbe dysbiosis. Our results suggest that, with the right analysis modules in place, changes in gut microbe profiles could be detected and followed up with additional tests. The limitations come, however, in the form species identification of peaks of interest, as well as in general knowledge of the 16S-23S IS-region; NCBI nucleotide database search yields only 58 871 items versus a 16S search which yielded 11 660 431 items at the time of this writing.

---

# ABBREVIATIONS

ACF: aberrant crypt focus  
CIMP: CpG island methylator phenotype  
CIN: chromosome instability  
CpG: cytosine-phosphate-guanine  
CRC: colorectal cancer  
CT: computer tomography  
DGGE: denaturing gradient gel electrophoresis  
EHEC: enterohemorrhagic *Escherichia coli*  
FOBT: fecal occult bleeding testing  
GI: gastro intestinal  
hDNA: human DNA  
IS-pro: inter-genic space profiling  
MMR: mismatch repair  
MSI: microsatellite instability  
NGS: next generation sequencing  
NTC: no template control  
OBL: off-board lysis  
PCR: polymerase chain reaction  
PM: Mobio PowerMag Microbiome RNA/DNA Isolation Kit  
Q: Qiagen Qiamp Stool DNA Mini kit  
QE: Qiagen Qiamp Stool DNA Mini kit with enzymatic pre-treatment  
QS: QIAsymphony extraction robot  
RFU: relative fluorescence units  
rRNA: ribosomal RNA  
RT PCR: real time PCR  
S: Stratec PSP Spin Stool DNA extraction kit  
SE: Stratec PSP Spin Stool DNA extraction kit with enzymatic pre-treatment  
SCFA: short chain fatty acid  
SE: Stratec PSP Spin Stool DNA extraction kit with enzymatic pre-treatment  
UM: unlabeled IS-pro PCR primers  
WNT: wingless-related integration site

---

# 1. INTRODUCTION

Presented here are, first, some background pertaining to colorectal cancer (CRC), and the progression from normal intestinal epithelial cells to polyps to cancer. Morphology, terminology, molecular pathogenesis, and patient screening are considered as they pertain to neoplastic growth and CRC.

The second part of the introduction addresses the human intestinal microbiome. Specific examples of bacterial symbiosis and dysbiosis are presented, as well as species diversity and methods of evaluating such. Both bacterial and human DNA extraction from faeces samples is touched upon, as is this project's role within a larger research project (CRC).

Finally, the introduction closes with a summary of this project's aims.

## 1.1 POLYPS AND CANCER

### 1.1.1 COLORECTAL CANCER

Colorectal cancer is the third most common cancer in the world, with nearly 1.4 million new cases diagnosed in 2012<sup>1</sup>

Norway is among the world leaders of incidents of colorectal cancer. In 2014, there were 4166 new incidents. Statistics calculated during the period of 2010-2014 indicate that 10.7 women, and 17.1 men per 100 000 can expect to be inflicted by the disease<sup>2</sup>.

The increased survival rate has, however, made the situation slightly less grim than the numbers belie; five year survival was approximately 30% in the 1970's, but in 2014 it became as high as 60%. Better surgical treatment is at least partly responsible of this dramatic rise in survival. Just in the last ten years, recurrences of the disease have dropped from 30% to 10%<sup>2</sup>.

### 1.1.2 POLYPS

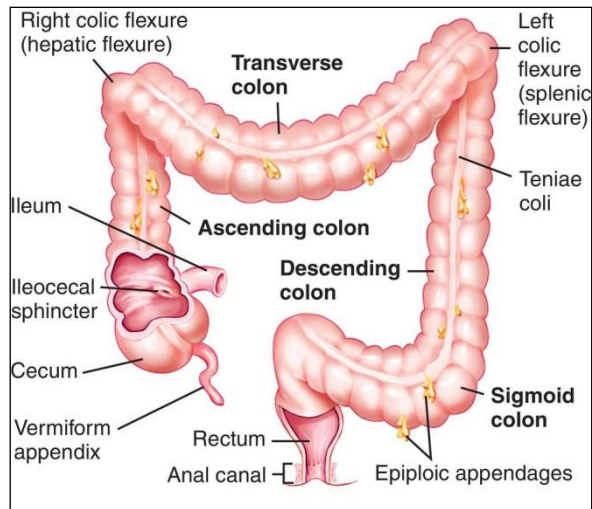
Polyps are defined as tumorous mass that protrudes into the lumen of the gut. Those with stalks are called pedunculated (figure 1.2); those without are known as sessile. Unless otherwise specified, the term polyp refers to lesions that arise from the epithelium of the mucosa.

The presence of polyps in the large intestine is quite common in Western society, as is the incidence of colorectal cancer (Eide, 1986). Polyps are roughly divided into two groups: neoplastic and non-neoplastic. Non-neoplastic polyps are generally benign, the exception being instances of polyposis syndrome (Absar and Haboubi, 2004). The great majority of intestinal polyps arises sporadically, and is increasingly more common with age. Non-neoplastic polyps comprise about 90% of polyps in the large intestine, and are found in more than 50% of people over the age of 60; most of these are hyperplastic polyps and have no malignant potential; the exception to this being some cases of hyperplastic polyps in the right side of the colon (figure 1.1), which may develop into colorectal carcinomas due to microsatellite instability, leading to cancer via the mismatch repair pathway (Kumar et al., 2003).

---

<sup>1</sup> <http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/colorectal-cancer-statistics>

<sup>2</sup> <http://www.kreftregisteret.no/no/Generelt/Fakta-om-kreft-test/Tykk--og-endetarmskreft> (3..3.2016)



**Figure 1.1:** Anatomy of the colon and rectum.

(Source:[http://medical-dictionary.thefreedictionary.com/\\_/viewer.aspx?path=MosbyMD&name=colon.jpg&url=http%3A%2F%2Fmedical-dictionary.thefreedictionary.com%2Fcolon](http://medical-dictionary.thefreedictionary.com/_/viewer.aspx?path=MosbyMD&name=colon.jpg&url=http%3A%2F%2Fmedical-dictionary.thefreedictionary.com%2Fcolon))

### 1.1.3 ADENOMAS

Adenomas are neoplastic polyps that arise from epithelial proliferation and dysplasia, which ranges from mild to a degree of severity that indicates the transformation of carcinoma. It is thought that single or multiple adenomas are present in over one third of the over 55 population (Eide, 1991). The majority of sporadic invasive colorectal adenocarcinomas are believed to arise from preexisting adenomas. According to Robbins Basic Pathology (7<sup>th</sup> ed.), the prevalence of colorectal adenomas is 20-30% before age 40, and rises to 40-50% after the age of 60. There is a four-fold greater risk for sporadic adenomas in those with first degree relatives that have adenomas; there is also a four-fold greater risk of colorectal carcinomas in any patient with adenomas.

There are three sub-types of adenomas (figure 1.2):

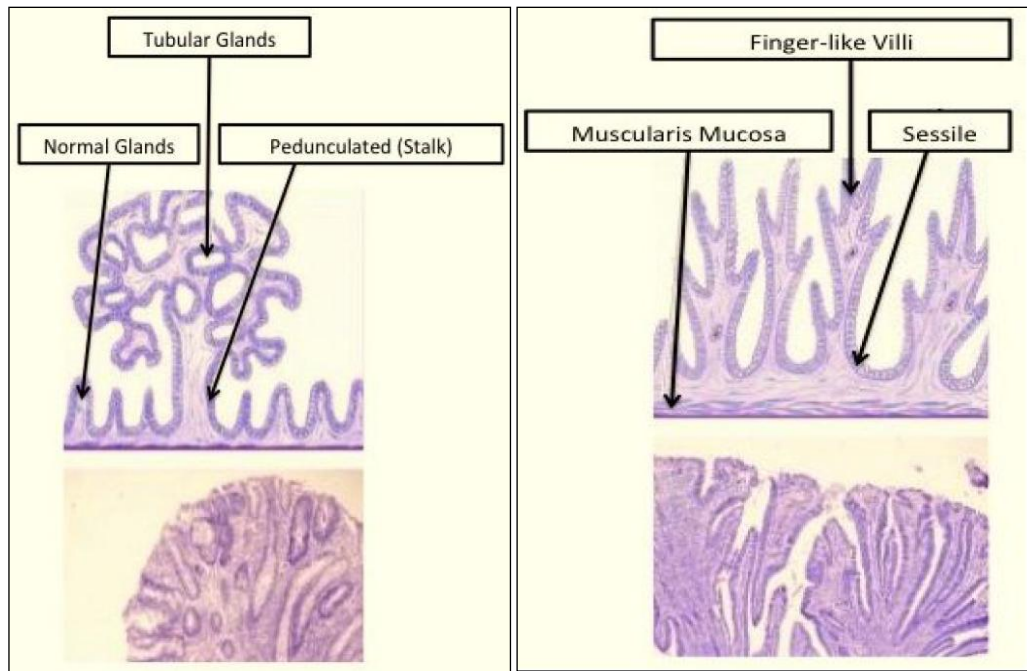
1. Tubular adenomas – mostly tubular glands; recapitulating mucosal topology
2. Villous adenomas – villous projections
3. Tubulovillous adenomas – a mixture of the two above types

Their potential for malignancy shown is in parentheses: tubular adenoma (5%), villous adenoma (41%), tubulovillous adenoma (23%), and adenocarcinoma (malignant polyp) (Absar and Haboubi, 2004). The potential for transformation to malignancy increases with the size, cellular atypia, and the presence of villous morphology of the adenoma (Muto et al., 1975).

Of these sub-types, tubular adenomas are most prevalent; 5-10% of adenomas are tubulovillous; and about 1% is villous. The vast majority of tubular adenomas are small and pedunculated. Although malignant risk is correlated with polyp size, histological architecture, and the severity of epithelial dysplasia, the principal determinant of an adenomas malignant potential is the polyp's maximum diameter.

All adenomas in the alimentary tract are potentially malignant, requiring expedient and adequate removal (Kumar et al., 2003).





**Figure 1.2:** Histological representations of pedunculated, tubular, and villous adenoma (Source: <http://cdn.intechopen.com/pdfs-wm/49127.pdf>)

#### 1.1.4 COLORECTAL CARCINOMA

About 98% of all cancer in the large intestine are adenocarcinoma, and are generally curable by resection. Thusly, early discovery is of paramount importance.

Adenocarcinomas constitute the bulk of colorectal cancers, comprising about 70% of all malignancies in the gastrointestinal (GI) tract (Kumar et al., 2003).

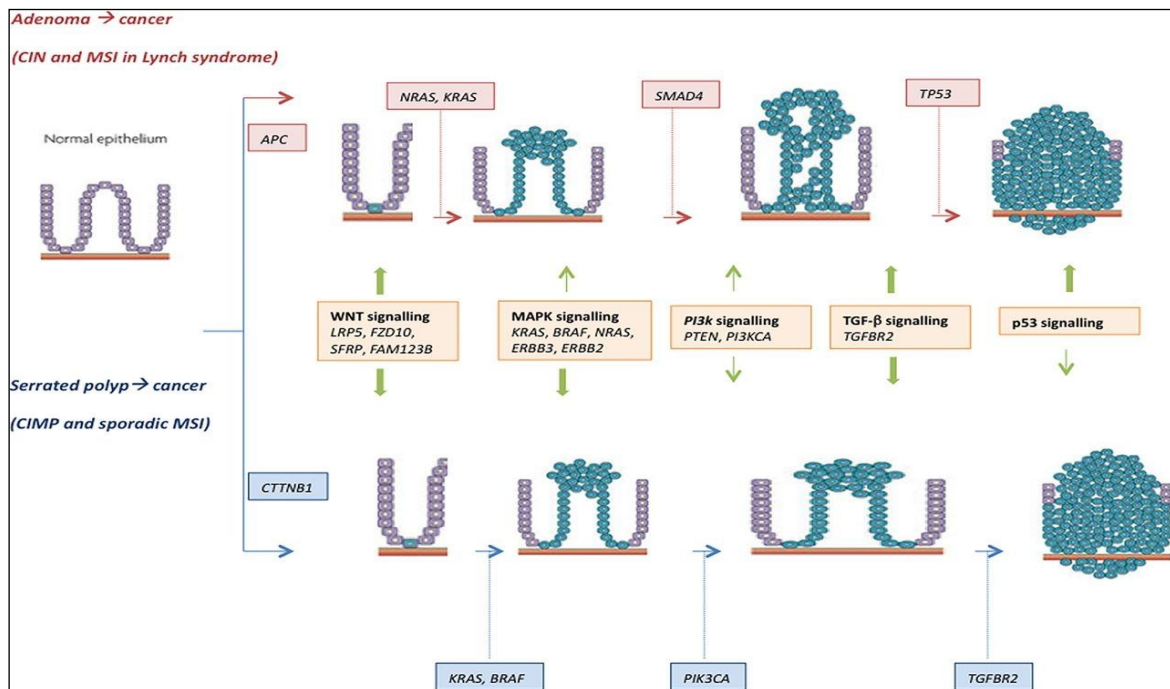
The development of colorectal cancer is multifarious process precipitated by an accumulation of somatic mutations, and epigenetic aberrations within the epithelial cells of the intestine. This process can take several years to develop into metastatic cancer, making a method early detection particularly desirable prospect in the reduction of the CRC mortality rate (Kumar et al., 2003).

#### 1.1.5 THE MOLECULAR PATHOGENESIS OF CRC

Current thought is that there are three pathologically distinct avenues that initiate the transformation of normal colonocytes to aberrant crypt foci (ACF) to polyp, ultimately leading to the development of colorectal cancer.

The first two pathways in the adenoma-carcinoma sequence, chromosomal instability (CIN) and microsatellite instability (MSI), affect adenomas and involve an accumulation of mutations, but the genes involved and the mechanisms by which they mutate are different (Kumar et al., 2003; Grady and Markowitz, 2014). The third pathway is known as CpG island methylator phenotype (CIMP) and is characterized by a high frequency of aberrantly methylated CpG loci which can lead to the development of CRC in a subset of polyps known as sessile serrated polyps (figure 1.3)(Grady and Markowitz, 2014; Dickinsen et al., 2015).

The adenoma-carcinoma sequence is also known as the APC/ $\beta$ -catenin pathway. It is believed that the first step in the in this pathway is the loss of the APC tumor suppressor gene; when both copies of this gene become mutated, it allows adenomas to develop; APC mutations are present in 85% of sporadic colorectal cancer cases. Next in the sequence is mutations of the WNT signal pathway. Mutations of the MAPK signal pathway follow; mutations here lead to a constitutively active state that signals mitosis, and prevents apoptosis. Next up is the loss of a putative cancer suppressor gene 18q21; a deletion here has been detected in 60-70% of colorectal cancers; the leading candidates here are DCC, DPC4/SMAD4, and SMAD2, but it is unclear which of these is instrumental in the progression to colorectal cancer. Finally, the loss of TP53 tumor suppressor gene completes the sequence (Kumar et al., 2003; Walther et al., 2009).



**Figure 1.3:** Illustration of the two pathways from normal epithelium to colorectal cancer: chromosomal instability (CIN)/ microsatellite instability (MSI), and CpG island methylator phenotype (CIMP)/sporadic MSI. Some of the more common mutations are depicted between adenoma representations. (Source: Dickinsen et al., 2015)

The MSI pathway accounts for about 15% of CRCs and involves inherited mutations in one of several mismatch repair (MMR) genes: MSH2, MSH6, MLH1, PMS1, and PMS2. Mutations in these genes that guard against mismatches during DNA replication results in a hypermutable condition affecting repetitive DNA sequences known as microsatellites. Some microsatellite sequences lie in the promotor region of genes such as type II TGF- $\beta$  receptor and BAX, which play roles in cell growth regulation and apoptosis respectively (Kumar et al., 2003; Grady and Markowitz, 2014).

Epigenetic alterations in the form of aberrant DNA methylation is a common thread in virtually all CRCs, but about 10-20% of cases have extremely high proportion of aberrantly methylated CpG loci; these CRCs are characterized as CIMP. CIMP tumors are thought to represent a distinct subclass of CRC, but there is a lack of consensus with regards to classification criteria. This is likely due to the lack of clear understanding of the underlying mechanisms (Grady and Markowitz, 2014).

DNA methylation of CpG rich regions in the 5' and of genes has the effect of changing chromatin structure and effectively silences the gene by making it inaccessible to transcription factor binding (Dickinsen et al., 2015). Some of the genes used to identify CIMP cancers include RUNX3, SOCS1, NEUROG1, CACNA1G, AND IGF2 (Grady and Markowitz, 2014).

### 1.1.6 SCREENING VIA COLONOSCOPY

Screening and prompt intervention in the form of excision polyps in their early stages, before metastatic spread of the disease, has most certainly played a role in the drop in mortality from colorectal cancer. Screening involves examination of those over 50 years old, despite the absence of symptoms. Screening via flexible sigmoid colonoscopy as a preventative measure has become common practice in the USA, and several European countries. Other popular screening methods include fecal occult bleeding testing (FOBT), and CT colonoscopy. (Schetter et al., 2012; Horiuchi and Tanaka, 2014)

In a quest for less invasive CRC screening methods, the increase in exfoliation from colonic neoplasms is currently being exploited to reveal molecular alterations associated with CRC in host DNA. Assays here include those to detect aberrant DNA methylation, gene mutations, and aberrant micro-RNA expression patterns (Schetter et al., 2012; Dickinsen et al., 2015).

## 1.2 THE HUMAN MICROBIOTA

The human microbiota refers to the collection of microorganisms present in various locations through out the human body. It is increasingly being acknowledged for its integral capacity as a homeostatic regulator of health. Examples of human microbiome symbiosis are evident in mutualistic role of *Bacteroides fragilis* in the development and regulation of the immune system (Troy and Kasper, 2011), the increased risk of atopic dermatitis in genetically predisposed individuals lacking in chronic early life exposure to antimicrobial antigens (Baker, 2006), and the beneficial effects of butyrate-producing bacteria which maintain healthy colonocytes, and regulate anti-inflammatory and anti-carcinogenic gene expression (Hamer, et al., 2008).

More microbes are present in the large intestine than anywhere else in the human body. The human intestinal microbiome is defined as the collective bacterial genome of all species present in the the gut. It is estimated at 4 million genes, and, together with viruses, fungi, and archaea, it is thought to be comprised of approximately 1000 bacterial species, divided predominately in 2 bacterial phyla: *Bacteroidetes* and *Firmicutes* (Dethlefsen et al., 2006; Dulal et al., 2014). These two phyla, together with *Actinobacteria*, are implicated in colonic health (Ng et al., 2013). Of these, *Bifidobacterium*, *Eubacterium*, *Bacteroides*, *Enterococcus*, and the family *Enterobacteriaceae* are the most common. It is estimated that more than  $10^{11}$  bacteria cells are present in one gram of feces, anaerobic species dominating by greater than 1000-fold. *E. coli*, having become possibly the most familiar to the lay population due to its role as a foodborne contaminant, represents less than 1% of human intestinal flora; it is, however, the bacteria responsible for more intra-abdominal disease than any other aerobic or facultative anaerobic species. The strict anaerobe most often responsible for intra-abdominal disease is *Bacteroides fragilis*. The two most abundant genera in the human intestine, *Eubacterium* and *Bifidobacterium*, are rarely pathogenic (Murray et al., 2002).

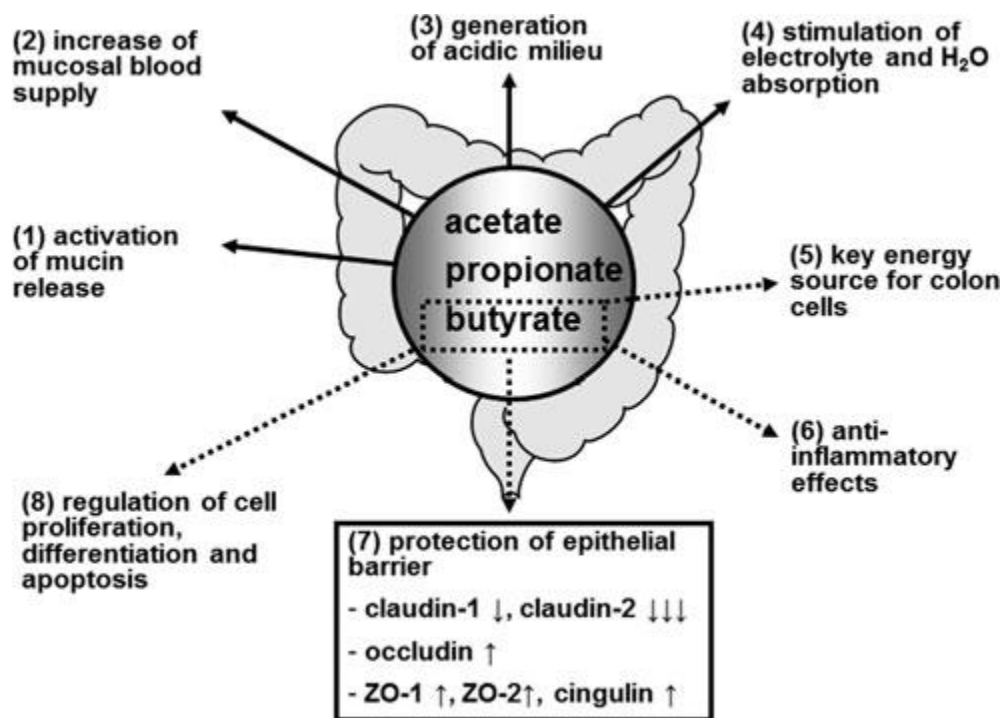
Alterations in the profile of one's intestinal microflora can result in the aftermath of antibiotic treatment, leading to morbidity. Antibiotic resistant strains of *Pseudomonas* and *Enterococci* are selected for and begin to proliferate. This shift in intestinal flora can cause the reduction of bacterial species such as *Lactobacilli* and *Enterococci* that resist colonization and hinder reproduction of *C. difficile*, a well documented pathogen. *C. difficile* thus begins to thrive in this post antibiotic environment (Murray et al., 2002; Tonna, 2005)

Exposure to other enteric pathogens such as *Shigella spp.*, enterohemorrhagic *E. coli* (EHEC), and *Entamoeba histolytica* are also known to disrupt one's normal gut flora, causing various intestinal diseases, not the least of which being cancer (Murray et al., 2002). Researchers have begun to postulate a succession of events beginning with the chronic intestinal inflammation leading to an increase in genotoxic *Proteobacteria*, progressing eventually to cancer (Arthur et al., 2012; Schwabe and Wang, 2012; Boleij et al., 2015)

Schwabe and Wang (2012) have implicated *E. coli* NC101 and its genotoxin colibactin, a polyketide synthase encoded by the genomic region known as *pks*, as having a role in both the development and invasiveness of CRC. The mechanism they propose begins with changes in the intestinal microbiota, possibly due to a deficiency in the anti-inflammatory cytokine interleukin-10 (IL-10), which in turn spurs a proliferation of *Proteobacteria* phylum and the *Enterobacteriaceae* family. This dysbiosis and the consequent increase in colibactin, and then leads to bowel irritation and adherence of the genotoxic species to the epithelial where the effects of immune cell response, oxidative stress, and DNA damage mediate the development of CRC.

Additionally, *Fusobacteria* over-abundance has been implicated in having a negative impact on human intestinal health. Although the exact role *Fusobacteria* play in the process remains unknown, *Fusobacterium nucleatum* presence correlates with both colorectal adenomas and advanced stage colorectal cancer (Bashir et al., 2014, Gao et al., 2015).

Conversely, several bacterial species of the gut have been thought to play a role in maintaining good health. Within these phyla there are beneficial bacteria species that cluster with *Eubacterium rectale*, *Eubacterium ramulus*, and *Roseburia cecicola*, during phylogenetic analysis of 16S rRNA sequencing. The findings of Wang (2012) suggest that *Rosburia* spp. may serve to protect the host from CRC. These are species that are able to digest carbohydrates that would otherwise be indigestible to a colonic microbiome lacking bacterial species with similar properties. These types of bacteria are known as butyrate producing bacteria due their employment of the butyryl coenzyme A-acetyl coenzyme A transferase pathway for butyrate production. Butyrate is a short chain fatty acid (SCFA) produced by the microbial fermentation of the digestive resistant carbohydrates from in certain forms of dietary fiber and starch (Barcenilla et al. 2000). Other SCFA that are produced via bacterial fermentation in the gut include acetate and propionate (Plöger et al., 2012), but it is butyrate that is of most interest in this consideration of the intestinal microbiome and its possible role in the development of colorectal cancer. Butyrate has an important role in the metabolism and normal development of colonic epithelial cells and is thought to protect individuals against cancer and ulcerative colitis (figure 1.4) (Hauge et al., 1997; Canani et al., 2011).



**Figure 1.4:** Representation of short chain fatty acid (SCFA) production in the human intestine, and the putative roles that each of these SCFA play in homeostasis. (Plöger et al., 2012).

### 1.2.1 DIET AS RISK FACTOR FOR COLORECTAL CANCER

It is here, in the dietary habits of individuals, that is considered the interface between change in an individual's microflora and their predisposition to colorectal cancer. A correlation has been shown between colorectal cancer and dietary factors such as low intake of insoluble vegetable fiber; high intake of refined carbohydrates; high saturated fat intake; and sub-par levels of vitamins A, C, and E. (Kumar et al., 2003).

It has been theorized that sub-optimal levels of insoluble dietary fiber lead to decreased fecal bulk, and, consequently, fecal retention and an altered intestinal bacterial profile. Oxidative byproducts of the refined carbohydrate metabolism are subsequently held in contact with the mucous membrane for longer than normal periods. Vitamins A, C, and E are known antioxidants, and their absence is thought to exacerbate the damage done by the aforementioned oxidative byproducts (Kumar et al., 2003).

Low intake of dietary fiber is also theorized to manifest itself as an increased risk of colorectal cancer via the relationship between it and the abundance of butyrate producing bacteria. Dennis P. Burkitt's epidemiological study, already in 1971, concluded by positing a correlation between colorectal health and the removal of dietary fiber (non-starch polysaccharides) and digestion-resistant starches. The metabolic product of the bacterial fermentation of such dietary elements has been shown, both *in vitro* and *in vivo*, to facilitate survival and health of normal colonocytes, as well initiation of apoptosis in colonic tumor cell lines (Hauge et al., 2004; Hinnebusch et al., 2002)

### 1.2.2 SPECIES DIVERSITY

Species diversity is the number of species, together with the relative abundance of each of these species in a given environment. Abundance refers to the number of members within a given species. Species richness, a term often confused with diversity, is related but distinct. Richness refers only to the number of different species in an environment; it does not take into account the proportional abundance of the members of an environment (Tuomisto, 2010). Three terms used to discuss different aspects of diversity, descending in scale, are gamma diversity, beta diversity, and alpha diversity.

Gamma diversity is defined as the total diversity of a dataset (Tuomisto, 2010), and, for reasons of practicality and relevance is not to be discussed in the context of this project.

Beta diversity is defined as the total species diversity across a landscape (Tuomisto, 2010); a landscape being, for example, an entire set of samples extracted with one method. To measure this quality, clustering analysis using the Pearson correlation can be carried out on each set of samples. The network length for resulting cluster analysis provides a value that can be compare across data sets to assess which of them possesses the greatest relative diversity.

Alpha diversity speaks to the species diversity in a single habitat eg. within each individual sample extracted by a single extraction method. The Shannon index is a method of evaluating the diversity in the individual samples. The Shannon index can be viewed as a measure of entropy in that it is a means to quantify the uncertainty that a species would be chosen at random in a given dataset selection (Tuomisto *et al.*, 2010).

### 1.2.3 METHODS FOR STUDYING SPECIES DIVERSITY

The sheer magnitude of species richness that comprises the microbiome of the gut makes certain analytical approaches better suited than others. Because of the specialized growth requirements of many

intestinal bacteria species, traditional methods of culturing and identifying bacterial species are laborious and often impractical (Walter *et al.*, 2000). Molecular methods seem a more appropriate choice if one is to portray a representative picture of the diversity of the intestinal microbiome.

Many different methods exist for studying species diversity. It is beyond the scope of this paper to discuss all of them. Two popular methods of molecular species diversity analysis include next generation sequencing (NGS) and denaturing gradient gel electrophoresis (DGGE).

NGS is rapidly becoming the tool of choice for microbiome analysis. It has the advantage of being capable of fast, massive parallel sequencing producing extraordinary amounts of data, and it is less sensitive to PCR biases (Dong *et al.*, 2015). One of its advantages, however, can also be a disadvantage, namely the massive amount of data that can be produced from one run. Despite the drop in cost per megabase, highly skilled bioinformaticians are compulsory, as are relatively costly methods of computation and storage of data (Souilmi *et al.*, 2015). The costs are still high for the application of small batches of samples, as is typical for clinical labs. Another disadvantage comes in the form of the initial investment. New technology is expensive, and NGS is no different in this respect.

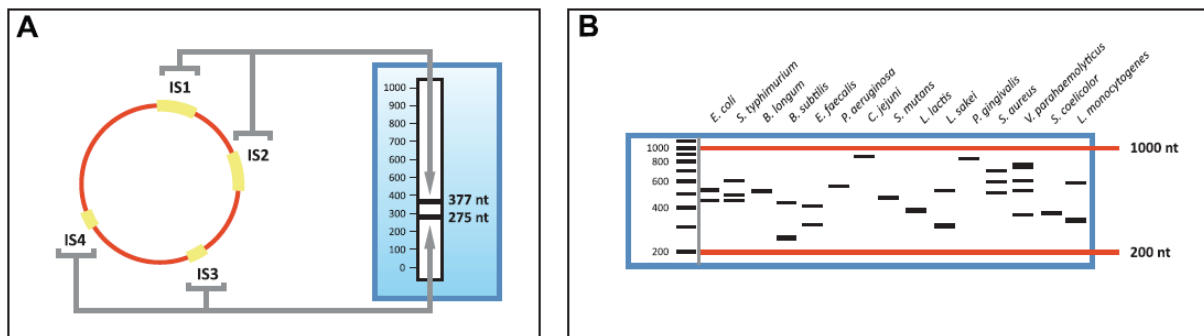
Fingerprinting techniques, such as denaturing gradient gel electrophoresis and IS-profiling are more cost efficient alternatives that offer rapid comparison data, and an acceptable compromise between high throughput and the amount of information acquired (Muyzer, 1999; Diez *et al.*, 2001; Budding *et al.* 2010).

DGGE presents itself as a potentially useful tool for intestinal bacterial diversity analysis. It exploits either the species variation in 16S denaturing, or both the denaturing variation and the length variation in the bacterial 16S-23S rRNA intergenic spacer region. Theoretically, DGGE can separate DNA of a one base pair difference in length. The method relies on a gradient of DNA denaturants (formamide and urea) over the length of a gel electrophoresis. It is purported to be a method which is reliable, reproducible, rapid, and relatively inexpensive (Kirk *et al.*, 2004; Tabatabaei *et al.*, 2009)

Drawbacks of DGGE include biases in PCR amplification and somewhat laborious sample handling. Results of this method can also be occasionally misleading due to DNA fragments of different sequences having the same mobility characteristics. Additionally, a species with insertion/deletion copy variants will give rise to multiple bands (Gelsomino *et al.*, 1999; Tabatabaei *et al.*, 2009)

#### 1.2.4 IS-PROFILING

To investigate the hypothesis that changes in one's profile of intestinal flora, a technique known as IS-profiling (Budding *et al.*, 2010) has been evaluated with the intent of possibly employing it as a microbiota profiling tool. This technique involves an amplification of the 16S-23S intergenic segment (IS region) (figure 1.5) of the bacterial rRNA genes. This area is known to be species specific in both length and sequence (Gurtler and Stanisich, 1996). The IS region is useful in discerning even closely related species, possessing highly conserved 16S rRNA genes, such as members of the Family *Enterobacteriaceae* (Cao *et al.*, 2009). IS-profiling of stool samples is a potentially attractive diagnostic method due to its relative simplicity and minimal financial investment.

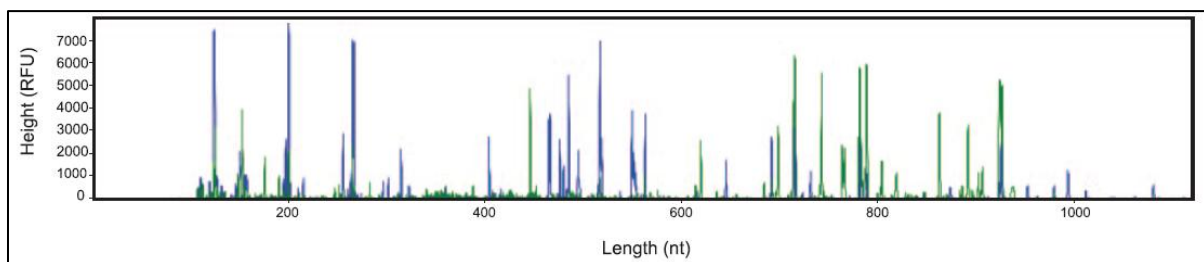


**Figure 1.5:** IS regions lengths. Inter-species differences of IS region length and sequence are exploited to produce intestinal flora profiles; some species have multiple IS regions that can vary in length. A) an illustration of the genome of *Enterococcus faecalis* (red), with its four IS regions (yellow); its four IS regions are of two lengths, resulting in two bands on gel electrophoresis. B) The bands pattern for a variety of species; thick bands are actually two bands that are very near to the same length (Budding *et al.*, 2010).

The method utilizes labeled forward primers, one, FAM-labeled, specific for both *Firmicutes*, and *Actinobacteria*; and the second, HEX-labeled, specific for *Bacteroidetes*. The primers bind to conserved areas present in the 16S (forward) and 23S (reverse) bacterial rRNA genes. These sequences and those of the three reverse primers designed to be specific to each of the three groups are described in Budding *et al.* (2010).

Additionally, at least some *Fusobacterium* species are also amplified with the *Firmicutes/Actinobacteria* primers; *F. nucleatum* (Budding *et al.*, 2010)

After amplification, the PCR product is analyzed via capillary gel electrophoresis. This fragment analysis of polybacterial samples exploits the species specific variation in IS region length to produce a profile or ‘fingerprint’ (figure 1.6) of a patient’s intestinal flora representing the target phyla of this project: *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, and *Fusobacteria*; and has been adapted from Budding *et al.* (2010).



**Figure 1.6:** Example of an IS-profile of colonic mucosa taken from Budding *et al.*, 2010; Blue (FAM) peaks are of *Firmicutes*, *Actinobacteria*, and *Fusobacterium*, Green (HEX) peaks are those of *Bacteroidetes*. Peak intensity (Y axis) is given in relative fluorescence units (RFU), and is semi-quantitative, reflecting the relative abundance of each fragment length (X axis) which is in nucleotides (Budding *et al.*, 2010).

IS-profiling is a relatively inexpensive method to analyze the intestinal microbiome. It can be set up in any reasonably equipped, small to medium size lab. It requires a PCR machine, genetic analyzer, and access to analysis software. It is superior to DGGE in that it offers a higher degree of automation and is less laborious than other methods. IS-pro offers, in addition, economic aspects more beneficial in comparison to NGS. Its attractiveness as a low-cost and compatible assay led to a technical evaluation of the method and its suitability for project at hand.

### 1.2.5 DNA-EXTRACTION.

Efficiency of DNA extraction impacts the results despite the chosen method of analysis. Sufficient yields of both bacterial and host DNA are important. Additionally, research suggests that different extraction methods and conditions can favor or disfavor specific types of bacteria, yielding possibly a deceptive species profile of the population in question (Wesolowska *et al.*, 2014; Nechvatal *et al.*, 2008). A DNA extraction that gives a true representative picture of the diversity of a feces sample is of great importance. Gram-positive bacteria cells can be more challenging to lyse, an extraction method that lacking in its ability to do so would lead to a skewed representation of landscapes diversity (Kirk *et al.*, 2004). Wesolowska *et al.* (2014) found that different DNA extraction methods yielded differences in both the taxonomic and functional distribution of genes, possibly impacting the interpretation of data from downstream applications.

Wesolowska *et al.* (2014) goes to discuss the challenges of heterogenic distribution of gut microbes, citing both uneven distribution of bacterial species with a stool sample, as well as differences between bacterial populations of the lumen versus mucosa. The paper goes on to downplay these potential differences citing that they are little importance when compared to inter-individual variation, admitting though, that intra-sample variation could have a larger impact upon longitudinal studies.

Thusly, caution must be employed when interpreting cross-study data where different extraction methods have been used.

## 1.3 CRC-PROJECT AT AHUS

The project discussed in this paper is but a small part of a larger project (CRC project) being undertaking in our laboratory. The objective of the CRC project is to investigate various biomarkers with the intent of finding a non-invasive means by which to facilitate the early detection of colorectal cancer.

A method of early detection born out of one of the hypotheses of this project is the tracking of changes in the patient's intestinal microbiome. The idea of characterizing a pattern of change in the patients' stool bacterial profile leading from healthy to polyp to cancer is a potentially useful and non-invasive tool that may lead to increased surveillance, and possibly earlier diagnoses

Three hypotheses constitute the core of the CRC project:

1. The methods of DNA extraction from stool samples have significant effects for PCR amplification of human and bacterial markers from stool samples.
2. Blood and stool human DNA (hDNA) methylation analysis provide sensitive method for early detection of CRC.
3. During the progression of CRC, the profile of intestinal microflora is altered and affects the progression of the cancer through interactions with host epithelial cells. Therefore, detection of a microbial dysbiosis can possibly be used for early detection of CRC.

Accordingly, a biobank has been designed to include samples from three groups of patients: those shown to have healthy large intestines; those with polyps; and finally, those patients where cancer has begun to develop.



## 1.4 SUMMARY OF PROJECT AIMS

The aim of this thesis is to evaluate IS-profiling as a screening tool upon development of polyps, uncovering, possibly, colorectal cancer in its early stages.

Three sub goals are defined:

### ***Technical evaluation of IS-pro analysis***

In order to standardize and apply IS-profiling for patient sample analysis, optimization and technical evaluation is necessary. Analytical sensitivity, specificity, and reproducibility will be assessed.

### ***Comparisons of fecal DNA extraction methods.***

Optimization of nucleic acid extraction to provide the best possible starting material for IS-profiling is paramount. With the aim of obtaining the most diverse bacterial profiles possible, several different DNA extraction methods will be evaluated. The criteria to be assessed are total 16S bacterial yield, bacterial species diversity, and host DNA yield.

### ***Comparison of the intestinal flora profiles among patients diagnosed with colonic polyps versus the profiles of healthy subjects***

The present study hypothesizes that changes in intestinal flora play a role in that progression from healthy to adenoma to malignancy. Stool samples from a healthy group of volunteers will therefore be compared to stool samples from patients with polyps. DNA will be extracted by the method found optimal in part two, and IS-profiling, 16S bacterial concentrations, and host DNA concentrations are to be compared between groups.

---

## 2. MATERIALS AND METHODS

### 2.1. MATERIALS

#### 2.1.1 BACTERIAL STRAINS

**Table 2.1:** Bacterial strains used in the IS-pro evaluation

PCR #	Species	CCUG*/Sample ID number
1	<i>Streptococcus agalactiae</i>	4209
2	<i>Staphylococcus epidermidis</i>	21989
3	<i>Clostridium perfringens</i>	1795T
4	<i>Enterococcus faecalis</i>	34289
5	<i>Enterococcus casseliflavus</i>	18657
6	<i>Propionibacterium anaerobius</i>	7835
7	<i>Staphylococcus lugdunensis</i>	25348
8	<i>Streptococcus milleri</i>	27298
9	<i>Enterobacter cloacae</i>	52947
10	<i>Bacteroides fragilis</i>	4856T
11	<i>Bacteroides thetaiotaomicron</i>	12297
12	<i>Streptococcus pyogenes</i>	33061
13	<i>Clostridium difficile</i>	4938T
14	<i>Shigella sonnei</i>	32079
15	<i>Escherichia coli</i>	11283
16	<i>Prevotella timonensis</i>	3489 6912**
17	<i>Staphylococcus aureus</i>	17621
18	<i>Listeria monocytogenes</i>	3456 1044**
19	<i>Enterococcus faecium</i>	542T
20	<i>Clostridium difficile</i>	54206
21	<i>Proteus vulgaris</i>	28449
22	<i>Campylobacter jejuni</i>	41359
23	<i>Proteus mirabilis</i>	26767
24	<i>Vibrio cholera</i>	33379
25	<i>Fusobacterium nucleatum</i>	332989

\*CCUG = Culture collection University of Gothenburgh

\*\*Patient sample isolates verified via 16S rRNA sequencing.

## 2.1.2 PATIENT AND SUBJECT STOOL SAMPLES

The volunteer subject stool samples used for extraction kit evaluation and method optimizing were donated voluntarily by workers at the Department for Multidisciplinary Laboratory Medicine, Department for Microbiology and Infection Control, and Department for Pathology at Akershus University Hospital.

The stool samples from the polyp patients scheduled for colonoscopy were collected by means of a kit mailed out to them some days before their appointments. Patients were instructed to contribute their samples the day before their scheduled appointments, and to deliver the samples at the time of their colonoscopy (appendix 8).

Both subject and patient samples were collected on 4 ml. of *RNAlater* RNA Stabilization Reagent in accordance with specific instructions which were included in each sample taking kit (appendix 10). These samples were homogenized, aliquoted and immediately frozen at -80°C.

**Table 2.2:** Age statistics of healthy subjects (H group) and polyp patients (P group).

Group	avg age	high	low	std dev
H group	40.95	56	30	8.20
P group	67.29	82	50	9.97

## 2.1.3 COMMERCIAL KITS, BUFFERS, CHEMICALS, MEDIA, AND ENZYMES

**Table 2.3:** Commercial kits

Name	Producer	Catalogue number
Quantifiler Human DNA Quantification Kit	Life Technologies	4343895
PSP Spin Stool DNA Kit	Stratec	1038/00200
QIAamp DNA Stool Mini Kit	Qiagen	51504
PowerMag Microbiome RNA/DNA Isolation Kit	Mobio	275-4-EP
QIASymphony Virus/Pathogen Kit	DSP Qiagen	90001297
SYBR Pre-mix ex <i>Taq</i>	Takara	RR420A
BigDye Terminator v3.1	Applied Biosystems	43 374 554 337 455
Platinum <i>Taq</i> DNA polymerase	Invitrogen	10966-018

**Table 2.4:** Chemicals, reagents and buffers.

<b>Name</b>	<b>Producer</b>	<b>Catalog number</b>
Prepman Ultra lysis buffer	Life Technologies	4318930
$\beta$ -Mercaptoethanol (1000X)	Gibco	21985-023
Phenolchloroform: isoamyl alcohol 25:24:1	Applichem	A0944,0250
POP-7 polymer	Applied Biosystems	<u>43 637 854 363</u> <u>785</u>
Buffer ACL	Qiagen	939015
Buffer ATL	Qiagen	939011
PBS buffer	Substrate section Ahus	8 g NaCl, 0.2 g KCl, 1.42 g Na <sub>2</sub> HPO <sub>4</sub> , 0.24 g KH <sub>2</sub> PO <sub>4</sub> , 1000ml ultra pure water
3130xl Running Buffer (10X)	Applied Biosystems	402824
TBE buffer	Substrate Section, Ahus	27.5 g borsyre, 54 g Sigma 7-9 Tris base, 20 ml EDTA 0.5 M pH 8, 1000 ml ultra pure water
LB medium	Substrate Section, Ahus	25 g LB – Broth Miller (Luria- Bertani), 15 g Bacto agar, 50 mg ampicillin 1000 ml ultra pure water.
RNAlater RNA	Applied Biosystems	AM7021
X-gal (40 mg/mL)	Invitrogen	15520-018
IPTG (100mM)	Invitrogen	15529-019
SeaKem LE agarose	Lonza	50004
SYBR® Safe DNA Gel Stain	Invitrogen	S33102

**Table 2.5:** Enzymes.

<b>Name</b>	<b>Maker</b>	<b>Catalogue number</b>
Mutanolysine	Sigma-Aldrich	M-9901-KU
Lysozyme	Sigma-Aldrich	12650-88-3
Lysostaphine	Sigma-Aldrich	L7386-1MG
Proteinase K 20 mg/ml	Qiagen	19133

## 2.1.4 OGLIONUCLEATIDES, REFERENCE DNA, PRIMERS, CLONING VECTORS, AND SIZE MARKERS.

**Table 2.6:** Oglionucleotides.

Name	Sequence (5' >3')	Description	Producer	Source
FirISf	FAM- CTGGATCACCT CCTTCTAWG	<i>Firmicutes/</i> <i>Actinobacteria</i> forward primer/ <i>Fusoba</i> <i>cterium</i>	Eurogentek	Budding <i>et al.</i> , 2010
BacISf	HEX- CTGGAACACCT CCTTCTGGA	<i>Bacteroidetes</i> forward primer	Eurogentek	Budding <i>et al.</i> , 2010
DuiSr1	AGGCATCCACC GTGCGCCCT	<i>Firmicutes</i> reverse	Eurogentek	Budding <i>et al.</i> , 2010
DuiSr2	AGGCATTCACC RTGCGCCCT	<i>Actinobacteria</i> reverse	Eurogentek	Budding <i>et al.</i> , 2010
DuiSr3	AGGCATCCRCC ATGCGCCCT	<i>Bacteroidetes</i> reverse	Eurogentek	Budding <i>et al.</i> , 2010
BAC-338	ACT CCT ACG GGA GGC AG	Broad range 16S, qPCR Forward	Eurogentek	Yu <i>et al.</i> , 2005
BAC-805	GAC TAC CAG GGT ATC TAA TCC	Broad range 16S, qPCR Reverse	Eurogentek	Yu <i>et al.</i> , 2005
TOPO 2.1 cloning vector	See figure XX	Cloning vector	Eurogentek	Purchased from producer
MapMarker X- Rhodamine Labeled DNA low mass ladder	NA	MM-1000XL- ROX	Invitrogen	Purchased from producer
M13-forward	GTAAAACGAC GGCCAG	primer from cloning kit	Invitrogen	Purchased from producer
M13-reverse	CAGGAAACAG CTATGAC	primer from cloning kit	Invitrogen	Purchased from producer
16S1	TGA AGA GTT TGA TCA TGG CTC AG	16S sequencing forward primer	Eurogentek	Perkin Elmer MicroSeq 16SrRNA Gene kit, primere 16S1-16S8.
16S2R	TAC CGC GGC TGC TGG CA	16S sequencing reverse primer	Eurogentek	Perkin Elmer MicroSeq 16SrRNA Gene kit, primer 16S1-16S8.
Non-methylated genomic <i>Escherichia</i> <i>coli</i> DNA	Genomic DNA	Cat. num. D5016	Zymo Research	Purchased from producer

**Table 2.7:** Size markers.

<b>Name</b>	<b>Manufacturer</b>	<b>Catalog number</b>
MapMarker 1000 ROX-labeled size standard	BioVentures	MM-1000-ROX
Low DNA Mass Ladder	Invitrogen	10068-013

### 2.1.5 INSTRUMENTS, SOFTWARE, AND DIVERSE MATERIALS.

**Table 2.8:** Instruments.

<b>Name</b>	<b>Producer</b>	<b>Catalog number</b>
Heraeus Pico 17	Thermo Scientific	10524723
NanoDrop 2000	Thermo Scientific	ND-2000
AB 7900 HT Real Time PCR instrument	Applied Biosystem	4329001
NucliSENS miniMAG	Biomerieux	200296
Rotor-Gene Q real time PCR instrument	Qiagen	9001560
3130xl genetic analyzer	Applied Biosystems	4315931
Duo Cyclor	VWR	VWRI732-1200
2720 Thermal Cyclor	Applied Biosystems	4359659

**Table 2.9:** Software.

<b>Program</b>	<b>Producer</b>	<b>Version</b>
QIASymphony SP	Qiagen	4.0
BioNumerics	Applied Maths	7.1
Excel Analyze-it	Microsoft	2010

**Table 2.10:** Diverse materials.

<b>Name</b>	<b>Producer</b>	<b>Catalogue number</b>
Fast Prep Bead tubes Lysing matrix E	MP Bio	6914-100

---

## 2.2. METHODS

### 2.2.1 PART 1: TECHNICAL EVALUATION OF IS-PROFILING

In an attempt to assess the practicality of using IS-profiling as a diagnostic tool, an optimization and standardization of IS-profiling was carried out, followed by an evaluation of the method's sensitivity, specificity, and reproducibility.

#### 2.2.1.1 DNA EXTRACTION OF BACTERIAL CULTURES

The 25 bacterial cultures used to optimize and evaluate IS-profiling are described in Table 2.1. DNA from cultures was extracted by transferring a small loop of a bacterial colony to 100  $\mu$ l of Prepman Ultra lysis buffer. The mixture was then heated to 95°C for ten minutes; after cooling, the tubes were centrifuged at 13000 RPM for 2 minutes in a bench-top micro-centrifuge. The supernatant was transferred to a new tube, the nucleic acid concentration measured via NanoDrop, and PCR suitable dilutions of 5 ng/ $\mu$ l were made.

#### 2.2.1.2 IS-PROFILING PROCEDURE

Amplification of IS-regions was performed in a multiplex PCR (IS-pro PCR) with primers specific for the phyla *Bacteroidetes* (HEX) and *Firmicutes*, *Actinobacteria*, and *Fusobacteria* (FAM) (table 2.6). The fluorescent dyes HEX and FAM were linked to the forward primers for the respective phyla.

The IS-pro PCR reaction was performed with the following content: 12.5  $\mu$ l SYBER Premix EX Taq, 2  $\mu$ l of each of two the forward and three reverse primers, FirISf\_(FAM), BacISf\_(HEX), DuiSr1, DuiSr2, and DuiSr4 (0.04  $\mu$ M reaction primer concentration); 12.15  $\mu$ l PCR grade water, 2.5  $\mu$ l Invitrogen 10X PCR buffer -MgCl, 1.75  $\mu$ l 50mM MgCl<sub>2</sub>, 0.5  $\mu$ l 10mM dNTP, 1  $\mu$ l BSA (1% solution), 0.1  $\mu$ l Platinum *Taq* polymerase, and 2  $\mu$ l template in a final reaction volume of 25  $\mu$ l.

Amplification was carried out on a Qiagen Rotor-Gene Q real time PCR instrument. The following thermal cycling conditions were 72°C for 2 min; 35 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 60 s; and 72°C for 5 min.

After PCR, optimization of the fragment analysis was performed by combining various dilutions of PCR product with different ratios of formamide and MapMarker 1000. The following combination was found to give the best result: 5  $\mu$ l PCR-product, 0.1  $\mu$ l Map Marker 1000 size standard and 19.9  $\mu$ l Hidi formamide.

DNA fragment analysis was performed on an ABI Prism 3130xlGeneticAnalyzer (Applied Biosystems) using a 50cm array and POP-7 polymer. The run module used for the 3130xl Genetic Analyzer was downloaded from Life Technologies: 1200LIZ size standard Module. The module was employed with the default settings unchanged.



Map Marker 1000 is a rhodamine (Rox) labeled size standard with bands ranging from 50 – 1000. The fragments are 50, 75, 100, 125, 150, 200, 250, 300, 350, 400, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 and 1000 base pairs.

### ***Data processing***

Data were analyzed by BioNumerics software 7.1. Raw .fsa files produced by 3130xl during fragment analysis were uploaded to BioNumerics and a database was constructed.

Curves were processed with the default values in BioNumerics. Maximum optical density dynamic range (OD range) was set to 7000 points. The OD range defines the highest peaks that can be detected in each channel without causing "bleed-through" to other channels. "Bleed-through" is caused by spectral overlap between fluorophores used and results in artificial pull-up peaks in adjacent channels. The OD range must therefore be defined for each fluorophore and bleed-through corrected for. This was done under "Curve processing settings", "Set Maximum OD value" and "Detect bleed through regions".

The minimum OD range for each channel defines the lower limit for which peaks that should be recorded. "Peak detection" was set at 2% of "OD range"; and 5% of "curve range".

The size standard ladder (ROX) was defined and then used for normalization of curves, using the "fit by pattern" option, thus aligning the sample profiles. The peaks were verified manually before proceeding.

Processing of the *Bacteroidetes* (FAM) and *Firmicutes* (VIC) peaks required the removal of "primer-dimers" which typically are short fragments. This was done under "Band filters" and "filter by fragment length". A minimum fragment length of 60 base pairs was defined.

By-products of the PCR-amplification also required removal. By-products may be caused by strand slippage, incorporation of an A-tail etc. They are typically one repeat shorter than the main peak and may cause stutter peaks or shadow bands. Threshold for filtering shadow bands should be relative to primary peak size, and maximum distance from the main peak should be given. Under the option "Remove shadow bands", "maximum relative size" was set to 60 and "maximum relative distance" was set to 1.2.

A "composite data set" was then made for further analysis.

All values used in the setting of the IS-pro analysis were those recommended by Applied Maths.

#### **2.2.1.3 SENSITIVITY AND SPECIFICITY OF IS-PROFILING**

The 25 pure cultures of *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* isolates described in table 2.1 were tested with IS-pro for sensitivity and specificity of the assay.

For selected isolates an additional PCR was run in parallel using unlabeled forward primers for the purpose of gel electrophoresis and sequencing. These primers were of identical sequence to the labelled IS-pro primers, differing only in the absence of a fluorescent dye on the 5' end.

The PCR product was run on a 1% agarose gel; the bands were cut from the gel and purified using Purelink Quick Gel Extraction Kit. Low DNA Mass Ladder was used to estimate fragment size (table 2.7). Selected gel bands were then cloned and sequenced for verification of band/peak identity.

#### 2.2.1.4 CLONING OF IS-PRO FRAGMENTS

The excised bands were then either sequenced directly, or cloned into pCR 2.1-TOPO vector; the recombinant plasmids were transferred into One Shot Top 10 Chemically Competent *E. coli* (Appendix 9).

The transformed cells were grown on LB plates that both contained ampicillin, and had been spread with X-Gal and IPTG. In addition to antibiotic resistance, ligation positive colonies were also selected for using blue/white screening; where by positive colonies are white due to the recombinant plasmid's inability to metabolize X-Gal, a result of the vector having been ligated into the plasmid at a location that disrupts  $\beta$ -galactosidase expression of the LacZ gene.

Positive colonies were selected, and the plasmids were isolated using Invitrogen Pure Link Quick Plasmid Mini Prep kit. The isolated plasmids were screened by PCR, using M13 primers, and agarose gel electrophoresis. Amplification was performed with an Applied Biosystems 2720 Thermal Cycler. Each reaction contained 16.15  $\mu$ l PCR grade water, 2.5  $\mu$ l Invitrogen 10X PCR buffer, 1.75  $\mu$ l 50 mM MgCl<sub>2</sub>, 0.5  $\mu$ l 10 mM dNTP, 1  $\mu$ l each of M13 forward and reverse primers (0.04  $\mu$ M reaction primer concentration), 0.1  $\mu$ l Invitrogen Platinum *Taq* polymerase, and 2  $\mu$ l template. PCR cycling conditions were 94°C for 7min; 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min; and a final extension of 72°C for 5 min.

#### 2.2.1.5 SEQUENCING OF RECOMBINANT PLASMIDS

Sequencing was carried out on an Applied Biosystems 3130xl Genetic Analyzer, using Applied Biosystems BigDye Terminator 3.1 and the M13 primers as described in table 2.6.

The Big Dye Terminator reaction contained 1  $\mu$ l BigDye v. 3.1 mix, 3  $\mu$ l 2.5X BigDye sequencing buffer, 1  $\mu$ l M13 primer (0.32  $\mu$ M reaction concentration), 3  $\mu$ l PCR grade water, for a final reaction volume of 10  $\mu$ l.

Amplification was carried out using a VWR Duo Cycler PCR instrument. Cycling conditions were 25 cycles of 96°C for 10 s, 50°C for 5s, and 60°C for 4 min.

Sequencing data were analyzed using Gene Codes' Sequencher 5.3, and NCBI BLAST, using the default settings for blastn program.

#### 2.2.1.6. REPRODUCIBILITY OF IS-PRO

Reproducibility was tested by running six technical replicates of the IS-pro, and aligning them with BioNumerics software. The replicates were run from PCR to IS-pro fragment analysis, all stemming from the same sample extraction eluate.

The raw data files were imported from the BBI 1310xl to BioNumerics, and a database was constructed. All files were aligned and the settings were adjusted as described in section 2.2.1.2.

However, due to very intense signals from pure culture fragment analysis, processing of the *Bacteroidetes* (VIC) and *Firmicutes* (FAM) peaks required the following additional adjustments of the program settings: “remove shadow bands” was set at 80% for “maximum relative size,” and 3 for “maximum relative distance”; “remove noise on curve was activated and set at 2; “filter by relative peak height” was set at 30% maximum; “filter bleed through bands” was set at 95%; “remove doublets was set at 5 base maximum with a 50% minimum “valley.” The problem of excessive pure culture peaks is discussed further in the results and discussion sections of this thesis.

Band matching was performed to align identical fragments, and the ROX size standard was used for normalization.

## 2.2.2. PART 2: COMPARISON OF DNA EXTRACTION KITS

Six different DNA extraction kits were evaluated to assure optimal DNA quality and quantity. The stool samples should yield both an acceptable concentration of bacterial DNA, and the most diverse profile possible of bacterial species. Additionally, host DNA was quantified for experiments which are to be carried out in the larger project.

The following parameters were evaluated:

- 16S qPCR, to determine the total bacterial DNA yield
- IS-profiling (IS-pro) for bacterial diversity
- Human DNA quantification kit (Quantifiler), used to determine the quantity of host DNA in samples.

### 2.2.2.1 FECAL SAMPLES

Samples were collected from 22 presumed healthy volunteers as described in section 2.1.2, and 200 µl of stool sample was extracted with each of the six different DNA extraction methods.

### 2.2.2.2 DNA EXTRACTION WITH POWERMAG MICROBIOME RNA/DNA ISOLATION KIT (ADAPTED FOR MANUAL PIPETTING)

This kit, having been designed for use with the Eppendorf epMotion® 5075 TMX robot, was adapted for manual pipetting in the robot’s absence, described in appendix 1.

### 2.2.2.3 DNA EXTRACTION WITH QIAAMP STOOL DNA MINI KIT

A) This kit was employed both as instructed by the manufacturer, and also with the addition of a bead-beating mechanical extraction step. After a comparison of human genomic DNA (hDNA) yield with and without bead-beating, it was decided best to include the mechanical lyses step in further evaluations. This was carried out in Fast Prep “Soil” (Lysing matrix E) bead tubes, shaken at 6.5 m/s for 3 X 60 seconds, using MP Bio’s FastPrep 24 instrument

B) The same protocol was also performed with the inclusion of enzymatic pretreatment. Immediately after the InhibitEX tablet step the supernatant was transferred to a new 1.5 ml tube and centrifuged at full speed for three minutes. 200 µl was transferred to a new tube with enzyme cocktail: 100 µl mutanolysine (1 U/ µl = 100 U), 100 µl lysozyme (20 mg/ml = 2 mg), 100 µl lysostaphine (5mg/ml) and incubated at 37°C for 30 minutes. 15 µl Proteinase K and 200 ul buffer AL were added before incubation at 70°C for 10 minutes. The kit protocol was followed precisely after these deviations.

#### 2.2.2.4 DNA EXTRACTION WITH STRATEC MOLECULAR’S PSP SPIN DNA KIT

A) Here, Protocol 1 was strictly adhered to, including the optional step of bacterial DNA enrichment.

B) The same protocol was also performed with the inclusion of enzymatic pre-treatment. Immediately after the InviAdsorb tube step the supernatant was transferred to a new 1.5 ml tube and centrifuged at full speed for 3 minutes. 400 µl was transferred to a new tube with enzyme cocktail: 100 µl mutanolysine (1 U/ µl = 100 U), 100 µl lysozyme (20 mg/ml = 2 mg), 100 µl lysostaphine (5mg/ml) and incubated at 37°C for 30 minutes. 25 µl Proteinase K was added before incubation at 70°C for 10 minutes. The kit protocol was followed precisely after these deviations.

#### 2.2.2.5 DNA EXTRACTION WITH QIAGEN’S QIASYMPHONY SP SYSTEM

QIASymphony extraction was carried out following the QIASymphony DSP Virus/Pathogen Kit

‘with off-board lysis’ protocol. This protocol was strictly adhered to save for the addition of a mechanical lysis step and a dilution step to compensate for any substances that might lead to PCR inhibition downstream: the 200µl of stool sample was added to 1ml of PBS buffer; the mixture was then shaken at 6.5 m/s for 45 seconds; after bead beating, 200 µl of the dilution was then added to 430 µl of Qiagen’s recommendations for “off board lysis buffer” (ATL, ACL, Proteinase K nd ), and incubated at 68° C for 15 minutes, as per Qiagen’s manual off-board lysis protocol. Nucleic acid purification was then performed on the QIASymphony extraction robot, using Qiagen’s virus/pathogen protocol.

#### 2.2.2.6 QUANTIFICATION OF BACTERIAL DNA BY 16S QPCR

Bacteria DNA yield was evaluated using 16S real time PCR, and quantified via a standard curve established using non-methylated genomic DNA from the bacterium *Escherichia coli* . The standard curve was comprised of serial dilution, giving seven points: 500pg, 50pg, 5pg, 0.5pg, 0.05pg, 0.005pg, 0,0005pg, and a no template control (NTC). Each dilution and the NTC were run in triplicate. Samples were run in duplicate and were diluted 100-fold so that the vast majority of them landed towards the middle portion of the standard curve.

The primers used to amplify the 16S ribosomal RNA bacteria gene, measuring yield of total bacteria DNA extracted (Yu *et al.*, 2005) are described in table 2.4. The resulting PCR product is 468 bp. in length:

The 16S qPCR reaction contained 12.5 µl SYBR Premix EX *Taq*, 2 µl each of BAC-338 (forward) and BAC-805 (reverse) primer (0.4 µM final concentration each), 5.5 µl PCR grade water, 0.5 µl ROX II reference dye (included in SYBR Premix E EX *Taq* kit), and 2 µl Template, for a final volume of 25µl.

The analysis was run on a Qiagen Rotor-Gene Q real time PCR instrument. Cycling conditions were 95°C for 10 s; and 30 cycles of 95°C for 3 s, 55°C for 20 s, 60°C.

Statistical analysis was carried out with a non-parametric analysis method due to the non-Gaussian distribution of the data. P-values were adjusted accordingly, having been multiplied by the number of tests performed (Bonferroni corrected).

### 2.2.2.7 QUANTIFICATION OF HUMAN DNA BY QUANTIFILER

Human genomic DNA (hDNA) was quantified with Applied Biosystems Quantifiler Human DNA Quantification Kit. An eight point standard curve was established using the human DNA standard included in the kit. A serial dilution was made resulting in concentrations of 16.7, 5.6, 1.9, 0.62, 0.21, 0.07, 0.023 ng/µl; a negative template control (NTC) made up the last point on the standard curve. Standards were run in three parallels. Subject samples were run undiluted, in duplicate; a negative sample, also run in duplicate, was included.

The Quantifiler PCR reaction contained 10.5 µl Quantifiler primer/probe mix, 12.5 µl Quantifiler PCR mix, and 2.5 µl DNA template for a final volume of 25 µl.

Quantifiler runs were performed on Applied Biosystem's AB 7900 HT Real Time PCR instrument. Thermal cycling conditions were 95°C for 10 min; 40 cycles of 95°C for 15 s and 60°C for 1 min.

### 2.2.2.8 BACTERIAL DIVERSITY ANALYSIS BY IS-PROFILING

Bacterial diversity was evaluated by first performing IS-profiling and BioNumerics analysis as described in section 2.2.1.2. DNA eluates were diluted 1:10 with PCR grade water prior to PCR. BioNumerics diversity indexing and clustering tools were then used to further characterize the data.

#### ***Alpha diversity***

Within sample diversity was calculated using the Shannon index.

Shannon Index (H) is a commonly used diversity index that takes into accounts both abundance and richness of species present in the community. It is explained by the formula:

$$H = -\sum_{i=1}^s (P_i * \ln P_i)$$

H = the Shannon diversity index

P<sub>i</sub> = fraction of the entire population made up of species i (proportion of a species relative to total number of species present, not encountered)

S = numbers of species encountered

Bacterial diversity was calculated per sample and then averaged for each sample within each of the phylum specific primer sets. Results were then quantified and exported to Excel.

These values were then grouped by kit and averaged to determine which kit, if any, had the greatest mean species diversity per sample.

### ***Beta diversity***

Dissimilarities between samples in each kit, or between-sample diversity, was estimated by calculating the Pearson correlation for the target phyla in each kit.

The advanced cluster analysis tool in the BioNumerics software was also employed in the search for the extraction kit with the greatest beta diversity. In doing so the software calculates the topscore UPGMA (Unweighted Pair Group Method with Arithmetic Mean), and subsequently produces a dendrogram whose measurement of the total network length ie. the total length of all the branches in the resulting dendrogram is considered to correspond with degree of diversity in a given kit. The total length values of the dendrograms' total network length for each dye in each kit were then compared and used to determine the kit with the greatest species diversity.

Bacterial diversity was calculated per phylum specific primer set. Results were then quantified and exported to Excel.

## **2.2.3 PART 3: COMPARISON OF BACTERIAL COMPOSITION AND HUMAN DNA CONCENTRATION IN FECAL SAMPLES FROM POLYP PATIENTS AND CONTROLS**

In the third part of this thesis, the healthy subject group's samples were compared to those of the polyp group. As with the DNA extraction kit comparisons, the following three criteria were assessed:

- 16S qPCR, to determine the concentration total bacterial 16S rRNA.
- IS-profiling (IS-pro) for bacterial diversity.
- Human DNA quantification kit (Quantifiler), used to determine the concentration of host DNA in samples.

### **2.2.3.1. SAMPLES**

Fecal samples from 22 controls and 24 patients with polyps were collected as described in section 2.1.2.

### **2.2.3.2. DNA EXTRACTION**

DNA was extracted with the Stratec PSP Spin Stool DNA Kit, based on results from comparison of the six DNA extraction protocols.

### 2.2.3.3 QUANTITATION OF BACTERIAL DNA PRESENT IN STOOL SAMPLES

Total 16S content was analyzed in the polyp group as in the control group. This experiment was carried out according to the parameters of the 16S qPCR described section 2.2.2.6.

### 2.2.3.4 QUANTIFICATION OF HUMAN DNA PRESENT IN STOOL SAMPLES

Polyp group host DNA concentrations were measured and compared to the healthy group using Quantifiler Human DNA Quantification Kit, as described above in section 2.2.2.7.

Statistical analysis was carried out for both hDNA quantification and 16S qPCR, as in the extraction kit evaluations, with a non-parametric analysis method (Mann-Whitney) due to the non-Gaussian distribution of the data.

### 2.2.3.5 IS-PRO

IS-pro fragment analysis was performed as described in section 2.2.1.2. The samples were diluted 1 part sample to 9 parts PCR grade water before PCR to compensate for any possible remaining PCR inhibiting substances.

### 2.2.3.6 RAW DATA PROCESSING

Raw .fsa files from the 3130xl Genetic Analyzer were imported to BioNumerics, a database was constructed and curves were processed as described in 2.2.1.2.

### 2.2.3.7 COMPARISON OF THE BACTERIAL COMPOSITION IN STOOL SAMPLES

#### ***Profile clustering***

To reveal any group-specific clustering, Pearson's correlation was performed on all versus all samples. A correlation matrix and a similarity dendrogram were constructed for this purpose.

#### ***Jackknife test***

Additionally, group analysis was carried out via a "jackknife test". This entails manually assigning each sample to one of the two groups, healthy and polyp in this case. The analysis program removes one sample at a time from its group, and considers it as an unknown; the program then attempts to identify the unknown sample against the defined groups, and calculates the percentage of correct and incorrect predictions, displaying them, ultimately, as a matrix.

#### ***Relative phylum abundance***

Relative fluorescence units (RFU) were summed for each of the fluorescent signals and the ratios for the two groups were then compared.

#### ***Alpha diversity analysis***

Within-sample diversity for each group was calculated using the Shannon diversity index as described in section 2.2.2.8.

#### ***Beta Diversity analysis***

Between-sample diversity in each group was estimated by the construction of a dendrogram (Pearson correlation) for each group as described in section 2.2.2.8.

### ***Identification of group specific bands***

To identify band classes specific to one group, several correlation plots were made, using the results table with band classes and relative values. An attempt to construct a representative composite profile of each group was made. This was done by taking the RFU values for the various band fragment lengths (band classes) across the group of samples, first using band class mean values; thereafter, in an attempt to describe group band profiles in a more representative fashion, band class median, and sums were calculated; additionally, band class values were also log 2 transformed before, again, calculating mean, median, sum values, which were then used to fabricate a second set of correlation plots.

Points that deviated greatly from  $R^2 = 1$  were investigated further to reveal the fragment length they represented, and were then subsequently evaluated to determine if the band was indeed representative of the group



---

## 3. RESULTS

### 3.1 TECHNICAL EVALUATION OF IS-PROFILING (IS-PRO)

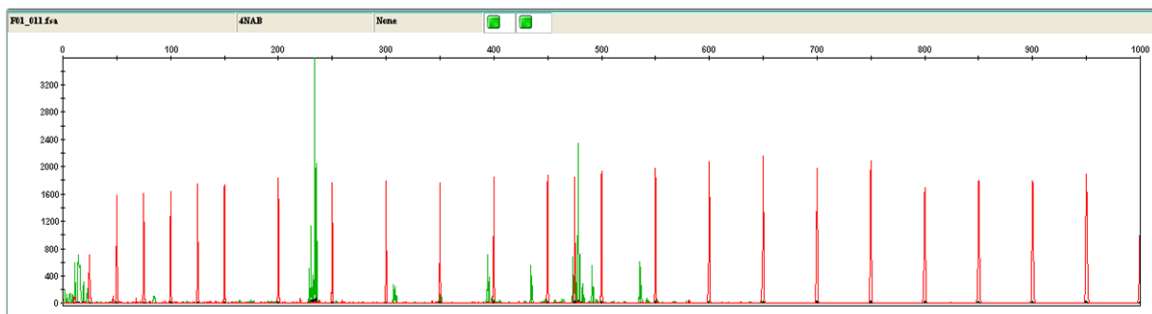
As outlined in the materials and methods section, several IS-pro parameters were evaluated, and testing was performed to ensure the method was reliable despite any dissimilarity between our laboratory and that used by Budding *et al.* (2010).

#### 3.1.1 OPTIMIZATION OF IS-PRO

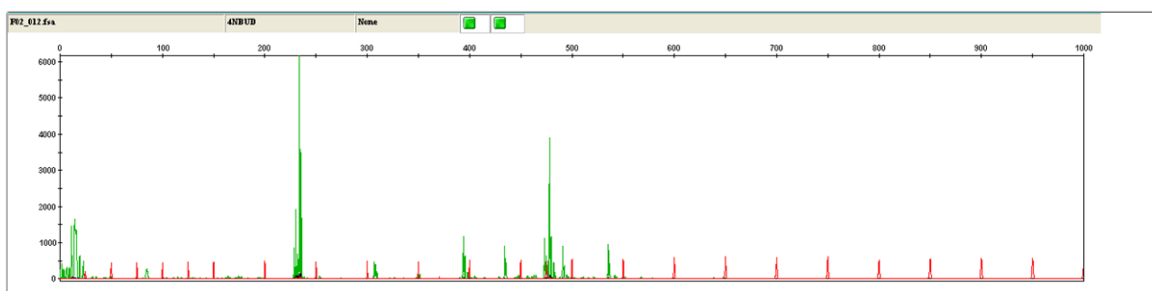
Optimization of the IS-fragment analysis was performed by varying the amount of PCR-product, Map Marker, and formamide. This test run (figure 3.1) was analyzed using the HEX labeled *Bacteroidetes* primer set. The red peaks are the ROX labeled Map Marker size standard. The sample used was a randomly chosen 1:10 dilution stool sample from the healthy group.

The recipe as described by Budding *et al.* (2010) described in section 2.2.1.2 resulted in both stronger sample peak signals, and, at the same time, a sufficient yet more economical use of the size standard (red peaks). All further IS-profiling was carried out using said mix.

A)



B)



**Figure 3.1:** Profile comparison of two IS-profiling master mixes. The same sample run first (A) with the suggested master mix recipe from Applied Biosystems, and then (B) with the recipe used in Budding, 2010. The green peaks are those of the *Bacteroidetes* phylum specific HEX-labeled primers, the red are those of the Map Marker size standard.

#### 3.1.2 SENSITIVITY AND SPECIFICITY OF IS-PROFILING

In the present work, I have chosen to define sensitivity as the proportion of positives that are correctly identified as such, and specificity as the proportion negatives correctly identified as such.

The IS-pro method is thoroughly validated by Budding *et al.* (2010). The authors performed theoretical validation such as *in silico* testing, as well as extensive *in vitro* testing: evaluations of reproducibility, primer specificity, and phylum bias. To ensure the method functions well in our hands, we chose to carry out a limited number of tests before proceeding further into our project.

In addition to *Firmicutes* and *Bacteroidetes* strains, several *Proteobacteria* strains were included in the list of pure bacterial strains to be tested. This test of specificity revealed that in the contrived context of being the only bacteria phylum present, certain *Protobacteria* strains were also amplified by the HEX-labeled *Bacteroidetes* primer set (table 3.1); they include *V. cholera*, *C.jejuni*, *E. coli*, *S. sonnei*, and *E. cloacea*. Detailed results of all species amplified under our validation can be found in tables 3.3 and 3.4.

**Table 3.1:** Overview of IS-pro primer sensitivity and specificity.

A)

<b>Firmicutes/Actinobacteria/Fusobacteria primers</b>	<b>True positives</b>	<b>True negative</b>	<b>Total</b>
IS-positive	15	0	15
IS-negative	0	10	10
Total	15	10	25

B)

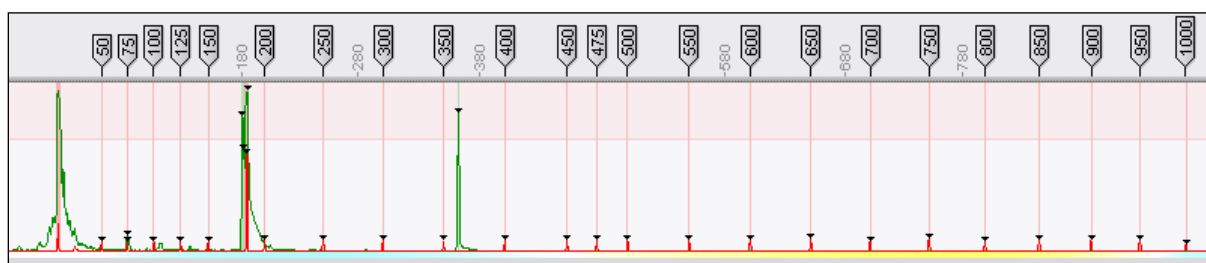
<b>Bacteroidetes-primers</b>	<b>True positives</b>	<b>True negative</b>	<b>Total</b>
IS-positive	3	5*	8
IS-negative	0	17	17
Total	3	22	25

\*All false positives were of the *Proteobacteria* phylum

A) Results for the primer sets designed to amplify the *Firmicutes*, *Actinobacteria*, and *Fusobacteria* phyla. B) Results for the *Bacteroidetes* primer set.

Due to the low number of samples in this test, no statistical analysis was carried out.

Additionally, Budding *et al.* (2010) alludes to the fact that the FAM-labeled primers also amplify *Fusobacterium* in the article's supplemental information, and it was deemed prudent to test this since it speaks to the specificity of the FAM labeled primers. Both gel electrophoresis and IS profiling results proved the *Firmicutes/Actinobacteria* primers effective in amplification of *Fusobacterium nucleatum*, resulting in fragment lengths at approximately 182 bp and 362 bp (figure 3.2).



**Figure 3.2:** BioNumerics IS pro chromatogram of *Fusobacterium nucleatum* showing two bands: circa 182bp and 362 bp.

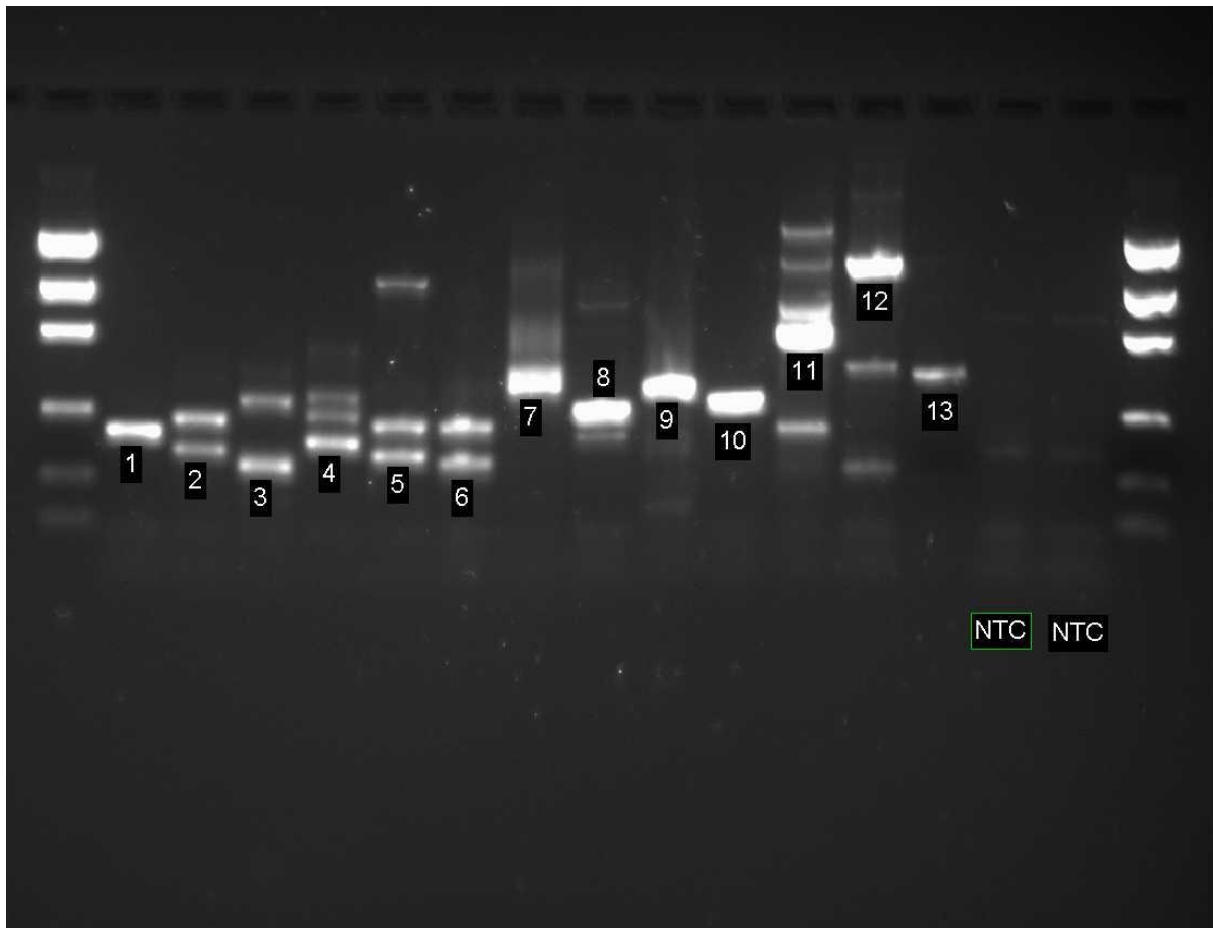
To summarize the specificity and sensitivity testing, all *Firmicutes* species, and *Fusobacterium nucleatum* were amplified by the *Firmicutes/Actinobacteria/Fusobacteria* specific primer set, while none of the *Bacteroidetes* or *Proteobacteria* species were amplified. The *Bacteroidetes* specific primer set amplified all of the *Bacteroidetes*

species, none of the *Firmicutes* species, but five of the seven *Proteobacteria* species were amplified. The results suggest that the *Bacteroidetes* primer set is somewhat unspecific within the context of this testing. No *Actinobacteria* species were tested during this thesis' abbreviated validation of the method.

### 3.1.3 CLONING OF IS-PRO FRAGMENTS

Species confirmation on selected strains was performed in our laboratory.

A subset of the pure bacterial strains, described in table 3.2, was amplified using unlabeled IS-pro primers of identical sequence. They were subsequently analyzed by gel electrophoresis (figure 3.3), and the bands were excised, purified, cloned and sequenced. Sequences were submitted to a BLAST search. IS-pro analysis results of the parallel PCR fragment lengths with labeled primers are summarized in tables 3.3 and 3.4.



**Figure 3.3:** Gel electrophoresis results of the parallel PCR using unlabeled IS-pro primers. Bacterial species lane numbers are as follows 1. *Streptococcus agalactiae*, 2. *Enterococcus faecalis*, 3. *Clostridium perfringens*, 4. *Staphylococcus epidermidis*, 5. *Enterococcus casseliflavus*, 6. *Peptostreptococcus anaerobius*, 7. *Staphylococcus lugdunensis*, 8. *Streptococcus anginosus*, 9. *Bacteroidetes fragilis*, 10. *Streptococcus pyogenes*, 11. *Entrobacter cloacea*, 12. *Proteus vulgaris*, and 13. *Proteus mirabilis*. The first and last lanes contains low mass DNA ladder (2000, 1200, 800, 400, 200, 100 base pairs), and two no template controls (NTC) were included.

The results confirmed that the IS-pro peaks and gel bands indeed represented the correct bacterial species. Table 3.2 summarizes the sequencing results of the IS profiling fragments which were amplified using unlabeled IS pro primers of the same sequence as the ones used in IS profiling fragment analysis.

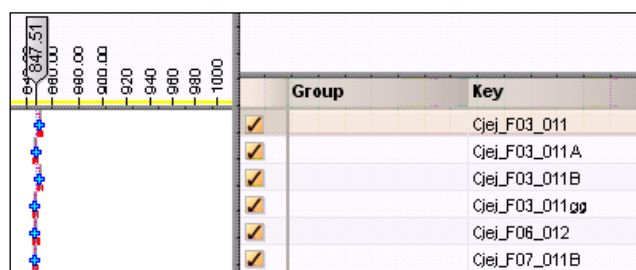
**Table 3.2:** Summary of cloning sequencing results.

Sequencing Results	GenBank Accession number	Percent Match	Purported Identity of Source Culture
1 <i>Streptococcus agalactiae</i>	AE009948.1	99	<i>Streptococcus agalactiae</i>
2 <i>Enterococcus faecalis</i>	CP002491.1	99	<i>Enterococcus faecalis</i>
3 <i>Clostridium perfringens</i>	EU334000.1	99	<i>Clostridium perfringens</i>
4 <i>Staphylococcus epidermidis</i>	AF269416.1	98	<i>Staphylococcus epidermidis</i>
5 <i>Enterococcus casseliflavus</i> EC20	CP004856.1	99	<i>Enterococcus casseliflavus</i> EC20
6 <i>Peptostreptococcus anaerobius</i> (clone LK51)	Z29059.1	90*	<i>Peptostreptococcus anaerobius</i> (clone LK51)
7 <i>Staphylococcus lugdunensis</i> HKU09-01	CP001837.1	94*	<i>Staphylococcus lugdunensis</i> HKU09-01
8 <i>Streptococcus anginosus</i>	JN181379.1	98	<i>Streptococcus anginosus</i>
9 <i>Bacteroides fragilis</i> 638R	FQ312004.1	97	<i>Bacteroides fragilis</i> 638R
10 <i>Streptococcus pyogenes</i> STAB901	CP007024.1	99	<i>Streptococcus pyogenes</i> STAB901
11 <i>Enterobacter cloacea</i>	CP011650.1	100	<i>Enterobacter cloacea</i>
12 No sequencing results	--	--	<i>Proteus vulgaris</i>
13 No sequencing results	--	--	<i>Proteus mirabilis</i>

\*Short sequence and poor quality data.

### 3.1.4 REPRODUCIBILITY OF IS-PRO

To evaluate the reproducibility of the method, we determined the number of times the procedure identified the same pattern in each sample. IS-profiling was performed on six technical replicates of each of the twenty-five bacterial strains listed in table 2.1. Subsequently, BioNumerics was used to analyze the data. Settings were adjusted so that the fragments would be interpreted as the same band. Bands that appeared at first to vary slightly in fragment length were then aligned during the ‘band matching’ phase of the analysis. An exact theoretical value was computed for the fragment length; the height of the curve was also computed. As an example, the settings used in these analyses consistently aligned the same bands from all six replicates of *C. jejuni* at 847.51 bp, even though there is clearly some spatial variation of the bands (figure 3.4).



**Figure 3.4:** Bionumerics software comparison of the reproducibility testing done during the initial evaluation of IS-profiling. Illustrated here, is the one IS fragment of the species *Campylobacter jejuni*. Depiction of a gel band simulation of each of the six replicates illustrates how the program settings allow for the band to be interpreted as a single band at about 847.51 base pairs despite the slight variation in spatial placement of the fragments in the actual individual replicate analyses.

Manual correction with the following criteria was judged necessary in some cases:

1. Overload of template causing the 3130xl genetic analyzer's laser to switch off at the peaks' highest point of intensity during the capillary electrophoresis, thus causing the splitting of a curve. This necessitated some subjective analysis of the resulting curves. Peaks within 2 base pairs of each other were judged to be a phenomenon of this template overload. Thusly, a peak position in the middle of the two peaks registered was assumed to be a single band within the IS-profile for a single bacterial species.
2. Bands of low intensity were treated with skepticism, and consequently ignored if they were the only deviation in an otherwise matching peak profile. The random signals encountered intermittently that are not consistent across the six replicates of a given species are often of drastically lower intensity and are, quite possibly, due to signal background.

The different band patterns were categorized using the criteria set forth above. The band pattern representing the majority of the isolates for each bacterial species tested was categorized as the main profile, and all band patterns that did not match the main profile were categorized as deviating (table 3.4). The number of replicates matching the main profile were divided by the total number of profiles obtained, resulting in a IS-pro reproducibility rating of 88.82% for the pure bacteria cultures. Due to the technical problems described above, the RFU values were not used in any reproducibility calculations. IS fragment lengths for the pure cultures are shown in table 3.3.

**Table 3.3:** Summary pure culture IS-fragments.

Species	Fragment length in base pairs														Signal	
<i>Streptococcus agalactiae</i>	330															FAM
<i>Staphylococcus epidermidis</i>	298	302	311	313	314	406	503									FAM
<i>Clostridium perfringens</i>	225	235	236	471												FAM
<i>Enterococcus faecalis</i>	262	274	275	376	377	1035										FAM
<i>Enterococcus Casseliflavus</i>	264	276	277	285	376											FAM
<i>Propionibacterium anaerobius</i>	239	245	249	251	372	373										FAM
<i>Staphylococcus lugdunensis</i>	540	541	565	613	642											FAM
<i>Streptococcus milleri</i>	440	447	1097													FAM
<i>Enterobacter cloacae</i>	221															HEX
<i>Bacteroides fragilis</i>	534	536	540	541												HEX
<i>Bacteroides thetaiotaomicron</i>	575	576	577	583	645	651	652									HEX
<i>Streptococcus Pyogenes</i>	471															FAM
<i>Clostridium difficile</i> 4938T	227	230	237	238	268	270	330	331	493	497	552	558				FAM
<i>Shigella sonnei</i>	80															HEX
<i>Escherichia coli</i>	922	925														HEX
<i>Prevotella timonensis</i>	202	617	618	619	620	670	671	684	685							HEX
<i>Staphylococcus aureus</i>	369	415	432	474	476	510	522	526	538							FAM
<i>Listeria monocytogenes</i>	282	295	296	549												FAM
<i>Enterococcus faecium</i>	377	393	394	493												FAM
<i>Clostridium difficile</i> 54206	224	227	232	236	238	331	370	373	388	453	492	493	556			FAM
<i>Proteus vulgaris</i>																IA
<i>Campylobacter jejuni</i>	847															HEX
<i>Proteus mirabilis</i>																IA
<i>Vibrio cholera</i>	600															HEX
<i>Fusobacterium nucleatum</i> *	180	181	184	362												FAM

\**Fusobacterium nucleatum* was run in only one replicate

A summary of the IS fragments present in the pure bacterial cultures where values represent the length of each fragment in base pairs, and the color of the cell represents the number of replicates in which each of these fragments was present. The primer label is specified to the right.

**Table 3.4:** A summary of IS-pro reproducibility with pure bacteria cultures.

Species	Bands in main profile	Replicates containing main profile	Replicates with deviating profiles
<i>Streptococcus agalactiae</i>	1	6	0
<i>Staphylococcus epidermidis</i>	5	4	2
<i>Clostridium perfringens</i>	3	6	0
<i>Enterococcus faecalis</i>	3	5	1
<i>Enterococcus casseliflavus</i>	3	5	1
<i>Propionibacterium anaerobius</i>	3	4	2
<i>Staphylococcus lugdunensis</i>	4	4	0
<i>Streptococcus milleri</i>	1	5	1
<i>Enterobacter cloacae</i>	1	5	1
<i>Bacteroides fragilis</i>	2	6	0
<i>Bacteroides thetaiotaomicron</i>	4	6	0
<i>Streptococcus pyogenes</i>	1	6	0
<i>Clostridium difficile</i> 4938T	5	3	3
<i>Shigella sonnei</i>	1	5	1
<i>Escherichia coli</i>	2	6	0
<i>Prevotella timonensis</i>	4	5	1
<i>Staphylococcus aureus</i>	8	5	1
<i>Listeria monocytogenes</i>	3	6	0
<i>E. flavium</i>	3	4	2
<i>Clostridium difficile</i> 54206	7	5	1
<i>Proteus vulgaris</i>	0	6	0
<i>Campylobacter jejuni</i>	1	6	0
<i>Proteus mirabilis</i>	0	6	0
<i>Vibrio cholera</i>	1	6	0
<i>Total</i>		135	17

### 3.2 PART 2: COMPARISON OF DNA EXTRACTION KITS

Evaluation of the five nucleic acid extraction methods was based on analysis of bacterial DNA quantity, human DNA quantity, and bacterial diversity. Six parallels of each sample were extracted with the following DNA extraction kits: Qiagen QIAamp DNA Stool Mini Kit, with enzymatic pre-treatment;

Qiagen QIAamp DNA Stool Mini Kit, without enzymatic pre-treatment; Stratec PSP Spin Stool DNA Kit, with enzymatic pre-treatment; Stratec PSP Spin Stool DNA Kit, without enzymatic pre-treatment; Qiagen QIAasymphony, with off-board lysis, and Mobio's PowerMag Microbiome RNA/DNA Isolation Kit.

### 3.2.1 QUANTIFICATION OF TOTAL BACTERIAL DNA BY 16S QPCR

Detailed in section 2.2.27, a standard curve was established using unmethylated genomic *E. coli* DNA allowing a quantification of the total amount of bacterial 16S rRNA in each sample. The reaction efficacy was 95% and the line's linear regression was 0.99897 (figure 3.5).

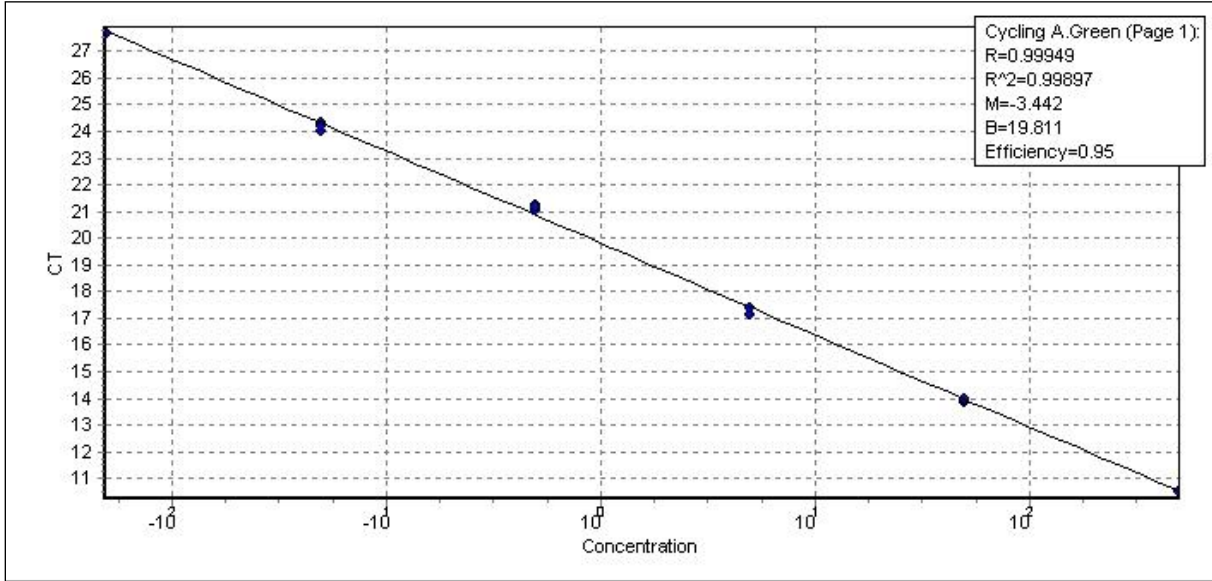
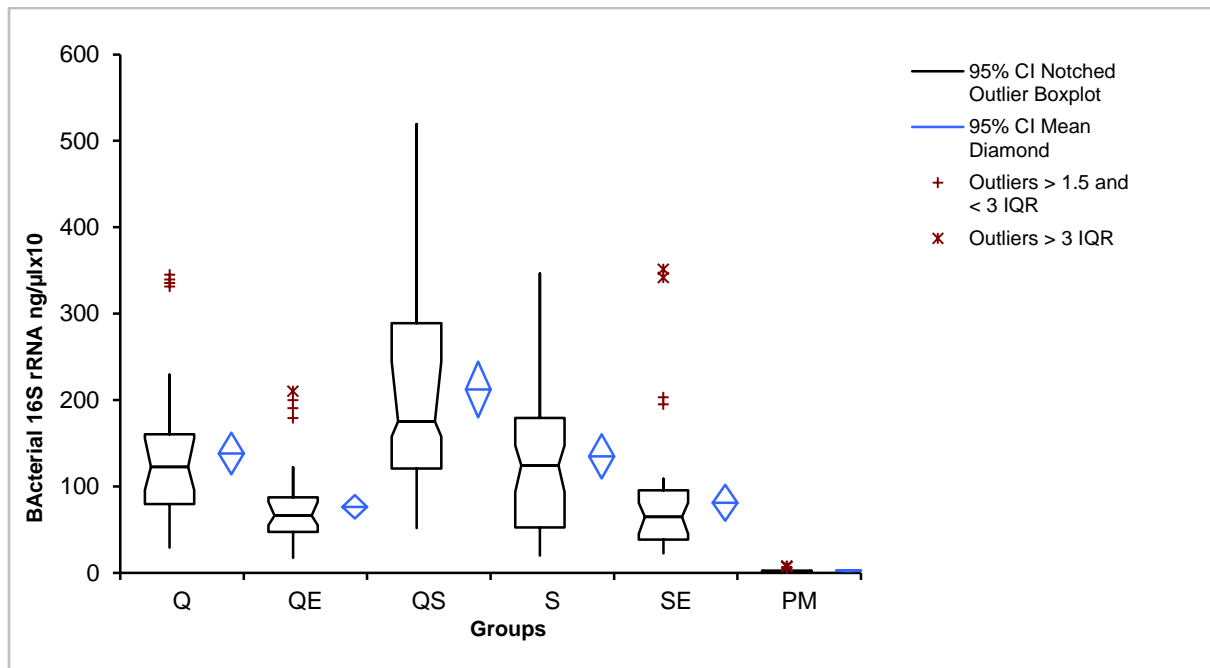


Figure 3.5: qPCR Curve established with unmethylated *E. coli* genomic DNA.





**Figure 3.6:** Side by side box plots of the 16S bacterial DNA yield of the five extraction kits. Kit abbreviations are Qiagen Stool (Q), Qiagen Stool with enzymatic pre-treatment (QE), QIA Symphony (QS), Stratec PSP Stool (S), Stratec PSP Stool with enzymatic pre-treatment (SE), Mobio PowerMag (PM).

Statistical analysis was carried out to determine the kit with the greatest bacterial DNA yield (figure 3.6, table 3.5). An ANOVA was performed on the data and it was determined that the difference in averages was statistically significant ( $P < 0.0001$ ). Primary analysis indicated clearly that the PowerMag kit had an extremely poor yield, therefore, an additional ANOVA was carried out omitting the PowerMag kit data, resulting in no difference in the P value. An ANOVA was then done on just the best three kits, resulting again in a statistically significant P value of  $< 0.0001$ .

**Table 3.5:** Results of ANOVA analysis of the Bacterial 16S qPCR.

Groups	n	Mean	SE	Pooled SE	SD
Q	44	137,88	12,23	11,18	81,12
QE	44	76,31	6,88	11,18	45,64
QS	44	212,16	16,44	11,18	109,07
S	44	134,69	13,04	11,18	86,48
SE	44	80,96	10,61	11,18	70,40
PM	44	2,30	0,29	11,18	1,94

Source of variation	Sum squares	DF	Mean square	F statistic	P
Groups	1115869,60	5	223173,92	40,58	<0.0001
Residual	1418911,51	258	5499,66		
Total	2534781,11	263			

Abbreviations are as follows: Q - Qiagen stool kit; QE - Qiagen stool kit and enzymatic treatment; QS - QiaSymphony with off-board lyses; S - Stratec PSP spin stool kit; SE - Stratec PSP kit and enzymatic treatment; PM - PowerMag.

The QIASymphony extraction robot resulted in the highest yield of total bacteria DNA yield.

Wilcoxon pair-wise assessments were then performed to evaluate the significance of the differences between the three kits that yielded the best qPCR averages: QIASymphony, Qiagen Stool DNA, and the Stratec kit. The QIASymphony yields differences were thus shown to be statistically significantly better than both the Stratec and Qiagen Stool DNA kits ( $P < 0.0001$ ). No statistic significance could be found between the Stratec and Qiagen Stool kits however ( $P = 0.2204$ ).

The poor performance of the PowerMag kit resulted in its exclusion from the any further kit comparisons. The 16S qPCR assay showed the kit to yield exponentially lower DNA concentrations compared to other DNA extraction kits. The kit yielded Nanodrop measurements that were, in some cases, better than those of the competing kits (Appendix 4). Sample DNA eluate from the PowerMag, the Qiagen Stool kit, and Stratec kits was run on 1% agarose gel electrophoresis in a further attempt to sort out this dilemma. The gel results showed smears in all three lanes (figure 3.7): Stratec's and Qiagen's smears began at a much higher molecular weight (larger than the 2000bp band of our ladder), while PowerMag started at around 1200bp. Thus from these gel results, it was concluded that the PowerMag yielded DNA had been severely degraded during the extraction process, supporting the decision to abandon further evaluation of this kit.

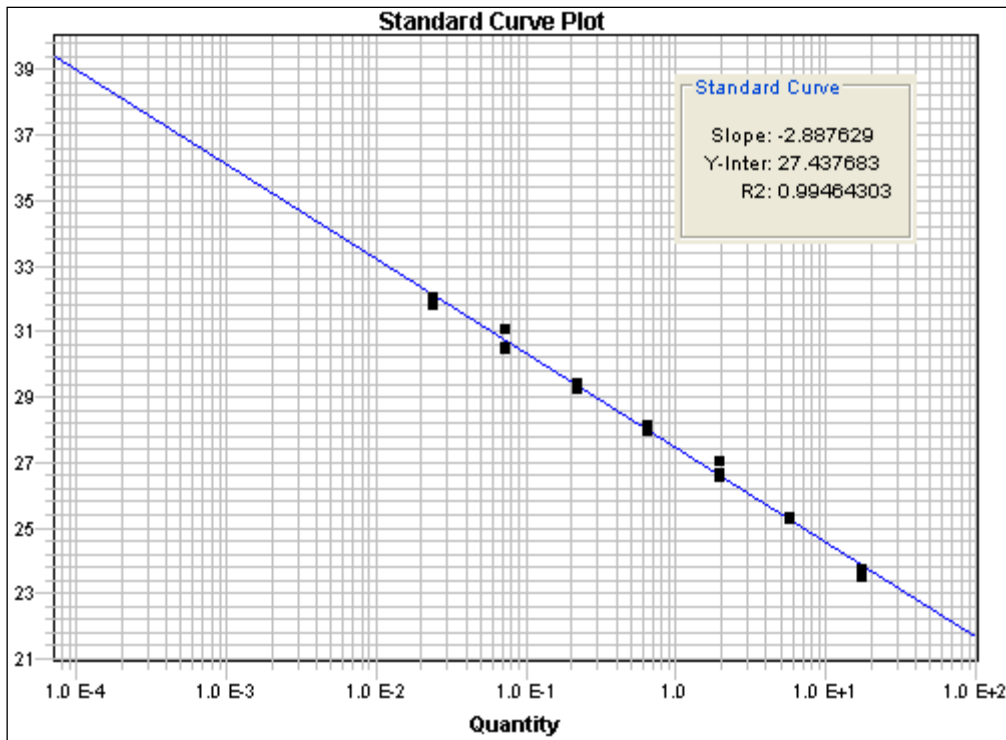


**Figure 3.7:** Comparison of the integrity of fecal DNA extracted by three different methods. Fecal DNA extracted using the PowerMag (lane 2), Qiagen Stool (lane 3), and Stratec PSP (lane 4) kits, analyzed on a 1% agarose gel. The DNA ladder's largest and next largest bands are 2000 and 1200 base pairs respectively (lane 1).

To summarize, results of the 16S qPCR assay indicate that the QIASymphony extraction robot yielded the highest bacterial genomic DNA yield.

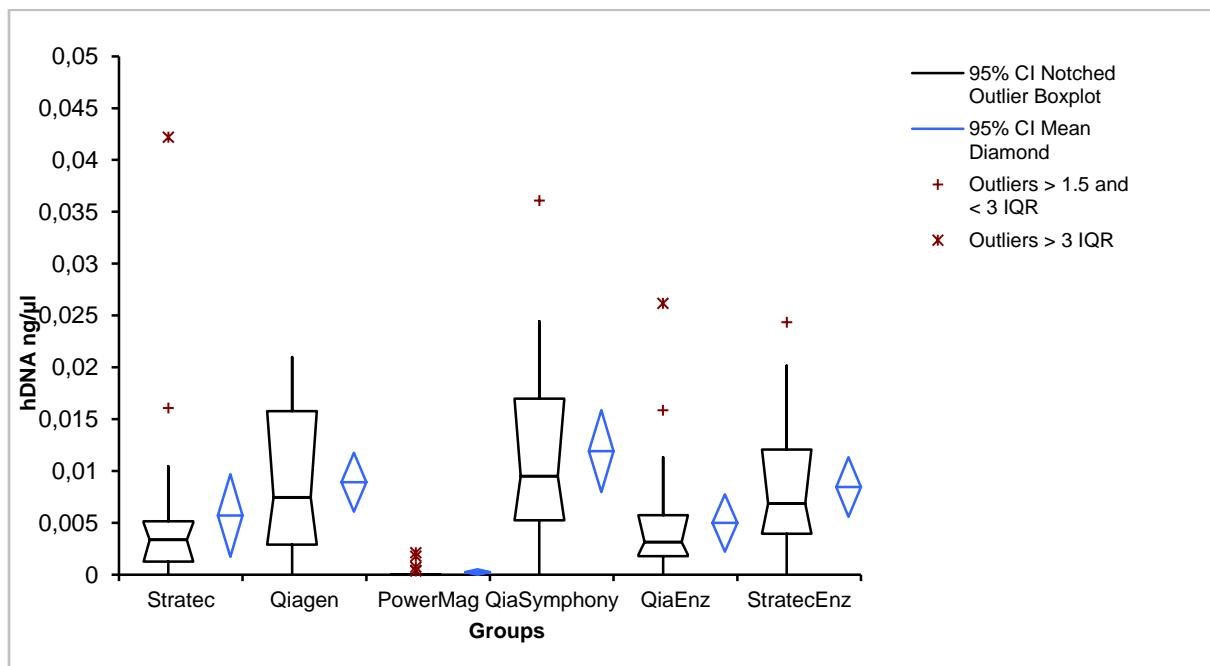
### 3.2.2 QUANTIFICATION OF HUMAN DNA BY QUANTIFILER

A standard curve was established using unmethylated human DNA (figure 3.8) resulting linear regression was 0.9946.



**Figure 3.8:** Standard curve established using Quantifiler genomic DNA quantification kit, used to evaluate human DNA concentrations via, as described in section 2.2.2.7 of the materials and methods section.

Results indicate the QIASymphony extraction yielded the highest mean value of human DNA (hDNA) among the six DNA extraction methods. The QIASymphony extraction robot yielded the most human DNA when each group was averaged. An ANOVA analyses (figure 3.9; table 3.6) was carried out to determine if the difference was statistically significant. The resulting P value (0.0001), even when multiplied by the number of assays, (0.0006) supports the assertion that the difference in the kits is likely real.



**Figure 3.9:** Boxplot of Quantifiler human genomic DNA quantification kit results.

**Table 3.6:** ANOVA analysis of human genomic DNA yield of the five extraction kits.

<b>Groups</b>	<b>n</b>	<b>Mean</b>	<b>SE</b>	<b>Pooled SE</b>	<b>SD</b>
Strattec	22	0,00570	0,00191	0,00146	0,00897
Qiagen	22	0,00892	0,00136	0,00146	0,00639
PowerMag	22	0,00024	0,00012	0,00146	0,00058
QIAsymphony	22	0,01191	0,00191	0,00146	0,00894
QiaEnz	22	0,00498	0,00133	0,00146	0,00623
StrattecEnz	22	0,00845	0,00138	0,00146	0,00647

<b>Source of variation</b>	<b>Sum squares</b>	<b>DF</b>	<b>Mean square</b>	<b>F statistic</b>	<b>p</b>
Groups	0,00178	5	0,00036	7,56	<0.0001
Residual	0,00592	126	0,00005		
Total	0,00770	131			

In addition to a visual inspection of the plotted data of the top three kits (Qiagen, QIAsymphony, and Strattec), a Shapiro-Wilk test was performed. It was determined that the distribution of the data sets was non-Gaussian, and a non-parametric test was employed to further compare the highest scoring kits.

A Wilcoxon pair-wise assessment was performed on the two kits with the highest average yields of hDNA, Qiagen Stool kit and QIASymphony, which resulted in a P value of 0.1375, making the statistical significance of the difference in the top two kits' mean values uncertain (table 3.7).

**Table 3.7:** Wilcoxon pair-wise analysis of A) QIASymphony - Qiagen Stool kit, and B) QIASymphony - Stratec hDNA yields.

A)

<b>Difference between pairs</b>	<b>n</b>	<b>Rank sum</b>	<b>Mean rank</b>
Positive	15	173,0	11,53
Negative	7	80,0	11,43
Zero	0		
Median difference	0		
95.4% CI	0	to 0	(exact)
Wilcoxon's statistic	173,0		
Z statistic	-		
2-tailed p	0,1375	(exact)	

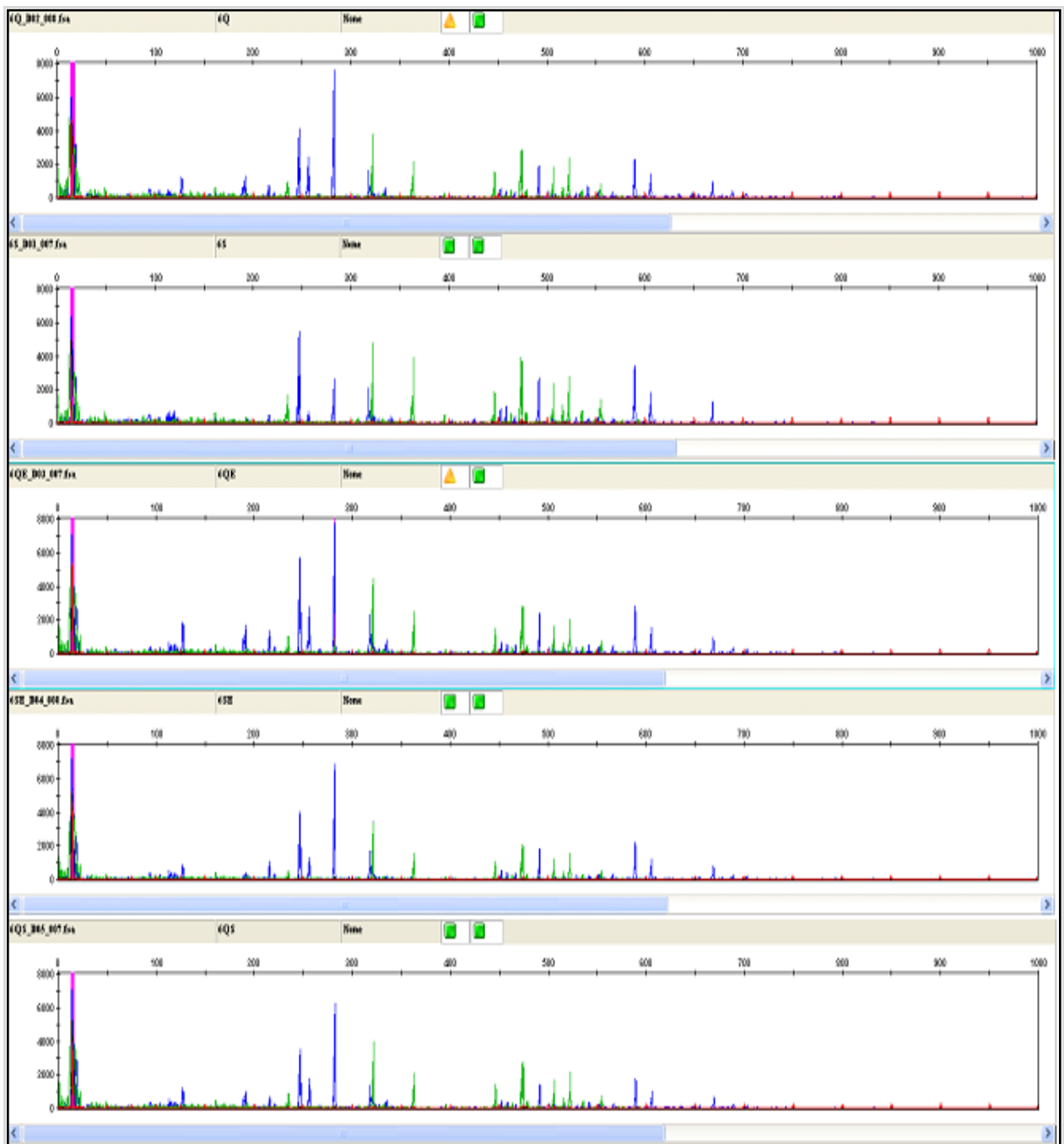
B)

<b>Difference between pairs</b>	<b>n</b>	<b>Rank sum</b>	<b>Mean rank</b>
Positive	18	224,0	12,44
Negative	4	29,0	7,25
Zero	0		
Median difference	0		
95.4% CI	0	to 0	(exact)
Wilcoxon's statistic	224,0		
Z statistic	-		
2-tailed p	0,0008	(exact)	

The results here, lead one to recommend the QIASymphony extraction robot over other DNA extraction methods when attempting to purify host DNA from stool samples. The average yield of the QIASymphony method, although not statistically significantly higher than the Qiagen Stool kit, was the best. Additionally, the automated nature and higher throughput capabilities of the automated system make it a more desirable choice to that of the more laborious manual procedure of the Qiagen Stool DNA Kit.

### 3.2.3 BACTERIAL DIVERSITY ANALYSIS BY IS-PROFILING

To evaluate the bacterial diversity in the DNA extracted with the various DNA extraction kits, IS-pro was performed on all samples. Parallel runs of IS-profiling were performed with DNA from the twenty-two stool samples extracted via the five different extraction methods. The five extraction methods were compared and contrasted using BioNumerics version 7.1 software by Applied Maths. Differences between in both the abundance of peaks, as well as the heights of common peaks was used in assessing the effectiveness of a given extraction kit. Figure 3.10 depicts the five IS-profiles of sample 6 as obtained from the five different extraction methods.



**Figure 3.10:** Parallel IS-pro analyses of a fecal sample extracted with the five extraction methods. The chromatograms presented here are, from top to bottom, Qiagen (Q), Stratec (S), QiaEnzyme (QE), StratecEnzyme (SE), and QIASymphony (QS).

Each of the peaks in figure 3.10 is a IS fragment. The RFU values (peak heights) are thought to be proportional to each IS fragment’s abundance. Both similarities and differences can be observed between the extraction methods’ chromatograms, and although no conclusion was drawn from these alone, it illustrates the need for a detailed comparison of sample diversity between methods, such is the focus of the next section.

### *Alpha diversity*

BioNumerics software was used to calculate Shannon indices, which assess the diversity within each sample using both peak intensity (abundance) and fragment length (species). The best kit should have the highest average diversity, but must also be consistent, resulting in as little variance as possible. The standard deviation of each kit was used as a measure of consistency. Results are shown below in table 3.8. The means and standard deviations varied somewhat, both between kits and between the two primer label colors within the same kit. The Shannon indices of intra-sample diversity were, therefore, weighted accordingly during the process of choosing an extraction method.

**Table 3.8:** A summation of the Shannon indices means and standard deviations from each group.

<b>Kit</b>	<b>Q</b>	<b>Q</b>	<b>QE</b>	<b>QE</b>	<b>QS</b>	<b>QS</b>	<b>S</b>	<b>S</b>	<b>SE</b>	<b>SE</b>
Label	FAM	HEX	FAM	HEX	FAM	HEX	FAM	HEX	FAM	HEX
Mean	2,83	2,65	2,91	2,57	2,83	2,60	2,84	2,77	2,90	2,58
SD	0.40	0.39	0.36	0.40	0.39	0.39	0.34	0.41	0.32	0.35

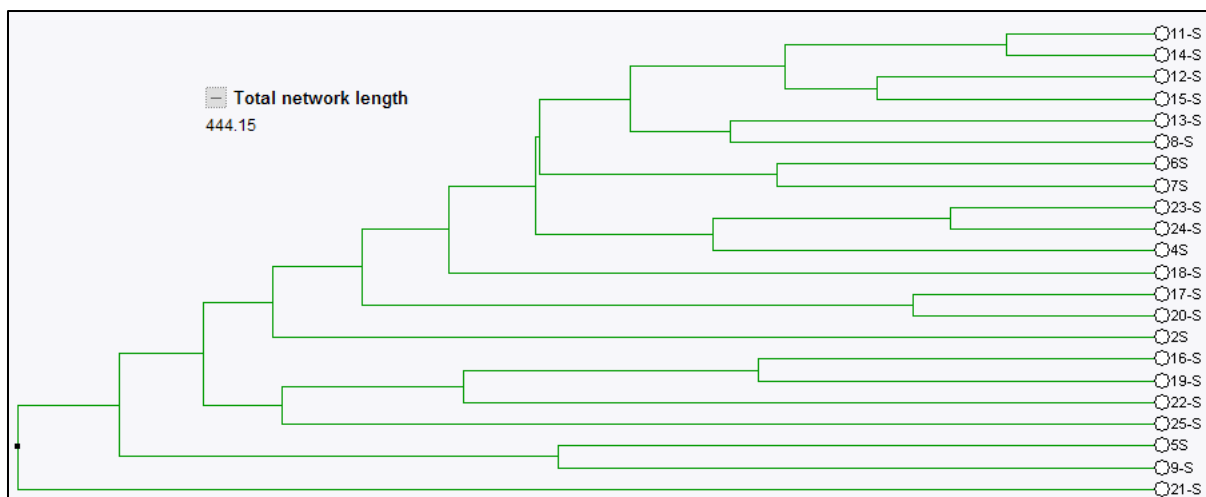
### *Beta diversity*

Between sample diversity for each extraction kit should be as high as possible, considering the samples are unrelated. A dendrogram was produced from the same dataset in BioNumerics and total network length of the dendrograms produced using Pearson correlation as a measure of diversity (figure 3.11). Pearson diversity indices were calculated for each kit, for both the FAM labeled fragments and the HEX labeled fragments.

As table 3.9 shows, the Stratec kit was decisively the most diverse of the HEX profiles, and it came in a close second among FAM profile diversity; Qiagen with enzyme treatment was marginally better. The QIASymphony diversity indices were in the middle of the pack for HEX, but greatly less diverse than all of the other FAM profiles.

**Table 3.9:** Beta diversity ratings calculated using the Pearson coefficient. The network lengths are given by kit and primer label, Qiagen (Q), Qiagen with enzyme treatment (QE), QIASymphony (QS), Stratec (S), Stratec with enzyme treatment (SE).

<b>Kit</b>	<b>Q</b>	<b>Q</b>	<b>QE</b>	<b>QE</b>	<b>QS</b>	<b>QS</b>	<b>S</b>	<b>S</b>	<b>SE</b>	<b>SE</b>
Label	FAM	HEX	FAM	HEX	FAM	HEX	FAM	HEX	FAM	HEX
Network length	313	350	367	329	291	343	355	444	347	339



**Figure 3.11:** Beta diversity of the HEX labeled peaks of the Stratec kit's samples. Depicted here, a dendrogram produced using BioNumerics analysis software and the Pearson coefficient.

As mentioned earlier, bacterial DNA yield was an important factor in choice of extraction method, but bacterial species diversity was of central importance. The higher peaks the QIA Symphony profiles were rejected in favor of the greater bacterial diversity obtained by the Stratec kit. Higher bacterial diversity could lead to more informative group profiling in later stages of the project when evaluating patient samples.

### 3.3 PART 3: COMPARISON OF THE BACTERIAL COMPOSITION AND HUMAN DNA CONCENTRATION IN FECAL SAMPLES FROM POLYP PATIENTS AND CONTROLS

After DNA extraction via the kit chosen in part 2, Stratec, the fecal samples from the 24 polyp patients alluded to in section 2.1.2 were then analyzed and compared to the healthy control group described in the same section. The polyp group was evaluated with the same assays used in choosing the best DNA extraction kit: bacterial 16S qPCR, human DNA quantification, and diversity analysis. Nanodrop measurements were also recorded, and are presented in appendix 7.

#### 3.3.1 QUANTITATION OF BACTERIAL DNA PRESENT IN STOOL SAMPLES

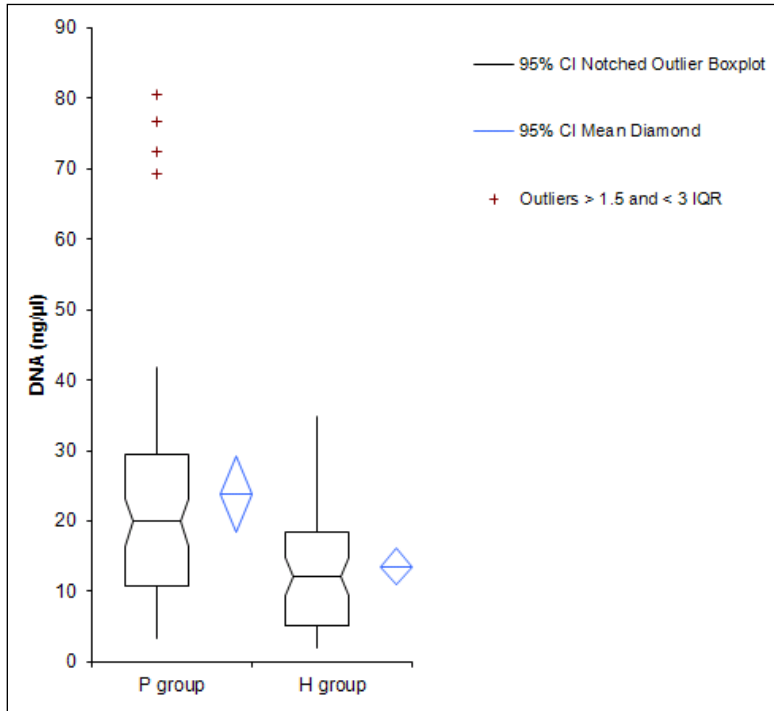
Bacterial DNA in stool samples were quantified with the 16S rRNA qPCR as described in the materials and methods (section 2.2.2.6).

16S rRNA qPCR analysis showed about a 10ng/μl higher mean bacterial DNA concentration in the polyp group (table 3.10, figure 3.12). Detailed measurements are presented in appendix 5.



**Table 3.10:** Average concentrations of bacterial 16S rRNA DNA .

16S qPCR average concentrations	
Healthy Group	13,51 ng/ $\mu$ l
Polyp Group	23,77 ng/ $\mu$ l



**Figure 3.12:** Boxplot and calculated values for the average Bacterial 16S rRNA DNA concentrations of the polyp (P) group and Healthy (H) group.

### 3.3.2 QUANTITATION OF HUMAN DNA PRESENT IN STOOL SAMPLES

Host DNA in stool samples was quantified as described in section 2.2.2.7. Quantifier human DNA quantifier kit analysis showed an average of greater than ten times the presence of polyp group host DNA (table 3.11, figures 3.13 and 3.14). Detailed measurements are presented appendix 6.

**Table 3.11:** Average human DNA concentrations

Average Host DNA	
Polyp Group	89.8 pg/ $\mu$ l
Healthy Group	5.6 pg/ $\mu$ l

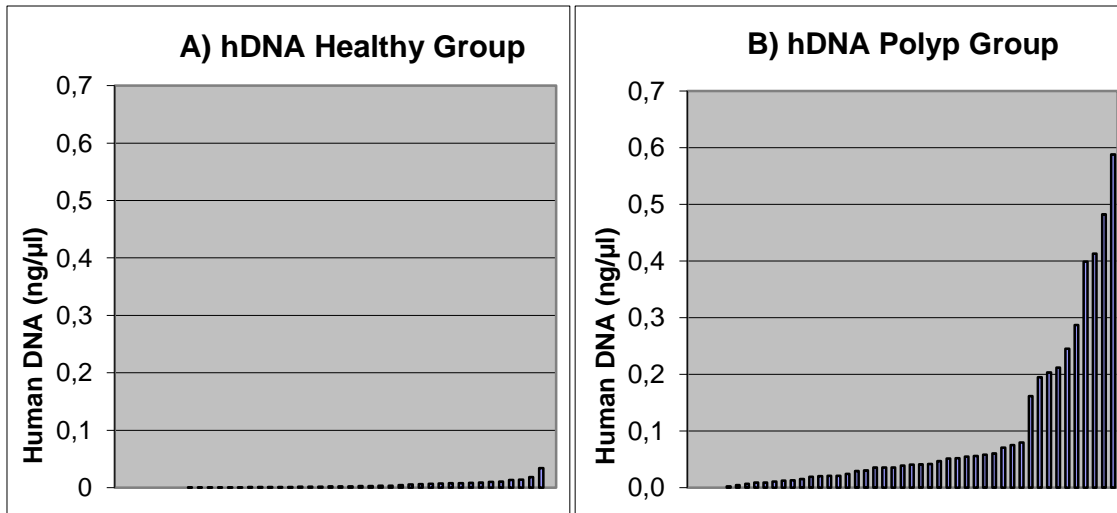


Figure 3.13: Graph of Quantifiler results: A) healthy group; B) polyp group.

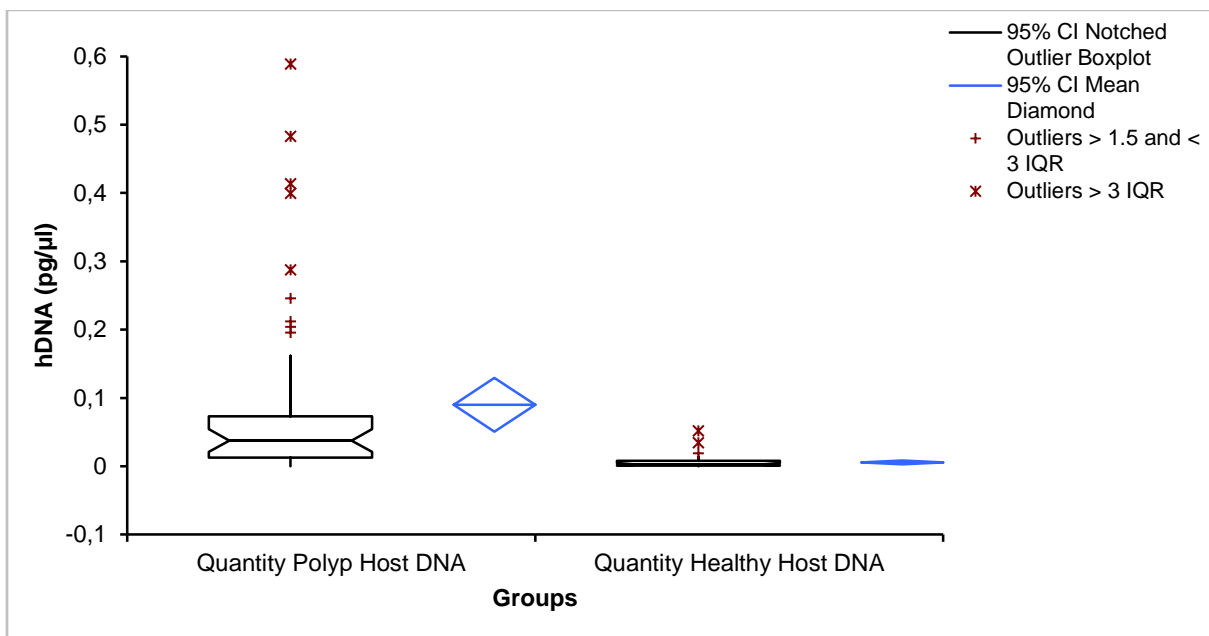


Figure 3.14: Boxplot of the Quantifiler results illustrating higher mean host DNA concentration in the polyp group.

### 3.3.3 COMPARISON OF THE BACTERIAL COMPOSITION IN STOOL SAMPLES

#### *Profile clustering*

A Pearson Clustering was performed on IS-pro data from all samples to determine if the samples clustered in the respective Polyp (P) and Healthy (H) group. The resulting dendrogram did not reveal any observable patterns of similarity, aside from small intermittent clusters of only a few samples.

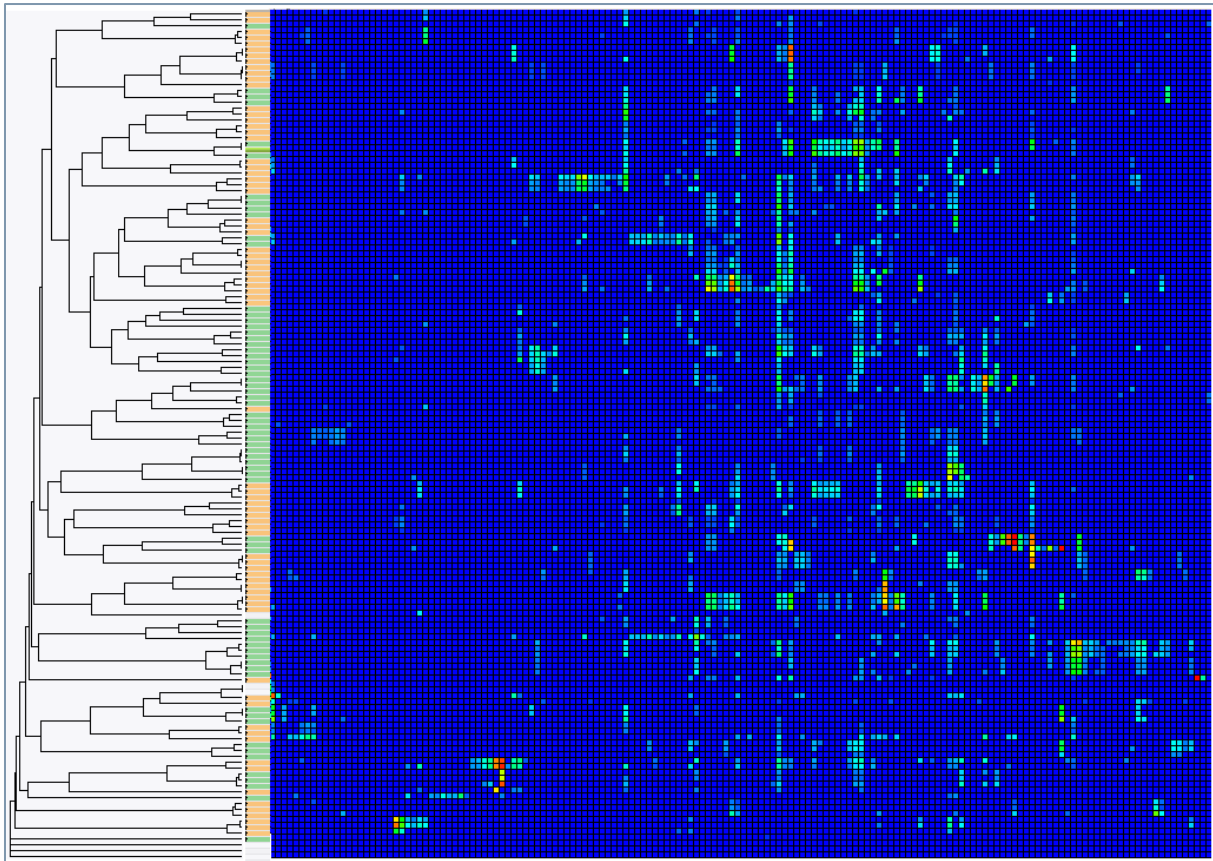


Figure 3.15: Pearson cluster analysis of the polyp group, the healthy group, and the negative samples. Colors of the matrix represent the intensity of IS-fragment intensity (RFU); they range from dark blue (absent) to red (high intensity). Cluster analysis was performed using the top-score UPMGA and the Pearson correlation resulting in a dendrogram to determine if the samples had sufficient common features to cluster into their original groups. Sample groups are color coded to aid viewing ease: the polyp group, orange; the healthy group, green. Dendrogram branches with no group color code are those of the negative samples.

### Jackknife test

Group separation analysis performed with Jackknife test indicates that the intestinal bacterial profile of the healthy group is more heterogeneous, and scores substantially more false identifications of random unknown samples from the same group. However, the polyp group scored significantly higher in this test, implying that the samples in this group are more similar to each other than they are to the samples of the healthy group (table 3.12)

**Table 3.12:** Average group separation using BioNumerics Jackknife test. The green cells indicate the percentage of the analysis program's correct identifications made on random samples treated as unknowns. The white cells indicate the false identifications made, and to which group the 'unknown' were incorrectly placed.

	Polyp	Healthy	Negative controls
Polyp	91.67	57.58	16.67
Healthy	8.33	39.39	0.00
Negative controls	0	3.03	83.33

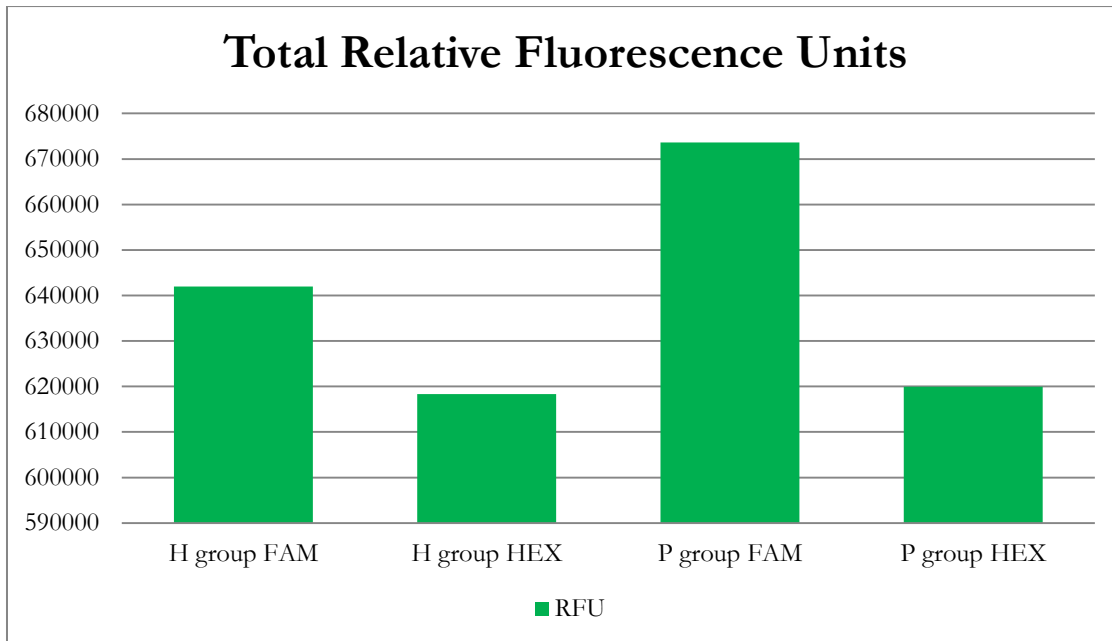
### Relative phylum abundance

The *Firmicutes* to *Bacteroidetes* ratio is commonly used to describe dysbiosis (Budding *et al.* 2010). Analysis of the ratio of *Bacteroidetes* vs *Firmicutes* in the two groups did reveal a slight difference due mostly to the greater total relative fluorescence units (RFU) of FAM in the healthy group, in other words a greater abundance of *Firmicutes*, *Actinobacteria*, *Fusobacterium*, or any combination thereof.

Healthy: total RFU FAM/total RFU HEX = 641988/618334 = 1.03

Polyp: total RFU FAM/total RFU HEX) = 673610/619952 = 1.08

Figure 3.16 depicts the values used to calculate these ratios, drawing attention to the greater total RFU of the Polyps group's FAM signal.



**Figure 3.16:** Total relative fluorescence units of IS-pro analysis. Each bar represents the summation of all the peak intensity values in RFU for each fluorescent label for each of the groups, Healthy (H) and Polyp (P).

### *Alpha Diversity*

Intra-sample bacterial diversity was assessed for all samples to evaluate differences between the two groups. Average Shannon alpha diversity indices (table 3.13) calculated for each sample show slight differences between the two groups.

**Table 3.13:** Computed Shannon indices for each signal of each group.

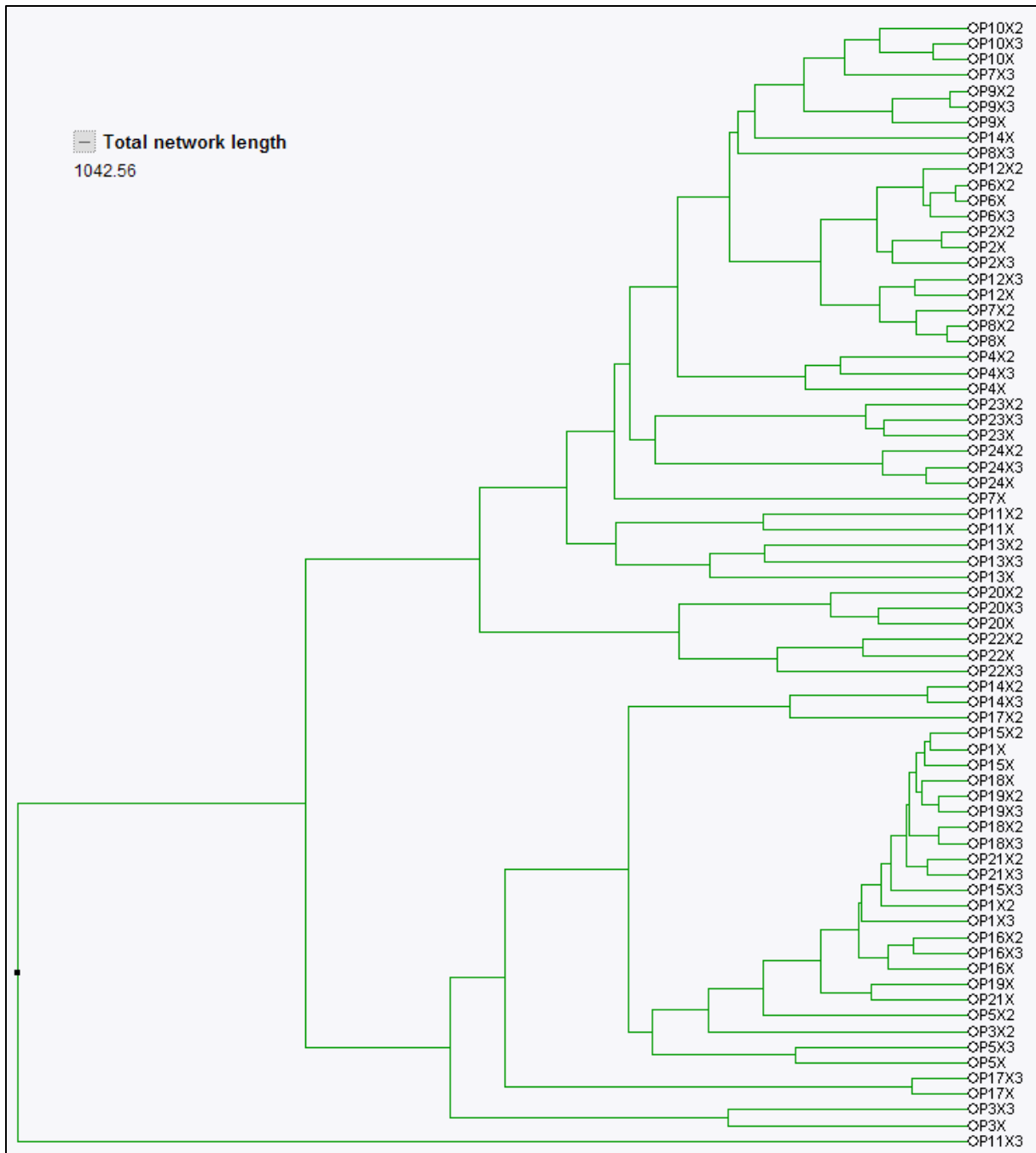
<b>Shannon index</b>	<b>HEX</b>	<b>FAM</b>
Median P group	1,87	1,81
Mean P group	1,77	1,84
Median H group	1,92	2,08
Mean H group	1,76	2,05

### ***Beta diversity***

Beta Diversity was assessed via BioNumerics analysis software. A dendrogram (figure 3.17) was created for each of the primer labels of each group. As shown in table 3.14, the polyp group shows slightly more beta diversity than the healthy group, both for FAM and HEX labels.

**Table 3.14:** Total network lengths of the dendrograms created using cluster analysis and the Pearson correlation.

<b>Total network length</b>		
Polyp	FAM	780.94
	HEX	1042.56
Healthy	FAM	767.66
	HEX	922.18

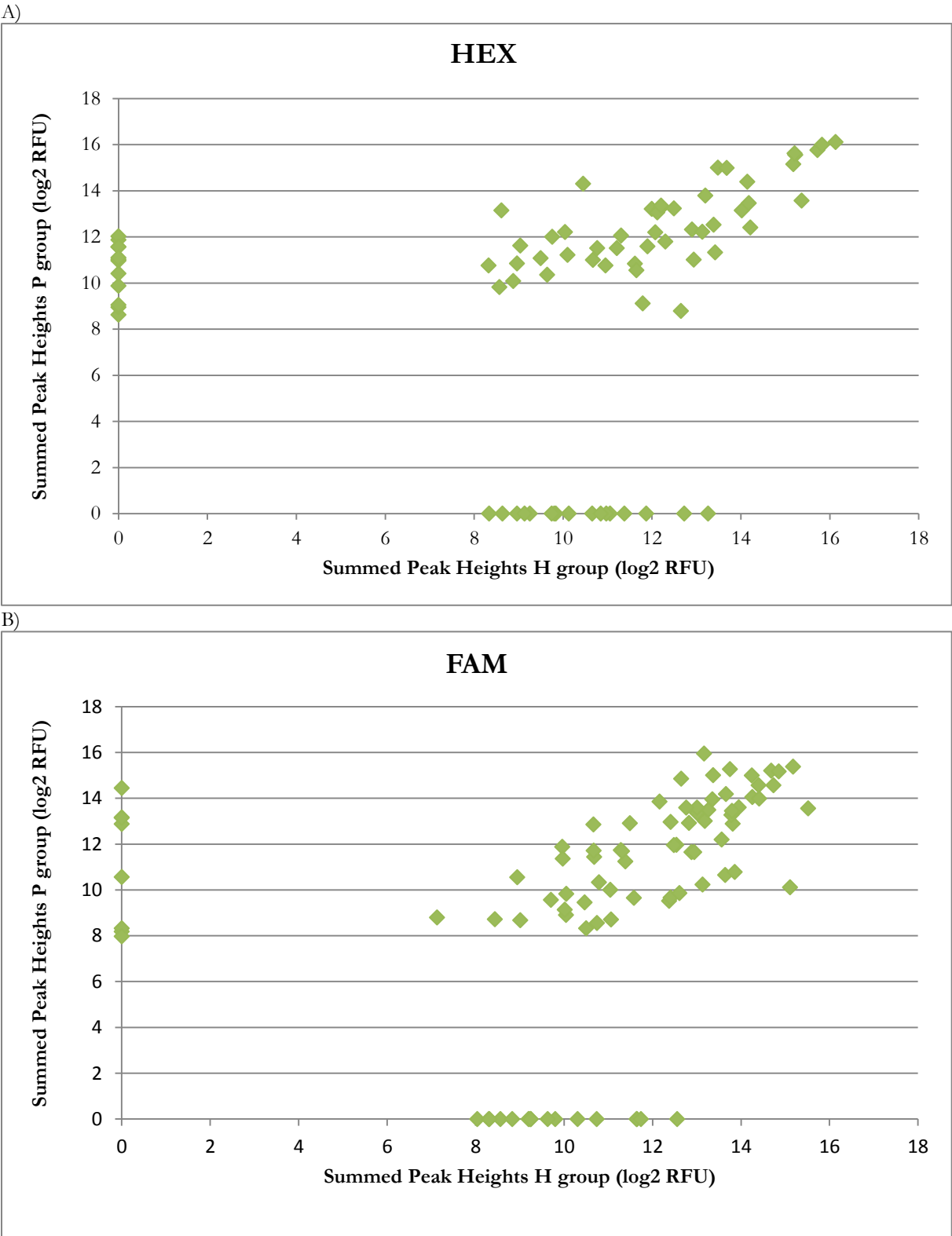


**Figure 3.17:** Dendrogram of the polyps group's HEX primer beta diversity estimation via the total network length of the dendrogram created using the Pearson correlation. Identical analysis was carried out for both colors of both groups.

### 3.3.4 IDENTIFICATION OF GROUP SPECIFIC BANDS

To identify band classes specific to one group, several correlation plots were made (figure 3.18). Plotting the correlation of the composite profiles obtained from IS-profiling of the healthy subjects' samples (H) against that of the samples of the polyp group resulted in a cluster of points near the origin, a result of the domination of small average peak values. Log2 transformation of the data provided a more clear understanding of that previously obscured by this tight cluster of small values. The points along the axes and those that deviated from  $R^2=1$  initially looked like promising candidates for peaks characteristic of

one group, but closer inspection revealed that they were just outliers present in very few samples. Hence, no peaks specific to one group were identified.



**Figure 3.18:** A correlation of composite profiles from the summation of RFU values, log2 transformed. Healthy (H) subjects' sample values are plotted against those from the polyp (P) group subjects. A) HEX peak RFU values Healthy (X) versus Polyp (Y) B) FAM peak RFU values Healthy (X) versus Polyp (Y). Summed values were log 2 transformed to compensate for the dominant cluster of small peaks, consequently also highlighting divergent peaks.

---

## 4. DISCUSSION

The discussion of this project's experiments and results is broken down in three parts, as in the other parts of this paper: Part 1, a technical evaluation of IS-profiling; Part 2, a discussion of the process of choosing a DNA extraction method; and Part 3, a discussion of the results as they specifically pertain to the project's polyp group and healthy group.

### 4.1 PART 1: TECHNICAL EVALUATION OF IS-PROFILING

#### **General discussion**

IS-profiling (IS-pro) is an attractive analysis method for a small to medium size laboratory due its relatively modest investment in both capital and person-hours. A thermal cycler, a genetic analyzer, and analysis software are all that are needed.

After a test run to compare the IS-pro master mix described by Budding *et al.*, 2010 and the fragment analysis master mix recommended by Applied Biosystems, the master mix recipe from the Budding article was chosen, and an abbreviated testing of IS-pro primer sensitivity and specificity was carried out. It was not the intention of the project to reproduce all the method evaluations performed by Budding *et al.* (2010). Our objective was simply to control that the method functioned as intended in our lab, despite any small variations that might exist.

The primer sets designed for *Firmicutes* and *Actinobacteria* readily amplified *Fusobacterium nucleatum*. This is in concordance with previous results (Budding *et al.*, 2010). It is also of interest in the larger CRC project, as it is theorized that *F. nucleatum* presence in the colon is correlated with colorectal cancer (Bashir *et al.*, 2014; Allen-Vercoe and Jobin, 2014). Plans for a more thorough investigation of this relationship are already in place.

Evaluation of whether a primer amplified a specific species or not were carried out with parallel PCRs, using both the IS-pro labeled primers, and sequencing primers which were identical in sequence, but lacked the FAM and HEX labels of the IS-pro forward primers.

The *Firmicutes/Actinobacteria/Fusobacteria* primer sets proved to be both specific and sensitive within the limited selection of bacteria species tested here, amplifying all of the target species and none of the *Bacteroidetes* or *Proteobacteria*. No *Actinobacteria* species were tested during this thesis' abbreviated testing, however Budding *et al.* (2010) carried out *in vitro* testing of such to satisfactory results.

The primer set designed to target *Bacteroidetes*, however, was somewhat less specific. Sensitivity was not an issue as all of the target species were amplified as expected. Specificity was perfect if one only considered the amplification of the two target phyla against one another ie. *Firmicutes/Actinobacteria/Fusobacteria* versus *Bacteroidetes*. The issues with specificity arose under testing with *Proteobacteria* species: three negative results of the expected seven negative indicates, that a good deal of that we see as HEX labeled peaks in the IS-profiles of subject sample could theoretically actually be *Proteobacteria* species, or possibly species from another phylum.

A consideration must, however, be made as to the completely contrived nature in which this test of specificity has been designed. In the author's own experience, as well as in other documented sources, primer sets in the absence of their intended targets will find sufficiently compatibility in less than perfect sequence matches, leading to unintended amplification during PCR (Kommedal *et al.*, 2012). This phenomenon is witnessed on a regular basis in our lab when attempts are made to carry out 16S rRNA sequencing of bacteria direct from patient sample material. In the absence of target species, group specific 16S primers cross-react with other non-target species, as well as somewhat similar sequences in the human genome, resulting in human matches in BLAST searches. Kommedal *et al.* concurs in their work of 2012, elaborating that unspecific amplification can occur despite multiple primer mismatches in samples with a low target : non-target ratio (Kommedal *et al.*, 2012).

One could speculate that the specificity of this *Bacteroidetes* primer set would perform to a better specificity rating in the polymicrobial environment of actual stool samples.



Sequencing of purified bands excised from gel electrophoresis, as well as those purified from clones resulted in a confirmation of the correct bacterial species. A match of 99% or better is a common requirement for species identification via 16S rRNA sequencing (CLSI, 2008) but since no such guidelines requirements could be found for identification using the 16S-23S intergenic region (ITS), a subjective appraisal was made whereby the distance percent-wise to the next bacterial species evaluated, and judged to be sufficient. Since there appears not to be a general consensus, authors carry out empirical testing of the species in which they are concerned, and, subsequently, establish their own criteria. Chang *et al.* (2005) found intra-species ITS sequence similarities ranged from 0.99 to 1.0, whereas interspecies similarities varied from 0.86 to 0.92. Tannock *et al.* (1999) determined that a similarity of 97.5% or greater was considered sufficient to provide identification to the species level.

Using thus the most conservative criteria set forth by the literature reviewed in this paper, a distance of 2.5% to the next nearest species should be sufficient to identify the cultures used here. Concerning the two *Firmicutes* species with poor sequencing results, *S. lungdunsis* gave a BLAST result with a distance of 14% from the next most similar species, as well as a large difference in score (1212 to 153) to the next best species match; *P. anaerobius* BLAST resulted in no other species, despite the poor (90%) match, and short length of the sequence submitted to the database search (section 3.1.3).

### **Reproducibility of IS-pro**

Six technical replicates of each of the bacteria cultures were analyzed to determine the method's reproducibility. The benefit of analyzing replicates was twofold. Firstly, it aided in determining the settings of BioNumerics; since there are few peaks in the IS-profile of a pure culture and the peak profiles should be identical, it allowed adjustment of the settings aiding in automated alignment of peaks from sample to sample. Secondly, it provided a means by which to evaluate and quantify the method's reproducibility.

Consistency in band matching settings is of paramount importance when performing inter-sample comparisons.

The reproducibility, in some cases, proved somewhat challenging to evaluate. Two to three peaks, close in length were appearing where one peak was expected. After several failed attempts to remedy the problem via band-matching and curve-finding settings, Applied Maths, the designers of BioNumerics software, were contacted. Following several consultations, it was decided that an overload of PCR product had caused adequately high signals at the tops of some peaks to cause the 3130xl Genetic Analyzer's laser to shut down at the peak of these extremely high intensity fragment signals during analysis of certain of the bacterial isolates' IS-profiling. This phenomenon resulted in an IS-profile whereby the intense peak of a IS-pro curve was absent, leaving only the sides leading up to the peak remaining, thus giving the effect of two peaks where only one should have been.

Much time and resources had already been used in the testing of these replicates, so it was deemed prudent to go further without running the tests over, interpreting the data accordingly: peaks of similar height within two base pairs of each other were interpreted as a single peak with its actual peak somewhere between the two side peaks; low intensity peaks not common to three or more of the replicates were dismissed as heightened background.

For reasons stated above, the height (abundance) of peaks in each replicate was not evaluated.

Admittedly, the compromises detailed above lend a degree of subjectivity to the interpretation of the bacterial culture data. This factor, while not ideal, has little effect on the outcome on the objectives at hand. Overload of PCR product during fragment analysis was not an issue with the actual feces samples, as the IS-fragments are distributed across the spectrum of species present in each sample. This is in contrast to the distribution of IS-fragments within the cultured strains, where the entirety of the sample's nucleic acid concentration is distributed in one to very few peaks, causing the extremely high, problematic signals. Given these technical problems, peak intensity data of the pure cultures were not used to calculate reproducibility. Instead an evaluation was made of what defined a main profile for a bacterial species. This main profile was then used to judge the reproducibility of the method. This, unquestionably, introduced a large degree of subjectivity, and the 88.82% reproducibility rating should be interpreted with caution.

## 4.2 PART 2: COMPARISON OF DNA EXTRACTION KITS

### ***General Discussion***

As mentioned earlier, the optimal DNA extraction is of paramount importance in this study. It is necessary to extract as diverse a selection of bacterial DNA as possible, while simultaneously yielding sufficient amounts of both host and bacterial DNA to carry out the intended experiments. DNA quality also became an issue under the kit evaluation portion of the project; a kit that yields damaged, un-amplifiable DNA is of no use, as in the case of the PowerMag kit.

After initial bacterial 16S qPCR results, the PowerMag kit, having yielded exponentially less bacterial DNA than the other kits, was scrutinized in an attempt to identify the problem. Since Nanodrop results were on par, and in some cases higher, compared to the other methods, DNA degradation became the most likely cause. As a consequent of this experience, Nanodrop measurements were not given any weight in the evaluation of the extraction methods.

The eluates from Qiagen Stool and Stratec Stool kits were run on a gel electrophoresis (section 3.2.1) along with PowerMag eluate, revealing a smear band that began somewhere between the 1200 and 2000 base pair bands. In strong contrast, both the Qiagen Stool kit, and the Stratec Stool kits produced smear bands that never migrated lower than the 2000 base pair band of our ladder. It was then concluded that DNA degradation during extraction had sheared the DNA to the point that it was not amplifiable and useless for the purposes of this study.

The results of the PowerMag kit in this project were most probably a consequence of user error, or something done when adapting the protocol for manual use. Originally designed to be used with an EpiMotion robot, it was necessary to make changes to the protocol in order to include it in the project's evaluation in the absence of said robot. Since it is a well regarded kit from a reputable biotech firm (MoBio), it is assumed that the kit functions properly under normal use. Time and resource restraints precluded trouble shooting, and the evaluation proceeded without further consideration of the PowerMag kit.

### ***DNA yields: Bacterial 16S rRNA and Human DNA***

Of the remaining methods, all gave usable bacterial DNA yields, and might have been acceptable on that criterion alone. The larger CRC project is, however, also interested in exploring changes of host epigenetics. This requires that sufficient Human host DNA (hDNA) is extracted from the feces samples in addition to bacterial DNA. Finally, maximum bacterial diversity, as stated earlier, was also desirable.

One method, QIA Symphony scored highest with regards to both total bacterial 16S rRNA and hDNA concentration. In addition, it is the method that is best suited for high sample throughput, due its automation. Results here indicate that the automated QIA Symphony is the clear choice where high throughput analysis of human DNA from feces takes precedent.

### ***Bacterial composition***

However, for the purposes of the tasks of the present project, the broadest bacterial diversity possible was considered of chief importance. Studies have shown differences in bacterial diversity when comparing different DNA extraction methods (Wesolowska-Andersen *et al.*, 2014; Sohrabi *et al.*, 2016). Sample pre-treatment with enzymes has been described as effective in lysing gram positive cell walls (Sohrabi *et al.*, 2016), and was therefore included in two of our extraction protocols. We did not find any positive effects of enzyme treatment, however. The method that scored best in our evaluation of method diversity was the Stratec PSP Spin Stool kit.

Diversity was evaluated on both a per-sample basis (alpha diversity), and a per-kit basis (beta diversity). The average alpha diversity differences between kits were negligible, whereby no one method came out as most diverse in both FAM and HEX. The Stratec (S) kit scored the most average alpha diversity among the HEX labeled IS-profiles, And the Qiagen Stool kit with enzyme (QE) scored best among those labeled with FAM. Again, the differences were slight, and by no means decisive.

Beta diversity analysis results were, however, less equivocal. It was assumed that all samples were unrelated; therefore the kit resulting in the largest difference between samples (highest beta diversity) was

avored. The Stratec kit was, by far, the most diverse among the HEX profiles. It scored second among FAM profiles, which had a much smaller spread of values. QIASymphony resulted in the lowest FAM beta diversity, and somewhere in the middle of the pack among the HEX scores. This raises the question, is perhaps the method somewhat lacking in its ability to lyse the thick walled gram positive bacteria of the *Firmicutes* and *Actinobacteria* phyla.

Thusly, the Stratec PSP Spin Stool DNA Kit was chosen to perform the DNA extraction duties of both the CRC project, as well as Part 3 of this project.

### 4.3 PART 3: COMPARISON OF THE BACTERIAL COMPOSITION AND HUMAN DNA IN FECAL SAMPLES FROM POLYP PATIENTS AND CONTROLS

#### ***General discussion***

The assays used to evaluate the differences between extraction kits were also used to compare the group of presumed healthy subjects and the group of polyp patients: total 16S rRNA, host DNA, and bacterial diversity. Additionally, several attempts were made to determine if there were any characteristics unique to one of the two groups.

The group referred to herein as the healthy group is, admittedly, not the ideal control group. The intended healthy group was to be patients that had undergone colonoscopy, and declared healthy due to the confirmed absence of colorectal polyps. This group took too long to assemble with regards to the time restraints of this master's thesis. The healthy control group of the present project is that which was used to evaluate the extraction methods discussed in Part 2. This has introduced confounds to this study that need to be addressed and considered when interpreting the results.

Firstly, an age bias exists. The average age of the healthy group (40.95) was about nine years under the 50 year old recommended age where screening begins for the general population. In fact, only four members of the group were above the 50 year mark. In contrast, the polyp group contains no subjects under the age of 50; the average age for the polyp group is 67.29 years old, well above the 50 year mark. Both the human intestinal microbiota and the incidence of intestinal polyps are shown to change with age (Mariat *et al.*, 2009; Hopkins *et al.*, 2002; Eide, 1991). It could very well be this age related change in intestinal microflora composition that leads to an increased risk of adenoma development.

The age bias confounds data interpretation, introducing uncertainty concerning whether inter-group differences are a result of a natural age related progression, or a pathological development caused by, or causing, the formation of neoplastic intestinal growth, or both.

Secondly, the presumed part of the presumed healthy group introduces a second possible confounding variable in that some members of this group may, unbeknownst to themselves, have intestinal polyps. It was unfortunate that there was not time enough to assemble the healthy group that had been proven healthy via colonoscopy. This group is scheduled to replace the presumed healthy group in the larger CRC project.

Additionally, possible spatial heterogeneity should be taken into consideration when one attempts to characterize the diversity of a landscape. Despite efforts to homogenize samples, problems with biased sample taking could affect all down stream analysis (Kirk *et al.*, 2004). In this thesis, the subjects are in control of sample taking, possibly introducing an additional confound. The ecology of the small amount of feces required for DNA kit extraction is not necessarily representative of the ecology of the feces sample as a whole. Biological replicates might help to reduce the impact of spatial heterogeneity on results (Kirk *et al.*, 2004).

#### ***16S bacterial rRNA concentrations***

The results of the total 16S rRNA concentration varied somewhat between groups. The healthy group had an average concentration that was about 10 ng/ $\mu$ l, lower than that of the polyp group. A disproportionate increase in one or more of the bacterial theorized to have a pathogenic effect could be

responsible for polyp development, but since no systematic difference in peak pattern or inter-phyla ratio between groups was detected, theories of dysbiosis remain pure speculation.

An alternate explanation can be found in the method in which the feces samples were extracted. Using 200µl of homogenized feces blended with transport buffer, lends itself to a slightly inconsistent amount of actual feces going into the extraction process due to issues of sample taking and sample viscosity. One could argue that the issues with sample taking are fairly random, but a systematic difference in the viscosity of a polyp patient's stool could account for the difference in total bacterial DNA between the groups. These possibly confounding factors should also be considered when interpreting the next parameter, human DNA concentrations.

### ***Human DNA concentrations***

Human DNA concentrations differed greatly between groups. The polyp group concentrations were in excess of ten times that of the healthy control group. This finding was as expected since studies have shown enhanced exfoliation from colonic neoplasms (Dickensen *et al.*, 2015; Ahlquist *et al.*, 2000). Other studies have, however, contradicted these findings, showing increased in hDNA feces concentrations in colorectal cancer groups, but not in polyp groups (Loktionov *et al.*, 1998, Teixeira *et al.*, 2015). It seems logical though that the increase in epithelial surface area of the intestine due to neoplastic growths, which can be 200 times greater than predicted by a polyp's gross dimensions (Dickensen *et al.*, 2015), would lead to a corresponding increase in exfoliation.

The presence of greater hDNA concentrations in polyp patients' stools thus lends itself to earlier non-invasive screening for molecular abnormalities associated with CRC, and possibly earlier intervention than had otherwise been possible. This will be an area of focus in the larger CRC project of which this project is a part.

### ***Bacterial composition***

Alpha diversity analysis comparison of the polyp group versus the healthy group resulted in a slightly higher average rating of diversity in the individual samples healthy group, the greatest difference being the diversity of *Firmicutes/Actinobacteria/Fusobacteria* phyla: means of 2.05 for the FAM phyla for the healthy group, versus the polyp group's mean of 1.84. Analysis of the alpha diversity ratios of the target FAM and HEX phyla to each other (FAM:HEX) resulted in similar scores: 1.03 in the healthy group, 1.08 in the polyp group. The healthy group possesses somewhat greater diversity within individual samples.

Beta diversity analysis using, as in the evaluations of Part 2, total network length of a dendrogram created via the BioNumerics clustering tool and the Pearson correlation indicates the polyp group contains greater inter-sample diversity. Beta diversity analysis revealed higher polyp group scores for both the HEX and FAM target phyla: 780.94 (polyp) to 767.66 (healthy) for FAM, and 1042.56 (polyp) to 922.18 (healthy) for HEX. While the difference in *Firmicutes/Actinobacteria/Fusobacteria* diversity differs only slightly, there are more diverse populations of *Bacteroidetes*, and possibly *Proteobacteria*, in the polyp group than in the healthy group. This could indicate a shift in microbiota that has led to the formation of intestinal polyps. Both *Bacteroidetes* and *Proteobacteria* phyla contain species that have been implicated in human morbidity (Curtis *et al.*, 2014; Arthur *et al.*, 2012; Schwabe and Wang, 2012).

Group analysis testing was also carried out in the quest to uncover within group commonalities. The first test, a jackknife test, randomly chooses one sample at a time, treats it as an unknown, and then attempts to assign it to either the healthy or polyp group base on some common characteristics. The jackknife test was surprisingly successful in assigning polyp samples to the polyp group, guessing right in 91.67% of attempts. The test was largely unsuccessful in correctly assigning healthy group samples, guessing right in only 39.39% of attempts, less than expected if one had just randomly guessed to which group they belonged.

This suggests that the polyp group samples have some common characteristic, whether it is a peak, or a set of peaks, or perhaps the common absence of certain peaks, maybe similar peak abundance. The Pearson clustering of the two groups together revealed clustering of neither the polyp group nor the healthy group. There were only some small clusters comprised of a few samples.

This led to attempts to identify group characteristic peaks manually. Several correlation plots were made in an attempt to find data points. Peak heights (RFU) were manipulated in vain, trying to find some sort of group distinguishing data point. RFU value means, sums, and medians of each peak were plotted group against group without uncovering what similarity the polyp group shared. Log2 transformations of the data were also carried out in an attempt to identify small variations in a landscape confused by the domination of the data points from small peaks. Several data points appeared as promising candidates at first, but closer inspection of the raw data revealed the peaks to be of large values in only one, or a few samples, outliers of sorts. The lack of any single characteristic top suggests that the trait unique to the polyp group is a more complex relationship between two or more peaks that is too difficult to discern by means of manual scrutiny.

The absence of any clearly defining single characteristic and the seemingly contradictory nature of the alpha and beta diversity data make the data interpretation difficult. The alpha diversity analysis indicates that each individual healthy group sample contains, on average, greater species diversity; the beta diversity analysis implies that the polyp group samples show a greater degree of diversity in relation to each other. These analytical paradigms take species abundance (profile peak height, RFU) into account in addition to species richness (amount of peaks). Ratios of the intragroup total RFU for FAM:HEX revealed virtually no intergroup differences, but the values for the polyp group were somewhat higher.

Attempts to reconcile these relationships within the two groups' data lead one to postulate that the possible higher average total 16S bacterial rRNA concentrations in the polyp group might be the underlying commonality leading to the very successful results of the jackknife test. The polyp group contained higher total 16S rRNA on average. The jackknife test's inability to correctly place the majority of the healthy group's samples might possibly be due to their higher intra-sample diversity as indicated by their higher Shannon index values. More heterogeneity would make them less easily characterized.

Alternatively, the fact that the healthy group placement scored so much worse than random could also suggest that there's possibly something common between the correctly placed 91.67% of polyp samples and the 57.58% wrongly placed samples of the healthy group. Without a more experimentally sound control group it's difficult to strongly argue one possibility over the alternative.

Applied Maths recommended their BioNumerics Dimensioning and Matrix Mining module for a more detailed analysis of the more complex relationship between the composite profiles. Our lab has neither access to this module, nor the resources available to expand our license to include it.

### ***Application of IS-profiling***

IS-profiling is to be employed as a means of detecting systematic differences in the intestinal microflora across two or more groups of subjects. Both characteristics and irregularities common to all samples ultimately fall into the background, and it is the opinion of the author that differences between groups will continue to be elucidated by the method if they are there to be discovered.

It is here; in its role as a tool to identify differences in patterns across two or more groups of individuals, that IS-profiling finds its niche. With the right software, characteristic peaks, or patterns of peaks unique to a group could be uncovered. IS-pro, in the author's opinion is poorly suited as a means of species identification in polymicrobial samples, a cumbersome application more suited to NGS.

To identify individual peaks, it would be necessary to find and excise the band from a gel. The sheer volume of bands of varying length makes the task, at best, extremely challenging. This would be, of course, followed by gel band purification and sequencing.

Other practical difficulties arise when adjusting the settings within the software. With polymicrobial samples especially, it is difficult to ascertain whether two IS-pro peaks in close proximity to each other are in fact the same band from a single species, two slightly different bands from a single species, or, in fact, two bands from two distinct species that happen to have IS regions close in length. Such possibilities complicate the choice of BioNumerics analysis settings with respect to both band detection, as well as band matching.

These settings govern how the analysis program calculates shadow bands, which arise when the same band presents itself as two bands varying by one, or two base pairs due to some irregularities in amplification during PCR; finding out which bands correspond in inter-sample comparisons is also a challenging aspect that can have an impact on the conclusions drawn.

A database of bacterial isolates' IS-regions is something that could aid in the interpretation of the microflora profiles. An attempt was made to compile one during this project, but the putative sample overload and resulting problems of the laser shutting down during high intensity signals, made the construction of such somewhat subjective. The effort was abandoned.

AE Budding has assembled such a database, but this is a commercial venture, and the resources needed to access it were not available to our lab.

When studying the bacterial composition of a community, the choice of method depends on the biological question to be addressed. Other methods are better suited for studying overall diversity or identification of predominant members of a community. In conclusion, IS-pro is a promising method for analysis of fingerprint data, where coarse deviations are to be identified. Commonalities, whether specific or nonspecific amplification products, will become a baseline of sorts; unique peaks will still emerge. Care must, of course, be taken in optimizing the analysis settings to avoid deceptive results. More detailed inquiries are best served by PCR assays, or NGS.

### ***Future Studies***

The difficulties in uncovering the common characteristics that led to the jackknife test's high degree of accuracy in placing the polyp group correctly suggest that subsequent studies are in order. Assembly of a more legitimate healthy group is essential.

The higher values of total RFU for the *Firmicutes/Actinobacteria/Fusobacteria* FAM labeled primers of the polyp group indicate that a similar study using more specific primers for *Fusobacteria* or *C. difficile* are in order.

A closer examination of specifically *Proteobacteria* populations across the two groups would also certainly be interesting, especially in light of the recent findings of Daniels *et al.* (2014), who, through IS-pro, showed the genus' disproportionately high presence in the stool of diverticulitis patients.

Additionally, more intricate forms of data analysis are appropriate. The BioNumerics Dimensioning and Matrix Mining module appears necessary if one is to uncover intra-group similarities that are a result of a complex relationship between multiple peaks. Deep sequencing of the groups' intestinal microbiota via NGS promises to be an even better choice to uncover compositional differences between the two groups where the resources are available.

---

## REFERENCES

- Absar, M. S., and N. Y. Haboubi. 2004. "Colonic Neoplastic Polyps: Biopsy Is Not Efficient to Exclude Malignancy. The Trafford Experience." *Techniques in Coloproctology*. doi:10.1007/s10151-004-0172-3.
- Allen-Vercoe, Emma, and Christian Jobin. 2014. "Fusobacterium and Enterobacteriaceae: Important Players for CRC?" *Immunology Letters*. doi:10.1016/j.imlet.2014.05.014.
- Arthur, Janelle C, Ernesto Perez-Chanona, Marcus Mühlbauer, Sarah Tomkovich, Joshua M Uronis, Ting-Jia Fan, Barry J Campbell, *et al.* 2012. "Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota." *Science (New York, N.Y.)* 338 (6103): 120–23. doi:10.1126/science.1224820.
- Baker, Barbara S. 2006. "The Role of Microorganisms in Atopic Dermatitis." *Clinical and Experimental Immunology* 144 (1): 1–9. doi:10.1111/j.1365-2249.2005.02980.x.
- Barcenilla, a, S E Pryde, J C Martin, S H Duncan, C S Stewart, C Henderson, and H J Flint. 2000. "Phylogenetic Relationships of Butyrate-Producing Bacteria from the Human Gut." *Appl.Environ.Microbiol.* 66 (4): 1654–61. doi:10.1128/AEM.66.4.1654-1661.2000.
- Bashir, Arif, Abid Y Miskeen, Ashaquallah Bhat, Khalid M Fazili, and Bashir A Ganai. 2015. "Fusobacterium Nucleatum: An Emerging Bug in Colorectal Tumorigenesis." *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP)* 24 (5): 373–85. doi:10.1097/CEJ.0000000000000116.
- Boleij, A., E. M. Hechenbleikner, A. C. Goodwin, R. Badani, E. M. Stein, M. G. Lazarev, B. Ellis, *et al.* 2015. "The Bacteroides Fragilis Toxin Gene Is Prevalent in the Colon Mucosa of Colorectal Cancer Patients." *Clinical Infectious Diseases* 60 (2): 208–15. doi:10.1093/cid/ciu787.
- Budding, A. E., M. E. Grasman, F. Lin, J. A. Bogaards, D. J. Soeltan-Kaersenhout, C. M. J. E. Vandenbroucke-Grauls, A. A. van Bodegraven, and P. H. M. Savelkoul. 2010. "IS-pro: High-Throughput Molecular Fingerprinting of the Intestinal Microbiota." *The FASEB Journal* 24 (11): 4556–64. doi:10.1096/fj.10-156190.
- Burkitt, D P. 1971. "Epidemiology of Cancer of the Colon and Rectum." *Cancer* 28 (1): 3–13. <http://www.ncbi.nlm.nih.gov/pubmed/8223063> \n<http://www.ncbi.nlm.nih.gov/pubmed/5165022>.
- Canani, Roberto Berni, Margherita Di Costanzo, Ludovica Leone, Monica Pedata, Rosaria Meli, and Antonio Calignano. 2011. "Potential Beneficial Effects of Butyrate in Intestinal and Extraintestinal Diseases." *World Journal of Gastroenterology*. doi:10.3748/wjg.v17.i12.1519.
- Cao, Boyang, Min Wang, Lei Liu, Zhemin Zhou, Shaoping Wen, Antoni Rozalski, and Lei Wang. 2009. "16S-23S rDNA Internal Transcribed Spacer Regions in Four Proteus Species." *Journal of Microbiological Methods*. doi:10.1016/j.mimet.2009.01.024.
- Chang, H. C., Y. F. Wei, L. Dijkshoorn, M. Vaneechoutte, C. T. Tang, and T. C. Chang. 2005. "Species-Level Identification of Isolates of the Acinetobacter Calcoaceticus-Acinetobacter Baumannii Complex by Sequence Analysis of the 16S-23S rRNA Gene Spacer Region." *Journal of Clinical Microbiology* 43 (4): 1632–39. doi:10.1128/JCM.43.4.1632-1639.2005.
- CLSI. Interperative Criteria for Identification of Bacteria and Fungi by DNA Target Sequencing; Appoved Guideline. CLSI document MM18-A. Wayne, PA: Clinical and Laboratory Stands Institute; 2008.
- Curtis, Meredith M., Zeping Hu, Claire Klimko, Sanjeev Narayanan, Ralph Deberardinis, and Vanessa Sperandio. 2014. "The Gut Commensal Bacteroides Thetaiotaomicron Exacerbates Enteric Infection through Modification of the Metabolic Landscape." *Cell Host & Microbe* 16 (6): 759–69. doi:10.1016/j.chom.2014.11.005.
- Daniels L, Budding AE, de Korte N, Eck A, Bogaards JA, Stockmann HB, Consten EC, Savelkoul PH, Boermeester MA. 2014. "Fecal Microbiome Analysis as a Diagnostic Test for Diverticulitis." *Eur J Clin Microbiol Infect Dis* 33: 1927–36. doi:10.1007/s10096-014-2162-3.
- Diez, Beatriz, Carlos Pedr??s-Ali??, Terence L. Marsh, and Ramon Massana. 2001. "Application of Denaturing Gradient Gel Electrophoresis (DGGE) to Study the Diversity of Marine Picoeukaryotic Assemblages and Comparison of DGGE with Other Molecular Techniques." *Applied and Environmental Microbiology*. doi:10.1128/AEM.67.7.2942-2951.2001.
- Dethlefsen, Les, Paul B. Eckburg, Elisabeth M. Bik, and David A. Relman. 2006. "Assembly of the Human Intestinal Microbiota." *Trends in Ecology & Evolution* 21 (9): 517–23. doi:10.1016/j.tree.2006.06.013.
- Dickinson, Brandon T, John Kisiel, David A Ahlquist, and William M Grady. 2015. "Molecular Markers for Colorectal Cancer Screening." *Gut* 64 (9): 1485–94. doi:10.1136/gutjnl-2014-308075.
- Dulal, S, and T O Keku. 2014. "Gut Microbiome and Colorectal Adenomas." *Cancer J* 20 (3): 225–31.

doi:10.1097/PPO.000000000000050.

- Eckburg, Paul B, Elisabeth M Bik, Charles N Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R Gill, Karen E Nelson, and David A Relman. 2005. "Diversity of the Human Intestinal Microbial Flora." *Science (New York, N.Y.)* 308 (5728): 1635–38. doi:10.1126/science.1110591.
- Eide TJ. 1986. "RISK OF COLORECTAL CANCER IN ADENOMA-BEARING INDIVIDUALS WITHIN A DEFINED POPULATION." *Int J Cancer* 38: 173–76.
- Eide, Tor J. 1991. "Natural History of Adenomas." *World Journal of Surgery* 15 (1): 3–6.
- Gurtler, Volker, and Vilma A. Stanisich. 1996. "New Approaches to Typing and Identification of Bacteria Using the 16S-23S rDNA Spacer Region." *Microbiology*. doi:10.1099/13500872-142-1-3.
- Gao, Zhiguang, Bomin Guo, Renyuan Gao, Qingchao Zhu, and Huanlong Qin. 2015. "Microbiota Disbiosis Is Associated with Colorectal Cancer." *Frontiers in Microbiology* 6 (February). doi:10.3389/fmicb.2015.00020.
- Gelsomino, a, a C Keijzer-Wolters, G Cacco, and J D van Elsas. 1999. "Assessment of Bacterial Community Structure in Soil by Polymerase Chain Reaction and Denaturing Gradient Gel Electrophoresis." *Journal of Microbiological Methods* 38 (1-2): 1–15. doi:10.1016/S0167-7012(99)00054-8.
- Grady, William M., and Sanford D. Markowitz. 2015. "The Molecular Pathogenesis of Colorectal Cancer and Its Potential Application to Colorectal Cancer Screening." *Digestive Diseases and Sciences* 60 (3): 762–72. doi:10.1007/s10620-014-3444-4.
- Hague, A, B Singh, and C Paraskeva. 1997. "Butyrate Acts as a Survival Factor for Colonic Epithelial Cells: Further Fuel for the in Vivo versus in Vitro Debate." *Gastroenterology* 112 (3): 1036–40. doi:10.1053/gast.1997.v112.agast971036.
- Hamer, H. M., D. Jonkers, K. Venema, S. Vanhoutvin, F. J. Troost, and R. J. Brummer. 2008. "Review Article: The Role of Butyrate on Colonic Function." *Alimentary Pharmacology and Therapeutics*. doi:10.1111/j.1365-2036.2007.03562.x.
- Hinnebusch, B F, S Meng, J T Wu, S Y Archer, and R A Hodin. 2008. "The Effects of Short-Chain Fatty Acids on Human Colon Cancer Cell Phenotype Are Associated with Histone Hyperacetylation." *J.Nutr.*
- Hopkins, M. J., and G. T. Macfarlane. 2002. "Changes in Predominant Bacterial Populations in Human Faeces with Age and with Clostridium Difficile Infection." *Journal of Medical Microbiology* 51 (5): 448–54. doi:10.1099/0022-1317-51-5-448.
- Horiuchi, Akira. 2014. "Improving Quality Measures in Colonoscopy and Its Therapeutic Intervention." *World Journal of Gastroenterology* 20 (36): 13027. doi:10.3748/wjg.v20.i36.13027.
- Kirk, Jennifer L., Lee A. Beaudette, Miranda Hart, Peter Moutoglis, John N. Klironomos, Hung Lee, and Jack T. Trevors. 2004. "Methods of Studying Soil Microbial Diversity." *Journal of Microbiological Methods*. doi:10.1016/j.mimet.2004.04.006.
- Kommedal O, Simmon K, Karaca D, Langeland N, Wike HG. 2012. "Dual Priming Oligonucleotides for Broad-Range Amplification of the Bacterial 16S rRNA Gene Directly from Human Clinical Specimens." *Journal of Clinical Microbiolog* 50 (4): 1289–94.
- Kumar V, Cotran RS, Robbins SL, Robbins basic pathology: 7<sup>th</sup> ed. (Saunders, 2003), 579-585.
- L., Dong, Wang W., Li A., Kansal R., Chen Y., Chen H., and Li X. 2015. "Clinical next Generation Sequencing for Precision Medicine in Cancer." *Current Genomics* 16 (4): 253–63. <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L605103661\http://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=18755488&id=doi:&atitle=Clinical+next+generation+sequencing+for+precision+medicine+in+cancer&stitle=Curr.+Genomics&title=>.
- Lahti, Leo, Jarkko Salojärvi, Anne Salonen, Marten Scheffer, and Willem M. de Vos. 2014. "Tipping Elements in the Human Intestinal Ecosystem." *Nature Communications* 5: 4344. doi:10.1038/ncomms5344.
- Loktionov, Alexandre, Ian K. O'Neill, Katharine R. Silvester, John H. Cummings, Steve J. Middleton, and Richard Miller. 1998. "Quantitation of DNA from Exfoliated Colonocytes Isolated from Human Stool Surface as a Novel Noninvasive Screening Test for Colorectal Cancer." *Clinical Cancer Research* 4 (2): 337–42. <http://clincancerres.aacrjournals.org/content/4/2/337>.
- Louis, Petra, Pauline Young, Grietje Holtrop, and Harry J. Flint. 2010. "Diversity of Human Colonic Butyrate-Producing Bacteria Revealed by Analysis of the Butyryl-CoA:acetate CoA-Transferase Gene." *Environmental Microbiology*. doi:10.1111/j.1462-2920.2009.02066.x.
- Mariat, D, O Firmesse, F Levenez, VD Guimaraes, H Sokol, J Doré, G Corthier, and J-P Furet. 2009. "The Firmicutes/Bacteroidetes Ratio of the Human Microbiota Changes with Age." *BMC Microbiology* 9 (1): 123. doi:10.1186/1471-2180-9-123.



- Murray P, Rosenthal k, Kobayashi G, Pfaller M, Medical Microbiology, 4<sup>th</sup> ed. (Mosby, 2002), 80
- Muto T, Bussey HJR, Morso BC. 1975. "THE EVOLUTION OF CANCER OF THE COLON AND RECTUM." *Cancer* 36: 2251–227.
- Muyzer, Gerard. 1999. "DGGE/TGGE a Method for Identifying Genes from Natural Ecosystems." *Current Opinion in Microbiology*. doi:10.1016/S1369-5274(99)80055-1.
- Nechvatal, Jordan M., Jeffrey L. Ram, Marc D. Basson, Phanramphoei Namprachan, Stephanie R. Niec, Kawsar Z. Badsha, Larry H. Matherly, Adhip P N Majumdar, and Ikuko Kato. 2008. "Fecal Collection, Ambient Preservation, and DNA Extraction for PCR Amplification of Bacterial and Human Markers from Human Feces." *Journal of Microbiological Methods*. doi:10.1016/j.mimet.2007.11.007.
- Ng, Siew Chien, Emma F C Lam, Tommy T Y Lam, Yawen Chan, Wendy Law, Pete C H Tse, Michael A. Kamm, Joseph J Y Sung, Francis K L Chan, and Justin C Y Wu. 2013. "Effect of Probiotic Bacteria on the Intestinal Microbiota in Irritable Bowel Syndrome." *Journal of Gastroenterology and Hepatology*. doi:10.1111/jgh.12306.
- Parekh, Parth J., and Sanjay K. Sikka. 2015. "Basic Endoscopic Findings — Normal and Pathological Findings." In *Screening for Colorectal Cancer with Colonoscopy*. InTech. doi:10.5772/61256.
- Plöger, Svenja, Friederike Stumpff, Gregory B. Penner, Jörg Dieter Schulzke, Gotthold Gäbel, Holger Martens, Zanming Shen, Dorothee Günzel, and Joerg R. Aschenbach. 2012. "Microbial Butyrate and Its Role for Barrier Function in the Gastrointestinal Tract." *Annals of the New York Academy of Sciences*. doi:10.1111/j.1749-6632.2012.06553.x.
- Schetter, Aaron J., Hirokazu Okayama, and Curtis C. Harris. 2012. "The Role of MicroRNAs in Colorectal Cancer." *The Cancer Journal* 18 (3): 244–52. doi:10.1097/PPO.0b013e318258b78f.
- Schwabe, R. F., and T. C. Wang. 2012. "Bacteria Deliver a Genotoxic Hit." *Science* 338 (6103): 52–53. doi:10.1126/science.1229905.
- Sohrabi, Mohsen, Raj G. Nair, Lakshman P. Samaranyake, Li Zhang, Abu Hasanat Md. Zulfiker, Adnan Ahmetagic, David Good, and Ming Q. Wei. 2016. "The Yield and Quality of Cellular and Bacterial DNA Extracts from Human Oral Rinse Samples Are Variably Affected by the Cell Lysis Methodology." *Journal of Microbiological Methods* 122 (March): 64–72. doi:10.1016/j.mimet.2016.01.013.
- Souilmi, Yassine, Alex K Lancaster, Jae-Yoon Jung, Ettore Rizzo, Jared B Hawkins, Ryan Powles, Saaïd Amzazi, Hassan Ghazal, Peter J Tonellato, and Dennis P Wall. 2015. "Scalable and Cost-Effective NGS Genotyping in the Cloud." *BMC Medical Genomics* 8 (1): 64. doi:10.1186/s12920-015-0134-9.
- Tabatabaei, Meisam, Mohd Rafein Zakaria, Raha Abdul Rahim, Andr<sup>π</sup>enis G. Wright, Yoshihito Shirai, Norhani Abdullah, Kenji Sakai, *et al.* 2009. "PCR-Based DGGE and FISH Analysis of Methanogens in an Anaerobic Closed Digester Tank for Treating Palm Oil Mill Effluent." *Electronic Journal of Biotechnology* 12 (3). doi:10.2225/vol12-issue3-fulltext-4.
- Tannock, G. W., A. Tilsala-Timisjarvi, S. Rodtong, J. Ng, K. Munro, and T. Alatosava. 1999. "Identification of Lactobacillus Isolates from the Gastrointestinal Tract, Silage, and Yoghurt by 16S-23S rRNA Gene Intergenic Spacer Region Sequence Comparisons." *Applied and Environmental Microbiology*.
- Teixeira, Yolanda, Jacqueline Miranda LIMA, Maria Luiza Almeida Prado Oliveira SOUZA, Pedro AGUIAR Jr, Tiago Donizetti SILVA, and Nora Manoukian FORONES. 2015. "Human DNA Quantification in the Stools of Patients with Colorectal Cancer." *Arquivos de Gastroenterologia* 52 (4): 293–98. doi:10.1590/S0004-28032015000400008.
- Tonna, I. 2005. "Pathogenesis and Treatment of Clostridium Difficile Infection." *Postgraduate Medical Journal* 81 (956): 367–69. doi:10.1136/pgmj.2004.028480.
- Troy, Erin B, and Dennis L Kasper. 2010. "Beneficial Effects of Bacteroides Fragilis Polysaccharides on the Immune System." *Frontiers in Bioscience (Landmark Edition)* 15: 25–34. doi:10.2741/3603.
- Tuomisto, Hanna. 2010. "A Consistent Terminology for Quantifying Species Diversity? Yes, It Does Exist." *Oecologia*. doi:10.1007/s00442-010-1812-0.
- Walter J, Tannock GW, Tilsala-Timisjarvi A, Rodtong S, Loach DM, Munro K, Alatosava T. 2000. "Detection and Identification of Gastrointestinal Lactobacillus Species by Using Denaturing Gradient Gel Electrophoresis and Species-Specific PCR Primers." *Applied and Environmental Microbiology* 66 (1): 297–303. doi:10.1128/AEM.66.1.297-303.2000.
- Walther, Axel, Elaine Johnstone, Charles Swanton, Rachel Midgley, Ian Tomlinson, and David Kerr. 2009. "Genetic Prognostic and Predictive Markers in Colorectal Cancer." *Nature Reviews Cancer* 9 (7): 489–99. doi:10.1038/nrc2645.
- Wang, Tingting, Guoxiang Cai, Yunping Qiu, Na Fei, Menghui Zhang, Xiaoyan Pang, Wei Jia, Sanjun Cai, and

- Liping Zhao. 2012. "Structural Segregation of Gut Microbiota between Colorectal Cancer Patients and Healthy Volunteers." *The ISME Journal* 6 (2): 320–29. doi:10.1038/ismej.2011.109.
- Wesolowska-Andersen, Agata, Martin Iain Bahl, Vera Carvalho, Karsten Kristiansen, Thomas Sicheritz-Pontén, Ramneek Gupta, and Tine Rask Licht. 2014. "Choice of Bacterial DNA Extraction Method from Fecal Material Influences Community Structure as Evaluated by Metagenomic Analysis." *Microbiome* 2: 19. doi:10.1186/2049-2618-2-19.
- Yu, Youngseob, Changsoo Lee, Jaai Kim, and Seokhwan Hwang. 2005. "Group-Specific Primer and Probe Sets to Detect Methanogenic Communities Using Quantitative Real-Time Polymerase Chain Reaction." *Biotechnology and Bioengineering* 89 (6): 670–79. doi:10.1002/bit.20347.

---

# APPENDICES

## APPENDIX 1

### MoBio PowerMag Microbiome RNA/DNA Isolation Kit protocol

The following details the adaptations made to the PowerMag kit

- Pre heated 650µl lysis solution per sample to 60° C for 15 minutes.
- 16.25 µl β-mercaptoethanol was added to the lysis solution in 2 ml tubes
- The glass beads were removed from the kit's 96 well plates, and transferred to the tubes.
- 100µl of phenolchloroform: isoamyl alcohol (25:24:1) was added to each tube.
- 200 µl of stool was added to each tube.
- The tubes were then shaken using a FastPrep 24 instrument for 40 seconds, at 6 m/s.
- The tubes were then centrifuged for 6 minutes at 4500G.
- The supernatant was then transferred to a new tube.
- 150 µl of Inhibitor Removal Solution was added, the tubes vortexed, and incubated at 4°C for 5 minutes.
- The tubes were then centrifuged for 6 minutes, at 4500G.
- The supernatant was transferred over to a new tube and centrifuged at 4500G for 6 minutes.
- 850µl of the supernatant was then transferred to a new tube.
- 870 µl of ClearMag Beads / ClearMag Binding Solution was added.
- The tubes were incubated at 55°C, with a mixing speed of 1000 RPM for 5 minutes.
- MiniMAG magnetic separation of ClearMag Beads, 10 minutes.
- The supernatant was removed and discarded.
- 500 µl of Wash Solution was added, mixed.
- MiniMAG magnetic separation of ClearMag Beads, 10 minutes.
- The supernatant was removed and discarded
- 500 µl of Wash Solution was added, mixed.
- MiniMAG magnetic separation of ClearMag Beads, 10 minutes.
- The supernatant was removed and discarded.
- 500 µl of Wash Solution was added, mixed.
- MiniMAG magnetic separation of ClearMag Beads, 10 minutes.
- The supernatant was removed and discarded.
- 100µl of PCR grade water was added to each tube.
- The tubes were at 1200 RPM for 10 minutes at room temperature.
- MiniMAG magnetic separation of ClearMag Beads, 10 minutes.
- 100 µl of eluate was then transferred to a new tube, and frozen at -20°C.

## APPENDIX 2

### Quantifiler data

Each sample value is the mean value of the duplicates. Mean hDNA yield from each kit lies at the bottom of each column.

	Stratec	Qiagen	PowerMag	QiaSymphony	QiaEnz	StratecEnz
2	0,0013	0,0016	0,0000	0,0092	0,0042	0,0067
4	0,0047	0,0158	0,0000	0,0098	0,0057	0,0077
5	0,0422	0,0210	0,0021	0,0167	0,0105	0,0136
6	0,0000	0,0183	0,0004	0,0071	0,0113	0,0092
7	0,0069	0,0181	0,0000	0,0225	0,0262	0,0243
8	0,0050	0,0086	0,0000	0,0163	0,0035	0,0099
9	0,0023	0,0159	0,0000	0,0185	0,0066	0,0125
11	0,0002	0,0001	0,0000	0,0021	0,0000	0,0011
12	0,0050	0,0070	0,0004	0,0081	0,0020	0,0050
13	0,0048	0,0127	0,0000	0,0045	0,0046	0,0045
14	0,0105	0,0101	0,0000	0,0244	0,0159	0,0201
15	0,0028	0,0026	0,0000	0,0059	0,0038	0,0048
16	0,0013	0,0079	0,0000	0,0107	0,0000	0,0053
17	0,0161	0,0060	0,0018	0,0213	0,0028	0,0120
18	0,0039	0,0047	0,0000	0,0023	0,0025	0,0024
19	0,0048	0,0112	0,0007	0,0168	0,0025	0,0097
20	0,0005	0,0016	0,0000	0,0140	0,0000	0,0070
21	0,0012	0,0063	0,0000	0,0024	0,0019	0,0022
22	0,0069	0,0171	0,0000	0,0361	0,0035	0,0198
23	0,0029	0,0045	0,0000	0,0079	0,0000	0,0040
24	0,0009	0,0026	0,0000	0,0053	0,0021	0,0037
25	0,0019	0,0029	0,0000	0,0000	0,0000	0,0000
AVG.	0,0057	0,0089	0,0002	0,0119	0,0050	0,0084

## APPENDIX 3

### 16S q PCR data

Table XX: Results of Bacterial 16S qPCR; all values are in picogram and diluted 1:100. Abbreviations are as follows: Q - Qiagen stool kit; QE – Qiagen stool kit and enzymatic treatment; QS – QIA Symphony with off-board lyses; S – Stratec PSP spin stool kit; SE – Stratec PSP kit and enzymatic treatment; PM – PowerMag.

Sample	Qiagen	QiaEnz	QiaSymphony	Stratec	StraEnz	PowerMag
2	99,37	45,06	174,97	107,18	31,49	1,38
2	100,86	43,43	173,86	101,32	31,13	1,30
4	80,07	17,42	104,46	45,63	22,87	0,73
4	81,72	17,89	109,27	47,82	23,18	0,70
5	154,17	42,72	121,77	278,68	34,10	3,97
5	160,34	55,07	131,37	266,45	35,12	4,16
6	155,46	47,97	184,84	117,98	34,70	2,01
6	155,50	45,67	169,80	154,13	45,17	1,88
7	215,83	90,91	323,66	209,54	98,29	2,00
7	205,13	86,70	328,94	188,02	96,12	2,07
8	331,28	178,88	519,30	326,23	351,31	7,71
8	344,93	190,37	489,04	346,28	341,79	6,65
9	155,92	61,83	218,58	111,89	40,12	1,18
9	149,39	70,26	224,01	114,90	40,64	1,05
11	122,60	31,32	51,71	19,80	22,60	0,35
11	133,96	27,51	92,19	20,37	22,33	0,33
12	29,13	52,94	286,71	194,09	95,94	1,21
12	29,31	52,95	263,99	174,93	93,19	1,21
13	108,95	76,40	82,78	37,32	70,28	0,95
13	108,83	81,92	84,81	43,63	59,87	0,91
14	335,32	210,29	391,58	270,46	203,30	7,00
14	339,36	199,96	392,94	268,26	195,06	7,22
15	63,07	62,45	149,25	43,10	67,30	0,93
15	54,48	48,02	157,84	45,90	57,91	1,07
16	135,97	71,11	326,29	130,31	101,65	3,73
16	126,66	84,60	324,04	109,26	81,30	3,97
17	88,68	92,17	216,14	136,64	105,76	1,11
17	78,76	87,32	206,68	139,98	104,76	1,12
18	69,09	56,22	103,74	172,98	38,32	1,45
18	69,78	58,23	92,51	168,31	38,51	1,52
19	95,76	37,85	174,09	53,16	73,42	1,99
19	92,24	27,19	163,69	50,64	64,70	2,02
20	180,40	112,96	293,24	86,78	61,35	2,56

20	159,21	107,74	295,28	93,46	76,25	2,45
21	185,05	122,06	276,71	133,65	108,86	1,48
21	160,29	120,02	301,13	124,37	92,59	1,29
22	217,31	72,48	244,27	230,85	62,15	1,91
22	229,17	70,02	252,01	260,28	77,78	1,89
23	59,04	66,06	118,00	43,67	58,53	1,18
23	61,19	60,59	116,94	46,62	56,52	1,33
24	67,72	83,94	173,21	64,03	43,87	5,45
24	67,12	87,19	172,64	59,81	32,57	4,88
25	102,24	59,65	139,50	140,18	95,21	0,98
25	106,17	42,20	117,42	147,36	74,56	0,92
MEAN	137,88	76,31	212,16	134,69	80,96	2,30

## APPENDIX 4

### Nanodrop measurements

Values are in ng/ $\mu$ l.

Sample	QS	Q	S	PM
2	33,4	15	25	20
4	33,7	13	13	16
5	44	30	50	31
6	65	41	37	30
7	130	34	35	31
8	107	53	64	36
9	70	19	25	24
11	20	5	6	25
12	119	14	35	21
13	9	12	8	20
14	269	69	59	22
15	39	11	11	23
16	53	29	27	51
17	57	15	32	35
18	63	12	38	32
19	31	9	14	28
20	54	27	24	36
21	47	28	34	29
22	28	26	52	30
23	12	9	13	32
24	71	13	17	37
25	10	15	29	38

## APPENDIX 5

16S rRNA qPCR data

16S Bacterial qPCR results for A) Healthy group, and B) Polyp group.

A)



Healthy Group	Calc Conc (pg/μl) (1:100fort.)	H Adjusted to ng/μl
1	108,29	10,83
1	102,89	10,29
2	46,06	4,61
2	47,63	4,76
3	283,76	28,38
3	273,35	27,33
4	118,02	11,80
4	154,49	15,45
5	211,83	21,18
5	189,37	18,94
6	330,14	33,01
6	349,46	34,95
7	111,78	11,18
7	116,90	11,69
8	20,06	2,01
8	20,28	2,03
9	195,16	19,52
9	175,97	17,60
10	36,95	3,70
10	43,43	4,34
11	272,58	27,26
11	274,36	27,44
12	43,41	4,34
12	45,87	4,59
13	132,49	13,25
13	109,20	10,92
14	138,66	13,87
14	141,16	14,12
15	173,69	17,37
15	169,56	16,96
16	53,12	5,31
16	50,06	5,01
17	87,18	8,72
17	93,46	9,35
18	137,03	13,70
18	124,47	12,45
19	231,57	23,16
19	258,81	25,88
20	43,48	4,35
20	46,11	4,61
21	62,87	6,29
21	59,18	5,92
22	138,87	13,89
22	147,23	14,72
neg	0,01	
neg	0,01	
<b>AVG P</b>		135,69
<b>Adjusted (ng/μl)</b>		13,51

B)

Polyp Group	Calc Conc (pg/ul)(1:100fort)	P Adjusted to ng/μl
-------------	------------------------------	---------------------

P1	235,68	23,57
P1	296,68	29,67
P2	96,00	9,60
P2	108,60	10,86
P3	238,47	23,85
P3	259,52	25,95
P4	693,38	69,34
P4	724,22	72,42
P5	70,41	7,04
P5	68,96	6,90
P6	107,11	10,71
P6	111,42	11,14
P7	33,88	3,39
P7	33,52	3,35
P8	70,89	7,09
P8	82,40	8,24
P9	121,88	12,19
P9	129,61	12,96
P10	172,55	17,26
P10	186,17	18,62
P11	177,21	17,72
P11	171,24	17,12
P12	189,39	18,94
P12	196,90	19,69
P13	412,24	41,22
P13	417,86	41,79
P14	202,00	20,20
P14	205,69	20,57
P15	67,96	6,80
P15	67,22	6,72
P16	59,10	5,91
P16	61,53	6,15
P17	767,66	76,77
P17	805,75	80,58
P18	218,48	21,85
P18	229,14	22,91
P19	229,44	22,94
P19	227,78	22,78
P20	216,79	21,68
P20	231,46	23,15
P21	359,91	35,99
P21	320,21	32,02
P22	164,63	16,46
P22	164,04	16,40

P23	401,36	40,14
P23	380,02	38,00
P24	290,98	29,10
P24	330,08	33,01
NTC	0,08	
NTC	0,02	
AVG H	237,65	
<b>Adjusted</b>	23,77	

## APENDIX 6

### Quantifiler data

Results of the Quantifiler human DNA concentration quantification assay.

151119 Quantifiler Polyp Group cn		151119 Quantifiler Healthy Group cn	
Sample Name	Quantity	Sample Name	Quantity
10P	0,008902	10a	0,009068451
10P	0,020974	10b	1,73E-04
11P	0,015465	11a	0,014003683
11P	0,020636	11b	0,007040814
12P	0,030743	12a	0,003658771
12P	0,019135	12b	0,001869335
13P	0,000000	13a	0,001076927
13P	0,002328	13b	0
14P	0,000000	14a	0,01339813
14P	0,000000	14b	0,018343758
15P	0,070933	15a	0,001338923
15P	0,058073	15b	0,00643306
16P	0,006542	16a	0,002638021
16P	0,010479	16b	0,007382775
17P	0,245186	17a	8,98E-04
17P	0,287175	17b	0
18P	0,004505	18a	0,001574202
18P	0,000000	18b	9,09E-04
19P	0,052022	19a	0,008082174
19P	0,020937	19b	0,005690496
1P	0,051277	1a	0,00156664
1P	0,041343	1b	0,001095275
20P	0,074964	20a	0,003193641
20P	0,054562	20b	0,002436943
21P	0,195132	21a	5,16E-04
21P	0,161642	21b	0,001166485
22P	0,035541	22a	0,00246551
22P	0,040839	22b	0,001347774
23P	0,041654	2a	0,008603189
23P	0,046742	2b	6,34E-04
24P	0,588431	3a	0,05124875
24P	0,482380	3b	0,033924628
2P	0,035610	4a	0
2P	0,039282	4b	0
3P	0,211695	5a	0,010877541
3P	0,203450	5b	0,002791656
4P	0,029443	6a	0,007930675
4P	0,035578	6b	0,00221252
5P	0,413053	7a	0
5P	0,399134	7b	0,004783299
6P	0,060474	8a	0
6P	0,013065	8b	4,68E-04
7P	0,009277	9a	0,009858999
7P	0,000000	9b	0
8P	0,024341	<b>AVG</b>	<b>0,005697753</b>
8P	0,012689		
9P	0,079492		
9P	0,055706		
<b>AVG</b>	<b>0,089809</b>		

## APENDIX 7

### Nano drop values P group versus H group

The Nanodrop measurements of the healthy (H) group versus the polyp (P) group. Values are given in ng/ $\mu$ l.

Sample	H group	Pgroup
1	NA*	54
2	25	14
3	NA*	49
4	13	135
5	50	9
6	37	15
7	35	4
8	64	9
9	25	21
10	NA*	21
11	6	18
12	35	41
13	8	67
14	59	118
15	11	10
16	27	11
17	32	150
18	38	51
19	14	29
20	24	30
21	34	38
22	52	22
23	13	71
24	17	NA*
25	29	NA*

\*Not applicable; sample does not exist

## APENDIX 8

Patient/subject sample taking instruction sheet

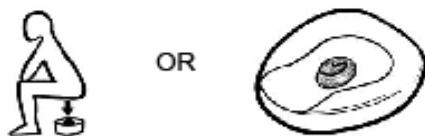
# Prøvetaking instruksjoner for kolorektal kreft prosjektet



1. Uriner først om nødvendig



2. Løft toalettsettet; legg ned plastfolje eller avispapir for å fange avføringen. Det er viktig at avføring ikke kommer i kontakt med urin eller vann.



3. Alternativt kan man bruke en plastboks eller en pott.



4. Tiltsett å skjeer avføring til hvert glass fra forskjellige områder.



5. Skru på korken godt.



6. Rist forsiktig.



7. Ikke skriv noe på rørene; alle prøvene skal være helt anonymt. Fyll ut samtykkeskjema.

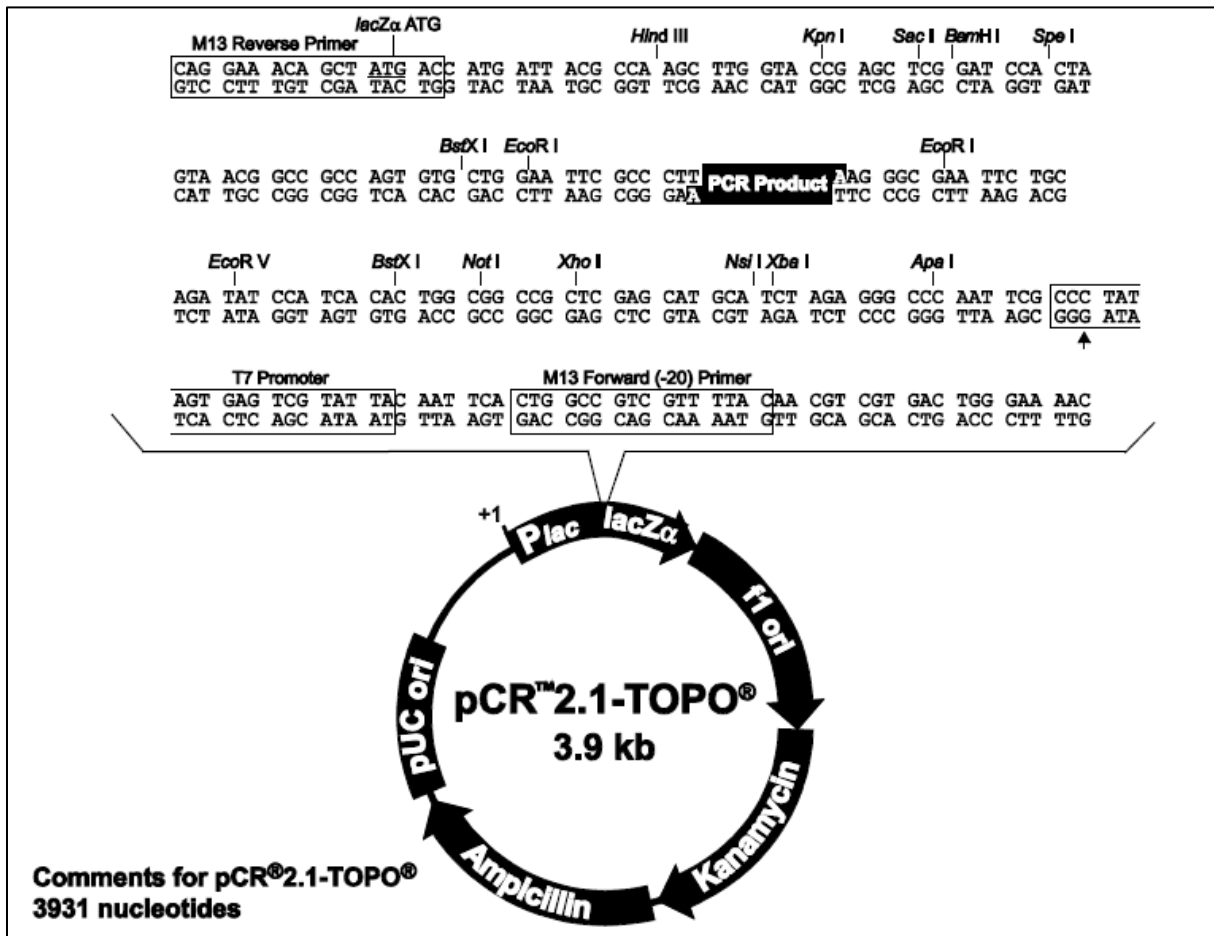


8. Ta prøven tilbake til laben. Prøvene skal settes i esken i den markerte skuffen i hovedlab 1, og skjema settes i konvolutten i samme skuff.

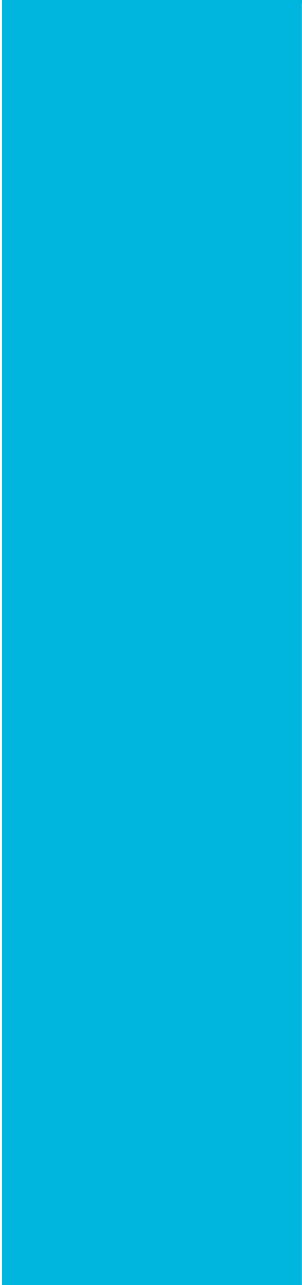
## APENDIX 9

Sequence, restriction map, and the location of the M13 primers of the pCR 2.1-Topo vector

Source [http://tools.invitrogen.com/content/sfs/manuals/topota\\_man.pdf](http://tools.invitrogen.com/content/sfs/manuals/topota_man.pdf)







Norges miljø- og biosciensbaserte universitet  
Norwegian University of Life Sciences  
NMBU

Postboks 8005  
N-1402 Ås  
Norway