Development of a gut microbiota diagnostic tool for pediatric inflammatory bowel disease based on GA-map™ technology platform

Dina Lilleseth Vangen

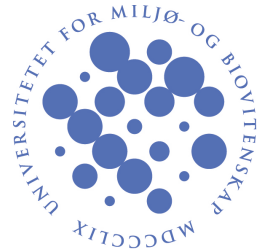
Norwegian University of Life Sciences

Department of Chemistry, Biotechnology and Food Science

Master Thesis 60 credits 2012

Development of a gut microbiota diagnostic tool for pediatric inflammatory bowel disease based on GA-map™ technology platform

DINA LILLESETH VANGEN



Norwegian University of Life Science

Department of Chemistry, Biotechnology and Food Science

Master Thesis

2011/2012

ABSTRACT

Inflammatory bowel disease (IBD) is an idiopathic, severe disease, which is characterized by chronic inflammation of the gastrointestinal tract. The incidence of IBD has increased through the last decades and specially among the pediatric population. The time from onset of symptoms to a final diagnose is made, is often related to delays and for many patients it is an emotionally demanding process. Early investigation in suspected cases may reduce the delay so that a treatment can begin as soon as possible. The involvement of intestinal microflora for pathogenesis of IBD is a link to further investigations to understand the disease, and to help people who suffer from IBD. The aim of the present work was to distinguish between pediatric IBD and non-IBD by identifying signatures in the microbiota. This was accomplished by use of a diagnostic tool based on GA-map^{TM} technology and the use of single nucleotide primer extension (SNuPE) probes to search for complementary bacterial 16S rRNA gene sequences. Seventy-four feces samples were collected from cohort and tested against 77 SNuPE probes. Statistical analysis was performed with Partial Least Squares – Discriminant Analysis and presented specificity by 82 % and sensitivity by 86 %. Classification error presented 16 % and indicated how many that was misclassified by the model. Inflammatory bowel disease is considered to include two major disorders where Crohn's disease is one of them, and best correlation was found between Crohn's disease and non-IBD through statistical analysis. Common bacteria that are normally shared among people with a healthy intestinal were obtained from the cohort, and as a conclusion seems an alteration of common bacteria to be involved in Crohn's disease.

TABLE OF CONTENT

1 PREFACE

1.1 Acknowledgements

I would like to express my gratitude to all who have supported me and contributed to the accomplishment of this thesis. I would like to thank my main supervisor Professor Knut Rudi (UMB, Genetic Analysis) and co-supervisor Caroline Jevanord Frøyland (MSc, Genetic Analysis) for excellent knowledge, kind supervision, ambitious plans and enthusiasm. Big thank to Akershus University Hospital by Morten H. Vatn for sharing of unique patient material – this thesis could not be completed without. I would like to thank Genetic Analysis by Morten Isaksen (PhD, Chief executive officer for Genetic Analysis) for giving me the opportunity to write this thesis. I would also like to thank the staff at Genetic Analysis for helping me and inspired me during the time I spent with them during my work. An especially thank to Monika Sekelja (PhD, Genetic Analysis) for her excellent statistical advices, patience and kind supervision, Selma Kreso (MSc, Genetic Analysis) for great helpful counseling in the laboratory work, and Heidi Vebø (PhD, Genetic Analysis) for excellent knowledge behind Luminex technology. Thank you all for sharing your broad knowledge within the field of GA-map™ technology, and I am so grateful for the support and patience Genetic Analysis have showed me during this thesis. I have enjoyed working with you all and I am forever thankful to have had the opportunity to perform this thesis. Finally, I would like to thank family and friends for support, love and joy.

Dina Lilleseth Vangen
Ås, May 2012

1.2 Glossary

Term	Explanation	Reference
Antigen	All substance or structures that is capable to be recognized of the immune system (e.g. bacterium or virus).	(Lea, 2006)
Biopsy	The removal and examination of tissue from the living body.	(Dorlands, 1981)
Colonoscopy	An elongated flexible endoscopy, which permits visual examination of the entire colon.	(Dorlands, 1981)
Commensal	An organism living on or within another, but not causing injury to the host.	(Dorlands, 1981)
Dysbiosis	A condition of microbial imbalance.	(Matarese & Kandil, 2012)
Endoscopy	Visual inspection of any cavity of the body by means of an endoscope.	(Dorlands, 1981)
Enteric	Pertaining to the small intestine.	(Dorlands, 1981)
Epithelium	The covering of internal and external surfaces of the body including the lining of vessels and other small cavities.	(Dorlands, 1981)
Idiopathic	Of the nature of an idiopathy; self-originated, of unknown causation.	(Dorlands, 1981)
Infection	Disease caused by pathogenic microorganisms.	(Granum, 2007)
Inflammation	A localized protective response elicited by injury or destruction of tissues, which serves to destroy, dilute, or wall of both the injurious agent and the injured tissue.	(Dorlands, 1981)
Luminal	Pertaining to the lumen of a tubular structure.	(Dorlands, 1981)
Mucosa	A mucous membrane.	(Dorlands, 1981)
NOD2/CARD15	CARD15, the gene product of the protein NOD2. Pattern recognition receptor, belonging to the human innate immunity.	(Joossens et al., 2011; Schreiber, P., Albrecht, Hampe, & Krawczak, 2005)
Pathogen	Any disease-producing microorganism.	(Dorlands, 1981)
Pathogenesis	The development of morbid conditions or of disease; more specifically the cellular events and reactions and other pathologic mechanisms occurring in the development of disease.	(Dorlands, 1981)
Pediatric population	Children up to 18 years.	(Perminow, Brackmann, et al., 2009)

1.3 Abbreviations

16S	16 Svedberg
Ahus	Akershus University Hospital
bp	Base pair
CARD15	Caspase recruitment domain family, member 15
CD	Crohn's disease
DNA	Deoxy Ribonucleic Acid
GA	Genetic Analysis AS
GI	Gastrointestinal
IBD	Inflammatory bowel disease
IBDU	Inflammatory bowel disease unclassified
IBSEN-II	Inflammatory Bowel South Eastern Norway-II
NOD2	Nucleotide binding oligomerization domain 2
Non-IBD	Patients not having IBD (systematic controls)
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PLS-DA	Partial Least Squares - Discriminant Analysis
rRNA	Ribosomal Ribo Nucleic Acid
SNuPE	Single nucleotide primer extension
T_m	Melting temperature
UC	Ulcerative colitis

1.4 Problem description and aim of the thesis

Crohn's disease (CD) and ulcerative colitis (UC) are known as the two main disorders for inflammatory bowel disease (IBD). IBD is characterized by relapse of chronic inflammation of the gastrointestinal (GI) tract (Baumgart & Carding, 2007; Mikhailov & Furner, 2009). The incidence among the pediatric population has increased during the last decades (Bousvaros et al., 2006; Escher et al., 2005), and debut during the first years of life is often associated with a widespread disease and presents severe and aggressive problems in the upcoming years (Nieuwenhuis & Escher, 2008; Perminow, Reikvam, et al., 2009). Why some people develop IBD is still an unanswered question, although it is fairly understood that there are complex interactions between immunological factors, bacterial factors and environmental triggers in genetic predisposed individuals (Baumgart & Carding, 2007; Bousvaros, et al., 2006; Mikhailov & Furner, 2009). The intestinal microflora is complicated and a complete survey is impossible. However, there are numerous studies of IBD that shows the involvement of intestinal microflora in the pathogenesis of IBD (Bousvaros, et al., 2006) and further investigations are important.

One fifth goes to their doctor with intestinal problems (Knobel, 2011) but IBD is not the cause for all intestinal problems that occurs. There are also several cases where a definitive diagnosis between CD and UC initially could not be completed and where re-evaluation of diagnosis are needed (Mikhailov & Furner, 2009; Ravikumara & Sandhu, 2006). To distinguish between IBD and non-IBD are in many situations a long and comprehensive task, and the disease may have evolved during the time, before the disease is revealed. Based on these observations there is need to develop a diagnostic tool that reveals signs in an early stage of the disease course and to preclude or confirm IBD.

By investigations performed by Genetic Analysis, there was found an overview of GI bacteria as potential biologic markers in diagnostics of IBD (Frøyland, 2010), and an evaluation where children feces samples highlighted the biodiversity of the gut microbiota (Nwosu, 2011). It is now feasible to take the next step to develop a diagnostic tool that reveals signs in the pathogenesis of pediatric IBD. The aim for this thesis was to distinguish between children with IBD and non-IBD by identifying signatures in the microbiota. This was accomplished by studying bacteria directed against core bacteria among healthy people, intestinal bacteria for infant, and bacteria correlated to IBD. A diagnostic tool related to bacteria correlating to IBD

was to be designed in this thesis, to see if this could reveal more bacteria related to IBD compared to core and infant bacteria. This was to be performed by the use of a diagnostic tool based on GA-map^{TM} technology and the use of single nucleotide primer extension (SNuPE) probes to search for complementary bacterial 16S rRNA gene sequences. By identifying signatures for bacteria that correlated for IBD in an early stage, this could be an opportunity to reveal a sign for IBD, and to further start a treatment as soon as possible to prevent the development of the disease.

2 THE HUMAN GUT MICROBIOTA

The majority of microbes reside in the gut have an influence on human physiology, nutrition and well-being. To understand the impact of gut microbes on human health it is necessary to decipher the content, diversity and function of the microbial gut community (Qin et al., 2010). The human gastrointestinal (GI) tract is essential for human life because of its involvement for supplying the human body with nutrients and energy sources through absorption of food (Zoetendal, Rajilic-Stojanovic, & Vos, 2008). Moreover, is microbes in the GI responsible for polysaccharide hydrolysis and fermentation, vitamin production, immune system stimulation, modulation of gut motility, and protection of human host from pathogen invasion (Agans et al., 2011).

The intestinal microflora constitute a complex ecosystem with four bacterial divisions who dominate (represents more than 99 %), these are Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria (Sartor, 2008). 10^{14} bacteria of more than 1,000 different species and subspecies colonize the mucosal surface of the human GI tract (Cucchiara, Iebba, Conte, & Schippa, 2009) and the task to survey the whole gut seems almost impossible. A stable bacterial flora is first established after the child has turned 3 years (Baumgart & Carding, 2007; Cucchiara, et al., 2009; Palmer, Bik, DiGiulio, Relman, & Brown, 2007) but the microflora of adolescent children seems to be different from that of adults (Agans, et al., 2011). An establishment is influenced by microbial and host interactions and by external and internal factors (Cucchiara, et al., 2009; Fanaro, Chierici, Guerrini, & Vigi, 2003) and can therefore be quite individually. Despite individual differences, humans share a common core of microbes (Zoetendal, et al., 2008). The given total number of microbes over the GI tract is hard to imagine. A complex community and individual specific composition among people, complicates establishing links between the microbiota and intestinal disorders that can occur in the GI tract (Zoetendal, et al., 2008).

There are numerous studies of inflammatory bowel disease (IBD) that confirms the involvement of intestinal microflora in the pathogenesis of IBD (Bousvaros, et al., 2006), but the pathogenesis remains poorly understood (Schwiertz et al., 2010; Shikhare & Kugathasan, 2010). Whether the observed dysbiosis is either a cause or a consequence of the disease remains unclear (Baumgart & Carding, 2007; Joossens, et al., 2011), even though much point

in the direction that an altered configuration of the microflora may be a consequence rather than a cause (Sartor, 2008, 2011).

3 INFLAMMATORY BOWEL DISEASE

Crohn's disease (CD) and ulcerative colitis (UC) are known as the two main disorders of inflammatory bowel disease. Why some people develop IBD is still an unanswered question, although it is fairly understood that there are complex interactions of various factors contributing to the disease (Baumgart & Carding, 2007; Bousvaros, et al., 2006; Mikhailov & Furner, 2009). Through the last decades there have been observed an increased incidence of IBD among the pediatric population (Bousvaros, et al., 2006; Escher, et al., 2005) and several epidemiological studies reports that particular CD has increased (Escher, et al., 2005). Debut during the early years of life is associated with a widespread disease and more aggressive and severe problems in the upcoming years (Nieuwenhuis & Escher, 2008; Perminow, Reikvam, et al., 2009). Children have fewer environmental cofounders and are considered as a more "clean" population, which is an advantage in terms of revealing the disease. Tracking IBD in the early years of life is therefor important, to reveal the natural history of the disease (Bousvaros, et al., 2006; Hait, Bousvaros, & Grand, 2005).

3.1 Epidemiology

Studying epidemiology can be important to get closer to a solution to why IBD develops in terms of where it occurs, whom the disease affects and when it occurs (Bousvaros, et al., 2006). Highest incidences of IBD are reported from developed countries (Baumgart & Carding, 2007; Ravikumara & Sandhu, 2006; Russel, 2000), and children with IBD accounts for nearly 30 % of the total cases (Escher, et al., 2005; Perminow et al., 2010; Ravikumara & Sandhu, 2006). However, the thought of a difference between developed and developing countries seems to be less than expected, based on results from recent studies (Bousvaros, et al., 2006; Ravikumara & Sandhu, 2006; Shikhare & Kugathasan, 2010). Conditions such as access to health care, genetic background, increased diagnostic accuracy due to advent of improved technology, increased awareness, and unavailable data due to limited reported cases, may be some of the reasons why differences occurs in different parts of the world (Baumgart & Carding, 2007; Ravikumara & Sandhu, 2006; Russel, 2000). In Norway, the total numbers of IBD patients are about 24,000. With a fairly realistic distribution, the

numbers of CD and UC cases are set to be 7,200 and 17,000, respectively. These numbers are based on statistic from Statistics Norway, and are a minimum number of cases. No separate statistics by age are worked out, but Statistics Norway is establishing an IBD registry in Norway that is handled by professional expertise, and a more complete list will be established during the upcoming years (Schatten, 2012).

The onset of IBD can occur at any age (Biank, Broeckel, & Kugathasan, 2007; Bousvaros, et al., 2006; Khor, Gardet, & Xavier, 2011) but the peak of onset is in adolescent and present before the age of 20 years in 25 % of all IBD patients (Escher, et al., 2005; Shikhare & Kugathasan, 2010). It is therefor often conflicted with the pubertal growth and development, which does not make the disease easier to handle.

3.2 Clinical presentation of pediatric IBD

To survey IBD from an early stage of the disease course can be an advantageous, to be able to start a treatment as soon as possible and further get control over the disease. Unfortunately in many pediatric (< 18 years) cases, the classical symptoms of IBD are absent and difficult to recognize and can in addition be reflected in a misinterpretation of pubertal development (Büller, 1997). UC and CD are different illnesses but are grouped together as IBD because they produce similar signs and symptoms (Shikhare & Kugathasan, 2010). In some situations there are difficult to see a clear distinction between the two diseases at their onset (Mikhailov & Furner, 2009; Ravikumara & Sandhu, 2006).

Initial symptoms for children are often abdominal pain, loss of appetite, psychological stress, pubertal delay and impaired growth (Bousvaros, et al., 2006; Büller, 1997; Nieuwenhuis & Escher, 2008). Some of these circumstances can be present for a healthy, young person, e.g. the variation in growth velocity that is often considered to be part of ongoing puberty rather than a symptom of an ongoing disease (Büller, 1997). 36-88 % of children with IBD have impaired growth caused by multiple factors related to nutrition and the immune system, but the precise mechanisms by which they act are not fully understood (Bousvaros, et al., 2006).

Ulcerative colitis is a relapsing disease and includes diffuse mucosal inflammation, extending from the rectum to the left or right colon (Figure 1) (Baumgart & Sandborn, 2007; Shikhare &

Kugathasan, 2010). Key features of UC include bloody diarrhea, abdominal pain, fever, and passage of pus and/or mucus (Sawczenko & Sandhu, 2003).

Crohn's disease is also a relapsing disease but can affect the entire gastrointestinal tract (Figure 1) from mouth to anus, normally patchy. Depending on the clinical presentation and disease location, patients with CD have diarrhea, abdominal pain, weight loss, fever, clinical signs of bowel obstruction, passage of blood and/or mucus, and persistent vomiting (Baumgart & Sandborn, 2007; Shikhare & Kugathasan, 2010). The three former symptoms are most common presenting of the disease but are not always presenting, and only 25 % of pediatric CD presents with the classic triad (Nieuwenhuis & Escher, 2008; Sawczenko & Sandhu, 2003). Relapse may be infrequent and occur once during one year but relapse can also be continuous with persistent symptoms without a period of remission (Stange et al., 2006).

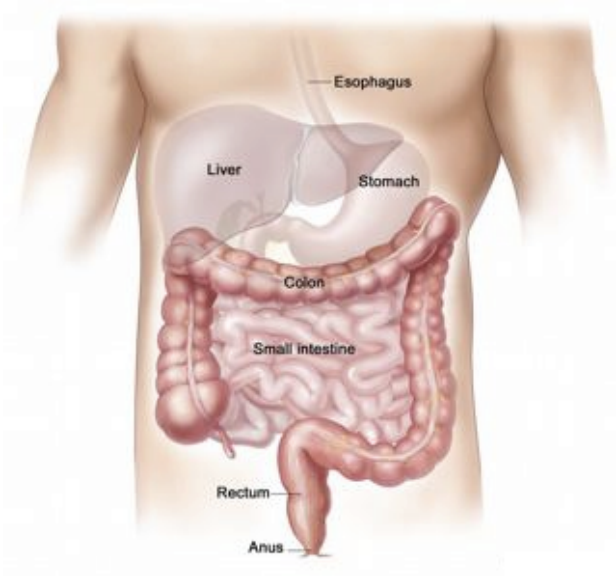


Figure 1: Illustration of the gastrointestinal tract ("progressivehealth," 2012). UC occurs from rectum to the left or right colon. CD can affect the entire gastrointestinal tract from mouth to anus.

It is important to note that there exists no hard scientific evidence about the differing etiology from child to adult but pediatric onsets differ from adult in many aspects. However, it is thought that the early onset of the disease represents a more aggressive phenotype than similar diseases in individuals older than 20 years (Biank, et al., 2007; Perminow, Reikvam, et al., 2009). Whether the disease phenotype is a result from the difference in investigation for diagnose or represents a true phenotype difference, is uncertain (Biank, et al., 2007).

3.3 Development of Inflammatory Bowel Disease

Inflammatory bowel disease is of unidentified origin, but different environmental factors and the intestinal flora can initiate immunological mechanisms in genetically predisposed individuals and cause IBD (Figure 2) (Baumgart & Carding, 2007; Mikhailov & Furner, 2009; Russel, 2000).

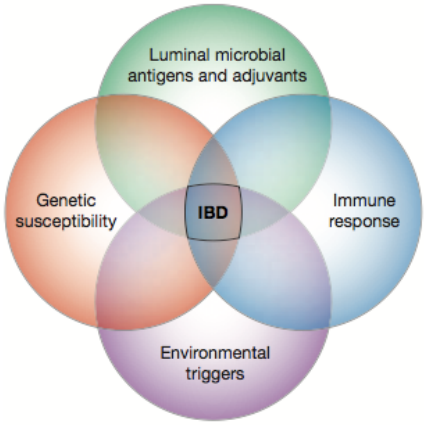


Figure 2: Interacting factors contributing to IBD are intestinal flora, environmental triggers and the immune system in genetically predisposed individuals (Sartor, 2006).

3.3.1 Environmental factors

Incidence of IBD has increased over the past decades (Ravikumara & Sandhu, 2006; Russel, 2000) and especially among children (Escher, et al., 2005; Mikhailov & Furner, 2009). Rapid changes in the incidence can best be explained by changes in environmental factors since changes in genetic factors do not occur as rapidly as the incidence of IBD has been noted (Koloski, Bret, & Radford-Smith, 2008; Mikhailov & Furner, 2009; Russel, 2000). There have been reported several studies where connection between environmental factors and IBD was questioned. Breastfeeding, passive smoking, nutrition, and birth weight are some of the investigated areas (Mikhailov & Furner, 2009), but due to conflicting results, nothing can be stated yet (Koloski, et al., 2008; Mikhailov & Furner, 2009). Hygiene hypothesis have been posted as an explanation for the high frequency of IBD during the last decades where the hygiene has been improved. The thought of a child's overprotection from exposure of common infections agents, and thereby omit developing of the immune system to later exposed infectious, may be a factor contributing to the pathogenesis of IBD. More

investigation to verify the hygiene hypothesis is needed (Koloski, et al., 2008; Sartor, 2008; Shikhare & Kugathasan, 2010).

3.3.2 Genetic factors

Genetic influence is considered to be an element for both UC and CD but is regarded to be strongest for CD (Baumgart & Carding, 2007; Joossens, et al., 2011; Mikhailov & Furner, 2009). Based on the relatively short exposure time to environmental factors concerning children, is genetic influence thought to play a larger role for how the onset of the disease occur. Numerous family and twin studies and the discovery of many susceptibility genes strongly support the concept that IBD is highly heritable (Baumgart & Carding, 2007; Biank, et al., 2007; Escher, et al., 2005; Khor, et al., 2011; Shikhare & Kugathasan, 2010). However, as previously mentioned, genetic influence seems most likely not to be the strongest explanation for incidence of IBD because of the rapid changes during the last decades (Koloski, et al., 2008; Mikhailov & Furner, 2009; Russel, 2000). The first gene identified in CD is the NOD2/CARD15 gene that plays a role in bacterial sensing and defense against organisms (Joossens, et al., 2011). There have been found strong associations for mutation of NOD2/CARD15 gene, but is neither sufficient nor necessary for developing the disease, and up to 70 % of CD patients do not carry a NOD2/CARD15 mutation (Biank, et al., 2007).

3.3.3 Immunobiology

It is believed that inflammatory events originate from an abnormal activation of the mucosal immune system and a defect immune response (Khor, et al., 2011). The first line defense mechanism of the mucosal immune system is the epithelial barrier where the intestinal epithelium is a single layer covered by mucus in which commensal microbes are embedded. People suffering from IBD have their epithelial barrier and antigen recognition disturbed, and causes lowered epithelial resistance (Baumgart & Carding, 2007). Whether IBD is induced by a pathogen – a transient infection trigger what initiate IBD and then perpetuated in susceptible hosts by resident (not pathogenic) commensal luminal bacteria, or by an abnormal immune response to normal enteric microflora, are still unresolved questions (Sartor, 1997). It has also been few studies dealing with the immunology of pediatric IBD and as a result, survey of the early stages of the development of IBD has been missing (Bousvaros, et al., 2006).

3.3.4 Pathogenesis

The pathogenesis of IBD remains poorly understood (Schwiertz, et al., 2010; Shikhare & Kugathasan, 2010), and to date, no specific microorganism is being directly associated with the pathogenesis (Marteau et al., 2004). In the course of time, there have been several bacteria identified among the pediatric IBD as an own distinct group. *Bifidobacterium* cell counts lower number in patients with active IBD compared with healthy controls (Schwiertz, et al., 2010), and decreased number of *Faecalibacterium prausnitzii*, a predominant species of the Clostridia cluster IV, is characterized by CD pediatric (Joossens, et al., 2011; Schwiertz, et al., 2010). An increased number of *E. coli* has been reported in pediatric IBD and especially for CD (Cucchiara, et al., 2009; Schwiertz, et al., 2010). There must be noted that most light has been shed on the pathogenesis role for CD and most bacteria are reported from this disease (Joossens, et al., 2011; Mondot et al., 2011; Schwiertz, et al., 2010).

3.4 Diagnostic procedures

Disease activity and propagation is essential for therapy selection. The main goal is to get control over the inflammatory process and for the long-term; prevent relapses, limit disease complication, and improve quality of life for the patient (Baumgart & Sandborn, 2007; Shikhare & Kugathasan, 2010). An IBD diagnose is based on clinical signs and symptoms, laboratory manifestations, radiographic, endoscopic, and histological studies (Baumgart & Sandborn, 2007; Escher, et al., 2005; Mikhailov & Furner, 2009) and is a comprehensive process. The period to a diagnosis is confirmed is often conflicted with time and can be a demanding process for the patient, and early investigation in suspected cases can reduce delays (Sawczenko & Sandhu, 2003).

During the 1960's and 70's, the endoscopic investigation was introduced for investigations in the gastrointestinal tract to be able to reveal cases of IBD. Eventually, methods were further developed and from the late 1980's and the early 1990's total colonoscopy was performed for IBD patients. For the pediatric population, the investigations were normally performed under general anesthesia up to the age of 16 but a strong indication for the disease had to be present. In later years, magnetic resonance imaging (MRI) was suggested in order to determine the extent of disease with a good accuracy (Perminow, 2010).

Mucosal biopsies from different segment of the colon and the ileum are normally accomplished to influence the accuracy of the histological diagnosis, and are collected during upper and lower endoscopy. Analyses of a full colonoscopic biopsy are performed for the initial diagnosis and further to confirm the diagnosis. Blood tests are investigated for a full blood count in patients with CD. The C-reactive protein is an indication of the acute phase response to inflammation and is correlated with disease activity (Stange, et al., 2006).

4 TECHNIQUES FOR EXPLORING GUT MICROBIOTA

The human gastrointestinal (GI) tract microbiota is vital for human health and is also implicated in several GI disorders (Qin, et al., 2010). Investigation and more attention are therefor important and several techniques have been used to explore the diversity in the bacterial community. This can lead to the discovery of biomarkers to understand and predict the microbial life in the human GI tract (Zoetendal, et al., 2008). Despite several breakthroughs on the composition and functionality, no information has yet been transformed into diagnostic procedures to reveal different diseases. The largest challenge in the gut microbiota diagnostics is the establishment of a correlation between microbiota patterns and disease. There are not many diseases that have been characterized with respect to specific dysbiosis of the microbiota, and this can be a challenge since a survey of the gut microbiota needs to further be explored to get more knowledge over the gut microbiota (Rudi & Isaksen, 2012).

4.1 Cultivation analysis

The gut microbiota is a complex system and requires technology that can handle and construe the content. Traditional cultivation based analysis cannot be used to reveal signatures for the gut microbiota because most of the growth conditions for gut bacteria are unknown (Rudi & Isaksen, 2012). The myriad of bacteria in the human gut makes cultivation analysis not so relevant because the analysis comes to brevity. A challenge is also related to growth conditions since many of the bacteria in the human gut are strictly anaerobic. Many GI tract microbes also develops intimate relation with the host and with each other, which makes microbes dependent of the metabolic activity of another member of the ecosystem. Based on important features that is essential for the gut microbes can it be almost impossible to perform cultivation analysis (Zoetendal, et al., 2008).

4.2 High-throughput sequencing

High-throughput sequencing is a synonymous term with next-generation sequencing. Next-generation sequencing can be used to open entirely new areas of biological inquiry and characterization of ecological diversity, which can be useful to get more understanding over the gut microbiota. The technology allows thousands of sequence reactions to performed in parallel. Primers are covalently attached to the template on solid surface through hybridization, and further DNA polymerase is bounded to the primer template configuration to initiate the next-generation sequencing reaction. DNA polymerase adds fluorescently modified nucleotides, which represent the complement of the template base, and can further be detected (Metzker, 2010). The major challenge by unprecedented amount of sequencing is the methods for data storages, transfer, and data analysis (Xiong, Zhao, Arnold, & Yu, 2010). It is also expected to take time and further development before sequencing of the gut microbiota can become a routine diagnostic tool (Rudi & Isaksen, 2012).

4.3 DNA microarray

The most commonly used high-throughput analytical method for the GI tract is DNA microarray (Zoetendal, et al., 2008). Microarray is an effectively powerful tool for its application for the analysis of intestinal samples, to provide novel insights into the relationship between microbiota in the human gut. Microarray encompassing over thousands of known sequences immobilized on a microscope slide and can further be subjected to a series of hybridization experiments that are performed simultaneously. The slide is hybridized with different labeled fluorescent probes and allows indication whether a particular gene is significantly expressed (Watson et al., 2008). This technology can target genes to measure abundance and diversity of the intestinal microbiota and to further reveal some signatures.

Phylogenetic microarrays is used for characterization of GI tract microbiota to gain insight into the structure and population (Zoetendal, et al., 2008). Phylogenetic microarrays contain probes derived from rRNA sequence information and are suited for analysis of the microbial community, as GI tract. The oligonucleotide probes can be designed in a phylogenetic framework to survey different levels of sequences from broad taxonomic level to grouping at genus and species level. One critical issue by the use of microarrays is detection specificity, because rRNA genes are highly conserved and present in all microorganisms (Zhou, 2003).

Another limitation by the use of phylogenetic microarrays is the dependency on the isolation of nucleic acids and subsequent polymerase chain reaction (PCR) amplification of the gene. However, this is generally a drawback of culture-independent technologies. Phylogenetic microarrays have a dynamic range that only covers the dominant microbes present in the GI tract, and can be another limitation by the use of this analysis (Zoetendal, et al., 2008).

4.4 Probe target 16S rRNA

The most widely applied targeted approach to describe the human gut microbiota is the use of probes targeting 16S rRNA gene (Rudi & Isaksen, 2012). A probe is a short (about 14-25 bp) unlabeled oligonucleotide that is labeled if the target bacterium (the probe identifies) is present in a reaction (Vebø et al., 2011). One important feature for 16S rRNA gene is that this gene is present in all bacteria and can be used as a universal target for bacterial identification. Throughout evolution, small changes have been seen for 16S rRNA gene and the gene is also large enough (1,500 bp) to contain statistically relevant sequence information (Patel, 2001).

4.5 SNuPE probe

For a more direct approach to get a signature of the gut microbiota is the use of highly specific single nucleotide primer extension (SNuPE) probes. The SNuPE probes are constructed so that the probes hybridize adjacent to discriminative gene positions (16S rRNA gene) and is labeled with fluorescence by a DNA polymerase if the target bacterium is present (Vebø, et al., 2011). Genetic Analysis AS (GA) is a Norwegian company commercialized GA-map™ (Genetic Analysis microbiota array platform) array technology within the field of molecular diagnostics of diseases related to gut imbalances ("Genetic Analysis," 2012). GA-map™ assay technology is based on use of highly specific SNuPE probes for target/non-target discrimination (Vebø, et al., 2011). Moreover, it can be performed on feces samples to get a profile over the patient's gut, which can be related to various health conditions and diseases.

4.6 GA-map™ assay

Genetic Analysis holds two GA-map™ assays that are not currently commercially available on the market. The first is GA-map™ Core assay and is based and constructed on probes established from microbial genomes that are shared from a healthy gut based on a study

performed in Europe. One hundred and twenty four individuals were collected in the study to determine a healthy microbial core to get a better understanding over the common microbes that have a profound influence on human physiology and are crucial for human life (Qin, et al., 2010). The second is GA-map™ Infant assay and is designed to compare the temporal development of the gut microbiota in IgE-sensitized and non-sensitized children during the first two years of life. The assay represents the minimum number of probes that covers the expected diversity of bacteria in the infant gut (Vebø, et al., 2011).

GA first started to perform analyzes on an array but has until recently switched to the use of assay with beads. The reason for the change was due to a time consuming process and less control of the samples, since the array was ordered and not produced by GA self. With current technology and the use of assay and beads, it is more likely that the SNuPE probe will capture the 16S rRNA in a reaction, instead of an immobilized probe on an array where the target-bacterium had to catch one spot (Sekelja, 2012).

5 THE 16S rRNA GENE AS A PHYLOGENETIC MARKER

16S rRNA gene is present in all bacteria and can be used as a universal target for bacterial identification (Patel, 2001). 16S rRNA gene comprises of highly conserved regions and other hyper-variable regions. The conserved regions are important for classification of higher taxa, while the hyper-variable regions can be used for differentiation between closely related species (Rudi, Zimonja, Trosvik, & Næs, 2007).

Species that share a common ancestor should be classified closer together than distantly related species (Fletcher, Hickey, & Winter, 2007) in an evolutionary connection. What is inferred from sequence data is an estimate of the phylogeny of a given gene but even without systematic errors there is intrinsic random errors in every phylogenetic tree. The best phylogenetic tree for a given molecule is not necessarily the correct phylogenetic tree for either the gene or the organism since the result is an inevitable consequence of the finite body of data used in the inference (Olsen & Woese, 1993).

The gut microbiome is extremely complicated (Bousvaros, et al., 2006) and classification and phylogenetic studies can both be incorrect and hard to perform. Functionality of bacteria in the gut correlated to phylogeny is still an ongoing issue. Much point in the direction that

functions cannot directly be inferred from the gut microbiota 16S rRNA gene analyses. However, phylogenetic framework can be used to deduce the probability of functions and establish correlations in the gut microbiota (Rudi & Isaksen, 2012).

6 LUMINEX

The company behind Luminex develops, manufactures and markets biological testing technologies with applications throughout the clinical diagnostic and life science industries. To be able to conduct biological tests (bioassays), the company uses technologies such as xMAP® and xTAG® ("Multiplexed Solutions for Life - Luminex Corporation," 2012). The original instruments behind xMAP technology are Luminex® 100/200™ and FLEXMAP 3D® hardware, based on the principle of flow cytometry, laser and Photo Multiplying Tubes.

MAGPIX® is a more recently introduced instrument and is based on fluorescent images where LED (light emitting diodes) and CCD (charge coupled device) camera analyses the samples. MAGPIX performs assays by photograph MagPlex microspheres (beads) and distinguishes 1 to 50 unique magnetic microspheres in a single sample. Microspheres is impregnated with dye mixture of red to infrared which makes it possible to identify each microsphere particle ("Multiplexed Solutions for Life - Luminex Corporation," 2012).

7 MATERIALS AND METHODS

7.1 Outline of the materials and methods used in this thesis

This thesis was worked out as collaboration between Norwegian University of Life Sciences and Genetic Analysis AS (GA) and started in September 2011. GA has established a collaboration with Akershus University Hospital (Ahus), which has given GA access to unique feces material of people with IBD and non-IBD. A flow chart (Figure 3) is presented for a better understanding of the process through the present study. The first step was to develop GA-map™ IBD assay, which included design of probes. Through second step, the patient samples were tested on three different diagnostic tools before they were analyzed on Magpix instrument.

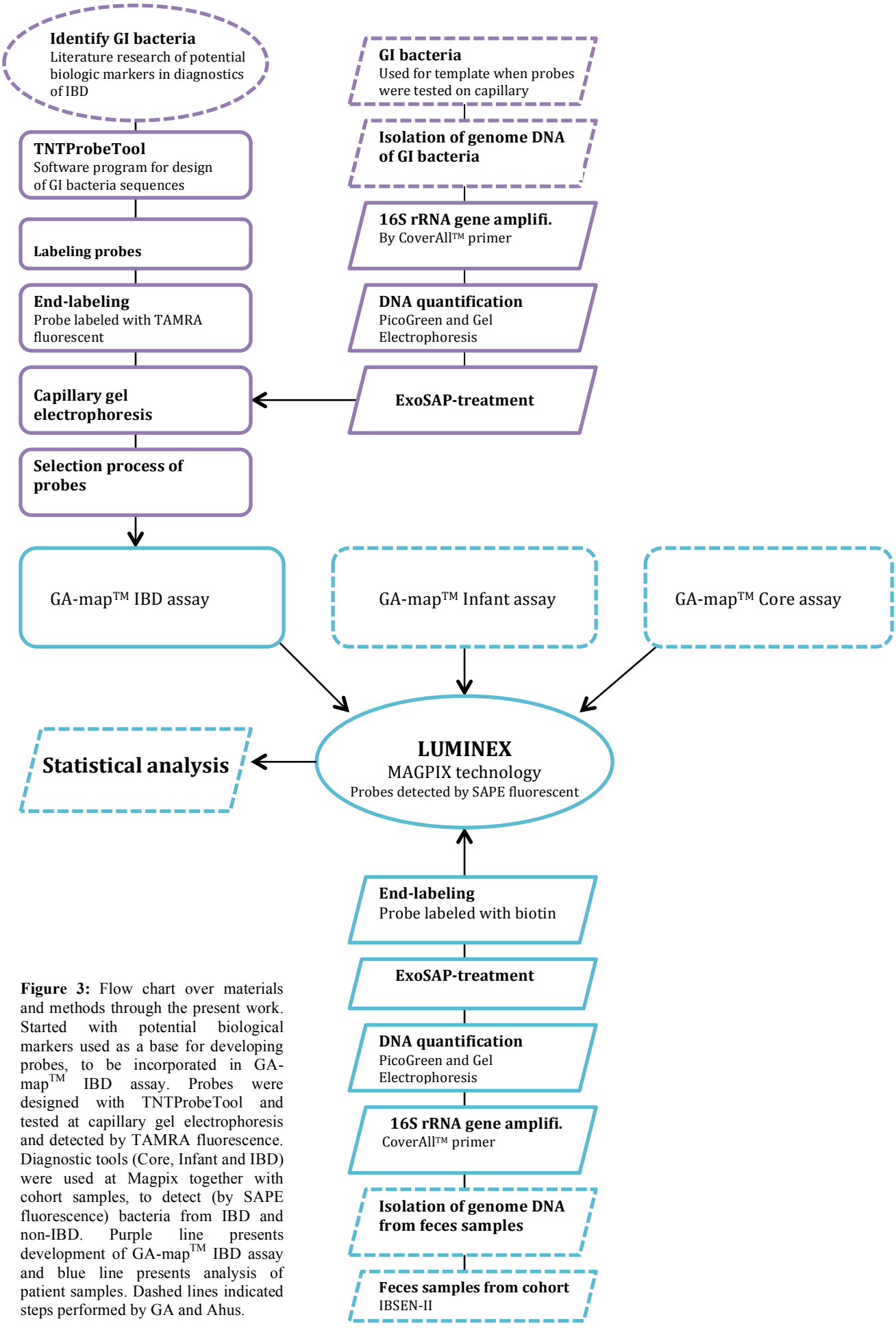


Figure 3: Flow chart over materials and methods through the present work. Started with potential biological markers used as a base for developing probes, to be incorporated in GA-map™ IBD assay. Probes were designed with TNTProbeTool and tested at capillary gel electrophoresis and detected by TAMRA fluorescence. Diagnostic tools (Core, Infant and IBD) were used at Magpix together with cohort samples, to detect (by SAPE fluorescence) bacteria from IBD and non-IBD. Purple line presents development of GA-map™ IBD assay and blue line presents analysis of patient samples. Dashed lines indicated steps performed by GA and Ahus.

7.2 Development of GA-map™ IBD assay

7.2.1 Identify gastro intestinal bacteria suitable for IBD diagnostics

Prior to this thesis there have been performed literature research by Genetic Analysis to identify gastro intestinal bacteria suitable for inflammatory bowel disease (IBD) diagnostics. By classifying amounts of bacteria from fecal and mucosa as increased, equal or decreased relative to controls (Frøyland, 2010) there was accomplished a list with 157 genus and species names to be included for GA-map™ IBD diagnostic. This list was the basis for the probe search performed in present work.

7.2.2 Design of probe in TNTProbeTool

TNTProbeTool version 1.0 software program (made by GA) was used for design of a probe sequence. The names of different species and genus bacteria found through literature research were the basis for the probe search. The user defined target bacterium among target group and resulted all other bacteria to be non-target. TNTProbeTool used 8-nucleotide as the length of matching area for probe search. This implied a start criterion for areas at target sequence and was not found among non-target bacteria. The matching area was the base for the probe but the 8-nucleotide area was extended to fulfill demanded melting temperature (T_m) before a final probe suggestion could be designed. The definition of T_m is the temperature (°C) at which 50 % of the oligonucleotide and its perfect complement is in duplex ("www.sigmaaldrich.com," 1998). The T_m was essential for probe design to succeed a good hybridization and several factors were included, for instance “nearest-neighbor” (Sekelja, 2012). More information behind TNTProbeTool is noted in Appendix C.

Probes were designed in TNTProbeTool with a minimum T_m of 55 °C by the nearest neighbor for the target group, and T_m < 30°C or absence of a cytosine for non-target group. After a probe was designed, it was tested for duplication (performed by GA). A duplicated probe is the existence of a copy that already is presented in GA library. Duplicated probes will not be deleted from the study but will be assigned the originally probe identification from the existing probe. Oligonucleotides (probe material) were ordered through Thermo Scientific, Germany. Arrived probes from Thermo Scientific were stored at 4 °C before they were diluted to 100 pmol/µl (stock sample) and 1 µM (working solution) and later stored at –20 °C.

7.2.3 Preparation of template

Isolated DNA was amplified by CoverAll™ primer and ExoSAP-treated. DNA quantification was performed with Gel Electrophoresis and PicoGreen.

7.2.3.1 16S rRNA amplification by the CoverAll™ primer

All bacterial 16S rRNA gene contains conserved region on the 5' and 3' ends. The CoverAll™ takes advantages of the two flanking conserved regions by using primers (forward and reverse) that anneal to this area of the DNA. 16S rRNA gene amplification by the CoverAll™ primer was performed as follows: 0.05 U HOT FIREPol® DNA Polymerase (Solis Biodyne), 1 x HOT FIREPol® buffer2 (Solis Biodyne), 2.5 mM Magnesium-dichloride (Solis Biodyne), 200 μM deoxynucleotide triphosphate (dNTP) (Applied Biosystems), 0.2 μM Mangala forward primer (Genetic Analysis), 0.2 μM 16S 1015U reverse primer (Genetic Analysis) and 5-100 ng bacterium template in a total reaction volume of 50 μl. 0.2 μM positive (*E. coli*) and negative controls (no template) were added. The amplification by the CoverAll™ primers included a 15 min initial denaturation stage at 95 °C, followed by 30 cycles with 30 sec denaturation at 95 °C, 30 sec annealing at 55 °C, and 80 sec elongation at 72 °C. The final elongation was performed at 72 °C in 7 min. Amplification by CoverAll™ primer was accomplished on Veriti 96 well Thermal Cycler (Applied Biosystems) (Nestestog, 2010a).

7.2.3.2 Gel Electrophoresis

Gel electrophoresis was used to visualize amplificated PCR product of bacterial 16S rRNA gene. Gel electrophoresis DNA quantification was performed as follows: 1.5% agarose powder (SERVA), Red Gel Nucleic Acid Strain (BIOTIUM), 1X TBE buffer, 6X Gel Loading Dye (New England BioLabs), 1-Kb DNA Ladder (New England Biolabs) and amplified 16S rRNA gene product. Positive and negative controls were added. Run gel: 100 V, 50 mA, in 40 min and visualized band by UV light used INGENIUS (SYNGENE BIO IMAGING) and GeneSnap Gel Analysis Software (SYNGENE) for image (Nestestog, 2010b). Results from gel electrophoresis are not shown in this thesis.

7.2.3.3 PicoGreen

PicoGreen was used for detection and quantitate amounts of DNA (in ng/µl). Quant-iT™ PicoGreen dsDNA reagent is an ultrasensitive fluorescent nucleic acid strain for quantitating double-stranded DNA (dsDNA). Concentration was measured with absorbance at 260 nm (A_{260}). PicoGreen DNA quantification was performed as follows: Quant-iT™ PicoGreen dsDNA (Invitrogen™), 20X TE buffer (Invitrogen™), Lambda DNA Standard (Invitrogen™) and amplified 16S rRNA gene product. Non-amplified master mix (NAMM) (Genetic Analysis) was included to serve as a control for PicoGreen measurement. Lambda was used for DNA standard curve where fluorescence intensity vs. DNA concentration was plotted so quantitative amount of DNA in samples could be measured. PicoGreen was performed with FLUOstar OPTIMA (BMG LABTECH) and OPTIMA software version 2.20R2 (BMG LABTECH) (Nestestog, 2010c). Referred to Invitrogen protocol for more detailed information behind PicoGreen (Invitrogen, 2008). Results from PicoGreen measurements are not shown in this thesis.

7.2.3.4 Exonuclease I (Exo I) and Shrimp Alkaline Phosphatase (SAP) –treatment

Exo I removed nucleotides from single-stranded DNA in 3' to 5' direction from a reaction mixture after PCR. SAP removed phosphate groups from the 5' end of DNA and hence dephosphorylated excess deoxynucleotide triphosphates (dNTP) remaining after PCR. PCR product was ExoSAP-treated with 3 U Exonuclease I (BioLabs®) and 8 U Shrimp Alkaline Phosphatase (ArcticZymes). The treatment included a 120 min enzyme activation step at 37 °C before 15 min inactivation step at 80 °C and was performed on Veriti 96 well Thermal Cycler (Applied Biosystems) (Kreso, 2010).

7.2.4 Probe analysis on capillary gel electrophoresis

DNA hybridize by forming hydrogen bonds between its nucleotides: Adenine (A) forms two hydrogen bonds with thymine (T), and guanine (G) forms three hydrogen bonds with cytosine (C) (Vebø, et al., 2011). A probe hybridized adjacent to discriminative gene position and was incorporated with fluorescence by DNA polymerase if the bacterium was presented (Vebø, et al., 2011). Labeling occurred at 3' end of the probe after the whole probe had hybridized to bacterium template.

Two hundred and seventeen probes were tested to identify probes that satisfied criteria of target detection and non-target exclusion. This was performed through end-labeling for each single probe in following order: self-hybridization (I), target bacteria (II), up to 10 non-target bacteria (III) and probe used in probe set (IV). To test a probe against target and non-target – bacteria there was used bacterial DNA from pure culture (template) as described in section 7.2.3. Template was replaced with water during self-hybridization (I) and probe tested in a probe set (IV).

7.2.4.1 End-labeling reaction

End-labeling reaction was as follows: 0.25 U HOT Termipol DNA polymerase (Solis Biodyne), 1 x HOT Termipol buffer C polymerase (Solis Biodyne), 4 mM Magnesium-dichloride (Solis Biodyne), 0.4 μM ddCTP-TAMRA (5-propargylamino-ddCTP – 5/6-TAMRA) (Jena Bioscienc), 0.1 μM probe and 5-100 ng/ml bacterium template in a total reaction volume of 10 μl. The end-labeling included a 12 min initial denaturation stage at 95 °C, followed by 10 cycles with 20 sec denaturation at 96 °C, 30 sec annealing at 60 °C, and 5 sec extension at 60 °C. Positive control was included with probe (Universal 4) and *E. coli* as template. Negative control included end-labeling reagents without probe or template (replaced with water). End-labeling was performed on Veriti 96 well Thermal Cycler (Applied Biosystems) (Frøyland, 2011b).

7.2.4.2 Shrimp Alkaline Phosphatase (SAP)-treatment

SAP-treatment was performed to remove phosphate groups from the 5' end of DNA. TAMRA labeled product was treated with 0.09 U SAP (ArcticZymes) in a total reaction volume of 11 μl, and incubated at 37 °C for 60 min and inactivated at 80 °C for 15 min on Veriti 96 well Thermal Cycler (Applied Biosystems) (Frøyland, 2011b).

7.2.4.3 Preparation for Genetic Analyzer

SAP-treated and end-labeled product was handled with Hi-Di^{TM} Formamid (Applied Biosystems) and GenScan 120 LIZ Size Standard (Applied Biosystems) in a total reaction volume of 10.5 μl and incubated at 95 °C for 5 min on Veriti 96 well Thermal Cycler (Applied Biosystems) (Frøyland, 2011b). This step was performed for preparation for the

Genetic Analyser (capillary gel electrophoresis) where Hi-Di™ Formamid kept template and probe denatured through the analyses, and GenScan 120 LIZ Size Standard was present for size standard and presented different fragments in GeneMapper (software program for analyzing results from Genetic Analyser).

7.2.4.4 Testing probes on capillary gel electrophoresis

All probes were evaluated by using ABI Genetic Analyzer 3130xl sequencer (Applied Biosystems) and analyzed with the software program GeneMapper version 4.0 (Applied Biosystems). Probe tested for target detection was required signal up to 7,000 – 8,000. Concerning self-hybridization and probes tested for non-target detection, no signals were desired (see Appendix D for examples over how results could be seen). Self-hybridization probes were excluded by signals over 150. For target detection, probes were excluded if they got signals under 500 or double peaks. According to non-target detection, probes were excluded with signals over 300 or double peaks. An optimal probe gave signals over 1,000 and under 160 for target detection and non-target detection, respectively.

7.2.5 The election process of probes to GA-map™ IBD assay

A probe set (GA-map™ IBD assay) could hold 50 probes/bacteria (included control). An election process was therefor important for decision of which probes that should be included in the probe set. The election process was determined on the basis of five criteria. The first criterion was to choose optimal probes. An optimal probe does not self-hybridize nor captures non-target bacteria. An optimal probe was a specific probe that only captured the target bacterium (I). The next criterion was frequently reported bacteria. Even though all bacteria related to the intestinal microbiota according to IBD was of interest, there are some bacteria that have been more frequently reported and seems more important than others (II). In addition to the frequently reported, bacteria related to Crohn's disease are of more interest compared to ulcerative colitis. This is due to the fact that there are found less correlation between bacteria and ulcerative colitis (III). A probe must work with other probes (not hybridize) in a probe set. This was performed in TNTProbeTool (by GA) where all probe sequences were tested against each other (IV). A probe set should include at least one bacterium per phylum (Firmicutes, Actinobacteria, Bacteroides and Proteobacteria) related to the intestinal tract (V).

7.3 Analysis of patient samples

7.3.1 Cohort

Akershus University Hospital (Ahus) holds feces samples from people suffering from IBD before they have gone through medication and treatments. Feces samples were from a large population-based study in Norway focused on inflammatory bowel disease at all ages, and named Inflammatory Bowel South Eastern Norway – II (IBSEN-II). A diagnose was set according to Porto criteria (Appendix A) and based on histopathological verification in biopsis taken during upper and lower endoscopy. Endoscopic procedure in children (<16 years) was performed under general anesthesia (Perminow, Brackmann, et al., 2009), and stool samples were collected for microbiological investigations to exclude infectious causes, e.g. *Salmonella, Shigella, Yersinia* and *Campylobacter* (Escher, et al., 2005).

Seventy-four 74 feces samples, represented from the pediatric part of IBSEN-II study, and included 28 CD samples, 15 UC samples, 3 IBDU samples (unconfirmed prognosis), and 28 non-IBD samples (Appendix H). The samples were randomly distributed on three plates (due to other adults samples from IBSEN-II, not included in this thesis).

7.3.2 Preparation of template

Ahus isolated DNA from feces samples. DNA was amplified by CoverAll™ primer, DNA quantificated, and ExoSAP-treated as described in section 7.2.3.

7.3.3 Analysis on Magpix™ instrument

Cohort (IBSEN-II samples) was first analyzed with two diagnostic probe sets: GA-map™ Core assay (probes included are noted in Appendix L) and GA-map™ Infant assay (probes included are noted in Appendix M). The two diagnostic tools were together with GA-map™ IBD assay analyzed in a second round.

The first step for analyzing samples on Magpix instrument was end-labeling. During end-labeling was labeling probes tested against cohort samples and the probes were labeled if the bacterium the probe identified was present. The next step was hybridization between labeling

probe and capture probe (complementary sequence for labeling probe). Capture probe was coupled to microspheres (beads) and detected by Magpix instrument in the third step.

7.3.3.1 End-labeling of labeling probes

End-labeling reaction of probes for the Magpix platform was as follows: 0.125 U HOT Termipol DNA polymerase (Solis Biodyne), 1 X Buffer C (Solis Biodyne), 4 mM Magnesium-dichloride (Solis Biodyne), 0.5 μM Biotin-11-ddCTP (PerkinElmer^{TM} Precisely), 0.01 μM probe and ExoSAP-treated PCR-products (IBSEN-II samples) in at total reaction volume of 10 μl. End labeling included a 12 min initial denaturation stage at 95 °C, followed by 5 cycles with 20 sec denaturation at 96 °C, and 35 sec combined annealing, and elongation at 60 °C performed on Veriti 96 well Thermal Cycler (Applied Biosystems) (Vebø, 2012).

7.3.3.2 Capture probes analyzed on Magpix^{TM}

MagPlex microsphere (bead) was 6,2 microns in diameter, contained surface carboxyl groups, and superparamagnetic. Microsphere was internally labeled with fluorescent dye, and one specific dye per microspheres gave specific recognition and separation. The microsphere was coupled with capture probes (complementary sequence of labeling probe) to be able to capture a labeling probe that was labeled with fluorescent if the target bacterium was present from cohort sample. Capture probe was tagged with a 5' end amino modified C12 – a reactive primary amino group and facilitated the coupling to the carboxyl group on the microspheres (Frøyland, 2011a). Illustration for how labeling probe hybridized to cohort samples, and further was captured by capture probe, before the capture probe was detected by fluorescent at Magpix instrument, can be seen in figure 4.

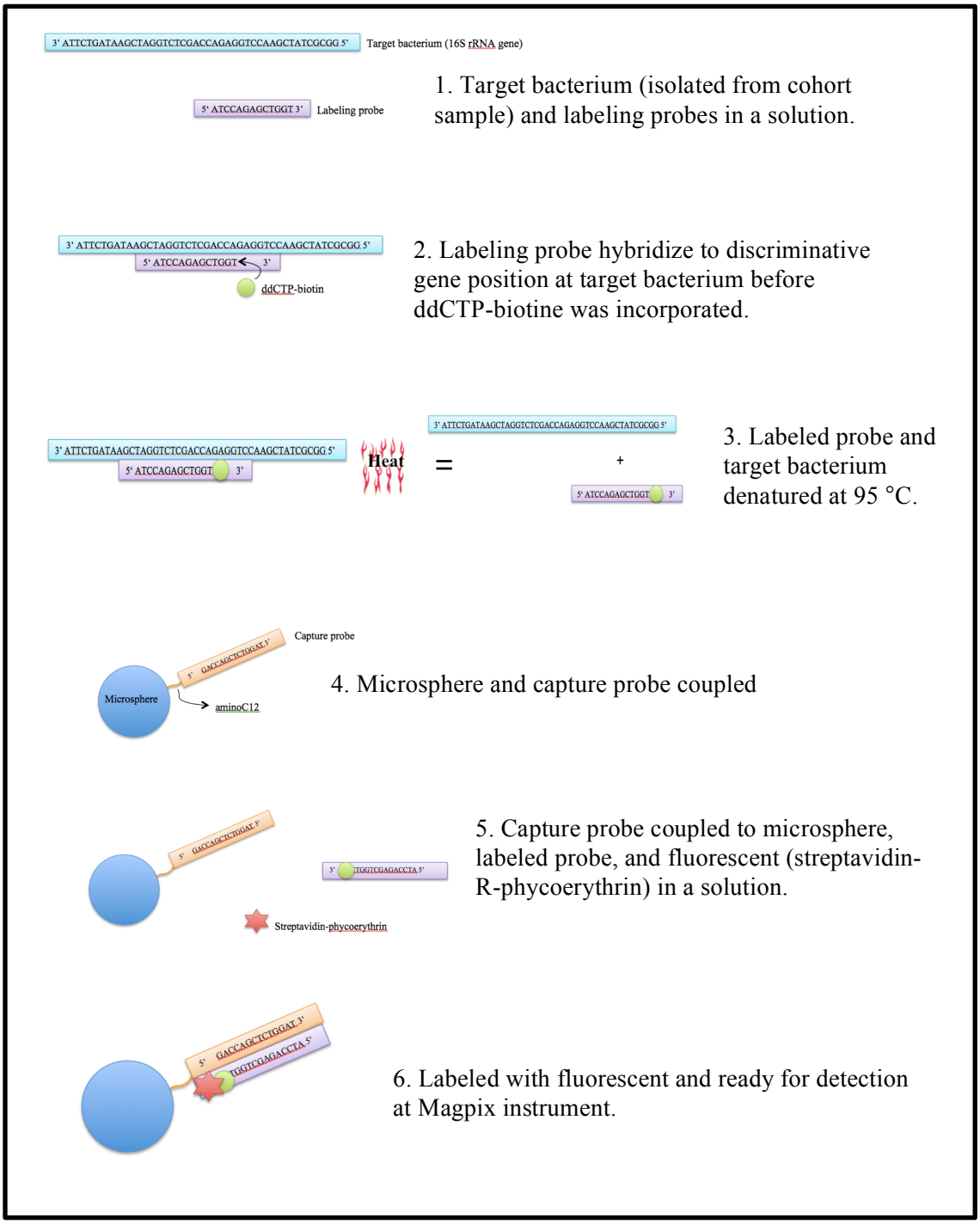


Figure 4: Labeling probe hybridizes to target bacterium (template) and is incorporated with ddCTP-biotine. Biotin-labeled probe and template is denatured before biotin-labeled probe hybridizes to capture probe (complementary sequence to labeling probe). To be detected by Magpix instrument, streptavidin binds to ddCTP-biotine before labeled with phycoerythrin. Phycoerythrin was the fluorescence and was a component of the complex streptavidin-R-phycoerythrin (Vebø, 2012).

One capture probe was coupled to one microsphere (performed by GA). A selected oligonucleotide-coupled microsphere set was vortexed and sonicated for 20 sec before the set was diluted around 1:20 in 1.5 X Tetramethylammonium chloride (TMAC) hybridization buffer (SIGMA Life Science). 1.5 X TMAC consisted of 4.5 M TMAC, 0.15 % Sarkosyl, 75 mM Tris-HCl (pH 8.0), 6 mM EDTA (pH 8.0) and H_2O. Sonication and vortexing were performed to prevent uncoupled microspheres to stick together and further to bring them in homogenized solution. A 1:4 ratio (labeled:unlabeled) mix of a hybridization control probe consisting of a pre-labeled (oligonucleotide probe modified with a 3' ddCTP-biotin, 0.1 μM HYC01_Biotin) and the same oligonucleotide probe without 3' ddCTP-biotin (0.1 μM HYC01) were added to each well. HYC01 was used to evaluate the hybridization step and to normalize differences in hybridization efficiencies. One well contained up to 50 different microspheres (coupled with capture probes), hybridization control, and end-labeled reaction and constituted a total reaction volume of 50 μl, before the hybridization was initiated. The hybridization included a 3 min denaturation stage at 95 °C for any secondary structure in the oligonucleotide probe, followed by 15 min hybridization of oligonucleotide probe to complementary capture probe at 45 °C performed on Veriti 96 well Thermal Cycler (Applied Biosystems) (Vebø, 2012).

The hybridized samples were then added 0.03 μg Streptavidin-R-phycoerythrin (SAPE) (Invitrogen) and 1 X TMAC Hybridization Solution (consists of 3 M TMAC, 0.1 % Sarkosyl, 50 mM Tris-HCl (pH 8.0), 4 mM EDTA (pH 8.0) and H_2O) in a total reaction volume of 75 μl (in each well), and the samples were incubated further at 45 °C for 15 min on Veriti 96 well Thermal Cycler (Applied Biosystems) (Vebø, 2012).

The samples were analyzed with Magpix instrument. In each well, totally 50 μl microspheres (mixture of the different microspheres) were absorbed and brought to a magnet that held the magnetic microspheres in place while the microspheres were illuminated and imaged by a red and green light-emitting diode (LED). The samples were first measured for internal dye fluorescence for classification and to distinguish the 50 different microspheres by the red LED, before the green LED excites surface and reports fluorescence (phycoerythrin). The magnetic microspheres were washed before illumination to reduce disturbance. At least 50 microspheres needed to be counted from each type to get the measure approved. There is also an upper limit for the magnet to hold the microspheres before disturbance destroys the signal.

Analyzes of the images were performed by Luminex xPONENT software (Luminex, 2011). How the Magpix instrument is build up and how the samples are run at the instrument are denoted in Appendix B.

7.4 Statistical analysis

There were several steps in the statistical analysis before results over probes and samples could be announced, and the steps were:

1. Pre-processing – reduce noise and irrelevant information.
2. Make classifications – healthy vs. sick people using probe signals as predictors.
3. Cross-validation – leave-one-out.
4. Present the results.

7.4.1 Multivariate statistic

When probes were tested against cohort, one was left with large amounts of data that included many variables. To find a connection between probes and samples, and to analyze more than two variables, multivariate data analysis was performed to represent a simplification of the data (Nortvedt, Brakstad, Kvalheim, & Lundstedt, 1996). The two multivariate data analyses used in the present work were Principal Component Analysis (PCA) and Partial Least Squares – Discriminant Analysis (PLS-DA).

PCA describes the variation in a data matrix by finding new variables called super-variables that maximize the explained variance in the data. PCA can therefore be useful in the visualization of the most important information that can be extracted from a data set with many variables. PLS-DA is a classical regression that decomposes a data matrix in a similar way as the PCA but uses additional information (diagnostic information from each patient) to find super-variables. Super-variables were found by using information in one or more response variables, to control separation of the samples. Score plot and loading plot were used to present super-variable results where score presented cohort samples and loading presented probes (Sekelja, 2012). The principle behind PCA and PLS-DA is described in Appendix E.

7.4.2 Pre-processing

Pre-processing was performed of raw data to reduce noise and irrelevant variation before the data was analyzed with respect to diagnosis. To reduce plate-to-plate variation, plate ID was used as response variable and probe signals as explanatory variable. Hybridization control probe (HYC01) was used to compensate for the time differences that occurred between the three plates, and all the signals from one well were divided at HYC01 signal. Raw data was sorted into score-plot and loading-plot. Score values presented IBSEN-II samples and their location in the new coordinate system, while the loading values presented probes that contributed to form the basis for the new coordinate system (Sekelja, 2012). Information behind pre-processing is described in Appendix F.

7.4.3 Classification and discriminant analysis

Classification was used to arrange predefined groups (IBD/non-IBD) and included sensitivity, specificity and classification error. The sensitivity explained the percentage of people classified as IBD, while specificity explained the percentage of people classified as non-IBD. Classification error (number between 0 and 1) described the amount of people misclassified, e.g. how many healthy people that were classified as sick (Sekelja, 2012). More information about classification can be seen in Appendix G. A discriminant analysis explained what distinguished different groups from each other by looking at the characteristics of groups (e.g. bacterial composition of the samples belonging to IBD/non-IBD group) (Sekelja, 2012).

7.4.4 Cross-validation

A cross-validation method called "leave-one-out" was used for the present work. The principle behind "leave one out" is simple; take one sample out at a time, and then build a classification model and compare the predicted value of the sample taken out with its observed value. The process is repeated until all the samples have been left out once (Browne, 2000). The correlation between the predicted and the actually observed values can be used as a measure of the model's usefulness and can give a value for how GA-map^{TM} assay corresponded to distinguish between IBD and non-IBD.

8 RESULTS

8.1 Probe designed in TNTProbeTool

Designed probes in TNTProbeTool are reported in Appendix I (species-probes) and Appendix J (genus/class/order-probes). Comments for the probe search are also noted in the mentioned appendices. Repeated problems during probe design were designated to no available bacteria, in addition to melting temperature (T_m). Total number of probes designed in TNTProbeTool is reported in table 1.

8.2 Capillary gel electrophoresis

Results from probes tested against target and non-target bacteria at capillary gel electrophoresis, are reported in Appendix K. Based on results from capillary gel electrophoresis, the probes were categorized and given color code, to be able to distinguish optimal probe and excluded probes. A summarized overview of number of probes tested on capillary gel electrophoresis is presented in table 1.

Table 1: Summarized results over number of probes designed in TNTProbeTool and probes tested on capillary gel electrophoresis. Color code represents probe within the different categories.

	Category		**Number**
TNTProbeTool	Probe designed with TNTProbeTool		471
		Color code	
Capillary gel electrophoresis	Tested at GA from earlier study (not tested at capillary)		22
	Excluded from study: self-hybrid., or unspecific probe		69
	Excluded from study: target bacterium not available		68
	Problem with signals at capillary		23
	Optimal probe		57
	Optimal species probe		53
	Optimal probe at higher taxonomic level		4
	Total number of probes tested on capillary		217

Totally 471 probes were designed in TNTProbeTool and further were 217 probes tested on capillary gel electrophoresis, as seen in table 1. Results after the probes were tested on capillary gel electrophoresis excluded 69 probes due to self-hybridization or non-specificity.

Fifty-seven probes could be categorized as an optimal probe where 53 probes were designed at species level. Results from Appendix K showed double peaks that occurred for some of the probes as an indication for a non-optimal probe.

8.3 GA-map™ IBD assay

Probe identification with its target bacterium used in GA-map™ IBD assay are presented in table 2. More comprehensive information concerning relatedness of bacteria is denoted in Appendix N.

Table 2: Bacteria and probe ID incorporated in GA-map™ IBD assay.

Target bacteria	Probe ID
Akkermansia muciniphila	AG0798
Catenibacterium mitsuokai	AG0894
Clostridium aminophilum	AG0899
Clostridium difficile	AG0907
Clostridium methylpentosum	AG0912
Clostridium ramosum	IG0013
Dialister invisus	AG0931
Fusobacterium periodonticum	IG0113
Haemophilus (genus)	AG1021
Lactobacillus johnsonii	AG1051
Lactobacillus ruminis	AG1058
Pseudomonas straminea	AG1133
Ruminococcus albus	AG1148
Veillonella (genus)	AG1207
Bifidobacteria	AG1219
Alistipes putredinis	AG0339
Bacteroides fragilis	AG0377
Bacteroides stercoris	AG0416
Clostridium leptum	AG0470
Coprococcus comes	AG0495
Coprococcus eutactus	AG0569
Dorea formicigenerans	AG0581
Enterococcus faecalis	AG0591
Eubacterium rectale	AG0620
Faecalibacterium prausnitzii	AG0651
Ruminococcus gnavus	AG0703
Subdoligranulum (genus)	AG0744
Bacillus licheniformis	AG0849
Bacteroides acidofaciens	AG0861

Enterococcus faecium	AG0962
Eubacterium biforme	AG0974
Helicobacter canis	AG1024
Helicobacter hepaticus	AG1030
Klebsiella pneumoniae	AG1033
Lactobacillus reuteri	AG1057
Mannheimia (genus)	AG1066
Blautia schinkii	AG1228
Proteobacteria (phylum)	IG0005
Haemophilus parainfluenza	IG0008
Firmicutes 1 (phylum)	IG0012
Firmicutes 2 (phylum)	IG0023
Gammaproteobacteria (class)	IG0056
Bacteroides (genus)	IG0060
E.coli/Shigella	IG0133
Blautia coccoides/Clostridium coccoides	AG0879
Roseburia hominis	AG1146
Ruminococcus bromii L263	AG1152
Bifidobacteria	IG0028
Enterococcus (genus)	IG0014
Universal probe	UNI05

8.4 Partial Least Squares – Discriminant Analysis (PLS-DA)

Through statistical analysis was there observed best correlation between CD and non-IBD, and best results were obtained from GA-map™ Core assay and GA-map™ Infant assay. Further results are based on the mentioned observations (Sekelja, 2012) and can be seen in analysis report (in section 8.4.2)

8.4.1 Pre-processing

Raw data is sorted into score and loading -plot in figure 5 as a result from plate variation. Score plot show cohort samples that are distributed among three plates. The two probe sets (core and infant) can be seen in loading plot. Score presents cohort samples and their location in the new coordinate system, while loading presents probes that contribute to form the basis for the new coordinate system. Loading and score plot can be compared to reveal which probe set that contributes to difference among the three plates.

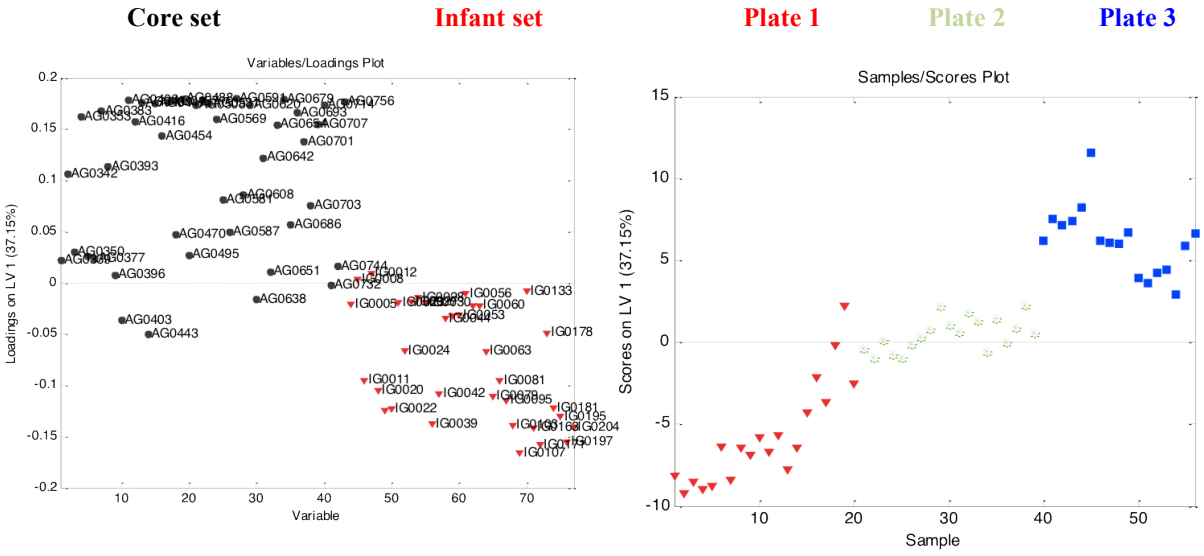


Figure 5: Reveals different signal strength between the plates. Left: core set and infant set presented in loading plot. Right: three plates of cohort samples distributed in scores plot.

Figure 5 presented the result of the raw data where a clear difference among the three plates could be seen. By comparing score and loading –plot horizontal, one could see that probes from core probe set (in loading plot) contributed to the difference for plate 3 (in score plot), while probes from infant contributed to the difference for plate 1.

Small differences that occurred naturally between plates are to be reduced through pre-processing. By using results from PLS-DA with plate ID as response variable and probe signals as explanatory variables, one can reduce plate-to-plate variation, as shown in figure 6.

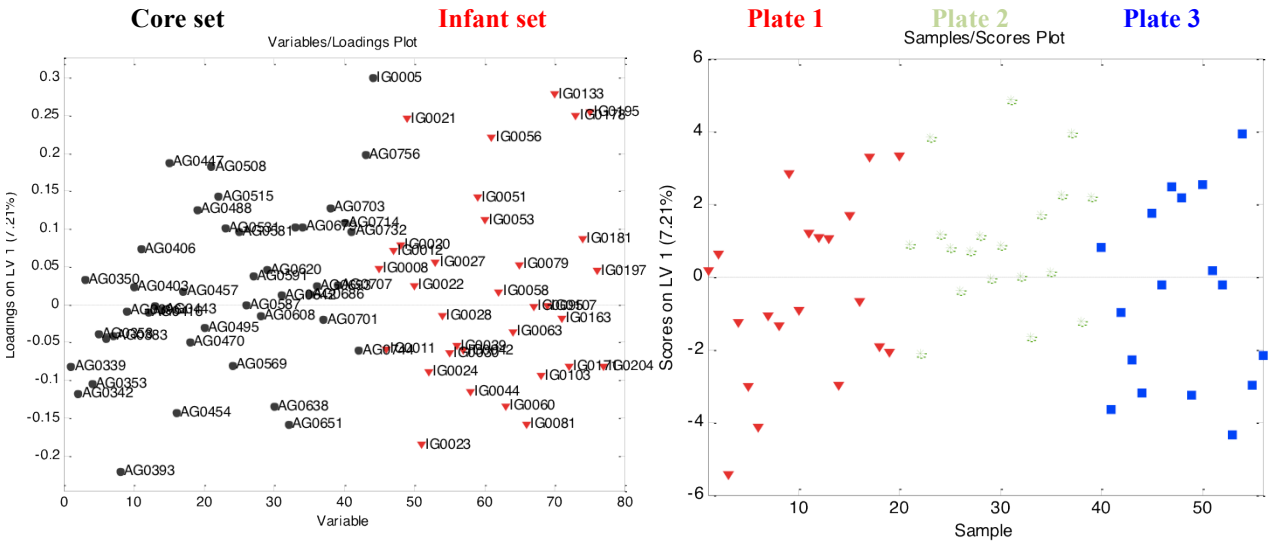


Figure 6: Results after correction of variation. Left: core set and infant set represented in loading plot. Right: Three plates of cohort samples represented in scores plot.

As a result from reduced plate-to-plate variation in figure 6, the plates were less clustered and further more distributed along the x-axis. Differences among plates and probes could no longer be observed.

8.4.2 Analysis report from PLS-DA

Statistically work was performed in two separated parts where first the GA-map^{TM} Core assay (core) and GA-map^{TM} Infant assay (infant) were included, and secondly where GA-map^{TM} Core assay, GA-map^{TM} Infant and GA-map^{TM} IBD (IBD001) were included. All samples were analyzed with the whole cohort (CD, UC, IBDU and non-IBD) but during statistically work, there was found best correlation for CD and non-IBD (Sekelja, 2012). This can be seen in table 3 as a result from analysis report through cross-validation in PLS-DA.

Table 3: Analysis report from PLS-DA over CD and non-IBD for core+ infant, and core+infant+IBD001. Specificity, sensitivity and accuracy (in percent) are shown for the two tests.

	CD/non-IBD	
	Core + Infant	Core + Infant + IBD001
Specificity	82	61
Sensitivity	86	46
Accuracy	84	54

Based on results from table 3, one could see that core+infant gave highest scores, which indicated best result. Specificity, sensitivity and accuracy for core+infant were 82 %, 86 % and 84 %, respectively. Accuracy for 84 % gave a classification error by 16 %.

8.4.3 Score plot

Score plot over subject diagnosed with CD and subject classified as non-IBD for cohort samples and their location in the new coordinate system, are presented in figure 7. Samples vs. predicted subjects are presented in figure 8 where an indication for misclassified samples (results from cross-validated results) is shown with a cutoff for 0.5. CD samples are to be present over the cutoff value and non-IBD under the cutoff value.

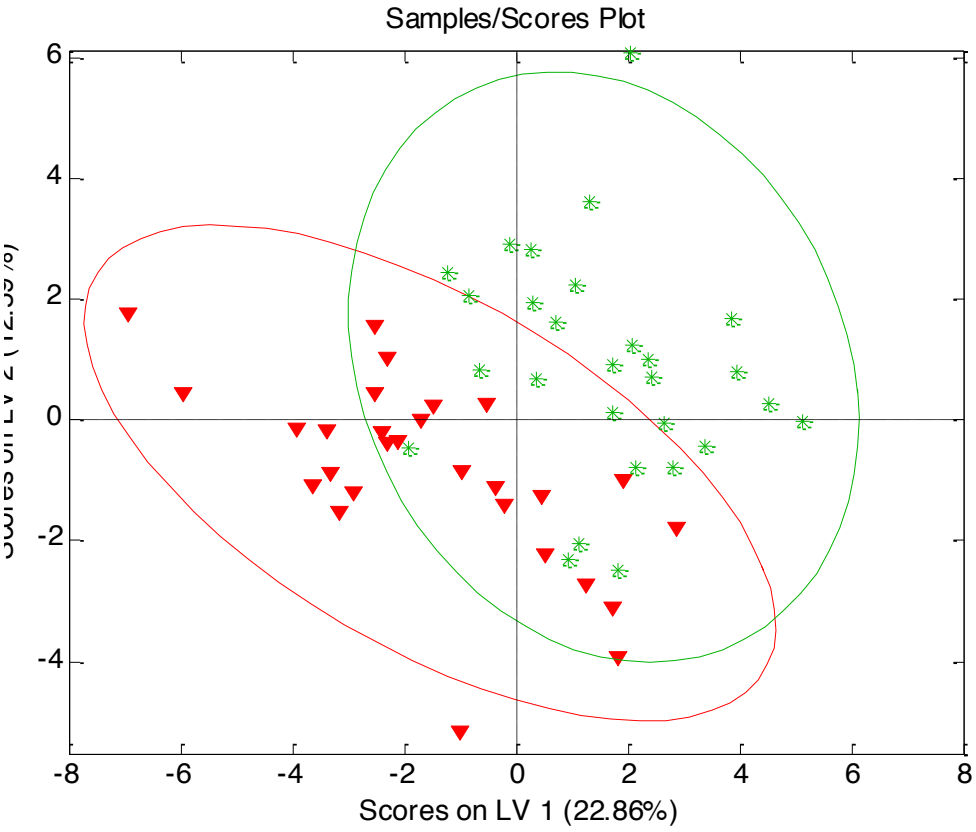


Figure 7: Subjects diagnosed with CD (red) and subject classified as non-IBD (green) presented in score plot.

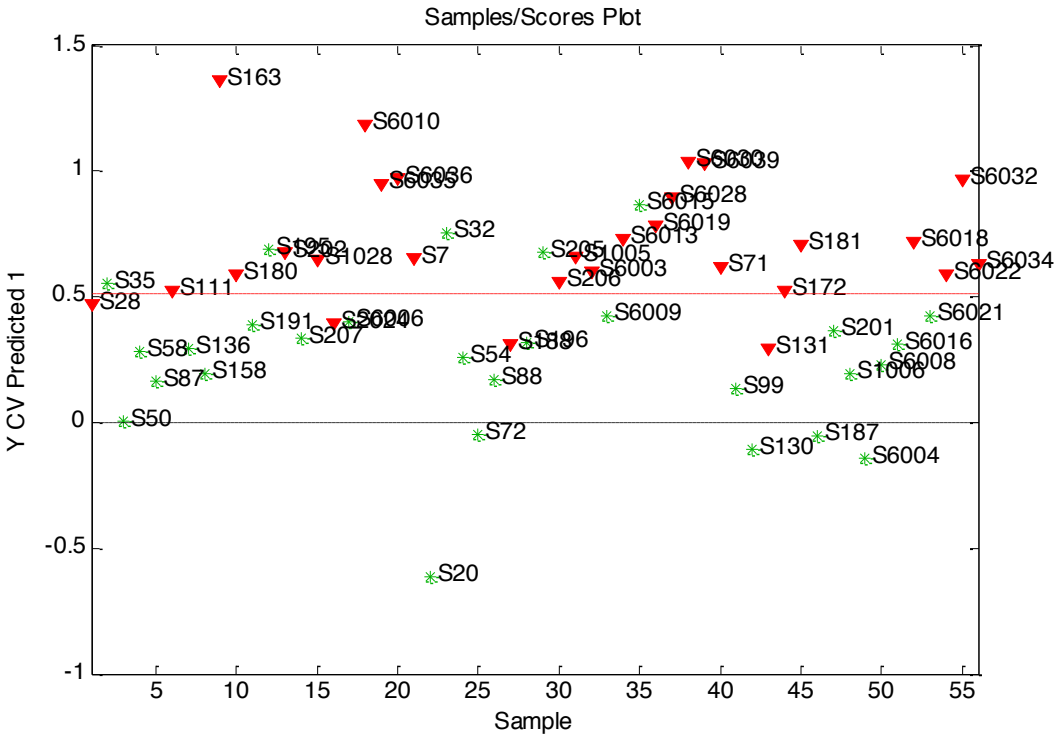


Figure 8: Predicted vs. samples over non-IBD (green) and CD (red). 9 of 56 were misclassified by the cutoff for 0.5.

Figure 7 showed separation between CD and non-IBD and their location in the new coordinate system. The result from cross-validation was presented in figure 8 and displayed 9 of 56 samples as misclassified by a cutoff for 0.5 as a result from non-IBD classified as IBD and IBD as non-IBD.

8.4.4 Loading plot

In figure 9, the probes are presented in loading plot to see which probes that are contributing to distinguish non-IBD from CD. The same loading plot is presented in figure 10 but probe ID is replaced with name of the bacteria. Probes that stretch far out in loading plot contributes to most separation between CD and non-IBD.

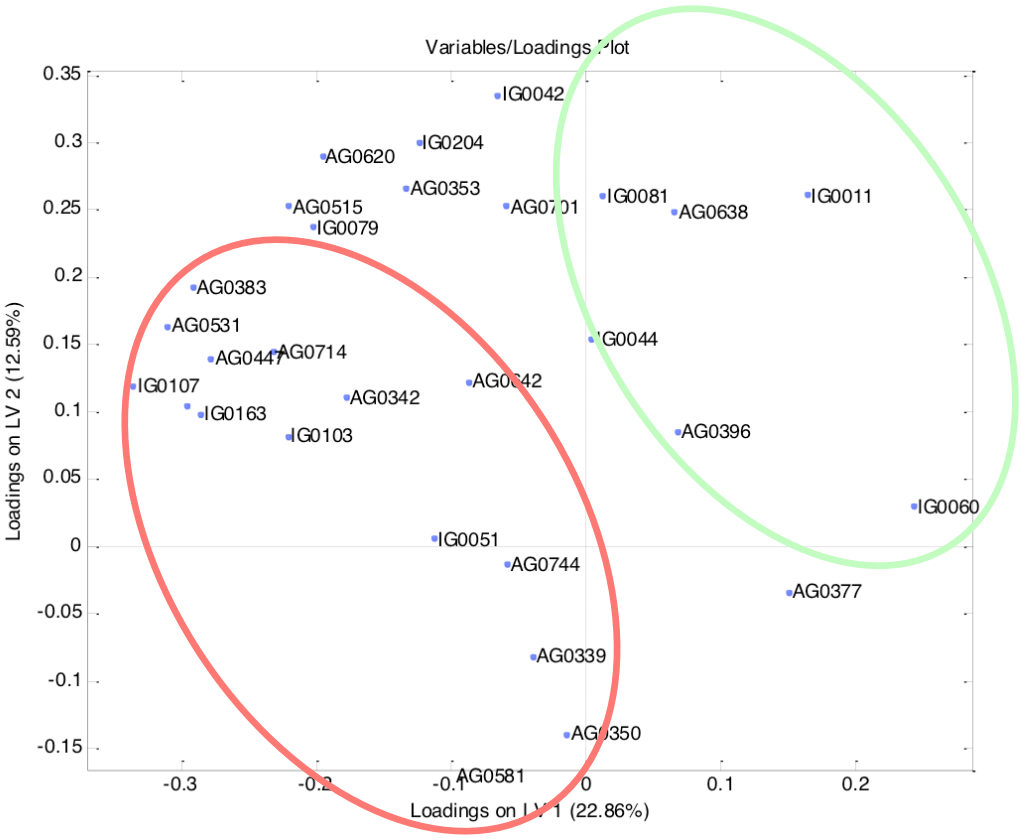


Figure 9: Loading plot for probe identification over probes that contributed to distinguish between CD (red circle) and non-IBD (green circle).

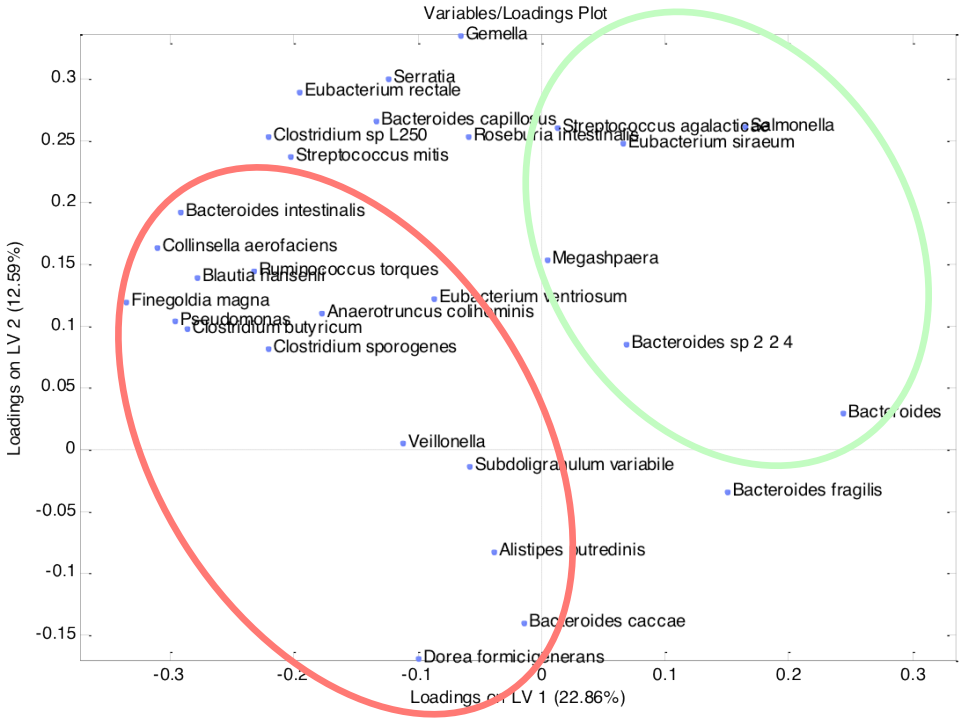


Figure 10: Loading plot over bacteria that contributed to distinguish between CD (red circle) and non-IBD (green circle).

As seen in figure 9 and 10 there were bacteria that contributed to distinguish between CD and non-IBD. Bacteria for non-IBD direction were: *Eubacterium siraeum*, *Bacteroides sp. 2 2 4, Bacteroides fragilis, Bacteroides*, IG0011 (*Salmonella, Citrobacter, Cronobacter, Enterobacter* and *Morganelle*), IG0081 (*Streptococcus agalactiae* and *Eubacterium rectale*). Bacteria directed for CD were: *Collinsella aerofaciens, Finegoldia magna, Bacteroides intestinalis, Ruminococcus hansenii* (*Blautia hansenii*), *Subdoligranulum variable*/*Faecalibacterium prausnitzii, Clostridium butyricum, Alistipes putredinis, Ruminococcus torques* and *Bacteroides caccae.*

9 DISCUSSION

9.1 Cohort samples

Non-IBD was a systematic control group for patients not having IBD. Those people who were included as non-IBD came initially to the doctor because they had intestinal problems. By further testing, they passed not the Porto criteria and were therefor not considered as IBD patients. To what extend this affected and weakened the result between IBD and non-IBD was uncertain, and can only be confirmed by performing investigations with another control group. However, the gut is affected by microbial and host interactions and by external and internal factors (Cucchiara, et al., 2009; Fanaro, et al., 2003). People from control group will always have individual compounds of microbiota, and non-IBD can be suitable enough as control group.

Material of detection in present work represented feces samples from cohort. In the literature there are many studies that collect samples from mucosal flora, to investigate the microbiota. Feces represents a sample from the bacterial community that reflects the bacteriology of the rectum, and do not offer much ecological knowledge of other regions of the GI tract. Mucosal biopsy provides samples, collected from regions of the intestinal tract where inflammation occurs (Tannock, 2008). It is conflicting evidence whether feces represents a sample correlated to the actually activity of the intestinal microbiota or not. Some believe that fecal flora is different from mucosal flora, and it is more important to study the latter because they appear to play a greater role in cases of IBD (Bousvaros, et al., 2006). In contrast, there have

been found high degree of similarity between bacterial collections associated with biopsy and feces (Bibiloni et al., 2008).

There is without doubt a more comprehensive task to undergo a biopsy compared to give a feces sample. From a person's point of view, a feces sample is more easily available and gives less discomfort compared to biopsy, which is important in terms of the person.

9.2 Probe designed in TNTProbeTool

In several cases one could not detect probes that satisfied the search criteria through TNTProbeTool. Problems by distinguish target bacterium from non-target bacteria could cause no probe design. Nearby non-target bacteria, surrounded target bacterium that disturbed the probe search, were deleted due to their sequence similarities. If many nearby non-target bacteria were deleted, the probe would be less specific. The meaning of a specific probe correlates to single annealing to the complementary sequence present on only target bacterium. An unspecific probe could capture more than the target bacterium and causing unwanted bacteria to be caught.

The melting temperature (T_m) was set to be 60 °C under normal conditions when probe search was performed in TNTProbeTool. There were several cases where probes could not be designed under the normal condition, and the T_m was changed to accomplish a probe design. A probe designed in TNTProbeTool was only a theoretical suggestion where the software program calculated T_m based on e.g. bases, hybridization and nearest-neighbor thermodynamic approach. Further test in vitro was performed with similar temperature for all probes, however, probed designed with another T_m could be an explanation to why probe did not annealed to the target bacterium.

TNTProbeTool could design several probes for one species/genus. It was no possibilities to test more than three probes through the present work due to limited time. The election process to reduce the number of probes was performed by calculations in TNTProbeTool (accomplished by GA). A probe could not be designed if the target bacterium was not present in TNTProbeTool. When the software program was made, some 16S rRNA gene sequences were not available and could therefor not be incorporated in TNTProbeTool. Probes at higher

taxonomic level could not be designed if just one bacterium for the taxonomic level was available.

Phylogeny serves to define and relate the diversity of environmental populations. Probes that were more directed to identifying phylogenetic groups of rank greater than species provided a phylogenetic overview to community structure. The complex microflora of the human gut (Bousvaros, et al., 2006) is difficult to study with probes on a species level due to the diversity. It was therefor more convenient to have probes specific for major genera and groups present in the gut, but there was also interesting to see if any of the reported species were present in the feces samples.

TNTProbeTool designs probes based on available sequence information, and in principle it is only bacteria incorporated in the software program. Although this could have been a desirable situation is this not the whole truth. Unknown bacterial sequences from gut microbiota are anonymous information for TNTProbeTool, and when probes were designed, these sequences could not be compared. Based on this, there might be unknown bacteria that anneal to the designed sequences (probes) in TNTProbeTool.

There are uncertainties related to the use of probes, and descriptions of probe applications are important to be able to compare probes from multiple research groups. Probe design, identification with different nomenclature and unpublished information according to probes (hybridization buffer compositions, target group specificity, etc.) can cause complications. Each probe should therefor include information as a standardized name, probe sequence, optimal hybridization conditions, intended target group and experimentally validated target group specificity (Alm, Oerther, Larsen, Stahl, & Raskin, 1996) to be able to compare different probes and hence bacteria.

9.3 Probe tested on capillary gel electrophoresis

Sixty-nine probes tested on capillary gel electrophoresis (hereafter capillary) were of different reasons excluded. Self-hybridized probes were caused by internal regions (intra-molecular homologies) of the probe that was used as target (Vebø, et al., 2011) because they formed secondary structures. Probe that did not hybridize to target bacterium was caused by either probe or template. There could be some errors in the probe sequence based on e.g. defect

oligonucleotides (ordered from Thermo Scientific) and will result in no capturing of bacteria. In terms of the template, it would be expected to get weak probe signals if the template gave weak results from PicoGreen and Gel Electrophoresis (Kreso, 2012). Information obtained from PicoGreen and Gel Electrophoresis was just an indication for DNA quantification and was not directly associated with the GA-map^{TM} assay, and were of this reason not reported.

If one probe captured non-target bacterium, which cannot be known before the probes are tested several times to confirm the catch, was an indication for an unspecific probe because the probe found discriminated sequences among non-target bacteria. The reason for double peaks from capillary is not known but is an indication for a non optimal probe (Vebø, 2012). Normally this was a recurrent problem for the probe because double peaks occurred in almost each case where the probes were tested against different bacteria template. Some problems occurred during capillary because the probe did not pass sizing. Software program GeneMapper could resolve the problem by identify peaks visually and perform a size standard but was unfortunately not always performed. This problem was mainly because of GeneMapper and had nothing to do with the probe or the template.

All factors affecting probe and annealing to template are not yet completely known but some explanations have been proposed. The sequence designed in TNTProbeTool was only a theoretical suggestion and the situation could be different in the real biological world. There could also be an error at the sequence as mentioned above. A sequence of 14-25 bp could be too short to be able to distinguish 100 % at 16S rRNA level and to further cover the identification of the bacterium. If this was the case, one could not be sure whether the sequence was the right identification for the bacterium. The T_m was important for the hybridization. However, there were no possibilities to spend time on each and every probe to treat it with the optimal T_m. This could be performed by the use of gradient PCA but has not been in focus by GA because of the amount of probes (Vebø, 2012).

Fifty-seven probes could be used in GA-map^{TM} IBD assay, and most probes were at species level. These probes correlated to an optimal probe and indicated no self-hybridization, annealed to discriminatory sequence on target bacterium, and were specific enough to not capture the non-target bacteria.

9.4 Partial Least Squares – Discriminant Analysis (PLS-DA)

In the literature there have been most attention on the pathogenic role of the gut microflora in CD (Joossens, et al., 2011; Mondot, et al., 2011; Schwiertz, et al., 2010) and can be related to that this disease correlates best to microbiota compared to UC. This could also be confirmed in the present work since statistical analysis, based on probe signals, found best correlation between CD and non-IBD. A probe signal was an indication for whether a group of bacteria was presented in the specific sample, or not. PLS-DA found super-variables who were found by using information in one or more response variables, to control separation of the samples (Sekelja, 2012), and only informative information was used. Pre-processed data was an important step where PLS-DA made calculations to reduce noise and irrelevant variation.

In terms of noise and irrelevant variation there were suggested some explanations. The cohort samples were randomly placed on three different plates and each plate was run separately. Each time one plate was placed in Magpix instrument there was introduced some small changes (noise) that were not a result from biological variation. Noise could be described as variation that occurred between the plate/sample as a result of e.g. temperature and pressure in the room, variation in pipetting etc. The main difference that occurred between samples, which was avoidable, was efficiency of hybridization. From the initial to the final sample was analyzed by Magpix instrument took approximately one hour. This introduced variability between each well with respect for how long the hybridization took place, hence the amount of fluorescence that was used in the hybridization reaction (Sekelja, 2012).

9.4.1 Misclassified samples

Classification error was based on sensitivity and specificity. Sensitivity reflected percentage of IBD patients who were correctly identified as having IBD, while specificity reflected the percentage of non-IBD who were correctly identified as not having IBD. Classification error indicated how many people that were misclassified by the model. The final model was based on GA-map^{TM} Core assay and GA-map^{TM} Infant assay because more samples were misclassified when GA-map^{TM} IBD assay was included (Sekelja, 2012). In the perfect situation would sensitivity be 100 % and specificity 100 %. Sensitivity and specificity were remarkably higher for GA-map^{TM} Core assay and GA-map^{TM} Infant assay, compared to GA-map^{TM} IBD assay.

To go further in diagnosis situation within the people that were misclassified by the model could reveal why some people were perceived as sick even if they were healthy, or opposed. This information could be used as an indication for weaknesses of the model so one could improve the model. Concerning diagnostic information one could ask some questions. Was there any doubt about the diagnosis? Did the person recover during a short time? Had healthy misclassified people an infection or another GI disorder? Because of confidentiality and assurance according to children included in the present work, no further diagnostic information existed.

Classification error can be used as a number that covers the diagnostic tool. In a hypothetical situation, the patient will go to the doctor to find out what causes the intestinal problems. From a doctor's point of view there will be questions whether the use of GA-map^{TM} assay will be reliably enough to be used as a tool to see if the symptoms is a result from CD or not. In a doubtful situation it can be advantageous to get a classification number that covers the GA-map^{TM} assay to decide if the patient should go further with the test.

GA-map^{TM} IBD assay gave weak results in analysis report from PLS-DA and was further excluded from the statically picture. One reason for this could concern bacteria presented in feces samples were not to be captured by probes incorporated in GA-map^{TM} IBD assay. Most likely, bacteria were present in the sample but could not be captured by the probes. The literature research ahead of the present work was to identify gastro intestinal bacteria suitable for IBD diagnostic. Most of these bacteria were based from adult samples and the situation could be different for children and adolescence. It has also been reported in the literature that microflora for children/adolescence differ from adult (Agans, et al., 2011). Another reason could be due to the probes, because, even though they worked at capillary, the situation could be different when probes were tested against cohort at Magpix instrument. There were different conditions for Magpix instrument as buffer, temperature, and beads, and the probe could respond differently. To get better results of GA-map^{TM} IBD assay one must optimized the probe set, by e.g. new or other probes.

9.4.2 Bacteria correlation to IBD and non-IBD

Repeated use of probes will give knowledge according to the expected target is the actually bacterium the probe captures. Based on this, was there more knowledge about infant probe set compared to core probe set.

PLS-DA found important bacteria that contributed to separating CD from non-IBD according to whether the probe was increased or decreased in relation to the opposite group (CD or non-IBD). *Eubacterium siraeum*, *Bacteroides sp. 2 2 4, Bacteroides fragilis* and *Bacteroides* were bacteria that contributed to non-IBD. These bacteria were obtained from core probe set and are largely shared among healthy individuals.

Based on observations from previously work performed by GA are there some probes that capture more than one species/genus. Probe IG0011 could basically capture *Salmonella, Citrobacter, Cronobacter, Enterobacter* and *Morganella*. This was an unspecific probe since the probe was expected to identify bacteria at a wide range. There was no possibility to decide whether the probe captured one of the mentioned genera or several. These genera could basically be observed in both CD and non-IBD direction. In present work, contributed probe IG0011 for non-IBD group, but based on weak specificity and identification would there be no certainty in relation to the real catch. *Salmonella* is a known pathogenic bacterium that cause local intestinal infection and Salmonellosis is essentially a food-borne infectious disease (Granum, 2007). However, the probe would not capture Salmonella since feces samples were collected for microbiological investigations in advance, to exclude infectious causes, and would have been revealed during this process.

IG0081 was named with *Streptococcus agalactiae* and *Eubacterium rectale* and contributed to non-IBD group. *Eubacterium rectale* is clustered in Clostridium cluster XIVa and reported as more prevalent in CD (Cucchiara, et al., 2009) and should therefor not be headed in the direction against non-IBD.

Clostridium butyricum, Finegoldia magna and *Collinsella aerofaciens* were in present work directed to contribute for CD. These bacteria is basically considered as prevalent intestinal in healthy individuals (Wang, Beggs, Erickson, & Cerniglia, 2004) and the results were not as expected.

Several bacteria that contributed in direction for CD were bacteria obtained from the core probe set. *Bacteroides caccae, Bacteroides intestinalis, Ruminococcus/Blautia hansenii*, and *Subdoligranulum variable*/*Faecalibacterium prausnitzii* are common bacteria found among a healthy human gut (Qin, et al., 2010) and uncertainties were related to why they were obtained and contributed to CD. However, *Faecalibacterium prausnitzii* is reported as decreased in CD (Joossens, et al., 2011). *Alistipes putredinis* and *Ruminococcus torques* are two common intestinal bacteria (Qin, et al., 2010) but are reported for CD patients as decreased (Mondot, et al., 2011) and increased (Martinez-Medina, Aldeguer, Gonzalez-Huix, Acero, & Garcia-Gil, 2006), respectively. *E. coli* is a bacterium that frequently are reported, especially in CD. Even though the bacterium was not presented in loading plot, the bacterium could be presented in feces samples but did not distinguish during statistical evaluation. More information about reduced or increased ratio should have been included to get more information. All one could say was that these bacteria contributed more in one or the other direction. More knowledge about the probe captures should however be explored before one could go further to see if the bacteria have decreased or increased.

There were different reasons why bacteria directed for CD and non-IBD were obtained among each other. A reason could be that the bacteria collected from cohort were taken in an early stage of the disease course and that these bacteria would increase or decrease in a later course of the disease. The time one intestinal goes from normal to more headed for CD is most likely to be performed during a period. This is also to be noticed when different studies have been reported different bacteria. Most of the bacteria are reported later in the course of the disease and after medication and treatments are performed at the patients and bacteria can be affected by the treatments. Even though bacteria revealed in present study could not be compared with previous studies, there was still possible to distinguish between people suffering from CD and non-IBD.

Based on observed differences between CD and non-IBD from present study, many of the bacteria were obtained from a common human intestinal and no specific pathogenic bacterium was obtained. This might indicate that there was an alteration of the already presented bacteria in the intestinal, because genetically predisposed individuals may appear to lose the normal tolerance to commensal bacteria, which leads to an elevated inflammatory response.

10 RECOMMENDATION FOR FURTHER WORK

For further work one could include more cohort samples to strengthen the results. Replace non-IBD with healthy subjects could also be a recommendation since non-IBD may have a diagnostic potential, and moreover defect the results. For detection of bacteria through the pathogenesis of IBD, there could be interesting to test the cohort after a time period to see if the same bacteria were presented. There should also be included information regarding the ratio between decreased and increased.

GA-map^{TM} IBD assay need to be optimized in the future to get better signals by the probes. This could be accomplished by trying different fluorescence to strengthen the probe signals, or change hybridization conditions. Further needs are based on knowledge behind the probes (what they captures), to compare the results from other studies.

As there were more probes designed than could be evaluated, there could be missing information. Further testing of several probes to optimize GA-map^{TM} IBD assay could be an opportunity to reveal more bacteria correlated to IBD. Moreover, strengthen the identification of the gut bacteria by using different genes, other than 16S rRNA gene, by extend the assay to investigate the biodiversity of the gut microbiota.

11 CONCLUSION

A tool based on GA-map^{TM} technology was designed and revealed differences between non-IBD and patients who suffer from CD, and the model gave a classification error by 16 %. An altered composition of the common bacteria in the intestinal tract seems responsibly for CD.

12 REFERENCES

Agans, R., Rigsbee, L., Kenche, H., Michail, S., Khamis, H. J., & Paliy, O. (2011). Distal gut microbiota of adolescent children is different from that of adults. ***Federation of European Microbiological Societies, 77*****, 404-412.**

Alm, E. W., Oerther, D. B., Larsen, N., Stahl, D. A., & Raskin, L. (1996). The oligonucleotide probe database. ***Applied and Environmental Microbiology, 62*****, 3557-3559.**

Baumgart, D. C., & Carding, S. R. (2007). Inflammatory bowel disease: cause and immunobiology. ***The Lancet, 369*****, 1627-1640.**

Baumgart, D. C., & Sandborn, W. J. (2007). Inflammatory bowel disease: clinical aspects and established and evolving therapies. ***The Lancet, 369*****, 1641-1657.**

Biank, V., Broeckel, U., & Kugathasan, S. (2007). Pediatric Inflammatory Bowel Disease: Clinical and Molecular Genetics. ***Inflamm Bowel Dis, 13*****, 1430-1438.**

Bibiloni, R., Tandon, P., Vargas-Voracka, F., Barreto-Zuniga, R., Lupian-Sanchez, A., Rico-Hinojosa, M. A., et al. (2008). Differential clustering of bowel biopsy-associated bacterial profiles of specimens collected in Mexico and Canada: what do these profiles represent? ***Journal of Medical Microbiology, 57*****, 111-117.**

Bousvaros, A., Sylvester, F., Kugathasan, S., Szigethy, E., Fiocchi, C., Colletti, R., et al. (2006). Challenges in Pediatric Inflammatory Bowel Disease. ***Inflamm Bowel Dis, 12*****, 885-913.**

Browne, M. W. (2000). Cross-Validation Methods. ***Journal of Mathematical Psychology, 44*****, 108-132.**

Büller, H. A. (1997). Problems in diagnosis of IBD in children. ***The Netherlands Journal of Medicine, 50*****, 8-11.**

Cucchiara, S., Iebba, V., Conte, M. P., & Schippa, S. (2009). The Microbiota in Inflammatory Bowel Disease in Different Age Groups. ***Digestive Diseases, 27*****, 252-258.**

Dorlands. (1981). ***Dorland's Illustrated Medical Dictionary*** **(26 th ed.): W. B. Saunders company.**

Escher, J. C., Dias, J. A., Bochenek, K., Buderus, S., Mesquita, M. B., Bujanover, Y., et al. (2005). Inflammatory bowel disease in children and adolescents: Recommendations for diagnosis - The Porto criteria. ***Journal of Pediatric Gastroenterology and Nutrition, 41*****(1), 1-7.**

Fanaro, S., Chierici, R., Guerrini, P., & Vigi, V. (2003). Intestinal microflora in early infancy: composition and development. ***Acta Pædiatr Suppl, 441*****, 48-55.**

Fletcher, H., Hickey, I., & Winter, P. (2007). ***Genetics*****: Taylor & Francis.**

Frøyland, C. J. (2010). ***Diagnostics of Inflammatory Bowel Disease using Fecal Microbiota: Diagnostic Markers and Commercial Potential.*** **Hedmark University College, Hedmark.**

Frøyland, C. J. (2011a). ***Coupling of oligonucleotide probes to MagPlex microspheres*****. Oslo.**

Frøyland, C. J. (2011b). ***Protocol on probe testing for the GA-map IBD assay*****. Oslo.**

Genetic Analysis. (2012). Retrieved 27.04.12, from http://www.genet-analysis.com

Granum, P. (2007). ***Matforgiftning: Næringsmiddelbårne infeksjoner og intoksikasjoner*****. Kristiansand: Høyskoleforlaget.**

Hait, L., Bousvaros, A., & Grand, R. (2005). Pediatric Inflammatory Bowel Disease: What Children Can Teach Adults. *Inflamm Bowel Dis, 11*, 519-527.

Invitrogen. (2008). Quant-iT PicoGreen dsDNA Reagent and Kits. from probes.invitrogen.com/media/pis/mp07581.pdf

Joossens, M., Huys, G., Cnockaert, M., Preter, V. D., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the feacal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut Online First*.

Khor, B., Gardet, A., & Xavier, R. J. (2011). Genetics and pathogenesis of inflammatory bowel disease. *Nature, 474*, 307-317.

Knobel, R. (2011). Interview with R. Knobel. Oslo.

Koloski, N. A., Bret, L., & Radford-Smith, G. (2008). Hygiene hypothesis in inflammatory bowel disease: A critical review of the literature. *World J Gastroenterol, 14*, 165-173.

Kreso, S. (2010). *ExoSAP treatment of 16S PCR products*. Oslo.

Kreso, S. (2012). Interview with S. Kreso.

Lea, T. (2006). *Immunologi og immunologiske teknikker*. Bergen: Fagbokforlaget.

Luminex. (2011). MAGPIX: Hardware Installation and User Manual. Retrieved 27.04.2012, from http://www.luminexcorp.com/prod/groups/public/documents/lmnxcorp/89-00002-00-237.pdf

Marteau, P., Lepage, P., Mangin, I., Suau, A., Doré, J., Pochart, P., et al. (2004). Review article: gut flora and inflammatory bowel disease. *Aliment Pharmacol Ther, 20*, 18-23.

Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D., & Garcia-Gil, J. (2006). Abnormal Microbiota Composition in the Ileocolonic Mucosa of Crohn's Disease Patients as Revealed by Polymerase Chain Reaction-Denaturing Gradient Gel Electrophoresis. *Inflamm Bowel Dis, 12*, 1136-1145.

Matarese, L. E., & Kandil, H. M. (2012). Preventing Bacteria-Induced Disease With Bacteria. *Nutrition in Clinical Practice, 27*, 242-246.

Melting Temperature. (1998). Retrieved 27.04.2012, from http://www.sigmaaldrich.com/img/assets/17240/meltingtemp.pdf

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature reviews, 11*, 31-46.

Mikhailov, T. A., & Furner, S. E. (2009). Breastfeeding and genetic factors in the etiology of inflammatory bowel disease in children. *World J Gastroenterol, 15(3)*, 270-279.

Mondot, S., Kang, S., Furet, M. S., Aguirre de Carcer, D., McSweeney, C., Morrison, M., et al. (2011). Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflamm Bowel Dis, 17*, 185-192.

Multiplexed Solutions for Life - Luminex Corporation. (2012). Retrieved 27.04.2012, from http://www.luminexcorp.com/index.htm

Nestestog, R. (2010a). *16S rRNA amplification by the CoverAll primers*. Oslo.

Nestestog, R. (2010b). *Gel Electrophoresis and DNA quantification*. Oslo.

Nestestog, R. (2010c). *Pico Green DNA quantification*. Oslo.

Nieuwenhuis, E. E. S., & Escher, J. C. (2008). Early onset IBD: What's the difference? *Digestive and Liver Disease, 40*, 12-15.

Nortvedt, R., Brakstad, F., Kvalheim, O. M., & Lundstedt, T. (1996). *Anvendelser av Kjemometri innen forskning og industri*.

Nwosu, F. C. (2011). ***Identification of gut microbiota markers for Inflammatory Bowel Diseases in children: Early diagnostic potentials.*** **Hedmark University College, Hedmark.**

Olsen, G. J., & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. ***FASEB, 7*****(1), 113-123.**

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the Human Infant Intestinal Microbiota. ***Plos Biology, 5*****(7), 1556-1573.**

Patel, J. B. (2001). 16S rRNA Gene Sequencing of Bacterial Pathogen Identification in the Clinical Laboratory. ***Molecular Diagnosis, 6*****(4), 313-321.**

Perminow, G. (2010). ***IBD in Children: Epidemiology and Mucosal Immunopathology. Inflammatory Bowel South Eastern Norway-II (IBSEN-II)*****. Oslo: University of Oslo.**

Perminow, G., Beisner, J., Koslowski, M., Lyckander, L. G., Stange, E., Vatn, M. H., et al. (2010). Defective Paneth cell mediated host defense in pediatric ileal Crohn's disease. ***The Americal Journal of Gastroenterology, 105*****, 452-459.**

Perminow, G., Brackmann, S., Lyckander, L. G., Franke, A., Borthne, A., Rydning, A., et al. (2009). A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from South-Eastern Norway, 2005-07, showing increased incidence in Crohn's disease. ***Scandinavian Journal of Gastroenterology, 44*****, 446-456.**

Perminow, G., Reikvam, D. H., Lyckander, L. G., Brandtzaeg, P., Vatn, M. H., & Carlsen, H. S. (2009). Increased Number and Activation of Colonic Macrophages in Pediatric Patients with Untreated Crohn's Disease. ***Inflamm Bowel Dis, 15*****, 1368-1378.**

progressivehealth. (2012). Retrieved 24.04.2012, from blog.progressivehealth.com/wp-content/uploads/Colon12.jpg

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. ***Nature, 464*****, 59-65.**

Ravikumara, M., & Sandhu, B. K. (2006). Epidemiology of Inflammatory Bowel Disease in Childhood. ***Indian Journal of Pediatrics, 73*****, 717-721.**

Rudi, K., & Isaksen, M. (2012). Gut Microbiota in Disease Diagnostics. ***New Advances in the Basic and Clinical Gastroenterology*****.**

Rudi, K., Zimonja, M., & Næs, T. (2006). Alignment-independent bilinear multivariate modelling (AIBIMM) for global analyses of 16S rRNA gene phylogeny. ***International Journal of Systematic and Evolutionary Microbiology, 56*****, 1565-1575.**

Rudi, K., Zimonja, M., Trosvik, P., & Næs, T. (2007). Use of multivariate statistics for 16S rRNA gene analysis of microbila communities. ***International Journal of Food Microbiology, 120*****, 95-99.**

Russel, M. G. V. M. (2000). Changes in the incidence of inflammatory bowel disease: what does it mean? ***European Journal of Internal Medicine, 11*****, 191-196.**

Sartor, R. B. (1997). Enteric Microflora in IBD: Pathogens or Commensals? ***Inflammatory Bowel Disease, 3*****(3), 230-235.**

Sartor, R. B. (2006). Mechanisms of Disease: pathogenesis of Crohn's disease and ulcerative colitis. ***Nature Clinical Practice Gastroenterology & Hepatology, 3*****, 390-407.**

Sartor, R. B. (2008). Microbial Influences in Inflammatory Bowel Diseases. *Reviews in Basic and Clinical Gastroenterology, 134*, 577-594.

Sartor, R. B. (2011). Key questions to guide a better understanding of host-commensal microbiota interactions in intestinal inflammation. *Mucosal Immunology, 4*, 127-132.

Sawczenko, A., & Sandhu, B. K. (2003). Presenting features of inflammatory bowel disease in Great Britain and Ireland. *Arch Dis Child, 88*, 995-1000.

Schatten, A. (2012). IBD incidence in Norway.

Schmidt, R. (2012). Testing a Test: Beyond Sensitivity and Specificity (Power Point Presentation ed.).

Schreiber, S., P., R., Albrecht, M., Hampe, J., & Krawczak, M. (2005). Genetics of Crohn Disease: An archetypal inflammatory barrier disease. *Nature, 6*, 376-388.

Schwiertz, A., Jacobi, M., Frick, J., Richter, M., Rusch, K., & Köhler, H. (2010). Microbiota in Pediatric Inflammatory Bowel Disease. *The Journal of Pediatrics, 157*, 240-244.

Sekelja, M. (2012). Interview with M. Sekelja (GA). Oslo.

Shikhare, G., & Kugathasan, S. (2010). Inflammatory bowel disease in children: current trends. *J Gastroenterol, 45*, 673-682.

Stange, E., Travis, S., Vermeire, S., Beglinger, C., Kupcinskas, L., Geboes, K., et al. (2006). European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *Gut, 55(Suppl 1)*, i1-i15.

Tannock, G. W. (2008). The search for disease-associated compositional shift in bowel bacterial communities of humans. *Trends in Microbiology, 16*(10), 488-495.

Vebø, H. C. (2012). Interview with H. Vebø (GA). Oslo.

Vebø, H. C., Sekelja, M., Nestestog, R., Storrø, O., Johnsen, R., Øien, T., et al. (2011). Temporal Development of the Infant Gut Microbiota in Immunoglobulin E-Sensitized and Nonsensitized Children Determined by GA-Map Infant Array. *Clinical and Vaccine Immunology, 18*, 1326-1335.

Wang, R., Beggs, M. L., Erickson, B. D., & Cerniglia, C. E. (2004). DNA microarray analysis of predominant human intestinal bacteria in fecal samples. *Molecular and Cellular Probes, 18*, 223-234.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losick, R. (2008). *Molecular Biology of the Gene*: Pearson International Edition.

Xiong, M., Zhao, Z., Arnold, J., & Yu, F. (2010). Next-Generation Sequencing. *Journal of Biomedicine and Biotechnology*.

Zhou, J. (2003). Microarrays for bacterial detection and microbial community analysis. *Current Opinion in Microbiology, 6*, 288-294.

Zoetendal, E. G., Rajilic-Stojanovic, M., & Vos, W. M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut, 57*, 1605-1615.

13 APPENDICES

APPENDIX A - Diagnostic criteria according to Porto criteria

Due to lack of consistent data from different studies there is important to have a common understanding for inflammatory bowel disease. Porto criteria was therefor developed where an agreement for diagnostic criteria was established for children (Escher, et al., 2005). Children from IBSEN-II were classified as CD, UC or IBD unclassified (IBDU) according to Porto criteria with following criteria (Escher, et al., 2005).

The presence of two or more of the following criteria were diagnosed CD

- Clinical features including abdominal pain, diarrhoea and weight loss.
- Macroscopic appearance at operation or endoscopy: segmental, discontinuous, and/or patchy lesions with or without rectal involvement, discrete or apthous ulcerations, fissuring or penetrating lesions, cobblestone or strictures.
- Radiological evidence of stenosis in the small bowel, segmental colitis or findings of fistulae.
- Histologic evidence of transmural inflammation or epithelial granulomas with giant cells.

The presence of three or more of the following criteria were diagnosed UC

- A history of diarrhoea and or blood/pus in stool.
- Macroscopic appearance at endoscopy, with continuous mucosal inflammation affecting the rectum in continuity with some of the entire colon.
- Microscopic features on biopsy compatible with UC.
- No suspicion of CD on small bowel roentgenography, ileocolonscopy, or biopsy.

Patients diagnosed with IBDU

- Patients with inconclusive or divergent endoscopy and histopathology according to CD or UC criteria were classified as IBDU.

Children who did not fulfilled the diagnostic criteria for IBD and who displayed no evidence of IBD, were included as a patient control group (non-IBD) (Perminow, Brackmann, et al., 2009).

APPENDIX B - MAGPIX instrument

The heart of a MAGPIX instrument is build up of a sample chamber, magnet, light-emitting diodes (LEDs) and CCD chamber (Figure A1). The instrument analyzes 96 samples in a microtiter plate where one and one sample is analyzed at a time. The whole process takes approximately one hour. One well contains a probe set in addition to one sample from cohort. In each well there are at least 50 microsphere per one bacterium and normally the number of microspheres are more than those who will be depicted. A sample probe (name of the instrument that retrieves the mixtures) goes into each well on the microtiter plate and absorbs totally 50 µl microspheres (mixture of the different microspheres). The microspheres will further be transported in the optics module where a magnet holds the magnetic microspheres in place while a red and green LED illuminate them. The red laser is first out to measure internal dye fluorescence for classification and distinguish the 50 different microspheres. Next, the green laser excites surface and reports fluorescence. When the magnet holds the magnetic microspheres, the magnetic microspheres are washed before illumination to reduce disturbance. During each illumination, the magnetic microspheres are imaged. After the images are recorded, the magnet withdraws and releases the microspheres (Luminex, 2011).

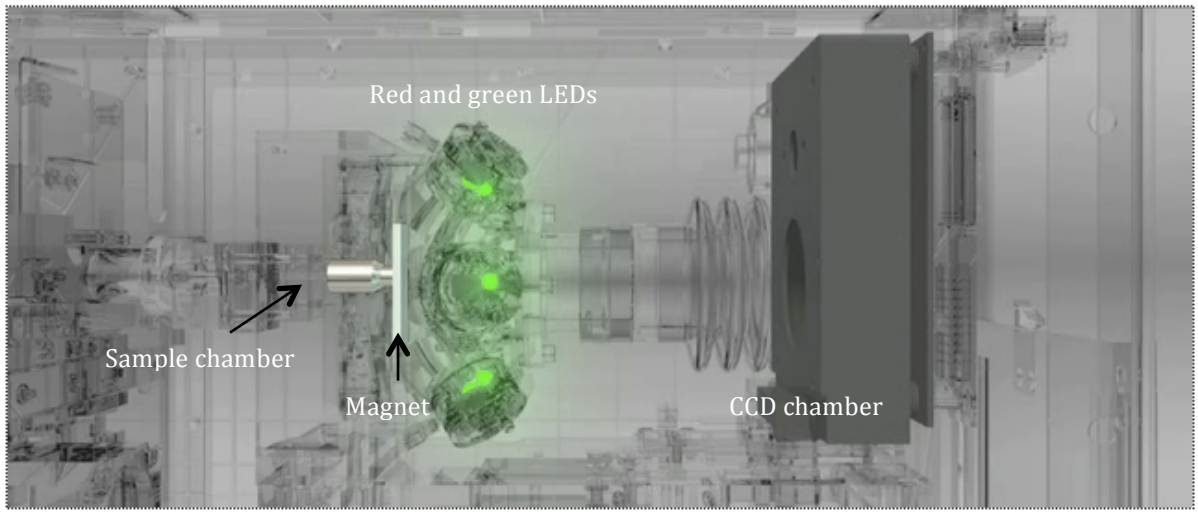


Figure A1: The heart of MAGPIX instrument is build up by a sample chamber, magnet, LED and CCD chamber ("Multiplexed Solutions for Life - Luminex Corporation," 2012) where the microspheres is illuminated and analyzed.

APPENDIX C – TNTProbeTool

TNTProbeTool is based on principal component analysis (PCA) for discrimination and classification of bacteria. PCA is a multivariate statistical method that visualizes the main structure in data with redundancies, and is well suited to build models that can be used for classification (Rudi, Zimonja, & Næs, 2006). One bacterium and its related sequence is represented as one spot in the PCA-plot in TNTProbeTool. Nearby sequences will fall in the same area of the PCA-plot and forms clusters on the basis of 16S rRNA phylogeny as seen for the four phyla, Actinobacteria, Firmicutes, Protebacteria and Bacteroides (Figure A2). In order to classify sequences for phylogeny, the sequences are transformed to pentamer frequency and then performed PCA on the pentamer frequency. TNTProbeTool uses the pentamer frequency to make a PCA-plot but the program has access to the real sequence for each bacterium at about 1,500 bp (16S rRNA gene) (Sekelja, 2012). The user defines target bacterium among target group and leaves all other bacteria as non-target (Figure A3). When the target bacterium is chosen, the program makes a probe suggestion.

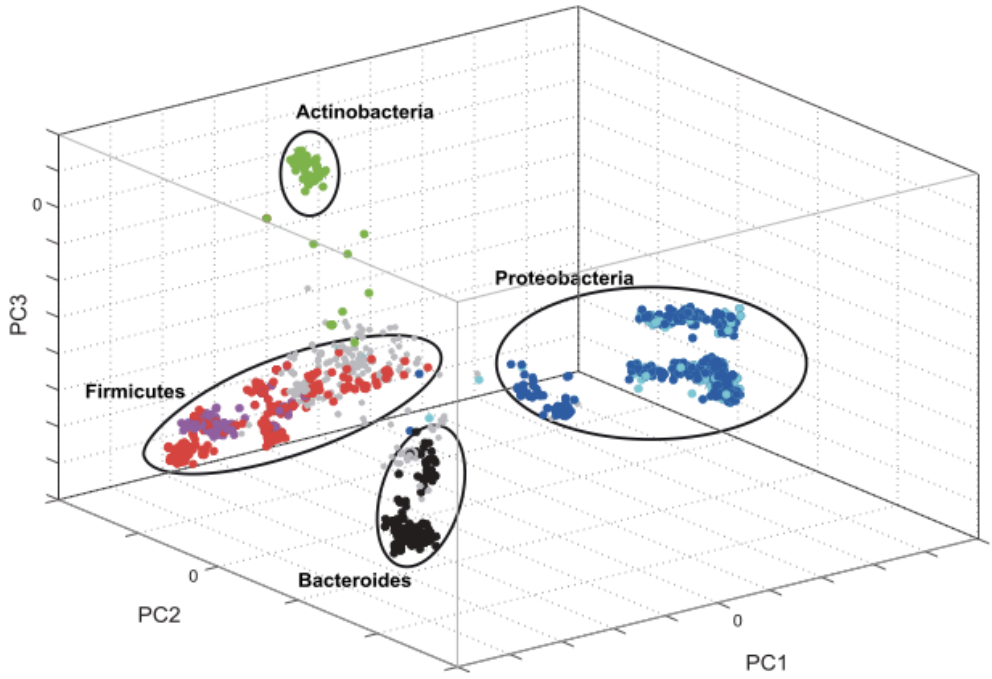


Figure A2: Score plot of the four phyla presented from the human GI tract where they form clusters (Vebø, et al., 2011).

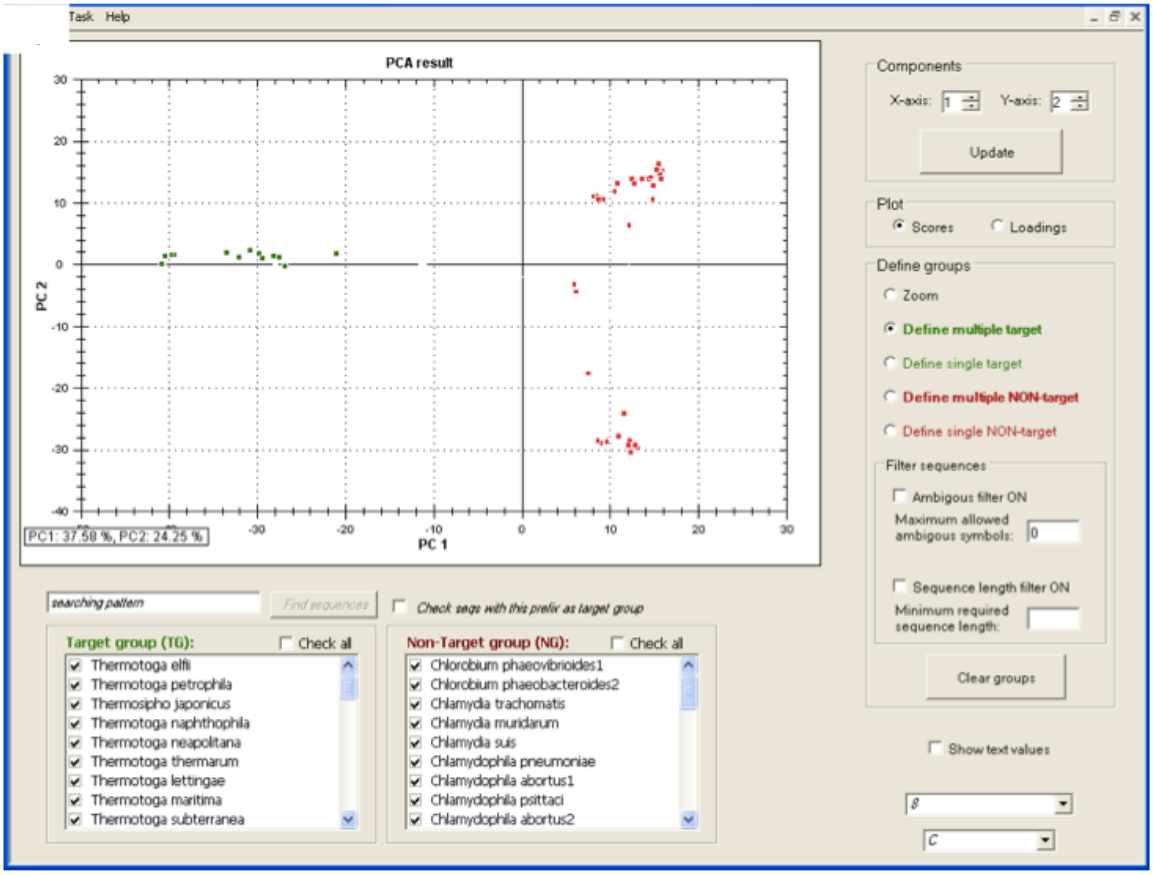


Figure A3: Illustration of TNTProbeTool software program where bacteria are represented in a PCA-plot based on their pentamer-frequencies. Defined target and non-target groups are shown in green and red, respectively (Vebø, et al., 2011).

To be able to design a probe set for identifying bacteria related to inflammatory bowel disease it is a prerequisite to have DNA sequences (16S rRNA gene) of all bacteria expected to find in the human gut, incorporated in TNTProbeTool. For this purpose there have been used several sources: The Human Microbiome Project, articles on the human intestinal microbiota published in international journals, sequence databases (NCBI and RDP10) and results from Genetic Analysis's own research (Vebø, et al., 2011).

APPENDIX D – Examples on result from capillary gel electrophoresis

Probes are tested against self-hybridization, target bacterium, and non-target bacteria. Examples for results from gel electrophoresis can be seen in Figure A4. Hybridization requires signals up to 7,000 – 8,000. If no signal is present one can observe noise that reflect signals up to 20.

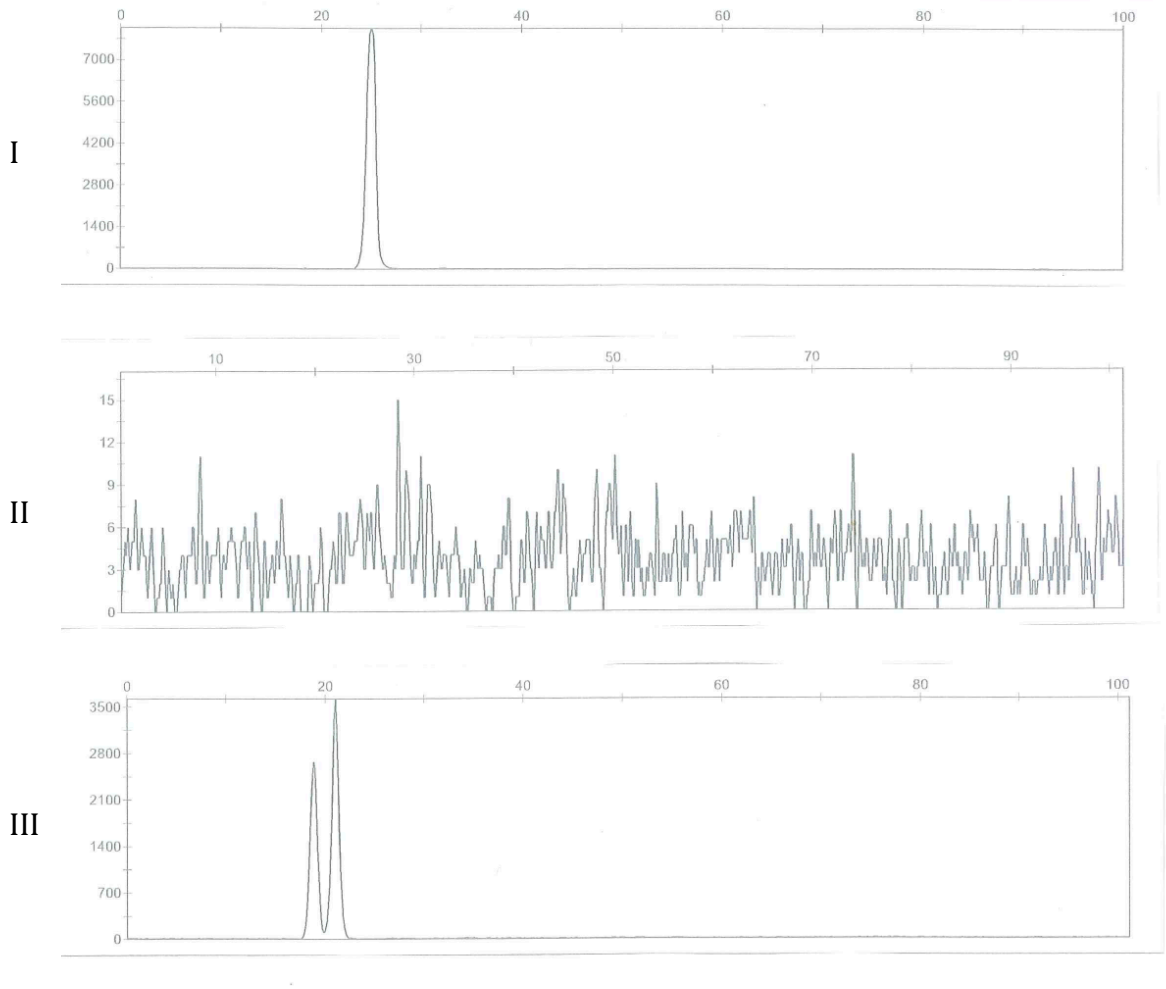


Figure A4: Examples on results from gel capillary electrophoresis. I – Hybridization, e.g. a probe has hybridized to target bacterium, II – No hybridization, e.g. probe tested for self-hybridization, III – Double peak

APPENDIX E – Multivariate data analysis

E.1 Principal Component Analysis (PCA)

PCA is used to compress large amount of data without loosing significant information. PCA can be used to identify patterns in data, and expressing the data in such a way, to highlight the structures, similarities, trends, outliers (describes a spot that lies outside the overall pattern of observations), differences or relationship between samples and variables (Nortvedt, et al., 1996). PCA describes the variation among the data by finding super-variables that maximize the explained variance in the data (Sekelja, 2012).

Samples form a swarm in a 3-dimentional space. A line seeks through the swarm and describes most of variance after the data are centralized. The meaning of centralized data is that the origin in the coordinate is moved to the average of all the samples. The first principal component (PC1) has the largest possible variance (that is, accounts for much of the variability in the data) in the length of the swarm. The second principal component has the highest variance possible under the constraint that it is orthogonal (perpendicular) to the preceding component (PC1) and explains the direction with second most variance (Nortvedt, et al., 1996). Error or noise that cannot be modeled as reasonable structures is called residual (Sekelja, 2012).

To get an overview of the data, it is often informative to plot two or three objects/variables in a coordinate system (usually this is the main variation in the data), to see how they are placed in relation to each other (Nortvedt, et al., 1996). This is performed by use of super-variables on the basis of the original variables. Two different plots (score plot and loading plot) are used to present super-variable results, and can later be compared.

E.2 Partial Least Squares – Discriminant Analysis (PLS-DA)

PLS-DA is a classical regression that decomposes a data matrix in a similar way as the PCA but uses additional information (such as diagnose) to find super-variables. PLS-DA has also much in common with the classical discriminant analysis method LDA (Linear discriminant analysis). LDA finds linear combinations of variables that characterize or separates two or more samples, based on one or more response variables, and the resulting combination can be used as a linear classifies. However, LDA can become unstable if the correlation between variables (probes) is high. The main difference between PLS-DA and LDA is that PLD-DA

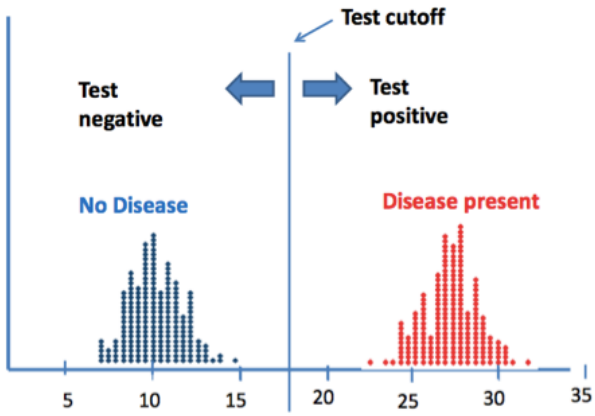
uses super-variables that are found by using information in one or more response variables, to control the separation of the samples (Sekelja, 2012).

APPENDIX F – Pre-processing and plate variation

Samples are randomly distributed into different plates. Each time one plate is placed on Magpix instrument there will be introduced some small changes (noise) that are not a result of biological variation. Noise can be described as variations that occurs between the samples as a result of e.g. temperature and pressure in the room, the time each plate is deferred before it is analyzed, variation in pipetting, etc. The small differences that occur naturally between the plates are reduced trough pre-processing, which means that signals (as a consequence from the noise), not relevant to the analysis, is deleted. In each well there is a hybridization control probe that is used to control the efficiency of hybridization, that is, the amount of used fluorescence during hybridization reaction. The Magpix instrument use less than one minute for the derivation of signal in each well, which introduces variability between each well (from well 1A to well 12H = approximately one hour) with respect to how long they will wait and hence, how long the hybridization takes place in each well (Sekelja, 2012).

APPENDIX G – Classification

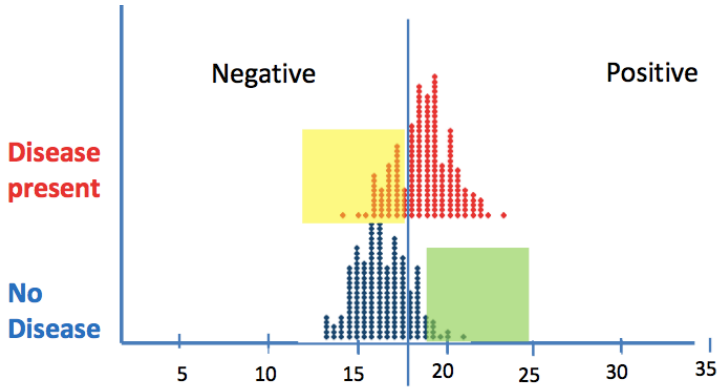
In a perfect situation there is possible to distinguish 100 % between healthy and sick people, as seen in Figure A5 where the two groups are placed at each side for the test cutoff.



	Disease		
Test	Present	Absent	Total
Positive	200	0	200
Negative	0	200	200
Total	200	200	

Figure A5: 100 % distinguish between healthy and sick people, presented with 200 people in each group. Illustration (Schmidt, 2012) is a made example and do not represent the present work.

Unfortunately, this is not the reality. A more realistic model would look more like figure A6. In this model, 40 healthy people (green area) are classified as sick (false negative), while 52 sick people (yellow area) are classified as healthy (false positive). Classification error is in this case 0,23 ((40+52)/400) or 23 %.



	Disease		
Test	Present	Absent	Total
Positive	148	40	188
Negative	52	160	212
Total	200	200	

Figure A6: Healthy and sick people together with misclassified people (yellow and green area). Illustration (Schmidt, 2012) is a made example and do not represent the present work.

APPENDIX H – Cohort

Cohort identification and diagnosis (UC, CD, IBDU and non-IBD).

Cohort ID	Diagnosis: 1= UC, 2=CD, 3= IBDU, 4=NonIBD	Sick=1, Healthy=2
28	2	1
35	4	2
50	4	2
58	4	2
79	1	1
87	4	2
104	3	1
111	2	1
136	4	2
158	4	2
163	2	1
180	2	1
191	4	2
192	1	1
195	4	2
202	2	1
207	4	2
1013	1	1
1028	2	1
2024	2	1
6006	4	2
6010	2	1
6023	1	1
6033	1	1
6035	2	1
6036	2	1
7	2	1
20	4	2
32	4	2
54	4	2
72	4	2
88	4	2
119	1	1
188	2	1
196	4	2
198	1	1
205	4	2
206	2	1
1005	2	1
2025	3	1
6001	1	1
6003	2	1
6007	1	1
6009	4	2
6013	2	1
6015	4	2
6017	1	1
6019	2	1

6026	1	1
6028	2	1
6030	2	1
6038	1	1
6039	2	1
71	2	1
99	4	2
126	1	1
130	4	2
131	2	1
172	2	1
181	2	1
187	4	2
201	4	2
1006	4	2
1030	2	1
6004	4	2
6005	1	1
6008	4	2
6011	1	1
6014	3	1
6016	4	2
6018	2	1
6021	4	2
6022	2	1
6032	2	1
6034	2	1

APPENDIX I – List of species-probes designed in TNTProbeTool

Overview of probes designed at species level in TNTProbeTool. Presents bacterium (target) name and number of designed probes, and comments to probe search.

Bacteria – species level	Number of probe designed	Comment to probe search
Acidimicrobidae bacterium Ellin7143	12 probes	
Acinetobacter junii	15 probes	
Actinobacterium GWS-BW-H99		Not found in target group
Actinomyces oxydans		Not found in target group
Akkermansia muciniphila	22 probes	
Alistipes onderdonkii		Not found in target group
Alistipes putredinis		Probe exists at GA
Bacillus licheniformis	2 probes	
Bacillus megaterium	4 probes	
Bacterium mpn-isolate group 18		Not found in target group
Bacterium mpn-isolate group 19		Not found in target group
Bacterium mpn-isolate group 5	4 probes	Nearby probes deleted
Bacteroides acidofaciens	3 probes	
Bacteroides distasonis		Probe exists at GA
Bacteroides fragilis		Probe exists at GA
Bacteroides ovatus		Probe exists at GA
Bacteroides stercoris		Probe exists at GA
Bacteroides thetaiotaomicron	4 probes	
Bacteroides uniformis		Probe exists at GA
Bacteroides vulgatus		Probe exists at GA
Bacteroides zoogleoformans	2 probes	
Bifidobacterium adolescentis	3 probes	
Bifidobacterium bifidum	2 probes	
Bifidobacterium infantis	5 probes	
Bifidobacterium longum		Probe cannot be designed, neither 55 °C
Blautia coccoides	2 probes	
Butyrate-producing bacterium L2-7	3 probes	
Butyrate-producing bacterium SR1/1		Not found in target group
Butyrivibrio fibrisolvens	6 probes	
Catenibacterium mitsuokai	9 probes	
Clostridium aminophilum	3 probes	
Clostridium clostridiiforme	2 probes	
Clostridium dificile	6 probes	
Clostridium leptum		Probe exists at GA
Clostridium limosum		Not found in target group
Clostridium litorale		Not found in target group
Clostridium lituseburense	2 probes	
Clostridium methylpentosum	2 probes	
Clostridium nexile		Probe exists at GA
Clostridium orbiscindens	2 probes	
Clostridium ramosum	3 probes	

Clostridium sp. SS2/1	14 probes	
Clostridium spiroforme		Not found in target group
Clostridium thermosuccinogenes		Not found in target group
Clostridium viride	5 probes	
Clostridium xylanolyticum	2 probes	
Dialister invisus	7 probes	
Dorea formicigenerans		Probe exists at GA
Dorea longicatena		Probe exists at GA
Drinking water bacterium Y7	19 probes	
Enterobacter cowanii	2 probes	
Enterococcus durans	1 probe	
Enterococcus faecalis		Probe exists at GA
Enterococcus faecium	2 probes	Nearby probes deleted
Enterococcus saccharolyticus	3 probes	
Eubacterium biforme	10 probes	
Eubacterium cylindroids	12 probes	
Eubacterium desmolans		Not found in target group
Eubacterium rectale		Probe exists at GA
Faecalibacterium prausnitzii		Probe exists at GA
Gamma proteobacterium DD103	13 probes	Named Gammaproteobacteria in further testing, but this probe includes only Gamma proteobacterium DD103
Helicobacter bilis	1 probe	
Helicobacter canis	1 probe	
Helicobacter ganmani	5 probes	
Helicobacter hepaticus	1 probe	Nearby probes deleted
Helicobacter trogontum	2 probes	Nearby probes deleted
Klebsiella pneumonia	1 probe	Nearby probes deleted
Lachnospira pectinoschiza	2 probes	
Lachnospiraceae bacterium A4	8 probes	
Lachnospiraceae Incertae Sedis	1 probe	Probe designed at 50-55 °C, unspecific
Lactobacillus acidophilus	1 probe	
Lactobacillus fermentum	3 probes	
Lactobacillus johnsonii	7 probes	
Lactobacillus reuteri	1 probe	
Lactobacillus ruminis	4 probes	
Listeria innocua	1 probe	Design is not very specific and capture other Listeria species
Listeria ivanovii	1 probe	Design is not very specific and capture other Listeria species
Listeria monocytogenes	1 probe	Design is not very specific and capture other Listeria species
Listeria seeligeri	1 probe	Design is not very specific and capture other Listeria species
Listeria welshimeri	1 probe	Design is not very specific and capture other Listeria species
Nocardioides sp. NS/27	11 probes	
Novosphingobium sp. K39	12 probes	
Oscillibacter valericigenes	5 probes	
Oscillospira guilliermondii	1 probe	
Prevotella bivia	7 probes	
Prevotella intermedia	5 probes	
Prevotella nigrescens	4 probes	

Prevotella ruminicola	9 probes	
Proteus vulgaris	5 probes	Probe designed at 55 °C, not specific
Pseudomonas straminea	2 probes	
Roseburia cecicola	7 probes	
Roseburia hominis	2 probes	
Ruminococcus albus	4 probes	
Ruminococcus bromii	3 probes	
Ruminococcus callidus	13 probes	
Ruminococcus gnavus		Probe exists at GA
Ruminococcus obeum	1 probe	
Ruminococcus torques		Probe exists at GA
Sphingomonas paucimobilis	13 probes	
Sphingomonas sp. AO1		Not found in target group
Sutterella wadsworthensis	15 probes	
Vibrio campbellii	1 probe	Nearby probes deleted
Wolinella succinogenes	2 probes	

APPENDIX J – List of genus/class/order-probes designed in TNTProbeTool

Bacteria overview from probe designed in TNTProbeToll at genus/class/order level with bacteria (target) name, number of designed probes and comments to probe search.

Bacteria - genus/class/order level	Number of designed probes	Comment to probe search
Aeromonas (genus)	2 probes	
Alistipes (genus)		Not found in target group
Alphaproteobacteria (class)		Not found in target group
Asteroleplasma (genus)	32 probes	
Bacillus (genus)	1 probe	Includes Bacillus cereus and Bacillus anthracis, designed at 55 °C
Bacteroides (genus)		Probe exists at GA
Bifidobacterium (genus)	3 probes	
Blautia (genus)	1 probe	
Campylobacter (genus)	1 probe	
Citrobacter (genus)	1 probe	Includes Citrobacter amalonaticus and Citrobacter diversus, designed at 58 °C
Clostridium (genus)		Probe exists at GA
Dialister (genus)	5 probes	
Dorea (genus)	1 probe	Designed at 55 °C
Enterococcus (genus)	1 probe	Includes Enterococcus avium, Enterococcus durans, Enterococcus faecium, Enterococcus raffinosus and Enterococcus saccharolyticus
Erysipelotrichi (class)		Not found in target group
Escherichia (genus)	6 probes	
Eubacterium (genus) gr 1	1 probe	Includes Eubacterium rectale, Eubacterium rectale and Eubacterium ventriosum
Eubacterium (genus) gr 2	8 probes	Includes Eubacterium biforme, Eubacterium cylindroides and Eubacterium dolichum
Fusobacterium (genus)	1 probe	Designed at 55 °C
Haemophilus (genus)	2 probes	Includes Haemophilus influenzae and Haemophilus aggregatibacter
Helicobacteraceae (family)	1 probe	Includes Helicobacter billis, Helicobacter canadensis, Helicobacter canis,Helicobacter cinaedi, Helicobacter ganmani, Helicobacter hepaticus, Helicobacter pullorum, Helicobacter pylori, Helicobacter trogontum and Helicobacter winghamensis
Lachnospiraceae (family)		Probe cannot be designed
Lactobacillus (genus)	1 probe	Includes Lactobacillus fermentum, Lactobacillus ultunensis, Lactobacillus ruminis, Lactobacillus salvivarius, Lactobacillus acidophilus, Lactobacillus helveticus, Lactobacillus amylolytic, Lactobacillus reuteri, Lactobacillus antri, Lactobacillus johnsonii and Lactobacillus paraplantarum
Listeria (genus)	2 probes	
Mannheimia (genus)	5 probes	Includes Mannheimia glucosida and Mannheimia haemolytica
Parabacteroides (genus)	2 probes	
Prevotella (genus)	2 probes	
Roseburia (genus)	4 probes	
Ruminococcus (genus)	2 probes	Includes Ruminococcus lactaris, Ruminococcus lactaris ATTCC, Ruminococcus torques and Ruminococcus torques ATCC
Shigella (genus)	1 probe	Designed at 55 °C, nearby probes deleted
Sphingomonas (genus)	1 probe	No genus probe designed, only Sphingomonas paucimobilis included. Nearby probes deleted
Subdoligranulum (genus)		Probe exists at GA
Veilonella (genus)	8 probes	

APPENDIX L – Bacteria for core probe set (GA-map™ Core assay)

CORE - probe set	
Bacteria	**Probe ID**
Alistipes putredinis (Parabacteroides distasonis)	AG0339
Anaerotruncus colihominis	AG0342
Bacteroides caccae?	AG0350
Bacteroides capillosus (Clostridium leptum)	AG0353
Bacteroides dorei group/Bacterioides sp. D4	AG0358
Bacteroides fragilis	AG0377
Bacteroides intestinalis?	AG0383
Bacteroides pectinophilus (Suboligranulum variable)	AG0393
Bacteroides sp 2 2 4/Bacteroides ovatus/Bacteroides sp. D1 (Bacteroides thetaiotaomicron)	AG0396
Bacteroides sp 4 3 47FAA/Bacteroides vulgatus (Parabacteroides distasonis/johnsonii)	AG0403
Bacteroides sp 9 1 4?	AG0406
Bacteroides stercoris?	AG0416
Bacteroides vulgatus/Bacteroides caccae/Bacteroides intestinalis +++	AG0443
Blautia hansenii	AG0447
Butyrivibrio crossotus	AG0454
Clostridium asparagiforme?	AG0457
Clostridium leptum (Ruminococcus bromii)	AG0470
Clostridium sp M63 Butyrate Producing	AG0488
Clostridium scindens	AG0508
Clostridium sp L250	AG0515
Collinsella aerofaciens?	AG0531
Coproccoccus comes	AG0495
Coprococcus eutactus	AG0569
Dorea formicigenerans (Dorea longicatena)	AG0581
Dorea longicatena (Ruminococcus bromii)	AG0587
Enterococcus faecalis	AG0591
Eubacterium hallii	AG0608
Eubacterium rectale?	AG0620
Eubacterium siraeum	AG0638
Eubacterium siraeum?	AG0439
Eubacterium ventriosum	AG0642
Faecalibacterium cf prausnitzii/Subdoligranulum variabile (Eubacterium siraeum)	AG0651
Holdemania filiformis	AG0654
No signal?	AG0756
Parabacteroides distasonis?	AG0679
Parabacteroides johnsonii/Parabacteroides merdae	AG0686
Parabacteroides merdae	AG0693
Roseburia intestinalis (Clostridium sp. SS2)	AG0701
Ruminococcus gnavus (Eubacterium ventriosum)	AG0703
Ruminococcus lactaris	AG0707
Ruminococcus torques	AG0714
Streptococcus thermophilus/Coprococcus comes/Clostridium nexile	AG0732
Subdoligranulum variabile	AG0744
Universal	UNI04

APPENDIX M – Bacteria for infant probe set (GA-map™ Infant assay)

INFANT - probe set	
Bacteria	**Probe ID**
Actinobacteria	IG0028
Anaerococcus prevotii	IG0095
Bacteroides	IG0060
Bifidobacterium (not B. breve)	IG0027
Bifidobacterium (not B. longum)	IG0030
Clostridiales and Veillonella	IG0058
Clostridiales and Veillonellaceae	IG0044
Clostridium butyricum	IG0163
Clostridium sporogenes	IG0103
Enterococcus faecalis	IG0171
Finegoldia magna	IG0107
Firmicutes (separates on Listeria, Veillonella and some Clostridiales)	IG0023
Firmicutes (separates on Streptococcus and some Clostridiales)	IG0012
Gamma-proteo group	IG0178
Gamma-proteobacteria subgroup	IG0056
Gemella	IG0042
Haemophilus	IG0008
Klebsiella pneumonia/Aeromonas	IG0181
Klebsiella oxytoca/Pantoea agglomerans	IG0195
Lactobacilli	IG0053
Listeria sp. (and E. coli/Shigella)	IG0021
Proteobacteria	IG0005
Pseudomonas/Morganella morganii	IG0039
Salmonella/Citrobacter/Cronobacter/Enterobacter/Morganella	IG0011
Serratia marcescens	IG0204
Shigella and E.coli	IG0133
Staphylococcus and Gemella	IG0024
Staphylococcus epidermidis, Eubacterium rectale, Clostridium SS2, Streptococcus agalactiae	IG0063
Streptococcus (α-haemolytic)	IG0197
Streptococcus (β-haemolytic groupA and B, α-haemolytic)	IG0079
Streptococcus (β-haemolytic groupB, α-haemolytic)	IG0022
Streptococcus agalacticae and Eubacterium rectale	IG0081
Streptococcus sanguinis and thermophilus	IG0020
Universal	UNI04
Veillonella	IG0051

APPENDIX K - Probe overview from capillary gel electrophoresis (target-bacteria)

Color code: Blue - optimal probe. Yellow - excluded due to self-hybrid, double top, weak signal. Green - tested at GA before. Grey - excluded, bacterium not available. Pink - problem with LIZ or signal under cut-off

Phyla: Genus(g), Species(s), Class(c)

B. - Bifidobacterium

Color code	ID	Target info	Phyla	Self-hybrid (Yes/No)	Comment self-hybrid	Target 1 bacterium	Sign. 1	Target 2 bacterium	Sign. 2	Target 3 bacterium	Sign. 3	Target 4 bacterium	Sign. 4	Target 5 bacterium	Sign. 5	Target 6 bact.	Sign. 6	Target 7 bact.	Sign. 7	Target 8 bact.	Sign. 8	Target 9 bact.	Sign. 9	Target 10 bact.	Sign. 10	Target 11 bacterium	Sign. 11
	AG0770	Acidimicrobidae bacterium Ellin	s	N		Target bacterium not available																					
	AG0774	Acidimicrobidae bacterium Ellin	s	N		Target bacterium not available																					
	AG0775	Acidimicrobidae bacterium Ellin	s	N	small peak at >100 bp	Target bacterium not available																					
	AG0777	Acinetobacter junii	s	N	small peak at >100 bp	Acinetobacter junii	7434																				
	AG0784	Acinetobacter junii	s	N	small peak at >100 bp	Acinetobacter junii	4273																				
	AG0783	Acinetobacter junii	s	N	small peak at >100 bp	Acinetobacter junii	1792/1483																				
	AG0792	Aeromonas	g	Y	>200 peak at 100 bp + >1000 peak at 200 bp																						
	AG0793	Aeromonas	g	N	small peak at >100 bp	Aeromonas veronii	1251	Aeromonas caviae	1714																		
	AG0815	Akkermansia muciniphila	s	N		Akkermansia muciniphila	2890																				
	AG0798	Akkermansia muciniphila	s	N		Akkermansia muciniphila	3119																				
	AG0801	Akkermansia muciniphila	s	N		Akkermansia muciniphila	7331																				
	AG1226	Alistipes	g	N		Genus probe excluded																					
	AG1225	Alistipes onderdonkii	s	N		Alistipes onderdonkii	8975																				
		Alistipes putredinis	s																								
	AG0819	Asteroleplasma	g	N		Target bacterium not available																					
	AG0827	Asteroleplasma	g	N	small peak at >100 bp	Target bacterium not available																					
	AG0829	Asteroleplasma	g	N		Target bacterium not available																					
	AG0848	Bacillus	g	N	small peak at >100 bp	Genus probe excluded																					
	AG0849	Bacillus licheniformis	s	N	small peak at >100 bp	Bacillus licheniformis	1153	Bacillus licheniformis	no signal																		
	AG0850	Bacillus licheniformis	s	N		Bacillus licheniformis	511	Bacillus licheniformis	no signal																		
	AG0851	Bacillus megaterium	s	N		Bacillus megaterium	2591																				
	AG0852	Bacillus megaterium	s	N		Bacillus megaterium	840																				
	AG0854	Bacillus megaterium	s	N		Bacillus megaterium	6612																				
	AG0856	Bacterium mpn-isolate group 5	s	N		Target bacterium not available																					
	AG0857	Bacterium mpn-isolate group 5	s	N	small peak at >100 bp	Target bacterium not available																					
	AG0858	Bacterium mpn-isolate group 5	s	N	small peak at >100 bp	Target bacterium not available																					
	AG0859	Bacteroides acidofaciens	s	N	700 peak at >100 bp	Bacteroides acidofaciens	39																				
	AG0860	Bacteroides acidofaciens	s	N		Bacteroides acidofaciens	76/94																				
	AG0861	Bacteroides acidofaciens	s	N		Bacteroides acidofaciens	2736																				
		(Para)Bacteroides distasonis	s																								
		Bacteroides fragilis	s																								
		Bacteroides ovatus	s																								
		Bacteroides stercoris	s																								
		Bacteroides thetaiotaomicron	s																								
		Bacteroides uniformis	s																								
		Bacteroides vulgatus	s																								
	AG0862	Bacteroides zoogleoformans	s	N		Bacteroides zoogleoformans	2474																				
	AG0863	Bacteroides zoogleoformans	s	N		Bacteroides zoogleoformans	7274																				
	AG0864	Bifidobacterium	g	N	small peak at >100 bp	Bifidobacterium adolescentis		B. angulatum	1700	B. animalis	2200	B. bifidum	2000	B. breve	2400	B. catenulatum	1600	B. dentium	2500	B. gallicum	double top	B. infantis	1500	B. longum	1800	B. pseudocatenulatum	1000
	AG0865	Bifidobacterium	g	N		Bifidobacterium adolescentis	5955	B. angulatum	7559	B. animalis subsp lactis	7339	B. bifidum	8333	B. breve	8195	B. catenulatum	4443	B. dentium	6245	B. gallicum	7065	B. infantis	7830	B.longum	7977	B. pseudocatenulatum	7022
	AG0867	Bifidobacterium adolescentis	s	N		Bifidobacterium adolescentis	1157/1900/984	B. adolescentis																			
	AG0868	Bifidobacterium adolescentis	s	N	small peak at 22 bp	Bifidobacterium adolescentis	720/1681	B. adolescentis																			
	AG0870	Bifidobacterium bifidum	s	N		Bifidobacterium bifidum	4430																				
	AG0871	Bifidobacterium bifidum	s	N		Bifidobacterium bifidum	2109																				
	AG0872	Bifidobacterium infantis	s	N		Bifidobacterium infantis	30																				
	AG0873	Bifidobacterium infantis	s	N		Bifidobacterium infantis	46																				
	AG0876	Bifidobacterium infantis	s	N		Bifidobacterium infantis	462																				
	AG1224	Blautia	g	N		Genus probe excluded																					
	AG1230	Blautia	g	N		Genus probe excluded																					
	AG0878	Blautia coccoides	s	N		Blautia coccoides	690																				
	AG0879	Blautia coccoides	s	N		Blautia coccoides	2576																				
	AG1227	Blautia schinkii	s	N		Blautia schinkii	1977/2137																				
	AG1228	Blautia schinkii	s	N		Blautia schinkii	6943																				
	AG1229	Blautia schinkii	s	N		Blautia schinkii	8366																				
	AG0880	Butyrivibrio fibrisolvens	s	N		Butyrivibrio fibrisolvens	726/262	Butyrivibrio fibrisolvens																			
	AG0881	Butyrivibrio fibrisolvens	s	N		Butyrivibrio fibrisolvens	668/239	Butyrivibrio fibrisolvens																			
	AG0884	Butyrivibrio fibrisolvens	s	N		Butyrivibrio fibrisolvens	914/365	Butyrivibrio fibrisolvens																			
	AG0886	Campylobacter	g	N		Campylobacter coli	427	Campylobacter concisus	2216	Campylobacter curvus	266	Campylobacter hominis	no signal	Campylobacter jejuni	2303	Campylobacter upsaliensis	4287										
	AG0893	Catenibacterium mitsuokai	s	N		Catenibacterium mitsuokai	2236/2377																				
	AG0895	Catenibacterium mitsuokai	s	N		Catenibacterium mitsuokai	7846																				
	AG0894	Catenibacterium mitsuokai	s	N		Catenibacterium mitsuokai	3791																				
	AG0896	Citrobacter	g	N		Citrobacter braakii	555	Citrobacter freundii	15	Citrobacter youngae	480																
	AG0898	Clostridium gr 2	g	N		probe excluded																					
	AG0899	Clostridium aminophilum	s	N		Clostridium aminophilum	3048																				
	AG0900	Clostridium aminophilum	s	N		Clostridium aminophilum	4357																				
	AG0902	Clostridium clostridioforme	s	N		Clostridium clostridioforme	31																				
	AG0903	Clostridium clostridioforme	s	N		Clostridium clostridioforme	no signal																				
	AG0905	Clostridium difficile	s	N		Clostridium difficile	1441																				
	AG0907	Clostridium difficile	s	N		Clostridium difficile	3494																				
	AG0908	Clostridium difficile	s	N		Clostridium difficile	6342																				
	AG0897	Clostridium gr 1	g	N		probe excluded																					
		Clostridium leptum	s																								
	AG1231	Clostridium limosum	s	N		Target bacterium not available																					
	AG0910	Clostridium lituseburense	s	N		Clostridium lituseburense	1164/653/377	Clostridium lituseburense																			
	AG0911	Clostridium lituseburense	s	N		Clostridium lituseburense	1228/1042/466	Clostridium lituseburense																			
	AG0912	Clostridium methylpentosum	s	N		Clostridium methylpentosum	7876																				
	AG0913	Clostridium methylpentosum	s	N		Clostridium methylpentosum	1607																				

		Clostridium nexile	s																		
	AG0914	Clostridium orbiscindens	s	N		Clostridium orbiscindens	1165/1800/665	Clostridium orbiscindens													
	AG0915	Clostridium orbiscindens	s	N		Clostridium orbiscindens	no signal	Clostridium orbiscindens													
	AG0917	Clostridium ramosum	s	N		Clostridium ramosum	693/754														
	AG0916	Clostridium ramosum	s	N		Clostridium ramosum	5887														
	AG0918	Clostridium ramosum	s	N		Clostridium ramosum	4142														
	AG1221	Clostridium sp. SS2-1	s	N		Clostridium sp. SS2-1	no signal														
	AG1223	Clostridium sp. SS2-1	s	N		Clostridium sp. SS2-1	no signal														
	AG1219	Clostridium sp. SS2-1	s	-	no signal	Clostridium sp. SS2-1	2721														
	AG0922	Clostridium viride	s	N		Clostridium viride	385/306														
	AG0919	Clostridium viride	s	N		Clostridium viride	7320														
	AG0921	Clostridium viride	s	N		Clostridium viride	9221														
	AG0924	Clostridium xylanolyticum	s	N		Clostridium xylanolyticum	1314/1188														
	AG0925	Clostridium xylanolyticum	s	N		Clostridium xylanolyticum	194														
		Coprococcus	g																		
	AG0926	Dialister	g	N		Genus probe excluded															
	AG0928	Dialister	g	N		Genus probe excluded															
	AG0930	Dialister	g	N		Genus probe excluded															
	AG0932	Dialister invisus	s	N		Dialister invisus	1931														
	AG0931	Dialister invisus	s	N		Dialister invisus	6122														
	AG0938	Dorea	g	N		Genus probe excluded															
		Dorea formicigenerans	s																		
		Dorea longicatena	s																		
	AG0939	Drinking water bacterium Y7	s	N		Target bacterium not available															
	AG0942	Drinking water bacterium Y7	s	N		Target bacterium not available															
	AG0948	Drinking water bacterium Y7	s	N		Target bacterium not available															
	AG0958	Enterobacter cowanii	s	N		Enterobacter cowanii	148														
	AG0959	Enterobacter cowanii	s	N		Enterobacter cowanii	41/47														
	AG0960	Enterococcus	g	N		Genus probe excluded															
	AG0961	Enterococcus durans	s	N		Enterococcus durans	2976														
		Enterococcus faecalis	s																		
	AG0963	Enterococcus faecium	s	Y	>7500 peak																
	AG0962	Enterococcus faecium	s	N		Enterococcus faecium	6299														
	AG0964	Enterococcus saccharolyticus	s	N		Enterococcus saccharolyticus	no signal X 2														
	AG0965	Enterococcus saccharolyticus	s	N		Enterococcus saccharolyticus	no signal X 2														
	AG0966	Enterococcus saccharolyticus	s	N		Enterococcus saccharolyticus	190														
	AG0968	Escherichia	g	N		Citrobacter youngae		Cronobacter sakazakii	4248	Enterobacter cancerogenus	3031	Enterobacter cowanii	6152	Escherichia coli	2444	Escherichia fergusonii	3841	Salmonella enteritidis	2610/1260/872		
	AG0972	Escherichia	g	N		Citrobacter youngae		Cronobacter sakazakii	8392	Enterobacter cowanii	8399	Escherichia coli	8627	Escherichia fergusonii	8405	Serratia marcescens	6357				
	AG0967	Escherichia	g	N		Citrobacter diversus	4369	Citrobacter youngae	341/520/1343/832	Cronobacter sakazakii	5756	Enterobacter cancerogenus	2634	Enterobacter cowanii	no signal	Escherichia coli	2094	Escherichia fergusonii	2909		
	AG0975	Eubacterium biforme	s	N		Eubacterium biforme	365/776														
	AG0980	Eubacterium biforme	s	Y	436 peak																
	AG0974	Eubacterium biforme	s	N		Eubacterium biforme	794														
	AG0983	Eubacterium cylindroides	s	N		Eubacterium cylindroides	2289/919/872	Eubacterium cylindroides													
	AG0985	Eubacterium cylindroides	s	N		Eubacterium cylindroides	1377/784/476	Eubacterium cylindroides													
	AG0986	Eubacterium cylindroides	s	N		Eubacterium cylindroides	885/1130/745	Eubacterium cylindroides													
	AG0995	Eubacterium gr 1	g	N		Genus probe excluded															
	AG0996	Eubacterium gr 2	g	N		Genus probe excluded															
	AG0999	Eubacterium gr 2	g	N		Genus probe excluded															
	AG1002	Eubacterium gr 2	g	N		Genus probe excluded															
		Eubacterium rectale	s																		
		Faecalibacterium prausnitzii	s																		
	AG1007	Fusobacterium	g	N		Fusobacterium gonidiaformans	157	Fusobacterium mortiferum	228	Fusobacterium necrophorum subsp funduliforme	800	Fusobacterium nucleatum	500	Fusobacterium ulcerans	794	Fusobacterium varium	no signal				
	AG1017	Gammaproteobacteria	c	N		Target bacterium not available															
	AG1018	Gammaproteobacteria	c	N		Target bacterium not available															
	AG1020	Gammaproteobacteria	c	N		Target bacterium not available															
	AG1021	Haemophilus	g	N		Haemophilus aphrophilus	3547	Haemophilus influenzae	2781	Mannheimia glucosida	3742	Mannheimia haemolytica	4550								
	AG1022	Haemophilus	g	N		Haemophilus aphrophilus	1369	Haemophilus influenzae	367	Mannheimia glucosida	1721	Mannheimia haemolytica	1585								
	AG1232	Helicobacter	g	N		Genus probe excluded															
	AG1023	Helicobacter bilis	s	N	100 peak at 100 bp	Helicobacter bilis	30														
	AG1024	Helicobacter canis	s	N		Helicobacter canis	3483														
	AG1026	Helicobacter ganmani	s	N		Target bacterium not available															
	AG1028	Helicobacter ganmani	s	N		Target bacterium not available															
	AG1029	Helicobacter ganmani	s	N		Target bacterium not available															
	AG1030	Helicobacter hepaticus	s	N		Helicobacter hepaticus	2359														
	AG1031	Helicobacter trogontum	s	N		Helicobacter trogontum	2650														
	AG1032	Helicobacter trogontum	s	N		Helicobacter trogontum	6919														
	AG1033	Klebsiella pneumoniae	s	N		Klebsiella pneumoniae	1975														
	AG1034	Lachnospiraceae Incertae Sedis	-	N		Clostridium aminophilum	3794	Coprococcus eutactus	2309	Eubacterium hallii	1383	Eubacterium rectale	2082	Roseburia hominis	1680	Clostridium aldenense	890	Clostridium saccharolyticum	2504	Coprococcus comes	2502
	AG1035	Lachnospira pectinoschiza	s	N	100 peak at 100bp	Target bacterium not available															
	AG1036	Lachnospira pectinoschiza	s	N		Target bacterium not available															
	AG1038	Lachnospiraceae bacterium A4	s	N	small peak at >100 bp	Target bacterium not available															
	AG1039	Lachnospiraceae bacterium A4	s	N		Target bacterium not available															
	AG1042	Lachnospiraceae bacterium A4	s	N		Target bacterium not available															
	AG1045	Lactobacillus	g	N	130 peak at >100 bp	Genus probe excluded															
	AG1046	Lactobacillus acidophilus	s	N		Lactobacillus acidophilus	2404														
	AG1047	Lactobacillus fermentum	s	Y	8298 peak at 12 bp + 79 bp																
	AG1048	Lactobacillus fermentum	s	N	peak 1000 at >100 bp	Lactobacillus fermentum	1171/379														
	AG1049	Lactobacillus fermentum	s	N	peak 150 >100 bp	Lactobacillus fermentum	1966/1068														
	AG1051	Lactobacillus johnsonii	s	N		Lactobacillus johnsonii	3852														
	AG1052	Lactobacillus johnsonii	s	N		Lactobacillus johnsonii	1779														
	AG1056	Lactobacillus johnsonii	s	N		Lactobacillus johnsonii	242														

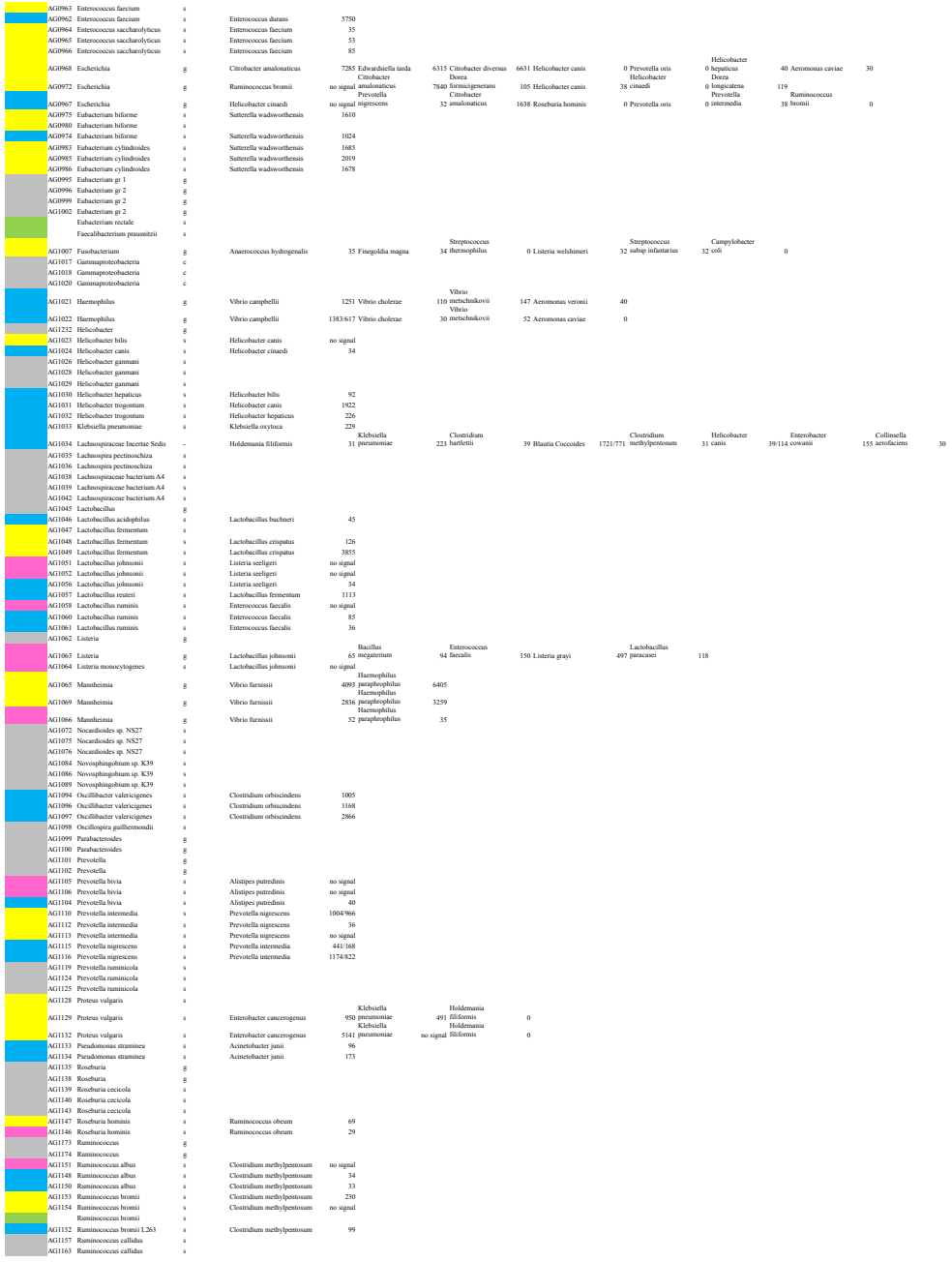
AG1057	Lactobacillus reuteri	s	N		Lactobacillus reuteri	972								
AG1058	Lactobacillus ruminis	s	N		Lactobacillus ruminis	3859								
AG1060	Lactobacillus ruminis	s	N		Lactobacillus ruminis	2289								
AG1061	Lactobacillus ruminis	s	N		Lactobacillus ruminis	2725								
AG1062	Listeria	g	N		Genus probe excluded									
AG1063	Listeria	g	Y	183 peak at 25 bp	Listeria innocua	8057	Listeria ivanovii	5728	Listeria seeligeri	4678	Listeria monocytogenes	6618	Listeria welshimeri	6078
AG1064	Listeria monocytogenes	s	N		Listeria monocytogenes	7234								
AG1065	Mannheimia	g	Y	3000 peak at 30 bp	Mannheimia glucosida	3430	Mannheimia haemolytica	8215						
AG1069	Mannheimia	g	Y	1000 peak at 30 bp	Mannheimia glucosida	8200	Mannheimia haemolytica	8563						
AG1066	Mannheimia	g	N		Mannheimia glucosida	7351	Mannheimia haemolytica	no signal						
AG1072	Nocardioides sp. NS27	s	N		Target bacterium not available									
AG1075	Nocardioides sp. NS27	s	N		Target bacterium not available									
AG1076	Nocardioides sp. NS27	s	N	two peaks at 80 bp and 90 bp	Target bacterium not available									
AG1084	Novosphingobium sp. K39	s	N		Target bacterium not available									
AG1086	Novosphingobium sp. K39	s	N		Target bacterium not available									
AG1089	Novosphingobium sp. K39	s	N		Target bacterium not available									
AG1094	Oscillibacter valericigenes	s	N		Oscillibacter valericigenes	1054								
AG1096	Oscillibacter valericigenes	s	N		Oscillibacter valericigenes	1471								
AG1097	Oscillibacter valericigenes	s	N		Oscillibacter valericigenes	2506								
AG1098	Oscillospira guilliermondii	s	Y	400 peak at 22 bp	Target bacterium not available									
AG1099	Parabacteroides	g	N		Genus probe excluded									
AG1100	Parabacteroides	g	N		Genus probe excluded									
AG1101	Prevotella	g	N		Genus probe excluded									
AG1102	Prevotella	g	N		Genus probe excluded									
AG1105	Prevotella bivia	s	N		Prevotella bivia	4073								
AG1106	Prevotella bivia	s	N		Prevotella bivia	3063								
AG1104	Prevotella bivia	s	N		Prevotella bivia	6727								
AG1110	Prevotella intermedia	s	N		Prevotella intermedia	958/869	Prevotella intermedia							
AG1112	Prevotella intermedia	s	N		Prevotella intermedia	1368/1213	Prevotella intermedia							
AG1113	Prevotella intermedia	s	N		Prevotella intermedia	32	Prevotella intermedia							
AG1115	Prevotella nigrescens	s	N		Prevotella nigrescens	951								
AG1116	Prevotella nigrescens	s	N		Prevotella nigrescens	1089								
AG1119	Prevotella ruminicola	s	N		Target bacterium not available									
AG1124	Prevotella ruminicola	s	N		Target bacterium not available									
AG1125	Prevotella ruminicola	s	N		Target bacterium not available									
AG1128	Proteus vulgaris	s	Y	Tested in MFA study										
AG1129	Proteus vulgaris	s	N	Tested in MFA study	P. mirabilis	no signal	P. vulgaris	16						
AG1132	Proteus vulgaris	s	N	Tested in MFA study	P. mirabilis	no signal	P. vulgaris	no signal						
AG1133	Pseudomonas straminea	s	N		Pseudomonas straminea	8504								
AG1134	Pseudomonas straminea	s	N		Pseudomonas straminea	3723								
AG1135	Roseburia	g	N		Genus probe excluded									
AG1138	Roseburia	g	N		Genus probe excluded									
AG1139	Roseburia cecicola	s	N		Target bacterium not available									
AG1140	Roseburia cecicola	s	N		Target bacterium not available									
AG1143	Roseburia cecicola	s	N		Target bacterium not available									
AG1147	Roseburia hominis	s	N		Roseburia hominis	369								
AG1146	Roseburia hominis	s	-	no signal	Roseburia hominis	5829								
AG1173	Ruminococcus	g	N		Genus probe excluded									
AG1174	Ruminococcus	g	N		Genus probe excluded									
AG1151	Ruminococcus albus	s	N		Ruminococcus albus	5655								
AG1148	Ruminococcus albus	s	N		Ruminococcus albus	5405								
AG1150	Ruminococcus albus	s	N		Ruminococcus albus	2142								
AG1153	Ruminococcus bromii	s	N		Ruminococcus bromii	224								
AG1154	Ruminococcus bromii	s	N		Ruminococcus bromii	148								
	Ruminococcus bromii	s												
AG1152	Ruminococcus bromii L263	s	N		Ruminococcus bromii	7871								
AG1157	Ruminococcus callidus	s	N	70 peak at 70 bp	Target bacterium not available									
AG1163	Ruminococcus callidus	s	N	50 peak at 25 bp	Target bacterium not available									
AG1167	Ruminococcus callidus	s	N		Target bacterium not available									
	Ruminococcus gnavus	s												
	Ruminococcus obeum	s												
	Ruminococcus torques	s												
AG1175	Shigella	g	N	150 peak at >100 bp	Escherichia coli	1500	Escherichia fergusonii	1500	Salmonella bongori	no signal	Shigella flexneri	1156	Shigella sonnei	1450
AG1176	Sphingomonas	g	Y	peak 6700	Genus probe excluded									
AG1180	Sphingomonas paucimobilis	s	N		Sphingomonas paucimobilis	37								
AG1182	Sphingomonas paucimobilis	s	N		Sphingomonas paucimobilis	41/33								
AG1184	Sphingomonas paucimobilis	s	N		Sphingomonas paucimobilis	46								
AG1235	Sphingomonas sp. AO1	g	N		Target bacterium not available									
AG1236	Sphingomonas sp. AO1	g	N		Target bacterium not available									
AG1237	Sphingomonas sp. AO1	g	N		Target bacterium not available									
	Subdoligranulum	g												
AG1199	Sutterella wadsworthensis	s	N		Sutterella wadsworthensis	356/674/1351	Sutterella wadsworthensis							
AG1201	Sutterella wadsworthensis	s	N		Sutterella wadsworthensis	4271/1767/1986	Sutterella wadsworthensis							
AG1203	Sutterella wadsworthensis	s	N		Sutterella wadsworthensis	1673/900/464	Sutterella wadsworthensis							
AG1211	Veillonella	g	Y	peak 200 at >100 bp										
AG1207	Veillonella	g	N		Veillonella dispar	3791								
AG1209	Veillonella	g	N		Veillonella dispar	1500								
AG1214	Vibrio campbellii	s	N		Vibrio campbellii	1195/1822/1636	Vibrio campbellii							
AG1215	Wolinella succinogenes	s	N	peak 90 at 90 bp	Wolinella succinogenes	1413/1598/1030	Wolinella succinogenes							
AG1216	Wolinella succinogenes	s	N		Wolinella succinogenes	1846/1377/741	Wolinella succinogenes							

APPENDIX K – Probe overview from capillary gel electrophoresis (non-target)

Color code: Blue - optimal probe. Yellow - excluded due to self-hybrid, double top, weak signal. Green - tested at GA before. Grey - excluded, bacterium not available. Pink - problem with LIZ or signal under cut-off

Phyla: Genus(g), Species(s), Class(c)

Color code	ID	Target info	Phyla	Non-target 1	Sign non-t. 1	Non-target 2	Sign. 2	Non-target 3	Sign. 3	Non-target 4	Sign. 4	Non-target 5	Sign.. 5	Non-target 6	Sign. 6	Non-target 7	Sign. 7	Non-target 8	Sign. 8	Non-target 9	Sign. 9	Non-target 10	Sign. 10	Non-target 11	Sign. 11
	AG0770	Acidimicrobiidae bacterium Ellin	s																						
	AG0774	Acidimicrobiidae bacterium Ellin	s																						
	AG0775	Acidimicrobiidae bacterium Ellin	s																						
	AG0777	Acinetobacter junii	s	Pseudomonas straminea	53																				
	AG0784	Acinetobacter junii	s	Pseudomonas straminea	40																				
	AG0783	Acinetobacter junii	s	Pseudomonas straminea	62																				
	AG0792	Aeromonas	g																						
	AG0793	Aeromonas	g	Dorea longicatena	45	Campylobacter coli	no signal																		
	AG0815	Akkermansia muciniphila	s	Clostridium sp. SS2-1	no signal																				
	AG0798	Akkermansia muciniphila	s	Clostridium sp. SS2/1	51																				
	AG0801	Akkermansia muciniphila	s	Clostridium sp. SS2/1	66																				
	AG1226	Alistipes	g																						
	AG1225	Alistipes onderdonkii	s	Alistipes putredinis	32																				
		Alistipes putredinis	s																						
	AG0819	Asteroleplasma	g																						
	AG0827	Asteroleplasma	g																						
	AG0829	Asteroleplasma	g																						
	AG0848	Bacillus	g																						
	AG0849	Bacillus licheniformis	s	Bacillus alcalophilus	31																				
	AG0850	Bacillus licheniformis	s	Bacillus alcalophilus	94 (7bp)																				
	AG0851	Bacillus megaterium	s	Listeria monocytogenes	124 (7bp)																				
	AG0852	Bacillus megaterium	s	Listeria monocytogenes	50																				
	AG0854	Bacillus megaterium	s	Listeria monocytogenes	80																				
	AG0856	Bacterium mpn-isolate group 5	s																						
	AG0857	Bacterium mpn-isolate group 5	s																						
	AG0858	Bacterium mpn-isolate group 5	s																						
	AG0859	Bacteroides acidofaciens	s	Bacteroides stercoris	48																				
	AG0860	Bacteroides acidofaciens	s	Bacteroides stercoris	30																				
	AG0861	Bacteroides acidofaciens	s	Bacteroides coprophilus	44																				
		(Para) bacteroides distasonis	s																						
		Bacteroides fragilis	s																						
		Bacteroides ovatus	s																						
		Bacteroides stercoris	s																						
		Bacteroides thetaiotaomicron	s																						
		Bacteroides uniformis	s																						
		Bacteroides vulgatus	s																						
	AG0862	Bacteroides zoogleoformans	s	Bacteroides coprophilus	37																				
	AG0863	Bacteroides zoogleoformans	s	Bacteroides coprophilus	125 (7 bp)																				
	AG0864	Bifidobacterium	g																						
	AG0865	Bifidobacterium	g	Megamonas hypermegale	37	Clostridium saccharolyticum	48	Desulfovibrio piger	52	Clostridium ramosum	34	Actinomyces odontolyticus	45	Collinsella aerofaciens	57	Collinsella tanakaei	53	Gordonibacter pamelaeae	44	Campylobacter concisus	59	Vibrio metschnikovii	75	Helicobacter canis	45
	AG0867	Bifidobacterium adolescentis	s	Bifidobacterium infantis	37																				
	AG0868	Bifidobacterium adolescentis	s	Bifidobacterium infantis	99																				
	AG0870	Bifidobacterium bifidum	s	Bifidobacterium dentium	no signal																				
	AG0871	Bifidobacterium bifidum	s	Bifidobacterium dentium	1504																				
	AG0872	Bifidobacterium infantis	s	Bifidobacterium animalis	32																				
	AG0873	Bifidobacterium infantis	s	Bifidobacterium animalis	38																				
	AG0876	Bifidobacterium infantis	s	Bifidobacterium animalis	814																				
	AG1224	Blautia	g																						
	AG1230	Blautia	g																						
	AG0878	Blautia coccoides	s	Ruminococcus gnavus	39																				
	AG0879	Blautia coccoides	s	Ruminococcus obeum	35																				
	AG1227	Blautia schinkii	s	Collinsella tanakaei	39																				
	AG1228	Blautia schinkii	s	Collinsella tanakaei	no signal																				
	AG1229	Blautia schinkii	s	Collinsella tanakaei	no signal																				
	AG0880	Butyrivibrio fibrisolvens	s	Coprococcus comes	36																				
	AG0881	Butyrivibrio fibrisolvens	s	Coprococcus comes	71																				
	AG0884	Butyrivibrio fibrisolvens	s	Coprococcus comes	48																				
	AG0886	Campylobacter	g	Vibrio metschnikovii	no signal	Helicobacter canis	no signal	Subdoligranulum variabile	35	Streptococcus thermophilus	41	Helicobacter bilis	30	Wolinella succinogenes	1346/355										
	AG0893	Catenibacterium mitsuokai	s	Lactobacillus ultunensis	31																				
	AG0895	Catenibacterium mitsuokai	s	Lactobacillus ultunensis	no signal																				
	AG0894	Catenibacterium mitsuokai	s	Lactobacillus ultunensis	38																				
	AG0896	Citrobacter	g	Ruminococcus torques		Proteus vulgaris	79																		
	AG0898	Clostridium gr 2	g	Coprococcus comes		Coprococcus eutactus																			
	AG0899	Clostridium aminophilum	s	Ruminococcus obeum	51																				
	AG0900	Clostridium aminophilum	s	Ruminococcus obeum	38																				
	AG0902	Clostridium clostridioforme	s	Clostridium asparagiforme	33																				
	AG0903	Clostridium clostridioforme	s	Clostridium asparagiforme	no signal																				
	AG0905	Clostridium difficile	s	Pediococcus acidilactici	37																				
	AG0907	Clostridium difficile	s	Pediococcus acidilactici	32																				
	AG0908	Clostridium difficile	s	Pediococcus acidilactici	33																				
	AG0897	Clostridium gr 1	g	Coprococcus comes		Coprococcus eutactus																			
		Clostridium leptum	s																						
	AG1231	Clostridium limosum	s	Clostridium histolyticum	no signal																				
	AG0910	Clostridium lituseburense	s	Clostridium scindens	91																				
	AG0911	Clostridium lituseburense	s	Clostridium scindens	39																				
	AG0912	Clostridium methylpentosum	s	Clostridium hiranonis	36																				
	AG0913	Clostridium methylpentosum	s	Clostridium hiranonis	32																				
		Clostridium nexile	s																						
	AG0914	Clostridium orbiscindens	s	Bacteroides capillosus	210																				
	AG0915	Clostridium orbiscindens	s	Bacteroides capillosus	30																				
	AG0917	Clostridium ramosum	s	Catenibacterium mitsuokai	no signal																				
	AG0916	Clostridium ramosum	s	Catenibacterium mitsuokai	no signal																				
	AG0918	Clostridium ramosum	s	Catenibacterium mitsuokai	no signal																				
	AG1221	Clostridium sp. SS2-1	s	Coprococcus eutactus	37																				
	AG1223	Clostridium sp. SS2-1	s	Coprococcus eutactus	no signal																				
	AG1219	Clostridium sp. SS2-1	s	Coprococcus eutactus	45																				
	AG0922	Clostridium viride	s	Clostridium orbiscindens																					
	AG0919	Clostridium viride	s	Clostridium orbiscindens	1261																				
	AG0921	Clostridium viride	s	Clostridium orbiscindens	1524																				
	AG0924	Clostridium xylanolyticum	s	Clostridium bolteae	no signal																				
	AG0925	Clostridium xylanolyticum	s	Clostridium bolteae	65																				
		Coprococcus	g																						
	AG0926	Dialister	g																						
	AG0928	Dialister	g																						
	AG0930	Dialister	g																						
	AG0932	Dialister invisus	s	Catenibacterium mitsuokai	no signal																				
	AG0931	Dialister invisus	s	Catenibacterium mitsuokai	31																				
	AG0938	Dorea	g																						
		Dorea formicigenerans	s																						
		Dorea longicatena	s																						
	AG0939	Drinking water bacterium Y7	s																						
	AG0942	Drinking water bacterium Y7	s																						
	AG0948	Drinking water bacterium Y7	s																						
	AG0958	Enterobacter cowanii	s	Ruminococcus torques	no signal																				
	AG0959	Enterobacter cowanii	s	Ruminococcus torques	33																				
	AG0960	Enterococcus	g																						
	AG0961	Enterococcus durans	s	Vibrio cholerae	4331																				
		Enterococcus faecalis	s																						



	AG1167	Ruminococcus callidus	s										
		Ruminococcus gnavus	s										
		Ruminococcus obeum	s										
		Ruminococcus torques	s										
	AG1175	Shigella	g	Ruminococcus torques	42	Citrobacter amalonaticus	223/274	Klebsiella pneumoniae	0	Citrobacter youngae	413/455/1 047	Helicobacter trogontum	0
	AG1176	Sphingomonas	g										
	AG1180	Sphingomonas paucimobilis	s	Corynebacterium ammoniagenes	30								
	AG1182	Sphingomonas paucimobilis	s	Corynebacterium ammoniagenes	60								
	AG1184	Sphingomonas paucimobilis	s	Corynebacterium ammoniagenes	44								
	AG1235	Sphingomonas sp. AO1	g										
	AG1236	Sphingomonas sp. AO1	g										
	AG1237	Sphingomonas sp. AO1	g										
		Subdoligranulum	g										
	AG1199	Sutterella wadsworthensis	s	Faecalibacterium prausnitzii	34								
	AG1201	Sutterella wadsworthensis	s	Faecalibacterium prausnitzii	53								
	AG1203	Sutterella wadsworthensis	s	Faecalibacterium prausnitzii	48/77								
	AG1211	Veillonella	g										
	AG1207	Veillonella	g	Dialister invisus	no signal								
	AG1209	Veillonella	g	Dialister invisus	no signal								
	AG1214	Vibrio campbellii	s	Vibrio mimicus	40								
	AG1215	Wolinella succinogenes	s	Campylobacter curvus	33								
	AG1216	Wolinella succinogenes	s	Campylobacter curvus	31								

APPENDIX L – Relatedness of bacteria presented in GA-map™ IBD assay

Target group - species	Probe ID	Phylum	Class	Order	Family	Genus
Akkermansia muciniphila	AG0798	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Akkermansia
Catenibacterium mitsuokai	AG0894	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Catenibacterium
Clostridium aminophilum	AG0899	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
Clostridium difficile	AG0907	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
Clostridium methylpentosum	AG0912	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
Clostridium ramosum	IG0013	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
Dialister invisus	AG0931	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Dialister
Fusobacterium periodonticum	IG0113	Fusobacteria	Fusobacteriia	Fusobacterales	Fusobacteriaceae	Fusobacterium
Haemophilus (genus)	AG1021	Proteobacteria	Gammaproteobac.	Pasteurellales	Pasteurellaceae	Haemophilus
Lactobacillus johnsonii	AG1051	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
Lactobacillus ruminis	AG1058	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
Pseudomonas straminea	AG1133	Proteobacteria	Gammaproteobac.	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Ruminococcus albus	AG1148	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcus
Veillonella (genus)	AG1207	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Veillonella
Bifidobacteria	AG1219	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium
Alistipes putredinis	AG0339	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
Bacteroides fragilis	AG0377	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
Bacteroides stercoris	AG0416	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
Clostridium leptum	AG0470	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
Coprococcus comes	AG0495	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus
Coprococcus eutactus	AG0569	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus
Dorea formicigenerans	AG0581	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Dorea
Enterococcus faecalis	AG0591	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus
Eubacterium rectale	AG0620	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Eubacterium
Faecalibacterium prausnitzii	AG0651	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
Ruminococcus gnavus	AG0703	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcus

Subdoligranulum (genus)	AG0744	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Subdoligranulum
Bacillus licheniformis	AG0849	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Bacteroides acidofaciens	AG0861	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
Enterococcus faecium	AG0962	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus
Eubacterium biforme	AG0974	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Eubacterium
Helicobacter canis	AG1024	Proteobacteria	Epsilonproteobact.	Campylobacterales	Helicobacteraceae	Helicobacter
Helicobacter hepaticus	AG1030	Proteobacteria	Epsilonproteobact.	Campylobacterales	Helicobacteraceae	Helicobacter
Klebsiella pneumoniae	AG1033	Proteobacteria	Gammaproteobac.	Enterobacteriales	Enterobacteriaceae	Klebsiella
Lactobacillus reuteri	AG1057	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
Mannheimia (genus)	AG1066	Proteobacteria	Gammaproteobac.	Pasteurellales	Pasteurellaceae	Mannheimia
Blautia schinkii	AG1228	Firmicutes	Clostridia	Clostridiales	-	Blautia
Proteobacteria (phylum)	IG0005	Proteobacteria				
Haemophilus parainfluenza	IG0008	Proteobacteria	Gammaproteobac.	Pasteurellales	Pasteurellaceae	Haemophilus
Firmicutes 1 (phylum)	IG0012	Firmicutes				
Firmicutes 2 (phylum)	IG0023	Firmicutes				
Gammaproteobacteria (class)	IG0056	Proteobacteria	Gammaproteobac.			
Bacteroides (genus)	IG0060	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
E.coli/Shigella	IG0133	Proteobacteria	Gammaproteobac.	Enterobacteriales	Enterobacteriaceae	Escherichia/ Shigella
Blautia coccoides/Clostridium coccoides	AG0879	Firmicutes	Clostridia	Clostridiales		Blauti/ coccoides
Roseburia hominis	AG1146	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia
Ruminococcus bromii L263	AG1152	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcus
Bifidobacteria	IG0028	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium
Enterococcus (genus)	IG0014	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	
Universal probe	UNI05					