Title: DIAGNOSIS OF CROHN'S DISEASE

Abstract: The invention provides an isolated oligonucleotide comprising the nucleotide sequence (i) ACGCTTGACCCCT (SEQ ID NO 1), (ii) AGGGTGCAACGT (SEQ ID NO 2), (iii) ACGCTTGACCCCTC (SEQ ID NO 3), (iv) GAGGGTGCAACGCT (SEQ ID NO 4), or (v) a nucleotide sequence capable of hybridising under conditions of high stringency to any one of SEQ ID Nos 1, 2, 3 or 4. The invention also provides methods of obtaining information relevant to a diagnosis of Crohn's Disease in a subject and methods of diagnosing Crohn's Disease in a subject in which the oligonucleotides of the invention are used. Kits and oligonucleotide arrays comprising the oligonucleotides of the invention are also provided.
— with sequence listing part of description (Rule 5.2(a))
Diagnosis of Crohn’s Disease

The present invention relates to oligonucleotides and their use in the diagnosis of Crohn's Disease (CD). The oligonucleotides selectively hybridise to a sequence found in certain Proteobacteria and Flavobacteriaceae and so the oligonucleotides may be used as oligonucleotide probes or oligonucleotide primers to determine the levels of such bacteria in a sample from the gastrointestinal (GI) tract of a subject, elevated levels of such bacteria in the sample being indicative of CD.

Inflammatory Bowel Disease is a collection of chronic inflammatory disorders that affect the GI tract. The most commonly seen of the IBD disorders are Crohn's Disease and Ulcerative Colitis (UC). The symptoms can be diarrhoea, rectal bleeding, abdominal pain, weight loss, fever, fatigue, and/or anaemia. Extraintestinal manifestations such as joint, skin and eye disorders might also be present. IBD can also cause lethal conditions such as intestinal perforation, non-responsive rectal bleeding, toxic megacolon and colon cancer. No cure for IBD exists. The cause of IBD has not been found, but most evidence points to a combination of genetic predisposition, immunological factors, environmental triggers and conditions, and GI microbes.

Currently IBD diagnostics are mainly based on medical history, clinical evaluation, laboratory tests, endoscopy, radiology and histology. Endoscopy is usually required to make a confident diagnosis. However, because the symptoms of IBD are not uncommon in other diseases or conditions, several conditions can be confused with IBD, e.g. Irritable Bowel Syndrome (IBS). To expose subjects suspected of having IBD to standard invasive tests (e.g. endoscopy) when the majority of these have a functional disease such as IBS is unsatisfactory from a patient welfare and cost perspective.

In addition, while Crohn’s Disease and Ulcerative Colitis have several clinical and pathological differences, many similarities exist and so, even if IBD is confirmed, it is difficult to distinguish CD from UC. As a result, approximately 10% of IBD patients are misclassified. Increased mortality has been shown in CD, but not in UC, and so an accurate differential diagnosis is essential.
To further complicate matters, the symptoms of IBD, and CD and UC within IBD, vary depending on the localization and severity of the disease. Diarrhoea, abdominal pain and/or weight loss are the main symptoms of CD. In UC bloody diarrhoea is the primary symptom, but it might also be present in CD. Patients with UC might also have symptoms as rectal bleeding and rectal urgency. Passage of mucus and pus are more common in UC than CD. Additionally symptoms of malaise, loss of appetite and/or fever are common in CD but could be a sign of severe attack in UC. Fistulæ are common in CD and may occur in UC.

CD affects the whole thickness of the walls of the GI tract. It can involve any part of the GI tract from the mouth to the anus, particularly the most distal part of the small intestine (ileum) as well as the proximal part of the large intestine (colon). Unlike UC, there may be unaffected tract between areas of active disease called skip lesions. Severe complications may include narrowing of parts of the intestine (strictures), abnormal tunnels that connect organs (fistulas) and cracks in the anal skin (fissures).

CD is subclassified in terms of the area of the gastrointestinal tract that it affects. Ileocolic Crohn's disease affects both the ileum and the large intestine; Crohn's ileitis, affects the ileum only; Crohn's colitis, affects the large intestine (and may be particularly difficult to distinguish from ulcerative colitis); gastroduodenal CD causes inflammation in the stomach and the duodenum and jejunoileitis effects the jejunum. CD can also be subclassified in terms of the presentation of the disease. There are three categories of disease presentation in CD: stricturing, penetrating, and inflammatory. Stricturing disease causes narrowing of the bowel that may lead to bowel obstruction or changes in the calibre of the faeces. Penetrating disease creates abnormal passageways (fistulæ) between the bowel and other structures such as the skin. Inflammatory disease (or non-stricturing, non-penetrating disease) causes inflammation without causing strictures or fistulæ.

UC is characterized by diffuse inflammation of the colonic mucosa, the innermost layer that is in direct contact with the faecal flow. The disease typically starts in the rectum and can extend as a continuous inflammation to the whole length of the colon. However, some patients develop inflammation in the ileum and in 10-15% of IBD cases CD or UC cannot be distinguished with conventional diagnostic techniques.

Although microbial pathogens have been postulated to cause CD and UC since their original descriptions, it is now generally accepted that IBD arises from,
and is perpetuated by, interactions between host genetic and immune factors, environmental triggers and GI microbes.

Studies of experimental animal models of IBD reveal that germ-free animals show few signs of inflammation and experimental colitis is exhibited only when the animal is exposed to natural microbial communities. Likewise human studies have shown a response of IBD patients to antibiotic and probiotic treatment. In CD patients inflammation most commonly appears in the gut locations where bacterial concentrations are high. Furthermore diversion of the faecal stream from sections of the GI tract lumen is associated with improvement of the inflammation in those sections, indicating a role for bacteria in the IBD pathogenesis.

The types of GI microbes which influence IBD pathogenesis have not yet been adequately described, however some bacterial species have been proposed to have a role. Relatively few studies report single organisms to be responsible for IBD, especially few for UC. Additionally, conflicting results are reported, thus complicating the picture. Whether it is reasonable to expect a single organism to be able on its own to produce the myriad of signs and symptoms associated with IBD, should be considered and it has been speculated that the difficulty of finding one organism responsible for IBD is that more than one organism is involved. However, some organisms including *Mycobacterium avium* spp. *paratuberculosis*, *Escherichia coli*, *Faecalibacterium prausnitzii*, *Saccharomyces cerevisiae* and *Candida albicans* have been proposed to have singularly a role in IBD pathogenesis.

In contrast to the interpretation of IBD pathogenesis by a single microorganism, certain changes in the composition of the overall intestinal microbiota has been proposed to have a role in IBD pathogenesis. Although some changes in the composition of microbiota could also be a result of the disease, discrepancies between IBD and control GI microbiota could be used in a diagnostic test.

Metagenomic studies have shown that most bacteria in the human GI tract are represented by only four phyla regardless of disease state: Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria. Nearly half of the sequences have been shown to belong to just two subgroups: The order Bacteroidales and the family Lachnospiraceae.

Healthy humans have relatively stable temporal compositions of GI microbiota, but in the case of CD patients in remission, the stability of the faecal
microbiota varies greatly over time. It has also been observed that even though microflora have a high degree of diversity in both situations, dominant microflora varies markedly between remission and flare. Additionally IBD patients have shown higher concentrations of bacteria in mucosal and faecal samples. In contrast, a notable decrease in operational taxonomic units (OTU) in IBD faecal and mucosal microbiota, mainly reflected by fewer types of bacteria in the Firmicutes phylum have been reported. Further studies have observed that Fusobacteria and Verrucomicrobia phyla are associated with healthy subjects, increased number of Bacteroidetes is associated with IBD.

Previous studies have therefore suggested that neither UC nor CD is characterized by a uniform, stereotypical microbiota. Hence, to date, it is unclear whether microbiota of the human intestine alone can be used to diagnose IBD (or UC or CD, or distinguish between the two), or whether human genetic, serologic and immunologic factors have to be included.

As there is no single clear pathogenic marker of IBD, a gold standard for making the diagnosis does not exist and distinguishing between IBD suffers with CD as opposed to UC is even more difficult. Several diseases presenting similar symptoms, e.g. IBS, also need to be ruled out. IBS differs from IBD in that it does not cause inflammation, ulcers or organic damage to the GI tract. IBS diagnostics are based on symptoms and evaluation of organic abnormality. Recent antibiotic use may indicate pseudomembranous colitis, travel abroad might cause infectious colitis and abdominal pain associated with bowel movements could represent IBS. Loose stools for more than six weeks normally differentiate IBD from infectious diseases.

The recommended diagnosis of CD and UC is therefore based on clinical evaluation, laboratory findings and medical history and is confirmed by negative microbiological test, endoscopy and histological findings. Blood and fecal testing can yield some helpful information but is rarely conclusive, other testing is more invasive and/or costly.

The development of a non-invasive test that can accurately screen for IBD and separate CD from UC, preferably without the need for parallel invasive testing would be highly desirable. The lack of a simple and reliable test means that further tools to provide diagnostic information are still required.

Increasing amounts of evidence show disturbed GI microbiota in IBD patients. A comprehensive culture-independent study revealed significant
differences between microbiota of small intestine from CD, UC and non-IBD samples. However, as a whole, the literature is inconclusive and clear correlations between changes in microbiota and disease have not yet emerged. Moreover, the identification of nucleotide sequences characteristic of candidate bacteria which can be used as molecular markers to follow the levels of such correlative bacteria have not been identified.

Some studies using faecal samples indicate differences in the microbial communities between IBD patients and controls, but the relevance of faecal bacteria in IBD pathogenesis and diagnostics remains unclear as controversy exists as to whether mucosal or faecal microbiota are more suitable for IBD diagnostics.

The inventors have surprisingly found that in GI tract samples levels of certain Flavobacteriaceae and Proteobacteria, e.g. those from at least the taxonomic families Aeromonadaceae, Campylobacteraceae, Desulfovibrionaceae, Enterobacteriaceae, Helicobacteraceae, Pasteurellaceae and Pseudomonadaceae (i.e. those bacteria that contain a complementary target sequence for the oligonucleotides of the invention) are positively correlated with CD. In particular, subjects with CD have elevated levels of these Proteobacteria and Flavobacteriaceae in GI tract samples compared with samples from subjects without CD (i.e. non-IBD subjects or subjects with UC). The inventors have also identified oligonucleotides which are surprisingly effective in monitoring the levels of these bacteria in GI tract samples which permits or assists with diagnosis of CD.

Proteobacteria include bacteria from the classes Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria and Zetaproteobacteria, e.g. from the orders Aeromonadales, Campylobacterales, Desulfovibrionales, Enterobacterales, Pasteurellales and Pseudomonadales, e.g. from the families Aeromonadaceae, Campylobacteraceae, Desulfovibrionaceae, Enterobacteriaceae, Helicobacteraceae, Pasteurellaceae and Pseudomonadaceae and e.g. the genera Aeromonas, Campylobacter, Chryseobacterium, Citrobacter, Desulfovibrio, Edwardsiella, Enterobacter, Escherichia, Haemophilus, Hafnia, Helicobacter, Klebsiella, Morganella, Proteus, Providencia, Pseudomonas, Salmonella, Serratia, Shigella, Yersinia and Yokenella. The species Aeromonas sobria, Campylobacter jejuni, Chryseobacterium gleum, Citrobacter youngae, Desulmononas pigra, Edwardsiella tard, Enterobacter aerogenes, Enterobacter cancerogenus, Enterobacter hormaechei, Enterobacter hormaechei oharae, Enterohemorrhagia
Escherichia coli, Escherichia coli, Haemophilus parainfl, Hafnia alvei, Helicobacter pylori, Klebsiella oxytoxa, Klebsiella pneumoniae, Morganella morganii, Proteus mirabilis, Proteus penneri, Proteus vulgaris, Providencia rustigianii, Providencia stuartii, Pseudomonas gessardii, Pseudomonas gessardii, Salmonella bongori, Salmonella enterica subsp. Enterica, Salmonella enteritidis, Serratia marcescens, Shigella dysenteriae, Shigella sonnei, Yersinia pestis and Yokenella regensburgei contain a complementary target sequence for the oligonucleotides of the invention and are detectable therewith.

Flavobacteriaceae include bacteria from the genus Chryseobacterium, e.g. the species Chryseobacterium gleum, which contains a complementary target sequence for the oligonucleotides of the invention and is detectable therewith.

In a first aspect therefore the invention provides an isolated oligonucleotide comprising the nucleotide sequence ACGCTTGCACCT (SEQ ID NO 1) or the nucleotide sequence complementary thereto (AGGGTGCAAGCGT; SEQ ID NO 2) or a nucleotide sequence capable of hybridising under conditions of high stringency to SEQ ID NO 1 or SEQ ID NO 2. In some embodiments the isolated oligonucleotide will comprise the nucleotide sequence ACGCTTGCACCTC (SEQ ID NO 3) or the nucleotide sequence complementary thereto (GAGGGTGCAAGCGT; SEQ ID NO 4) or a nucleotide sequence capable of hybridising under conditions of high stringency to SEQ ID NO 3 or SEQ ID NO 4.

In the following, references to "the nucleotide sequence of SEQ ID NO 1/2/3 or 4" also include nucleotide sequences capable of hybridising under high stringency conditions to SEQ ID NO 2, 1, 4 or 3, respectively, unless the context dictates otherwise.

The oligonucleotides of the invention may be up to 100 nucleotides, preferably up to 80, 60, 50, 40, 30, 25, 24, 23, 22, 21, 20, 19 or up to 18 nucleotides in length. The oligonucleotides of the invention may be at least 9, preferably at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 60 or at least 80 nucleotides in length. The nucleotides of the oligonucleotides can be any type of nucleotide so long as hybridisation specificity or efficiency and, if necessary, primer dependent nucleic acid amplification efficiency or nucleic acid polymerisation efficiency is not detrimentally affected. The oligonucleotides may therefore be a deoxyribonucleotide, a ribonucleotide, modifications thereof (e.g., PNA, morpholino-, LNA) and mixtures thereof. DNA oligonucleotides and LNA modified DNA oligonucleotides are preferred.
The nucleotide sequence of SEQ ID NO 1, 2, 3 or 4 may be found in any part of the oligonucleotide so long as the oligonucleotide can hybridise to the target sequence of GAGGTGCAAGCGT (SEQ ID NO 4) or its complement ACGCTTGACCCTC (SEQ ID NO 3) and, if required, can effect a nucleic acid extension reaction. In some embodiments the 3' nucleotide of SEQ ID NOs 1, 2, 3 or 4 is the 3' nucleotide of the oligonucleotide.

In other embodiments the oligonucleotides consist essentially of SEQ ID NOs 1, 2, 3 or 4. Thus, the oligonucleotide will have a nucleotide sequence corresponding to SEQ ID NOs 1, 2, 3 or 4 and 1, 2, 3, 4, or 5 additional nucleotides. In other embodiments the oligonucleotide primers will consist of SEQ ID NOs 1, 2, 3 or 4, preferably SEQ ID NOs 1, 3 or 4, more preferably SEQ ID NO 1 or SEQ ID NO 3, most preferably SEQ ID NO 1.

Unless otherwise stated, or dictated by specific context, all nucleotide sequences are recited herein 5' to 3' in line with convention in this technical field.

The oligonucleotides may be labelled with a moiety to assist with detection or manipulation. A large number of suitable moieties and labelling methods are known in the art and described in the literature. Many moieties can perform both functions. Any detectable or signal-generating molecule or reporter molecule may be used. Convenient labels include colorimetric, chemiluminescent, chromogenic, radioactive and fluorescent labels, but enzymatic (e.g. colorimetric, luminescent, chromogenic) or antibody-based labelling methods or signal-generating systems may also be used. Thus the term "label" as used herein includes not only directly detectable signal-giving or passive moieties, but also any moiety which generates a signal or takes part in a signal generating reaction or that may be detected indirectly in some way. For instance the moiety may be biotin and detection may be indirect via streptavidin carrying a colorimetric, chemiluminescent, chromogenic, radioactive or fluorescent moiety.

The label can, in some embodiments, comprise a plurality of moieties that contributes to the overall detectable output of the label. By varying the identity and/or the relative proportions of these moieties, a wide palette of unique labels can be constructed. For instance, a plurality of dyes, e.g. luminescent (e.g. bioluminescent, chemiluminescent, photoluminescent, radioluminescent, sonoluminescent, etc.) which combine to give a unique electromagnetic spectral signature upon excitation may be used. By varying the proportions of the selected dyes further differentiation in the spectral signature can be achieved. Signatures
based on the absorption of certain wavelengths of electromagnetic radiation are also envisaged.

Fluorescein or other fluorescently labelled nucleotides are particularly suitable for incorporation into the primers, and allow detection directly by fluorescence or indirectly by antibody interactions. These are commercially available. Primers can be labelled by e.g. $^{35}$S, $^{3}$H or $^{32}$P as described in Syvänen, A.C. et al. Genomics 8, [1990], 684-692. Any binding moiety may be used as a label, for instance an antibody fragment, His-tag, biotin or streptavidin. These may be incorporated in the form of labelled nucleotides.

The oligonucleotides can also be immobilised on solid supports. Suitable immobilising supports to which the oligonucleotides can be attached are known in the art and include any of the well known supports or matrices which are currently widely used or proposed for immobilisation, separation etc. of oligonucleotides. These may take the form of particles, sheets, gels, filters, membranes, fibres, capillaries, chips or microtitre strips, slides, tubes, plates or wells etc. Methods of immobilising or attaching oligonucleotides to solid supports are likewise known in the art. Particularly preferred are DNA chips (microchips, glass chips) now common in molecular biology procedures. In other embodiments membrane strips on to which the oligonucleotides may be spotted and then UV cross-linked may be used.

Also particularly preferred are particles, e.g. beads on to which the oligonucleotides have been attached. In certain embodiments the particles also carry a label to permit their detection. Suitable particles are described in more detail later.

Preferably the support is magnetic (preferably paramagnetic or superparamagnetic), e.g. magnetic particles, for instance magnetic beads.

High stringency conditions for hybridisation are defined as 2x SSC/50% formamide at 50°C for binding conditions and 2 x SSC at 65°C for washing conditions (where SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.2).

In preferred embodiments the nucleotide sequences that can hybridise to SEQ ID NOs. 1, 2, 3 or 4 under high stringency conditions will hybridise to all, or substantially all, e.g. at least 10, 11, 12, 13 or 14 contiguous nucleotides of the nucleotide sequences of SEQ ID NOs. 1, 2, 3 or 4, respectively.

Viewed alternatively, nucleotide sequences that can hybridise to the nucleotide sequences of SEQ ID NOs. 1, 2, 3 or 4 under high stringency conditions
may be those nucleotide sequences that correspond to the nucleotide sequence of
SEQ ID NOs. 2, 1, 4 or 3, respectively but with up to 8 bases (adenine,
thymine/uracil, guanine, or cytosine) in the nucleotide sequences of SEQ ID NOs.
2, 1, 4 or 3, being substituted with a different base. Preferably there will be up to 7,
6, 5, 4, 3 or 2 substituted bases or only a single base substitution. The base being
substituted into the sequence can be any standard or non-standard, naturally
occurring or synthetic base.

Nucleotide sequences that can hybridise to SEQ ID NOs. 1, 2, 3 or 4 under
high stringency conditions will preferably be 10, 11, 12, 13 or 14 nucleotides in
length, and consist of a contiguous part of the nucleotide sequence of SEQ ID NOs.
2, 1, 4 or 3, respectively, with the above described substitutions.

Preferably the base substitution(s) occur at or near the 5’ end of the
nucleotide sequence, e.g. in the final 15, 10 or 5 5’ nucleotides in the sequence.
Put differently, the base substitution(s) preferably do not occur at or near the 3’ end
of the nucleotide sequence, e.g. in the final 2, 3, 4, 5, 10 or 15 3’ nucleotides. In
other embodiments the 3’ nucleotide will not have a substituted base.

As mentioned above, the oligonucleotides of the invention selectively
hybridise to a sequence found in certain Flavobacteriaceae and Proteobacteria and
so the oligonucleotides may be used as oligonucleotide probes or oligonucleotide
primers to determine the levels of such bacteria in a sample from the
gastrointestinal (GI) tract of a subject. The inventors have found surprisingly that
levels of such bacteria are increased in the GI tract of subjects with CD as opposed
to subjects without CD. Accordingly, the oligonucleotides of the invention can be
used as molecular tools to determine the levels of these bacteria in the GI tract of a
subject suspected of having CD (or IBD in general) in order to diagnose CD and
differentiate CD from UC and/or to differentiate between a patient who has CD (i.e.
a form of IBD) from a patient who has a different disease or medical condition
associated with the gastrointestinal microbiota (e.g. IBS) or a healthy person.

In a further aspect, the present invention provides a method of diagnosing
Crohn’s Disease in a subject, said method comprising:

(i) contacting a sample from the GI tract of said subject with an
oligonucleotide as defined above;

(ii) subjecting the sample and oligonucleotide to conditions which allow
hybridisation of the oligonucleotide to its target sequence within nucleic acid
molecules in said sample, and
(iii) determining the amount of target nucleic acid in said sample.

The amount of target nucleic acid present in the sample is indicative of Crohn's Disease in said subject and can therefore be used for making a diagnosis of CD.

The target nucleic acid incorporates the target sequence. The target nucleotide sequence may be GAGGGTGCAAGCGT (SEQ ID NO 4) or the nucleotide sequence complementary thereto (ACGCTTGCACCTC; SEQ ID NO 3). These sequences are characteristic of certain Flavobacteriaceae and Proteobacteria and so the amounts of these target sequences in a sample are proportional to the amounts of these bacteria in the sample. The amount of the oligonucleotides of the invention that hybridise to these target sequences is in turn proportional to the amount of target sequence and so is proportional to the amount of these bacteria in the sample.

The results of the analysis of samples from the subject under investigation may be compared with previously prepared results or standards from samples from the GI tract of subjects without CD or from samples taken from confirmed CD patients. Preferably the comparison will involve results from corresponding sample types that have been collected analogously. The determination may be compared with earlier results from the same subject. The standards may be a combination of results obtained from multiple subjects, i.e. an average value. These standards may have been produced some time prior to the analysis of samples from the subject under investigation and may be provided to the practitioner digitally, e.g. on digital media or via electronic transfer to the user. In other embodiments a system may be in place in which the results obtained from the subject under test contributes to the development of the standard profile.

"Diagnosis of Crohn's Disease" refers to determination of the presence or existence of the disease or stage thereof in the subject. For the purposes of the invention the term is also intended to cover the monitoring of CD in a patient. "Monitoring" refers to establishing the extent of, or possible changes in, a subject's CD, for example to monitor the effects of treatment or the development of a disease or condition, e.g. to determine the suitability of a treatment, to provide a prognosis and/or to determine if a patient is in remission or relapse. The term also covers assessing the risk of a subject developing CD. "Assessing the risk of developing CD" refers to the determination of the chance or the likelihood that the subject will
develop the disease. This may be expressed as a numerical probability in some embodiments.

The amount of target nucleic acid can be determined by any convenient means and many such means will be familiar to the skilled man. This can be a partially, semi- or fully quantitative measurement, but can also be a qualitative (or relative) measure in which results from a sample from subjects with or without CD are simply compared to results from a sample from the subject under investigation, with any differences between the two being noted without numerical values being affixed.

If the measurement is quantitative, statistical analysis can be used to ascertain whether or not differences between the results obtained from the sample under investigation and the results obtained from a control sample or a prepared standard value are significant. In such embodiments the amount of target nucleic acid in the sample under investigation will preferably be statistically significantly increased as compared to the amount of target nucleic acid in the control sample or compared to standard values. The skilled man would be aware of suitable statistical techniques to use in these embodiments. Preferably the statistical technique will provide a "P value" as an indication that the trend being observed is not a random trend. A statistically significant result, i.e. a result that is not attributable to random variation when compared to its control, will have a P value of <0.05, preferably <0.01, <0.005 or <0.001. Thus, in these embodiments of the invention, a diagnosis of CD may be made if the increased amount of hybridisation in a GI tract sample under investigation as compared to a control sample, or a prepared standard value, has a P value of <0.05, preferably <0.01, <0.005 or <0.001. Merely by way of example, suitable techniques for measuring statistical significance in the methods of the invention are ANOVA, Mann-Whitney-Wilcoxon (MWW) Test, Kruskal-Wallis Test and Tukey's Honestly Significant Differences (HSD) Test. Many others would be familiar to the skilled man.

In a further aspect, the present invention provides a method of obtaining information relevant to a diagnosis of Crohn's Disease in a subject, said method comprising:

(i) contacting a sample from the GI tract of said subject with an oligonucleotide as defined above;
(ii) subjecting the sample and oligonucleotide to conditions which allow hybridisation of the oligonucleotide to its target sequence within nucleic acid molecules in said sample, and

(iii) determining the amount of target nucleic acid in said sample.

Thus, in such an embodiment information about the presence of target nucleic acid molecules in the sample is used together with one or more other clinical or laboratory investigations in order to provide a diagnosis of CD or to dismiss CD as an explanation for the symptoms presented by the subject. For example, methods of the invention may be used as an alternative or additional diagnostic measure to diagnosis using imaging techniques such as Magnetic Resonance Imaging (MRI), ultrasound imaging, nuclear imaging, X-ray imaging or endoscopy or CD serological markers, e.g. anti-Saccharomyces cerevisiae antibodies (ASCA) and peri-nuclear anti-neutrophil cytoplasmic antibodies (pANCA). An oligonucleotide of the invention could be included in an array, i.e. as part of a set of probes, which is used to diagnose CD.

In one embodiment the amount of target nucleic acid in the sample from the subject under investigation is determined by using the oligonucleotide of the invention with a label attached thereto that will allow detection by direct means or indirect means. In other words, the oligonucleotide of the invention is used simply as a conventional oligonucleotide probe. Suitable labels are described above. After contact of such a probe with the sample under conditions which allow hybridisation, and typically following a step (or steps) to remove unbound labelled oligonucleotide and/or non-specifically bound oligonucleotide, the strength of the signal from the label of the probe emanating from the sample under investigation (i.e. the amount of label bound to the sample) will be proportional to the amount of hybridised oligonucleotide and thus to the amount of target nucleic acid. In preferred embodiments the label is selected such that it is detectable only when the probe is hybridised to its target. In such embodiments, the need to remove the unbound probe is lessened.

Any convenient means may be used to remove any unbound or non-specifically probes, for instance with one or more washing steps (e.g. with water or a buffered solution which may contain formamide and/or a detergent), electrophoresis, centrifugation, capture onto solid supports, chromatography or any combination thereof. Suitable solid supports are described above. In another embodiment the probe may carry a binding moiety, or the label may be a binding
moiety, that will allow manipulation of the probe and any part of the sample hybridised thereto. Suitable binding moieties are discussed above.

Thus, the invention provides a method of diagnosing Crohn's Disease in a subject and a method of obtaining information relevant to a diagnosis of Crohn's Disease in a subject, said methods comprising:

(i) contacting a sample from the GI tract of said subject with an oligonucleotide as defined in above, wherein said oligonucleotide has a label attached thereto; and

(ii) determining the amount of said label bound to said sample by determining the strength of the signal from the label emanating from the sample.

The amount of label bound to the sample is indicative of Crohn's Disease in said subject.

In a preferred embodiment the method will comprise a step between steps (i) and (ii) in which unbound oligonucleotide and/or non-specifically bound oligonucleotide is removed.

In another embodiment the amount of target nucleic acid in the sample from the subject under investigation is determined by using an oligonucleotide of the invention as a probe which is labelled only when hybridised to its target sequence. In some embodiments the probe may already carry a label that is different to the label used to selectively label the probe. The strength of the signal from the selectively labelled probe emanating from the sample under investigation (i.e. the amount of label bound to the sample) will be proportional to the amount of hybridised oligonucleotide and in turn the amount of target sequence. Determining the amount of selectively labelled probe thereby enables the user to diagnose Crohn's Disease.

As mentioned previously, depending on the conditions employed, this can be a partially, semi- or fully quantitative measurement, but can also be a qualitative (or relative) measure in which results from a sample from a subject with or without CD are simply compared to results from a sample from the subject under investigation, with any differences between the two being noted without numerical values being affixed.

Conveniently, selective labelling may be achieved using labelled nucleotides, i.e. by incorporation into the oligonucleotide probe of a nucleotide carrying a label. In other words, selective labelling may occur by chain extension of the oligonucleotide probe using a polymerase enzyme which incorporates a labelled
nucleotide, preferably a labelled dideoxynucleotide (e.g. ddATP, ddCTP, ddGTP, ddTTP, ddUTP) more preferably labelled ddCTP, most preferably a fluorescently labelled, e.g. TAMRA labelled, ddCTP or a biotin labelled ddCTP. This approach to the detection of specific nucleotide sequences is sometimes referred to as primer extension analysis. Suitable primer extension analysis techniques are well known to the skilled man, e.g. those techniques disclosed in WO99/50448, the contents of which are incorporated herein by reference. Suitable labels are described above. Fluorescent labels and biotin are mentioned in particular.

In the case of oligonucleotide probes terminating with SEQ ID NO. 1 at their 3', the label will preferably be a labelled ddCTP.

More preferably, in this embodiment the probe will be an oligonucleotide terminating with SEQ ID NO. 1 at its 3' and the label will be a labelled ddCTP, e.g a TAMRA or biotin labelled ddCTP. Most preferably in this embodiment the probe will be an oligonucleotide consisting of SEQ ID NO. 1 and the label will be a labelled ddCTP, e.g. a TAMRA or biotin labelled ddCTP.

Detection of the labelled probe can be by any means convenient for the label being used. The skilled man would be able to devise suitable methods based on his selection of label. In preferred embodiments, the label is a fluorescent label (e.g. TAMRA) and in such embodiments the fluorescently labelled probed can be detected and, if required, quantified using a device that can measure the intensity (or strength) of fluorescent signals. A biotin label may be detected indirectly by exposing the label to streptavidin, or another biotin-binding molecule, which carries a detectable moiety, e.g. a colorimetric, chemiluminescent, chromogenic, radioactive or fluorescent label. In some embodiments, detection will occur after the labelled probe has undergone manipulation to remove, at least partially, contaminants (e.g. unlabelled probes, excess label, and other reagents used in the labelling reaction). Again, the skilled man would be very familiar with techniques which can achieve this, by way of example mention is made of electrophoresis (e.g. gel, e.g. capillary gel electrophoresis), centrifugation, chromatography and filtration based techniques, capture onto solid supports, or any combination thereof.

In other preferred embodiments the selectively labelled oligonucleotide probe is detected after a step in which the oligonucleotide probes from the selective labelling step (i.e. labelled and unlabelled), or the selectively labelled oligonucleotide probes only, are hybridised to a nucleotide sequence that is partially, or preferably fully, complementary to the oligonucleotide probe.
Conveniently, the complementary nucleotide sequence can be provided on a solid support, e.g. those described previously. In one embodiment that solid support can be a membrane strip on to which the complementary sequences may be spotted (e.g. in a defined spatial relationship) and UV cross-linked, or DNA chips (microchips, glass chips) now common in molecular biology procedures. In another convenient embodiment that solid support can be particulate, e.g. beads and microspheres. Preferably the complementary nucleotide sequence will be immobilised on a particle, e.g. a bead or a microsphere, having a label.

The particles may be labelled in any convenient way, e.g. using one or more of the labels described above. In one embodiment the particle label will not be or comprise an oligonucleotide, or a nucleic acid, or a labelled oligonucleotide or labelled nucleic acid. Conveniently the particulate solid support of these embodiments will be labelled with a dye, e.g. a luminescent (e.g. bioluminescent, chemiluminescent, photoluminescent, radioluminescent, sonoluminescent, etc.) dye, or a plurality of dyes (or proportions thereof) which combine to give a unique electromagnetic spectral signature upon excitation. Signatures based on the absorption of certain wavelengths of electromagnetic radiation are also envisaged.

Conveniently the dye will be fluorescent, e.g. comprise red or infrared fluorophores, e.g. phycoerythrin.

The label may be immobilised on and/or in the particle, e.g. by direct covalent binding to the substrate of the particle or it may be bound to another molecule which is in turn immobilised on and/or in the particle. The label may also be incorporated into and/or onto the particle by non-covalent means, e.g. by entrapment, absorption or adsorption of the molecules making up the label in or on the substrate of the particle, or by entrapment in void(s) within the substrate and/or on its surface.

In other embodiments the particle comprises nanoparticles on which and/or in which the label has been immobilised or incorporated.

The label can be applied to the particle after it is produced, or the label may be incorporated or immobilised into and/or onto the particle during its production, e.g. during the cross-linking of a polymeric substrate.

Preferably the label of the probe will be distinguishable from the label of the particle. In preferred embodiments the label of the particles will be detectable at the same time as the label of the probe. Preferably the labelled particles will also be magnetic, e.g. paramagnetic or superparamagnetic.

The invention therefore provides a method of diagnosing Crohn's Disease in a subject and a method of obtaining information relevant to a diagnosis of Crohn's Disease in a subject, said methods comprising:

(i) contacting a sample from the GI tract of said subject with an oligonucleotide as defined above;

(ii) selectively labelling the oligonucleotide when it is hybridised to its target nucleotide sequence within nucleic acid molecules in said sample; and

(iii) determining the amount of the labelled oligonucleotide produced in step (ii).

The amount of labelled oligonucleotide is indicative of Crohn's Disease in said subject. In some embodiments, step (iii) comprises hybridisation of the oligonucleotide from the labelling step to a nucleotide sequence complementary to the oligonucleotide.

In another embodiment the amount of target nucleic acid in the sample from the subject under investigation is determined by using an oligonucleotide of the invention as a primer in a nucleic acid amplification reaction. If the appropriate conditions are selected, such a reaction can be performed such that the amount of amplification product obtained will be proportional to the amount of target nucleic acid in the sample. Thus the amount of product the amplification reaction provides is proportional to the amount of the Flavobacteriaceae and Proteobacteria containing the target sequence in the sample. Accordingly, the amount of amplification product can be used to determine the levels of these bacteria in the GI tract of a subject suspected of having CD (or IBD in general) in order to diagnose CD, to differentiate CD from UC and to differentiate CD/IBD from a different disease
or medical condition associated with the gastrointestinal microbiota (e.g. IBS) or a healthy person.

As mentioned previously, depending on the conditions employed, this can be a partially, semi- or fully quantitative measurement, but can also be a qualitative (or relative) measure in which results from a sample from a subject with or without CD are simply compared to results from a sample from the subject under investigation, with any differences between the two being noted without numerical values being affixed.

Amplification can be achieved by any convenient primer-dependent nucleic acid amplification reaction. Most conveniently the polymerase chain reaction (PCR) will be used, although the skilled man would be aware of other techniques. For instance LAR/LCR, SDA, Loop-mediated isothermal amplification and nucleic acid sequence based amplification (NASBA)/3SR (Self-Sustaining Sequence Replication) may be used.

Many variations of PCR have been developed, for instance Real Time PCR (also known as quantitative PCR, qPCR), hot-start PCR, competitive PCR, and so on, and these may all be employed where appropriate to the needs of the skilled man.

In one basic embodiment of the invention using a PCR based amplification an oligonucleotide of the invention is contacted with a reaction mixture containing the sample, a suitable second primer to form a working primer pair and free nucleotides in a suitable buffer under conditions which allow hybridisation. Thermal cycling of the resulting mixture in the presence of a DNA polymerase results in amplification of the sequence characteristic of certain Flavobacteriaceae and Proteobacteria.

Optimal performance of the PCR process is influenced by choice of temperature, time at temperature, and length of time between temperatures for each step in the cycle. A typical cycling profile for PCR amplification is (a) 15 minutes of DNA melting at 95°C; (b) 30 seconds of primer annealing at 50-65°C; (c) 90 seconds of primer extending at 68-72°C; (d) 30 seconds of DNA melting at 95°C; and steps (b)-(d) are repeated as many times as necessary to obtain the desired level of amplification.

Modifications of the basic PCR method such as qPCR (Real Time PCR) have been developed that can provide quantitative information on the template being amplified. Numerous approaches have been taken although the two most
common techniques use double-stranded DNA binding fluorescent dyes or selective fluorescent reporter probes.

Double-stranded DNA binding fluorescent dyes, for instance SYBR Green, associate with the amplification product as it is produced and when associated the dye fluoresces. Accordingly, by measuring fluorescence after every PCR cycle, the relative amount of amplification product can be monitored in real time. Through the use of internal standards and controls, this information can be translated into quantitative data on the amount of template at the start of the reaction.

The fluorescent reporter probes used in qPCR are sequence specific oligonucleotides, typically RNA or DNA, that have a fluorescent reporter molecule at one end and a quencher molecule at the other (e.g. the reporter molecule is at the 5' end and a quencher molecule at the 3' end or vice versa). The probe is designed so that the reporter is quenched by the quencher. The probe is also designed to hybridise selectively to particular regions of complementary sequence which might be in the template. If these regions are between the annealed PCR primers the polymerase, if it has exonuclease activity, will degrade (depolymerise) the bound probe as it extends the nascent nucleic acid chain it is polymerising. This will relieve the quenching and fluorescence will rise. Accordingly, by measuring fluorescence after every PCR cycle, the relative amount of amplification product can be monitored in real time. Through the use of internal standard and controls, this information can be translated into quantitative data.

Thus, in another aspect the invention provides a method of diagnosing Crohn's Disease in a subject and a method of obtaining information relevant to a diagnosis of Crohn's Disease in a subject, said methods comprising:

(i) contacting a sample from the GI tract of said subject with an oligonucleotide as defined above;

(ii) performing a primer-dependent nucleic acid amplification reaction; and

(iii) determining the amount of amplification product produced from the oligonucleotide in said primer-dependent nucleic acid amplification reaction.

The amount of the product of the amplification reaction is indicative of Crohn's Disease in the subject from whom the sample was taken.

In a preferred embodiment step (i) will also comprise contacting the sample with a further oligonucleotide that is capable of functioning with the oligonucleotide of the invention in a nucleic acid amplification reaction, e.g. PCR, to produce an
amplification product. In this embodiment, when paired with a suitable second amplification primer, the oligonucleotides comprising SEQ ID NOs 1 and 3 will act as forward primers and oligonucleotides comprising SEQ ID NOs 2 and 4 will act as reverse primers.

The amplification product may be detected and amounts of amplification product can be determined by any convenient means. A vast number of techniques are routinely employed as standard laboratory techniques and the literature has descriptions of more specialised approaches. At its most simple the amount of amplification product may be detected or determined by visual inspection of the reaction mixture at the end of the reaction or at a desired timepoint. Typically the amplification product will be resolved with the aid of a label that may be preferentially bound to the amplification product. Typically a dye substance, e.g. a colorimetric, chromomeric fluorescent or luminescent dye (for instance ethidium bromide or SYBR green) is used. In other embodiments a labelled oligonucleotide probe that preferentially binds the amplification product, in particular a probe that binds preferentially to substantially all of the individual amplified nucleic acids in the amplification product, is used. A suitable probe might be based on the nucleotide sequence of SEQ ID NOs 1, 2, 3 or 4. Suitable labels for the probe are discussed above in relation to the oligonucleotide primers of the invention. In some embodiments the probe may be provided in an unlabelled form with labelling occurring after preferential binding to the amplification product, or preferential binding to substantially all of the individual amplified nucleic acids in the amplification product.

However, in some cases a nucleic acid precipitant (e.g. salt and/or alcohol) can simply be used to cause the amplification product to come out of solution and be visible without labelling.

To aid visualisation the components of amplification product can be dispersed in or on a solid support, for instance by electrophoresis (e.g. using agarose or polyacrylamide gels), chromatography (e.g. HPLC, TLC, affinity, gel filtration) or filtration, or a combination thereof, prior to or after contact with the label.

Depending on the label used detection can be made more accurate by using widely available detection technologies, e.g. radiation sensitive films and digital imaging technologies in combination with computer assisted image analysis, photometers, fluorometers, colorimeters, scintillation counters, and the like.
Preferably the amplification product is separated from the remainder of the amplification reaction before being contacted by the label, e.g. in the form of a labelled oligonucleotide probe. This may be by any convenient means, for instance with one or more washing steps (e.g. with water or a buffered solution which may contain formamide and/or a detergent), electrophoresis, centrifugation, capture onto nucleic acid binding solid supports, chromatography or any combination thereof. Conveniently, the probe can be provided on a solid support thereby effecting separation of the amplification product from the remainder of the amplification reaction in a single step. In another embodiment the probe may carry a binding moiety, or the label may be a binding moiety, that will allow manipulation of the probe and any amplification product hybridised thereto. Suitable binding moieties are discussed above.

Preferably any unbound label, e.g. in the form of a labelled oligonucleotide probe, will be separated from the amplification product before the detection step.

This can be by any convenient means, for instance with one or more washing steps (e.g. with water or a buffered solution which may contain formamide and/or a detergent), electrophoresis, centrifugation, capture onto solid supports, chromatography or any combination thereof. Suitable solid supports are described above.

If the amplification method used is itself quantitative, e.g. amplification methods in which internal standards and controls are incorporated (for instance qPCR) the method of this aspect of the invention can also provide quantitative data. In these embodiments the method can even affix a numerical value to the amount of target nucleic acid present in the sample and thus the amount of the bacteria containing said target nucleic acid in the sample. One such internal standard would be to amplify one or more (e.g. at least 2, 3, 5, or 10) samples containing a known amount of these bacteria or known quantities of target nucleic acid under the same conditions as the test sample to provide a standard curve plotting amount of amplification product against number of organisms or amount of target nucleic acid. The amount of amplification product obtained in the test sample can then be translated into a numerical value for the amount of these bacteria and/or amount of target nucleic acid in the sample.

In other embodiments, the progress of the amplification reaction can be followed in real-time and the amplification profile can be compared with amplification profiles from samples containing known quantities of target nucleic
acid or known amounts of bacteria containing said target nucleic acid. In other 5
embodiments the cycle threshold (C_T) can be used to calculate the amount of target sequence and therefore the amounts of certain Flavobacteriaceae and Proteobacteria in the sample. In all qPCRs there is a threshold at which the 10
fluorescence of the amplification product is detected above background. The cycle at which this threshold is crossed is the C_T. In the exponential phase of the reaction the quantity of DNA theoretically doubles every cycle and so relative amounts of DNA can be calculated between samples by comparing C_T values falling in the exponential phase. If the comparison is made with samples with a known quantity of template, the quantity of template in the test sample can be calculated and the amount of target nucleic acid present in the sample and thus the amount of the bacteria containing said nucleic acid in the sample can be determined.

While the methods of the invention focus on determining elevated levels of 15
particular target nucleic acids in GI tract samples in order to diagnose Crohn's Disease, i.e. target nucleic acids indicative of certain Flavobacteriaceae and Proteobacteria, it is contemplated that other correlations between Crohn's Disease and certain bacteria will be made and, as a result, the methods of the invention will further comprise the use of oligonucleotides specific to such newly correlated bacteria alongside the oligonucleotides of the invention. In such embodiments the attachment of distinct labels to the oligonucleotides and/or the attachment of the various oligonucleotides being used to distinctly labelled solid supports or spatially distinct regions thereof will permit the parallel use (multiplexing) of the various oligonucleotides in the methods of the invention. Suitable labels and solid supports are described above.

In view of the above the invention can be considered to provide a method of diagnosing Crohn's Disease in a subject and a method of obtaining information relevant to a diagnosis of Crohn's Disease in a subject, said methods comprising performing a nucleic acid detection procedure on a sample from the GI tract of said subject to determine the amount of a target nucleotide sequence of 25

GAGGGTGCAAGCGT (SEQ ID NO 4) or the nuclease sequence complementary thereto (ACGCTTGACCCTC; SEQ ID NO 3) in the sample. Typically the detection procedure comprises a step in which hybridisation of an oligonucleotide as defined above to said target sequence occurs. The amount of said target sequence in said sample is indicative of Crohn's Disease in said subject.
The nucleic acid detection procedure can be any convenient nucleic acid
detection procedure, e.g. procedures comprising any of the means described above
to confirm hybridisation of the oligonucleotide of the invention to its target
nucleotide sequence in said sample.

In a further aspect the invention provides a method of diagnosing Crohn's
Disease in a subject, said method comprising

(a) comparing the results of a method as described above to previously
prepared results or standards from samples from the GI tract of subjects without
Crohn's Disease or from samples taken from the GI tract of subjects confirmed as
having Crohn's Disease, and

(b) making a diagnosis of Crohn's Disease based on the relative amounts of
target nucleic acid in the samples being compared.

Preferably the results or standards to which the results of the subject under
test are compared to will be results or standards prepared in accordance with the
invention. This may be prepared results or standards, or could be results or
standards prepared at the same or substantially the same time as the sample under
investigation is being analysed. Preferably the comparison will involve results from
corresponding sample types that have been collected analogously. Further
discussion of such results and standards in given above.

The methods of the invention described above can also be considered as
methods of differentially diagnosing Crohn's Disease in a subject presenting with
unclassified IBD. All discussion of the various features of the methods of the
invention apply mutatis mutandis to this aspect of the invention. Patients with
unclassified IBD are patients in which the disease underlying the diagnosis of IBD
cannot be identified with confidence, in other words a differential diagnosis of
Crohn's Disease or Ulcerative Colitis or any of the other colitises which contribute to
IBD cannot be made with confidence. The methods of the invention can also be
used to subclassify CD patients as within samples showing elevated amounts of
hybridisation those samples from patients with ileal CD will have the most elevated
amounts of hybridisation. In a further aspect the oligonucleotides of the invention
may be used to discriminate or to provide information relative to a discrimination
between CD/IBD patients and those with a different disease or medical condition
associated with the gastrointestinal microbiota (e.g. IBS).

The subject may be any human or non-human animal subject, but more
particularly may be a vertebrate, e.g. an animal selected from mammals, birds,
amphibians, fish and reptiles. The animal may be a livestock or a domestic animal or an animal of commercial value, including laboratory animals or an animal in a zoo or game park. Preferably the subject is a human. The subject may be of any age, e.g. an infant, a child, a juvenile, an adolescent or an adult, preferably an adult. In humans, an adult is considered to be of at least 16 years of age.

The methods of the invention are in vitro methods performed using any sample taken from the GI tract. The GI tract, also referred to as the digestive tract or alimentary canal (and which terms may be used interchangeably with GI tract) is the continuous series of organs beginning at the mouth and ending at the anus. Specifically this sequence consists of the mouth, the pharynx, the oesophagus, the stomach, the duodenum, the small intestine, the large intestine and the anus. These organs can be subdivided into the upper GI tract, consisting of the mouth, pharynx, oesophagus, stomach, and duodenum, and the lower GI tract, consisting of the jejunum, the ileum (together the small intestine), the cecum, the colon, the rectum (together the large intestine) and the anus.

A GI tract sample of use in the invention may include, but is not limited to any fluid or solid taken from the lumen or surface of the GI tract or any sample of any of the tissues that form the organs of the GI tract. Thus the sample may be any luminal content of the GI tract (e.g. stomach contents, intestinal contents, mucus and faeces/stool, or combinations thereof) as well as samples obtained mechanically from the GI tract e.g. by swab, rinse, aspirate or scrape of a GI tract cavity or surface or by biopsy of a GI tract tissue/organ. Fecal samples are preferred.

The sample can also be obtained from part of a GI tract tissue/organ which has been removed surgically. The sample may be a portion of the excised tissue/organ. In embodiments where the sample is a sample of a GI tract tissue/organ the sample may comprise a part of the mucosa, the submucosa, the muscularis externa, the adventitia and/or the serosa of the GI tract tissue/organ. Such tissue samples may be obtained by biopsy during an endoscopic procedure. Preferably the sample is obtained from the lower GI tract, i.e. from the jejunum, the ileum, the cecum, the colon, the rectum or the anus. More preferably the sample is a mucosal or luminal sample.

Mucosal biopsies, unlike stool, provide samples collected from regions of the intestinal tract where inflammation occurs. In contrast, the faecal microbiota is thought to represent the microbiology of the distal large bowel and it has been
stated that it does not reflect the colon nor the terminal ileum. Since CD can occur anywhere in the digestive tract from mouth to anus it has been suggested that the use of mucosal samples from the precise areas of inflammation will be necessary for accurate diagnosis. Nevertheless, differences in the faecal microbiota between IBD patients and controls have been reported and so the analysis of the microbiota of faecal samples represents an opportunity to diagnose IBD and CD and UC without the need for a surgical procedure to be used for sample collection. The inventors have now shown that a correlation does exist between CD and the levels in faecal samples of certain Flavobacteriaceae and Proteobacteria, and which correlation is sufficiently robust to be harnessed in the diagnosis of CD. Thus, while the use of other GI tract samples is not excluded from the invention, the invention, for the first time, allows the skilled man to use faecal samples, with their easy availability, instead of mucosal samples obtained from biopsies or other surgical procedures. The faecal samples may be collected by the swab, rinse, aspirate or scrape of the rectum or anus or, most simply, the collection of faeces after defecation.

The sample may be used in the methods of the invention in the form in which it was initially retrieved. The sample may also have undergone some degree of manipulation, refinement or purification before being used in the methods of the invention. Thus the term "sample" also includes preparations thereof, e.g. relatively pure or partially purified starting materials, such as semi-pure preparations of the above mentioned samples. The term "sample" also includes preparations of the above mentioned samples in which the RNA of which, including the 16 S rRNA, has undergone reverse transcription.

The purification may be slight, for instance amounting to no more than the concentration of the solids, or cells, of the sample into a smaller volume or the separation of cells from some or all of the remainder of the sample. Representative cell isolation techniques are described in WO98/51693 and WO01/53525.

In other embodiments the invention uses a preparation of the nucleic acid from the above mentioned samples. Such preparations include reverse transcription products and/or amplification products of such samples or nucleic acid preparations thereof. Preferably the predominant nucleic acid of the nucleic acid preparation is DNA.

Techniques for the isolation of nucleic acid from samples, including complex samples, are numerous and well known in the art and described at length in the
literature. The techniques described in WO98/51693 and WO01/53525 can also be employed to prepare nucleic acids from the above mentioned samples. These preparations include relatively pure or partially purified nucleic acid preparations.

Preferably the amplification reaction performed on the sample will be universal, or substantially universal, in that the nucleic acid to be amplified, i.e. the region of 16S rRNA or 16S rDNA incorporating the above discussed target sequences, is amplified from all, or at least substantially all, prokaryotic cells that might be present in a sample. The term "amplification from substantially all prokaryotic cells present in a sample" refers to the number of different species of prokaryotic cells in the sample that will have the nucleic acid to be amplified, amplified. Thus, in this embodiment the nucleic acid to be amplified is amplified from at least one representative of substantially all species of prokaryotic cells in the sample.

By "prokaryotic cell" it is meant any organism that lacks a cell nucleus, i.e. any organism from the domains Bacteria and Archaea.

Conveniently this universal amplification is performed using a primer pair having the sequences TCC TAC GGG AGG CAG CAG (SEQ ID NO 5), also referred to as MangalaF-1, and CGG TTA CCT TGT TAC GAC TT (SEQ ID NO 6), also referred to as 16SU1510R. This primer pair is described in more detail in US 2011/0104692.

The target nucleotide sequence to be amplified in this embodiment is therefore present in 16S rRNA and the corresponding 16S rRNA gene (rDNA). Thus, reference to the amplification of this target nucleotide sequence is a reference to an increase in the number of nucleic acids that contain that sequence of nucleotides without limitation on the type of nucleic acids containing the nucleotide sequence. Typically, the nucleic acid that is formed as the amplification product is DNA, although the nucleotide sequence contained in that nucleic acid will still be the same as that of the target nucleotide sequence, or the complement thereof.

Conveniently, this embodiment of the invention will be performed with 16S rDNA, e.g. a 16S rRNA gene, as the template.

In other embodiments 16S rRNA may be the source of the target nucleotide sequence to be amplified. When a target nucleotide sequence from 16S rRNA is amplified in this embodiment of the method of the invention there will be a step in which an RNA-dependent DNA polymerase catalyses the formation of a DNA
molecule complementary to the 16S rRNA template (cDNA). This process is termed "reverse transcription". More specifically the RNA-dependent DNA polymerase catalyses the polymerisation of deoxyribonucleoside triphosphates in a sequence that is complementary (i.e. following Watson-Crick base pairing rules) to a primed template rRNA sequence.

Numerous enzymes have been identified that have the ability to catalyse this reaction and examples include, but are not limited to, HIV reverse transcriptase, AMV reverse transcriptase, M-MLV reverse transcriptase, C. therm. polymerase, and TTh polymerase. At its most basic a complete reverse transcription reaction mixture will contain a reverse transcription enzyme, the rRNA template, suitable primers that can bind to the template and from which the reverse transcriptase can begin polymerisation, dNTP’s and a suitable buffer. Incubation of the mixture at the working temperature of the reverse transcriptase results in cDNA production.

Upon completion of the reverse transcription reaction the cDNA can be used as the template in the embodiment of the method of the invention described above. The cDNA therefore has a nucleotide sequence that is complementary to the rRNA molecule that was its template. In addition the cDNA has a nucleotide sequence that is the same as a nucleotide sequence contained in one strand of the gene of the rRNA template and the cDNA is complementary to a nucleotide sequence contained in the other strand of the gene of the rRNA template.

As mentioned above, in embodiments of the method of the invention in which nucleic acid is amplified in a preceding step, if 16S rRNA is used as the source of the target nucleotide sequence (as opposed to 16S rDNA, e.g. a 16S rRNA gene) an initial reverse transcription step is required. Reverse transcription linked amplification reactions, in particular PCR, can be "one step" or "two step" processes. In a one step process the components of the reverse transcription reaction and the nucleic acid amplification reaction are present in a single reaction vessel and typically the early reaction conditions are selected to allow the reverse transcription reaction to proceed to completion and reaction conditions are then switched to conditions suitable to allow the nucleic acid amplification reaction to proceed.

In a two step process the components of the reverse transcription reaction are first combined and the reverse transcription reaction is performed. The reverse transcription product is then combined with the components of the amplification reaction and subjected to the amplification reaction. In a "one tube" two step
protocol the amplification reaction components are added to the same reaction vessel in which the reverse transcription reaction was performed. In a "two tube" two step protocol the amplification reaction is performed in a fresh reaction vessel. Preferably, in these embodiments, primers based on the nucleotide sequences of SEQ ID NOs 5 and 6 will make up a part of the components of the nucleic acid amplification reaction.

In a further aspect the invention provides kits comprising an isolated oligonucleotide comprising the nucleotide sequence ACGCTTGCACCCCT (SEQ ID NO 1) or the nucleotide sequence complementary thereto (AGGCTGCAAGCGT; SEQ ID NO 2) or a nucleotide sequence capable of hybridising under conditions of high stringency to SEQ ID NO 1 or SEQ ID NO 2.

The kits of the invention may be designed for use in the methods of the invention and may comprise further components. Each component may be provided in a separate compartment or vessel. Where convenient and practical, mixtures of components could be provided. The components may be provided in dry, e.g. crystallised, freeze dried or lyophilised, form or in solution, typically such liquid compositions will be aqueous and buffered with a standard buffer such as Tris, HEPES, etc.

The kit may also be provided with instructions for using the kit, or with directions for how instructions may be obtained.

The additional components can be any of the various components that may be used to put the methods of the invention into effect, e.g. any component discussed above. In a preferred embodiment the kit further comprises means for selective labelling of the oligonucleotide and/or nucleotide sequences complementary to the oligonucleotide, preferably immobilised on a solid support.

Further components might optionally be any or all of the means, e.g. buffers, enzymes etc. for performing an amplification and/or primer extension reaction with the oligonucleotides of the invention. For instance, the kits may optionally contain a PCR reaction buffer, nucleotide triphosphates, further oligonucleotide primers, or DNA polymerases, preferably a thermostable polymerase such as Taq polymerase.

Further components might optionally be any or all of the means, e.g. buffers, enzymes etc. for performing a reverse transcription reaction. For instance a reverse transcriptase, RNA specific primers, an RT reaction buffer, and nucleotide triphosphates.
Further components might optionally be any or all of the means to take the sample. For instance such means might include dipsticks, biopsy apparatus, swabbing devices, pouches or vessels. Preferably these means will be provided in sterile form.

Further components might optionally be any or all of the means to purify or refine the sample. For instance means to isolate or concentrate cells in a sample, e.g. cell binding solid supports or filtration devices. In other embodiments the means to purify or refine the sample might be any or all of the means for extracting nucleic acid from a sample. For instance cell lysis reagents (e.g. chaotropic salts, alcohols, detergents, membrane altering compounds), nucleic acid binding solid supports (e.g. as described above) or nucleic acid precipitating agents (e.g. salts, alcohols).

Further components might optionally be any or all of the means to detect amplified nucleic acid. For instance the labels described herein (e.g. double stranded DNA binding dyes, labelled oligonucleotide probes), apparatus to detect these labels, electrophoresis materials and apparatus, or chromatography materials and apparatus.

Further components might optionally be further oligonucleotides that selectively hybridise to target nucleic acids indicative of Crohn's Disease and/or any other disease or medical condition, particularly diseases or conditions associated with the gastrointestinal microbiota (e.g. IBS, UC, IBD) and which may accordingly be used in a manner similar to the oligonucleotides of the invention to provide information relevant to a diagnosis of Crohn's Disease and/or any other disease or medical condition, particularly diseases or conditions associated with the gastrointestinal microbiota (e.g. IBS, UC, IBD). In these embodiments the kit may be considered to be an oligonucleotide array (sometimes termed probe set).

The invention will be further described with reference to the following non-limiting Examples in which:

Figure 1 shows the distribution of RFU with the IG0005 probe in CD, UC, IBDU and control patients.
EXAMPLE 1

In order to identify oligonucleotide probes capable of diagnosing IBD, PCR amplified nucleic acid from faecal samples from 257 individuals (adults and children) were hybridized with five probes in separate reactions. The probes were extended by a fluorescence labelled ddNTP and detection was performed by capillary gel electrophoresis as described below.

The present study includes 152 patients and 105 controls.

Materials and methods

Patients and samples

The patient samples used in this study were a cohort from the Norwegian IBSEN study and were kindly distributed by Akershus University Hospital (Ahus). The samples comprise newly diagnosed untreated IBD patients and non-IBD patients. In total 257 patients were included in the study (Table 1). The inclusion criteria for the IBD cohort were abdominal symptoms including diarrhoea and/or blood in the stools for more than 10 days, endoscopic or radiological signs of inflammation as well as histological signs of chronic inflammation.

Exclusion criteria were infection of pathogenic gut bacteria (excluding Mycobacterium avium), findings of parasites, cysts and eggs. In addition, patients with cancer, haematological or hepatological disorders, significant cardiovascular, neurological or respiratory comorbidity, and chronic inflammation other than IBD were excluded from the study (both cases and controls).

Classification of patients

The IBD patients were further classified as CD, UC or IBD unclassified (IBDU). Patients were diagnosed as CD based on the presence of two or more of the following criteria:

- 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 aphthous ulcerations, fissuring or penetrating lesions, cobblestone or strictures
- 30 -

- Radiological evidence of stenosis in the small bowel, segmental colitis or findings of fistulae
- Histological evidence of transmural inflammation or epithelial granulomas with giant cells

The diagnosis of UC was based on the presence of three or more of the following criteria:
- 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 or the entire colon
- Microscopic features on biopsy compatible with UC
- No suspicion of CD on small bowel roentgenography, ileocolonoscopy, or biopsy

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

The non-IBD cohort was derived from patients referred to colonoscopy, but not suffering from IBD.

Table 1 Characteristics of patients

<table>
<thead>
<tr>
<th></th>
<th>Children</th>
<th>Adults</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC</td>
<td>13</td>
<td>50</td>
<td>63</td>
</tr>
<tr>
<td>CD</td>
<td>29</td>
<td>36</td>
<td>65</td>
</tr>
<tr>
<td>IBDU</td>
<td>3</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>Non-IBD</td>
<td>25</td>
<td>80</td>
<td>105</td>
</tr>
</tbody>
</table>

Universal polymerase chain reaction

Faecal samples were stored at -80°C. DNA was then extracted using QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). Using universal 16S rRNA primers (forward 5'-TCC TAC GGG AGG CAG CAG-3' [SEQ ID NO 5]; reverse 5'-CGG TTA CCT TGT TAC GAC TT-3' [SEQ ID NO 6]) PCR reactions were set up as
follows: HotFirePol 1.25U; B2 buffer, 1x; MgCl₂ 2.5mM; dNTP 200μM (Solis BioDyne, Tartu, Estonia); each primer 0.2μM; DNA template 10-100ng. Each reaction was done in a final volume of 25μl. The reactions were performed using a standard PCR machine and the following conditions: 95°C for 15 min, followed by 30 cycles of 95°C for 30s, 55°C for 30s, 72°C for 1min 20s, followed by a final elongation at 72°C for 7 min.

The PCR products were treated with Exonuclease (EXO I) 3U and Shrimp Alkaline Phosphatase (SAP) 8U in order to remove excess primers and unincorporated nucleotides respectively. Subsequently the samples were incubated at 37°C for 2h followed by 80°C for 15 min.

**Probes**

Probes with the sequences shown below were generated and respective specificities determined from among 262 bacterial strains as shown in Table 2.

Additionally a universal probe was used.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Sequence (SEQ ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG0005</td>
<td>ACGCTTGCACCC (SEQ ID NO 1)</td>
</tr>
<tr>
<td>IG0028</td>
<td>CGTGGCCTTCGTGC (SEQ ID NO 7)</td>
</tr>
<tr>
<td>AG0006</td>
<td>GGATAACGCTTGCACCCTCCGTATTACCGC (SEQ ID NO 8)</td>
</tr>
<tr>
<td>AG0012</td>
<td>CCAGGGCATAGGGGACTGACTTGAC (SEQ ID NO 9)</td>
</tr>
<tr>
<td>Universal</td>
<td>CGTATTACCGCGCTGCTGGCA (SEQ ID NO 10)</td>
</tr>
</tbody>
</table>
Table 2. Phylogenetic distribution of the probes. Probes are assumed to cover all species within a given family

<table>
<thead>
<tr>
<th>Probe</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG0005</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>Enterobacteriales</td>
<td>Enterobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Abermondales</td>
<td>Abermonadaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pseudomonadales</td>
<td>Pseudomonadaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pasteurellae</td>
<td>Pasteurellace</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Epsilonproteobacteria</td>
<td>Campylocacterales</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Campylobacteriales</td>
<td>Helicobacteraceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deltaproteobacteria</td>
<td>Desulfovibionales</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Desulfobacteriales</td>
<td>Campylobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bacteroidetes</td>
<td>Flavobacteriales</td>
</tr>
<tr>
<td></td>
<td>Actinobacteria</td>
<td>Actinobacteria</td>
<td>Actinomycetales</td>
<td>Actinomycetaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Corynebacteriaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Micrococcaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Streptomyctaceae</td>
<td></td>
</tr>
<tr>
<td>AG0006</td>
<td>Actinobacteria</td>
<td>Actinobacteria</td>
<td>Actinomycetales</td>
<td>Actinomycetaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Corynebacteriaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Micrococcaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Streptomyctaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bifidobacteriales</td>
<td>Bifidobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coriobacteridae</td>
<td>Coriobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clostridiales</td>
<td>Bacteroidaceae</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Bacteroidia</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prevotellaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flavobacteria</td>
<td>Flavobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sphingobacteriales</td>
<td>Sphingobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bacillales</td>
<td>Staphylococcus</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Clostridia</td>
<td>Clostridiales</td>
<td>Lactobacillaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clostridiales Family XI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incertae Sedis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Eubacteraeace</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lachnospiraceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Peptostreptococcaseae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ruminococcaseae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Erysipelotrichi</td>
<td>Erysipelotrichaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Verrucomicrobia</td>
<td>Verrucomicrobia</td>
</tr>
<tr>
<td>AG0012</td>
<td>Actinobacteria</td>
<td>Actinobacteridae</td>
<td>Actinomycetales</td>
<td>Corynebacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Streptomyctaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coriobacteridae</td>
<td>Coriobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bacteroidetes</td>
<td>Sphingobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clostridiales</td>
<td>Veillonellaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lactobacillaceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fusobacteriaceae</td>
<td>Fusobacteriaceae</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proteobacteria</td>
<td>Epsilonproteobacteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gammaproteobacteria</td>
<td>Pseudomonadaceae</td>
</tr>
</tbody>
</table>
Endlabelling

A bacterial strain specific probe was hybridized to complimentary sequences in the sample and extended with one base pair using fluorescent ddCTP (Tamra) in an end-labelling reaction (Single Nucleotide Extension, SNE). All samples were hybridized with the five probes in separate reactions. The reaction was performed with a final volume of 10μl. The conditions were as follows: HOT TermiPol DNA polymerase 0.25U, HOT TermiPol Reaction Buffer C 1x, MgCl₂ 4mM (Solis BioDyne, Tartu, Estonia), ddCTP Tamra 0.4μM (Jena Bioscience, Jena, Germany), probe 0.1 μM, EXOI-SAP treated template 2 μl. The samples were loaded on a thermal cycler: 95°C for 12 min, followed by 10 cycles at 95°C for 20s and 60°C 35s.

Residual nucleotides and phosphate groups from the 5' end were removed by SAP, 1U per 10μl end-label product. Further the samples were incubated in a thermal cycler at 37°C for 1h followed by 80°C for 15min.

Capillary gel electrophoresis

Next, 1μl of each SAP treated end-labelling product was transferred to a MicroAmp™ Optical 96-Well Reaction Plate and mixed with 9 μl Hi-Di™ formamide and 0.5μl GeneScan™ 12 LIZ™ Size Standard (Applied Biosystems, Foster City, CA, USA). Subsequently, the samples were denatured at 95°C for 5min. Capillary gel electrophoresis was performed using an AB 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Fluorescent signals were measured and represent semi-quantitative signal of endlabelled probe. The samples were analyzed using GeneMapper © Software v4.1 (Applied Biosystems, Foster City, CA, USA). Default settings were used including a cut off value of 100 relative fluorescent units (RFU).

Quantification of PCR products

Quantification of the PCR products was performed using Quant-iT PicoGreen dsDNA Reagent and Kits as previously described (Invitrogen, 2008) (Invitrogen, Carlsbad, CA, USA). In order to make a standard curve, Lambda DNA standard was diluted in 1x TE and Quant-iT PicoGreen Reagent to the following final concentrations: (1.0 ng/μl, 0.1 ng/μl, 0.01 ng/μl, 0.001 ng/μl and 0 ng/μl).

Furthermore, 2 μl of each sample was diluted in 100 μl Quant-iT PicoGreen Reagent and 1xTE. A control sample (Lambda DNA standard) with known
concentration was measured on each run. Each standard, control and sample was analyzed in triplicate. Fluorescent signals were detected using FluoStar Optima (BMG Labtech, Offenburg, Germany). The quantification of PCR product was done by the Mars Data Analysis software version 1.10 (BMG Labtech, Offenburg, Germany). Standard deviation (SD) and Coefficient of Variation (CV) were calculated for the RFU. Standards and controls were accepted at <10% CV and samples at <33% CV. Additionally the control was accepted at <3SD from the standard curve. The RFU of the blank was subtracted from the average RFU signal of each sample. Further the concentration of the sample PCR product was determined by the standard curve (y=53897x+27.082).

**Statistical Methods**

The patient samples were randomly distributed in three 96 well plates according to patient status. To further normalize potentially plate to plate variation, sample RFU values were transformed to percent value of mean RFU values per plate. Further, normalization of variable concentration of PCR product was done. The sample value was divided by the concentration of PCR product as measured by Quant-IT PicoGreen.

For probe signal between patient groups comparisons, 1-way analysis of variance (ANOVA) was performed. Non-parametric distributed data significances were confirmed using the Mann-Whitney-Wilcoxon- (MWW). Further, Tukey’s Honestly Significant Differences (HSD) Test was done for pairwise comparisons of the patient groups. The calculations were done using PASW Statistics 18 (SPSS Inc. Chicago, Illinois, USA). Statistical significance was accepted at P< 0.05.

Percent sensitivity (True positive / (true positive + false negative)) and specificity (true negative/(true negative + false positives) was calculated for each probe in patients versus controls. True positives being patients correctly identified as such, and false negative being patients diagnosed as non-IBD. Similarly true negative is non-IBD patients diagnosed as such, and false positives are non-IBD patients diagnosed as IBD.

**Results**

The probes hybridized to sequences in all cohorts. The signal intensity between probes varied greatly. While the IG0028 probe had a mean fluorescent
intensity of 289 the Universal probe had a mean intensity of 4116. Due to among other weak fluorescent signal, certain samples were removed from the study (Table 3).

<table>
<thead>
<tr>
<th></th>
<th>Universal Probe</th>
<th>IG0028</th>
<th>IG0005</th>
<th>AG0006</th>
<th>AG0012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>218</td>
<td>65</td>
<td>250</td>
<td>253</td>
<td>252</td>
</tr>
<tr>
<td>Removed</td>
<td>39</td>
<td>192</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Probe quantification**

Mean signal between group variance of the probes is presented in Table 4. In particular a significant P-value in probe IG0005 was observed. No significant differences were predicted using the Universal, IG0028, AG0006, or AG0012 probe. The MWV test showed similar significance for the IG0005 probe (Asymp Sig. (2-tailed) <0.0005). The Tukey HSD test further showed a significant increase in the IG0005 probe in CD patients compared to the control group (P=3.56E-4). Although not significant, low P-values was observed between CD-UC (P=0.089) and between CD-IBDU (P=0.065). No significant values among other patient cohorts were found in IG0028, AG0006 and AG0012 probes.

The distribution of IG0005 positive bacteria indicates two distinct peaks (Figure 1). Overall, no significant difference in RFU was found between adult and child (ANOVA). However, when analyzing the IG0005 probe signals separately (HSD), adult samples showed significant difference between CD and controls (p=0.003) while no significant value was found in the child cohort relative to controls (p=0.274). Accuracy data for the IG0005 probe was calculated (Table 5). According to this sample set the IG0005 used as a diagnostic marker in faeces, could detect 58.3% of the CD patients.
Table 4 The 1-way ANOVA between patient group comparisons. Normalized signals.

<table>
<thead>
<tr>
<th>Probe</th>
<th>P=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal</td>
<td>0.884</td>
</tr>
<tr>
<td>IG0028</td>
<td>0.773</td>
</tr>
<tr>
<td>IG0005</td>
<td>0.001</td>
</tr>
<tr>
<td>AG0006</td>
<td>0.444</td>
</tr>
<tr>
<td>AG0012</td>
<td>0.680</td>
</tr>
</tbody>
</table>

Table 5 Accuracy data for differentiating Crohn's disease from controls using the IG0005 probe with a cut off value of 12 of normalized values.

<table>
<thead>
<tr>
<th></th>
<th>CD patients</th>
<th>Control patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosed as CD</td>
<td>35</td>
<td>28</td>
</tr>
<tr>
<td>Diagnosed as non-CD</td>
<td>25</td>
<td>77</td>
</tr>
<tr>
<td>Total number of Patients</td>
<td>60</td>
<td>105</td>
</tr>
<tr>
<td>Sensitivity = 58.3%</td>
<td></td>
<td>Specificity = 73.3%</td>
</tr>
</tbody>
</table>

Discussion

Hybridisation of the IG0005 probe was found to be significantly increased in CD samples compared to controls. The IG0005 probe hybridizes to species in the taxonomic families of Enterobacteriaceae, Aeromonadaceae, Campylobacteraceae, Pseudomonadaceae, Helicobacteraceae, Desulfovibrionaceae and Pasteurellaceae, and species in the Flavobacteriaceae family (Table 2).

Two distinct peaks in the CD IG0005 histogram (Figure 1) were observed. It might indicate the existence of two sub groups of CD patients. One group containing increased amounts of IG0005 positive bacteria, the other group with similar amounts as controls. Higher quantities of mucosa-associated E. coli have previously been shown in subgroups of Crohn's ileitis. It is hypothesized that the CD
patients with increased amounts of the IG0005 in the faecal sample, reflect adherent invasive *E. coli* (AIEC) associated with ileal CD.

Diagnostic accuracy for the IG0005 probe was calculated (Table 5). The serologic markers anti-*Saccharomyces cerevisiae* antibodies (ASCA) and perinuclear anti-neutrophil cytoplasmic antibodies (pANCA), currently the most robust serological markers used in diagnosing IBD, have similar sensitivity values. A combination of ASCA(+) with pANCA(-) test in CD have been observed with 54.6%, 92.8% for sensitivity and specificity, respectively.

Actinobacteria have previously been found in decreased levels in faecal samples from CD patients. This was not confirmed in this study. The IG0028, AG0006 and AG0012 probe which comprise complementarities to Actinobacteria was not observed with significantly decreased levels. Only 25% of the samples hybridized with IG0028 probe had fluorescent signal intensity above the cut off value. In IBD, species within the Firmicutes and Bacteroidetes phyla has been observed in decreased and increased levels respectively. Especially Clostridia and Bacteroidia species have been observed with respectively decreased and increased levels in CD. The AG0006 and AG0012 probe both comprise sequences complementary to species including Actinobacteria, Bacteroidetes, and Firmicutes phyla (Table 2). It could be hypothesized that these probes covers too many bacterial species to reveal diagnostic relevant information.
CLAIMS

1. An isolated oligonucleotide comprising the nucleotide sequence
   (i) ACGCTTGCAACCCT (SEQ ID NO 1),
   (ii) AGGGTGCAAGCGT (SEQ ID NO 2),
   (iii) ACGCTTGCAACCCTC (SEQ ID NO 3),
   (iv) GAGGGTGCAAGCGT (SEQ ID NO 4), or
   (v) a nucleotide sequence capable of hybridising under conditions of
   high stringency to any one of SEQ ID NOs 1, 2, 3 or 4.

2. The isolated oligonucleotide of claim 1 wherein said oligonucleotide is
   labelled with a moiety to assist with detection or manipulation.

3. The isolated oligonucleotide of claim 2 wherein said moiety is colorimetric,
   chemiluminescent, chromogenic, radioactive, fluorescent, an enzyme, an antibody
   fragment, a His-tag, biotin or streptavidin.

4. The isolated oligonucleotide of any one of claims 1 to 3 wherein said
   oligonucleotide is immobilised on a solid support, preferably selected from particles,
   sheets, gels, filters, membranes, fibres, capillaries, chips, microtitre strips, slides,
   tubes, plates or wells.

5. The isolated oligonucleotide of claim 4 wherein said solid support, is a
   magnetic particle, preferably a magnetic bead.

6. The isolated oligonucleotide of claim 4 or claim 5 wherein said solid support
   is labelled with a dye or a plurality of dyes, preferably a luminescent dye.

7. A method of obtaining information relevant to a diagnosis of Crohn's
   Disease in a subject, said method comprising:
   (i) contacting a sample from the GI tract of said subject with an
   oligonucleotide as defined in any one of claims 1 to 6;
(ii) subjecting the sample and oligonucleotide to conditions which allow hybridisation of the oligonucleotide to its target sequence within nucleic acid molecules in said sample, and

(iii) determining the amount of target nucleic acid in said sample.

8. A method of diagnosing Crohn's Disease in a subject, said method comprising performing the method of claim 7, and (iv) making a diagnosis of Crohn's Disease based on the amount of target nucleic acid in said sample.

9. The method of claim 8, wherein step (iv) comprises comparing the results of step (iii) with previously prepared results or standards from samples from the GI tract of subjects without Crohn's Disease or from samples taken from the GI tract of subjects confirmed as having Crohn's Disease.

10. The method of any one of claims 7 to 9, said method further comprising a further diagnostic measure selected from magnetic resonance imaging, ultrasound imaging, nuclear imaging, X-ray imaging, endoscopy and the molecular analysis of serological markers of Crohn's Disease.

11. The method of any one of claims 7 to 10 wherein said oligonucleotide has a label attached thereto, and step (iii) comprises determining the amount of said label bound to said sample by determining the strength of the signal from the label emanating from the sample.

12. The method of any one of claims 7 to 10 wherein step (iii) comprises
(a) selectively labelling the oligonucleotide when it is hybridised to its target nucleotide sequence within nucleic acid molecules in said sample; and
(b) determining the amount of the labelled oligonucleotide produced in step (a).

13. The method of claim 12 wherein selective labelling occurs by chain extension of the oligonucleotide probe with a labelled nucleotide, preferably a labelled dideoxynucleotide.
14. The method of claim 13 wherein said labelled nucleotide is a labelled ddCTP, preferably ddCTP labelled with biotin.

15. The method of any one of claims 12 to 14 wherein step (b) comprises hybridisation of the oligonucleotide from labelling step (a) to a nucleotide sequence complementary to the oligonucleotide.

16. The method of claim 15 wherein said a nucleotide sequence complementary to the oligonucleotide is immobilised on a solid support, preferably selected from particles, sheets, gels, filters, membranes, fibres, capillaries, chips, microtitre strips, slides, tubes, plates or wells.

17. The method of claim 16 wherein said solid support is a magnetic particle, preferably a magnetic bead.

18. The method of claim 16 or claim 17 wherein said solid support is labelled with a dye or a plurality of dyes, preferably a luminescent dye.

19. The method of any one of claims 7 to 10 wherein step (iii) comprises
   (a) performing a primer-dependent nucleic acid amplification reaction; and
   (b) determining the amount of amplification product produced from the oligonucleotide in said primer-dependent nucleic acid amplification reaction.

20. The method of claim 19 wherein said primer-dependent nucleic acid amplification reaction is PCR.

21. A method of diagnosing Crohn's Disease in a subject, said method comprising
   (a) comparing the results of a method as defined in any one of claims 7 to 20 to previously prepared results or standards from samples from the GI tract of subjects without Crohn’s Disease or from samples taken from the GI tract of subjects confirmed as having Crohn's Disease, and
   (b) making a diagnosis of Crohn's Disease based on the relative amounts of target nucleic acid in the samples being compared.
22. The method of any one of claims 7 to 21 wherein said diagnosis of Crohn's Disease is
   (a) a differential diagnosis of Crohn's Disease in a subject presenting with unclassified IBD,
   (b) a subclassification of ileal Crohn's Disease in a subject with Crohn's Disease, or
   (c) a discriminatory diagnosis between Crohn's Disease/inflammatory bowel disease and a different condition associated with the gastrointestinal microbiota, preferably irritable bowel syndrome.

23. The method of any one of claims 7 to 22 wherein said sample from the GI tract is selected from
   (a) luminal contents of the GI tract, preferably stomach contents, intestinal contents, mucus and faeces/stool, or combinations thereof,
   (b) parts of the mucosa, the submucosa, the muscularis externa, the adventitia and/or the serosa of a GI tract tissue/organ,
   (c) nucleic acid prepared from (a) or (b), preferably by reverse transcription and/or nucleic acid amplification.

24. The method of any one of claims 7 to 23 wherein said GI tract sample is obtained from the jejunum, the ileum, the cecum, the colon, the rectum or the anus.

25. A kit comprising an isolated oligonucleotide comprising the nucleotide sequence ACGTCTGACCCCT (SEQ ID NO 1), the nucleotide sequence complementary thereto (AGGGTGCAAGCGT; SEQ ID NO 2) and/or a nucleotide sequence capable of hybridising under conditions of high stringency to SEQ ID NO 1 or SEQ ID NO 2.

26. The kit of claim 25, said kit further comprising at least one of
   (a) means for selective labelling of the oligonucleotide,
   (b) a nucleotide sequence complementary to the oligonucleotide, preferably immobilised on a solid support,
   (c) means for performing an amplification and/or primer extension reaction with the oligonucleotide,
   (d) means for performing a reverse transcription reaction,
(e) means to take a sample from the GI tract,
(f) means to purify or refine a sample from the GI tract,
(g) means for extracting nucleic acid from a sample from the GI tract,
(h) means to detect amplified nucleic acid,
(i) one or more oligonucleotides that selectively hybridise to target nucleic acids indicative of Crohn's Disease or any other disease or medical condition associated with the gastrointestinal microbiota, preferably inflammatory bowel disease, ulcerative colitis and irritable bowel syndrome.

27. An oligonucleotide array comprising the nucleotide sequence ACGTTGACCCCT (SEQ ID NO 1), the nucleotide sequence complementary thereto (AGGGTGCAAGCGT; SEQ ID NO 2) and/or a nucleotide sequence capable of hybridising under conditions of high stringency to SEQ ID NO 1 or SEQ ID NO 2.
**Fig. 1**

![Bar charts showing frequency distribution with mean, standard deviation, and sample size.](image-url)
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/68
ADD.

According to international Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, Sequence Search, BIOSIS, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>DATABASE Geneseq [Online] 8 March 2007 (2007-03-08), &quot;E. coli DNA PCR primer #24.&quot;, XP002670297, retrieved from EBI accession no. GSN:AEM30592 Database accession no. AEM30592 sequence ----- -/--</td>
<td>1,25</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents :
  * "A" document defining the general state of the art which is not considered to be of particular relevance
  * "E" earlier document but published on or after the international filing date
  * "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  * "O" document referring to an oral disclosure, use, exhibition or other means
  * "P" document published prior to the international filing date but later than the priority date claimed

* "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* "X" document of particular relevance; the claimed invention cannot be considered to be of particular relevance if the invention is taken alone
* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is taken together with one or more other such documents, such combination being obvious to a person skilled in the art.
* "Z" document member of the same patent family

Date of the actual completion of the international search

29 February 2012

Date of mailing of the international search report

13/03/2012

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk
Tel. (+31-70) 340-3040, Fax: (+31-70) 340-3016

Authorized officer

Ripaud, Leslie
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>CN 101 397 586 A (GUANGDONG CT FOR DISEASE CONTR [CN] GUANGDONG CT FOR DISEASE CONTROL A) 1 April 2009 (2009-04-01) claim 1 sequence 208</td>
<td>1,3-6, 25-27</td>
</tr>
</tbody>
</table>

Form PCT/ISA/210 (continuation of second sheet) (April 2008)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

Form PCT/R/SA/210 (continuation of second sheet) (April 2008)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>WO 99/50448 A2 (GENPOINT A S [NO]; DZIEGLEWSKA HANNA [GB]; RUDI KNUT [NO]; JAKOBSEN KJ) 7 October 1999 (1999-10-07) cited in the application pages 29-33</td>
<td>1-27</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>DE 19732086 A1</td>
<td>28-01-1999</td>
<td>NONE</td>
</tr>
<tr>
<td>CN 101397586 A</td>
<td>01-04-2009</td>
<td>NONE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 769566 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 3338999 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2325052 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 1308685 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE 69937447 T2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1068352 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 4481491 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2003516710 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO 20004962 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 6617138 B1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 9950448 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1815016 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2280085 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2009197249 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2006050479 A2</td>
</tr>
</tbody>
</table>