MOLECULAR MARKERS

166166
G00002

FIELD OF THE INVENTION

The present invention relates generally to nucleic acid based methods for detecting the presence of *E. coli* or *Shigella* or related microorganisms in a sample using one or more *E. coli* or *Shigella* species specific nucleotide sequences. More particularly, the present invention permits the identification of molecules capable of binding or otherwise facilitating abnormal cell growth or abnormal physiology such as found in cancer or cellular instability. The present invention further provides molecular probes for performing the nucleic-acid based methods of the invention and methods of testing and selecting nucleic acid sequences suitable for same. The methods and polynucleotides of the present invention are useful *inter alia* in the testing of food and water samples, for testing for genetic and cellular instability, and for testing for benign, pre-neoplastic and neoplastic disease in asymptomatic or symptomatic colorectal or gastric cancer patients or those at risk of the aforementioned conditions or those infected by Escherichieae and with other diseases or conditions.

BACKGROUND OF THE INVENTION

The identification of bacteria can be carried out using biochemical, cultural, antibody recognition and molecular biological tests (Feng PCS and Hartman PA: Fluorogenic Assays for Immediate Confirmation of *Escherichia coli*. 1982. Falkow S, Habermehl KO. ed: Rapid Methods and Automation in Microbiology and Immunology. Springer-Verlag, Berlin 1985: 30-33. AOAC Official Methods of Analysis 1995. Pepper IL, Gerba CP and Brendecke JW: Environmental Microbiology. A laboratory Manual. Academic Press 1995.)

Food and Water Hygiene

Biochemical test and culture medium

The most probable number (MPN) is the common method for the detection and quantitation of *E. coli* in foods. This method detects *E. coli* on the basis of the bacteria's ability to ferment lactose with the evolution of gas. Other non-*E. coli* organisms also ferment lactose and, therefore, several selective enrichment steps are required in order to

sequentially select for coliform bacteria and *E. coli*.

This widely used MPN method has several limitations. Many clinical *E. coli* isolates are lactose negative and thus are not detected using the MPN method. The MPN method requires a minimum of about four days to determine the absence of *E. coli* in food products and about seven days are required to get confirmed results. The growth of some *E. coli*, including the serotype 0157:H7 strains, is severely inhibited by the selectivity of the EC broth at 45.5°C and gas production in the MPN method is susceptible to interference by high levels of competitor organisms.

More rapid methods for detecting *E. coli* are needed because of the time and accuracy limitations of the MPN method. It has been reported that 94% to 97% of *E. coli* strains possess the B-D-glucuronidase that can be detected by specific hydrolysis of a synthetic substrate, 4-methylumbelliferyl-B-D-glucuronide (MUG), to a fluorescent end product. When MUG is incorporated into lauryl sulfate tryptose (LST) broth, 10^7 to 10^8 CFU/ml of *E. coli* will yield this fluorescent product which can be detected under longwave UV light. However, a number of enteropathogenic *E. coli* including serotype 0157:H7 strains, do not possess the B-D-glucuronidase enzyme, do not exhibit fluorescence in LST-MUG medium, and therefore yield false-negative results using the MUG method. In addition, the selectivity of the method is compromised by the fact that some *Shigella*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Salmonella*, and *Yersinia* species also produce B-D-glucuronidase and therefore yield false-positive results.

Another widely used test, the Analytical profile index (API) test strips, produced by BioMerieux (France), may be used to obtain test results quickly. These consist of a series of miniature capsules on molded plastic strips, each of which contains a sterile dehydrated medium in powder form. Addition of water containing a bacterial suspension simultaneously re-hydrates and inoculates the medium. A rapid reaction is obtained because of the small volume of medium and the large inoculum used. The identification of the unknown bacterium is achieved by determining a seven digit profile index number and consulting the API profile recognition system. However, there are strains of *E. coli* that yield a low discrimination value with the API strips.

When this occurs, further identification with sugar test is required for affirmation. Acid production from sugars such as D-Adonitol, Cellobiose, Lactose and D-Xylose are additional biochemical test for differentiation of *Escherichia* species and related species.

DNA Probes

The use of genetic probes in the detection of microorganisms is popular because they obviate the need for pure cultures, and are specific, sensitive, fast and reliable (Fred C. Tenover: DNA probes for infectious diseases. CRC Press, Inc. 1989). In DNA probe test, it is essential to know something about the nucleotide sequence of the microorganisms under investigation.

Bacteria belonging to different families or strains can be differentiated on the basis of heterogeneity in genetic sequences. One approach is the identification and use of specific toxin genes of disease causing strains to distinguish them from the normal flora. Another approach makes use of the conserved and polymorphic sites that are found in bacterial 16S ribosomal RNA (rRNA) sequences not present in human 18 rRNA or human mitochondrial 12S rRNA. The combination of the polymerase chain reaction technique for gene amplification, followed by sequencing of polymorphic regions and phylogenetic analysis of the resulting sequence information can also assist in strain identification. (Relman *et al. The New Engl J of Medicine*, **327**: 293-301, 1992, Kui *et al. FEMS Microbiology Letters* **57**:19-24, 1989. DeLong *et al. Science* **243**: 1360-1363, 1989).

The *E. coli* identification kit produced by gene-trak systems, Framingham, MA, USA, uses DNA oligonucleotides that complement the 16S rRNA. This assay uses hybridization techniques to detect *E. coli*, non-coli *Escherichia fergusonii* and *Shigella* species.

Another way of identifying bacteria specific DNA probes is by using randomly cloned chromosomal fragments. This involves the cloning of restriction enzyme cleaved genomic DNA of a bacteria, and selection of specific clones by determining their hybridization profiles by hybridization against its own species-sequences and other species-sequences. Only clones that hybridize to sequences from the same species but the clones were derived from will be selected (Tenover FC: DNA Probes for Infectious Diseases. CRC Press 1989).

Gastrointestinal Infection

Colorectal cancer is one of the top three cancer killers in the world. Factors implicated in its etiology include inappropriate diet, environmental factors and lack of reliable diagnostic markers. Recently, greater understanding of the genetic predisposition to colon cancer has been achieved through the identification of genes responsible for such susceptibility (Cowell JK, ed: In Molecular Genetics of Cancer. Dunlop MG: Molecular genetics of colon cancer. 1995. 113-134). Despite intensive research efforts, the mortality rate from colorectal cancer has not declined dramatically over the last 40 years.

Markers associated with cancer initiation or progression are important in patient care. Tumours diagnosed at an early stage can usually be cured by surgical excision or polypectomy (surgical excision cures 90% of patients with adenoma or carcinomas that are confined to the mucosa). Patients with advanced disease have a poor prognosis as mortality increases to more than 90% after metastasis takes place.

The gastrointestinal tract is often exposed to a range of microorganisms. When bacteria come into contact with a susceptible host, they can establish either a transient presence, colonize the individual, infect the individual or evolve with the host. The outcome can either be harmless, acute illness or a chronic condition that may lead to a serious outcome (Gibson GR and Macfarlane GT: Human Colonic Bacteria: Role in Nutrition, Physiology, and Pathology. CRC Press, Inc., 1995).

Bacteria have been associated with inflammatory bowel disease such as ulcerative colitis and Crohn's disease (Giaffer *et al. Gut* **33**:646-650, 1992, Cartun *et al. Mod Pathol* **6**:212-219, 1993; Liu *et al. Gastroenterology* **108**:1396-404, 1995). In addition, patients with pan-colitis of long duration are at risk of developing colorectal cancer (Wanebo HJ: In Colorectal Cancer. Lev R: Precursors of Colon Carcinoma 1993; 158-163). Although frequently implicated, the role of bacteria in colon related disease remains ill-defined and controversial. The identification of bacteria in physical proximity to diseased tissue does not provide definitive proof of a causal relationship between a bacterium and the diseased condition. This is especially so when the bacteria are commonly found surrounding the tissue (Swidsinski *et al. Gastroenterology* **115**:281-286, 1998), as is the case in the colon,

and there is no additional information to differentiate between bacteria. It is perhaps more convincing if the bacterium can be shown to be positioned *in-situ* in the diseased tissue and when isolated and characterized found to possess properties that will substantiate its presence within the tissue.

The bacterium *Helicobacter pylori* is an accepted Group 1 (definite) biological carcinogen for gastric cancer and causes of related gastric conditions such as duodenal ulcer, gastric ulcer and ulcer complications. *H. pylori* attaches to and thrives on the gastric mucosa resulting in a chronic immunological response from the host. (Marshall, B.J. *Gastroenterologist* **1**:241-247, 1993). It is not firmly established whether *H. pylori* has invasive properties. However, pathogenic strains have been identified that can cause epithelial cell damage and mucosal ulceration on an intragastric administration to mice (Telford *et al. J Exp Med* **179**:1653-1658, 1994) The question remains whether *H. pylori* is the only important factor in the development of gastric cancer because of its high infection/disease ratio. The current consensus is that there may be other factors other than *H. pylori* infection that are also important in gastric cancer risk (National Institutes of Health Consensus Development Panel on *Helicobacter pylori* in Peptic Ulcer Disease 1994). A separate study put forward the theory that a synergistic interaction between a non-invasive bacteria and other enteropathogens can facilitate invasion by the otherwise non-invasive bacteria (Geir Bukhowm and Georg Kapperud, *Infection and Immunity* **55**:2816-2821, 1987).

Numerous *in-vivo* and *in-vitro* studies have vividly shown that microorganism carry transmissible tumorigenic genetic information. Mutagenesis in such instances is either by transposition or site-specific recombination facilitated by conjugation, transformation and transduction. This information is constantly being exploited scientifically in creating mutants (Sherratt DJ (ed): Mobile genetic elements. Dale JW: Molecular genetics of bacteria. 2nd Edition. John Wiley and Sons Ltd. Oxford University Press 1995). In 1995, Couralin *et al.* showed that invasive strains of *Shigella flexneri* and *E. coli* can carry out gene transfer that are stably inherited and expressed by the mammalian cell progeny (Courralin *et al., C.R. Acad. Sci. Paris* **318**:1207-1212, 1995). Therefore, it is quite possible that the persistent presence of bacterial genetic sequences in the nucleus of mammalian cells can lead to genetic instability that may ultimately give rise to a tumour

cell.

Bacterial invasion can stimulate similar a pattern of protein phosphorylation to that induced by growth factor (e.g. EGF) and cellular proliferative responses may then be altered with consequences for disease progression. (Galan *et al. Nature* **357**:588-589, 1992). In addition, bacterial disruption of cell-cell interaction may affect cell proliferation patterns and differentiation (Epenetos AA and Pignatelli M (ed): Cell Adhesion Molecules in Cancer and Inflammtion; Pignatelli *et al.*: Adhesion molecules in neoplasia: An overview. Chapter 1:1-13. Harwood academic publishers 1995). Cytonecrotizing factors have been identified that can cause formation of large multinucleated cells and cells spreading in tissue cultures. (Denko *et al. Experimental Cell Research* **234**:132-138, 1997; Lemichez *et al.*, *Molec Microbiol* **24**:1061-1070, 1997; Machesky, L.M. and Hall, A, *TICB* **6**:304-310, 1996). Accordingly, the persistence presence of bacteria can cause cellular changes leading to cell disorientation, proliferation and changes in cell morphology.

One cancer causing effect of bacteria is when *Agrobacterium tumefaciens*, a soil phytopathogen, genetically transforms plant cells by the transfer of the tumour-inducing (Ti) plasmid to the plant genome where its integration and expression result in the crown gall phenotype. A crown gall is a tumorous proliferation of plant cells which are released from normal metabolic and reproductive controls (Hughes MA: Plant Molecular Genetics. Addison Wesley Longman Ltd. 1996).

People travelling across continents may suffer from traveler's diarrhoea as the bacteria they are exposed to are not common in their country. The assays/kits that are used for detecting microorganisms in the Asia-Pacific region are imported from other continents and these imported assays/kits may not be as sensitive or as specific for the bacteria in the Asia-Pacific region.

Microorganisms transmitted by water and food usually grow in the intestinal tract of man and animals and leave the body in the faeces. Bacteria are known to possess gene sequences that make them toxigenic, hemorrhagic, invasive and adherent to tissues. Acute bacterial infection is well documented but it is still not known that if bacteria that do not cause overt symptoms but persist and remain undetected in their host can cause diseases

with time. Therefore, it is important that the assays that are available are sensitive and specific for a wide range of pathogens.

The *E. coli* genetic sequence is published. (Blattner *et al. Science* 277:1453-1474, 1997). Some of its genetic sequence has homology to other bacteria (Janda JM and Abbott SL: The Enterobacteria. Lippincott-Raven Press 1998). The inventor, in accordance with the present invention, has identified *E. coli* DNA sequences which are unique to the Escherichae family and furthermore has shown that biochemical and cultural tests presently available are not adequate for detecting this family of bacteria. The present polynucleotide sequence in the genome of strains of the Escherichieae genus (*Escherichia* and *Shigella*), have proven to be more informative than the agar plates EMB, MacConkey and MUG. They can be used to detect *E. coli* that is either EMB, MacConkey or MUG negative. The sequences are also found in 0157:H7 and 029:NM strains of *E. coli.* Therefore, the present molecular markers provide improved tools for the detection and characterization of *E. coli.*

In addition, the invention permits the use of the sequence(s) to study the outcome of tissue infection in-situ. The present gene sequences are more specific than the gene-trak sequence (gene-trak systems) and the sequences can be amplified many-fold to increase their detection limit. This makes the present invention useful for studying the role of microorganism in gastrointestinal and other disease conditions. The presence of the polynucleotide sequence in cells can be located by the use of the polymerase-chain-reaction amplification technique *in-situ* followed by hybridization to the *in-situ* amplified signals with sequence specific DNA probe.

The identification of these specific polynucleotide sequence(s) that can be used to detect for the presence of strains of *E. coli* and *Shigella* and related microorganisms in food, water, fecal specimens, tissues, secretions and other biological, environmental and/or laboratory samples is important for health reasons as it enables one to check on the quality of food and water hygiene and monitor transmission of the microorganism. Sensitive detection techniques and methods for assessing the role of bacteria in clinical conditions will ultimately help in the control of harmful microorganisms.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

One aspect of the present invention provides a new use for the whole of Formula I and sequences within Formula I, as markers for species of bacteria within the Escherichieae family specifically *Escherichia coli* and *Shigella* species or related microorganisms.

Another aspect of the present invention relates to the use of the polynucleotide sequence of formula I to generate gene probes of smaller size which singularly or in combination have specificity for *E. coli* strains or related microorganisms but not necessarily specific for all the *Shigella* species.

A further aspect of the present invention provides Formula I and the smaller gene sequences within it as a means to detect the presence, in liquids, semisolids and solids combinations thereof or in aerosols or gases, of species of bacteria within the Escherichieae family specifically *E. coli* and some or all of the *Shigella* species so that high standard of sanitation can be achieved.

Yet another aspect of the present invention provides the aforementioned sequence(s) as a means to detect infection in a sample and/or a combination of samples by members of the Escherichieae family as aforementioned. Samples are defined in this invention as tissues or cells or explants of either human, animals or plant origin, such given examples being tissue/cells found in the colon, stomach, and other parts of the human or animal anatomy as well as in food, industrial and/or environmental samples.

A further aspect of the present invention provides a method for testing and identifying the various genes within formula I as new means to detect for changes in DNA content in cells infected or previously infected with the aforesaid Escherichieae family of bacteria (*E. coli* and *Shigella* species or related microorganisms). A cell is defined in this invention as a cell found in the animal and plant kingdom. Changes in DNA content in a cell in accordance

with this invention includes DNA sequences found in the cell which differs by one or more nucleotide substitutions, additions and/or deletions of existing DNA or by the introduction of a heterologous DNA.

Yet another aspect of the present invention provides the aforementioned Formula I within which polynucleotide sequences are a marker for use in recognizing early cellular DNA changes associated with any one or more members of the Escherichieae family (*E. coli* and *Shigella* species and related microorganisms) in the colonic epithelium before the histology criteria for such cellular changes are detectable. Early changes are defined in this invention by the presence of at least bacterial DNA sequences that are present in high, low copy numbers or present as single copies per haploid genome in a normal population.

Still yet another aspect of the present invention provides the aforementioned Formula I within which polynucleotide sequences are a marker for use in recognizing pre-malignant changes associated with any one or more members of the Escherichieae family (*E. coli* and *Shigella* species and related microorganisms) in the colonic epithelium as defined by histology criteria for such pre-malignant tissue. Pre-malignant changes are defined in this invention by the presence of at least bacterial DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population and are supported by histology criteria.

Even still another aspect of the present invention provides the aforementioned Formula I within which polynucleotide sequences can be a marker for use in recognizing malignant changes associated with any one or more members of the Escherichieae family (*E. coli* and *Shigella* species and related microorganism) in the colonic epithelium and malignant colonic tumours residing in other tissues. Malignant changes are defined by histology criteria. Malignant changes associated with any one member of the Escherichieae family are defined in this invention by the presence of at least bacterial DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population.

Another aspect of the present invention provides the aforementioned formula I within which polynucleotide sequences can be used as a marker for detecting pre-malignant

changes associated with any one or more members of the Escherichieae family in the gastric mucosa as defined by histology criteria for such pre-malignant tissues. Pre-malignant changes are defined in this invention by the presence of at least bacteria DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population and its histology criteria for the tissue defined.

Yet another aspect of the present invention provides the aforementioned Formula I sequence within which polynucleotide sequences can be used as a marker to recognize malignant changes associated with any one or more members of the Escherichieae family in the malignant gastric tumours and malignant gastric tumours residing in other tissues. Malignant changes are defined *inter alia* by histology criteria. Malignant changes associated with the Escherichieae family in this invention is defined by the present of at least DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population.

Still another aspect of the present invention provides the aforementioned Formula I within which sequences can be used as markers to recognize patients that are found harboring any one or more member of the Escherichieae family relative to normal patients not haboring the same and are thus identified as marker of infection of said family that are important in patient care.

Even still another aspect of the present invention provides the aforementioned Formula I sequence within which polynucleotide sequences can be used as markers being found in colorectal cancer patients relative to normal patients and thus identified as a marker of malignant disease that is important in patient care.

Even yet another aspect of the present invention provides the aforementioned marker that is found in gastric cancer patients relative to normal patients and is thus identified as a marker of malignant disease that is important in patient care.

Another aspect of the present invention provides a marker for cellular instability and therefore a marker for predisposition to cellular carcinogenesis. Cellular instability may occur as a forerunner to cellular carcinogenesis or other condition and is characterized

herein by changes in DNA content comprising one or more nucleotide substitutions, additions and/or deletions of existing DNA or by the presence of heterologous DNA.

Another aspect of the present invention provides a method of testing and selecting sequences in *E. coli* and *Shigella* species and related microorganisms as markers for use to detect changes in DNA content in cells in order to recognize cellular instability and, therefore, predisposition to cellular carcinogenesis, predisposition to colon and gastric cancer and as markers for use in recognizing benign, pre-malignant and malignant gastrointestinal tissues as optionally defined by histology criteria.

In accordance with the present invention, it is shown that Formula I comprises polynucleotide DNA sequence marker(s) for the Escherichieae family specifically *E, coli* species and *Shigella* species or related microorganisms. This Formula I and the various genes and sequence it contains allow the differentiation of the aforementioned members of the Escherichieae family from other bacteria families. In addition, the presence of such bacteria as indicated by the presence of the DNA sequences allows study of sanitation and health related matters such as infection, predisposition to cancer, cancer and cell instability.

Reference to "related microorganisms" includes microorganisms which are related at the immunological, biochemical, disease-causing, physiological or genetic levels. A derivative or mutant form of *E. coli* or *Shigella* species is an example of a related microorganism.

The present invention furthermore provides a method of testing and selecting other sequences in *E. coli* and *Shigella* species and related microorganisms as markers to test for their presence in cells with abnormal cell growth or physiology associated with cancer or a predisposition to the development of cancer.

Yet still another aspect of the present invention relates to a new use for the various polynucleotide sequences within Formula I as molecular probes in the determination of whether samples contain members of the Escherichieae family such as *E. coli* and *Shigella* species or related microorganisms.

In another aspect, the present invention provides methods for enhancement in the specificity and sensitivity of detecting the presence, among other bacteria of *E. coli* species and some of the *Shigella* species with some of the aforementioned sequences.

The presence of the polynucleotides sequences in food and water is evidence that they are contaminated with members of Escherichieae family such as *E. coli* species and probably some or all of the *Shigella* species. Thus the present molecular probes provide an alternative to microbiological and biochemical assays which are less specific, sensitive, reliable, often required for pure cultures, and are more time consuming.

A further related aspect of the present invention provides a new use for the sequences within the Formula I for determining whether tissue samples contain the DNA markers that originate from members of the *E. coli* and *Shigella* species. Both species within this family are known to have invasive, adherent and toxigenic properties. This aspect relates to the new use of polynucleotide sequence(s) within Formula I as marker(s) for detecting infection by identifying samples such as, for example, colonic and gastric mucosa tissues that contain them. The presence of the polynucleotide sequences in tissues is evidence that the tissues are infected by members of the *E. coli* and *Shigella* species.

A further aspect of the instant invention relates to a new use for the sequences within the Formula I for determining which cell type within tissues samples contain marker DNA sequences that originate from members of the Escherichieae family such as *E. coli* and the *Shigella* species and related microorganisms.

A further aspect of the instant invention relates to a new use for the aforementioned polynucleotide sequence(s) within Formula I as a marker(s) for detecting changes in cellular DNA composition in, for example, colonic and gastric mucosa cells before histology criteria for changes are detectable. The presence of the polynucleotide sequences in cells of tissues is evidence that the cells are infected by members of the *E. coli* and *Shigella* species. Changes in cellular DNA composition are defined in this particular aspect of the invention by the presence of at least bacteria DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population. The polynucleotide sequence of the present invention are only found in

Escherichieae family and, therefore, their presence in other species such as in eukaryotic cells, for an example, is a sign of an abnormal event. Accordingly, the present invention provides one or more molecular marker for screening patients to identify those who are at risk of having gastrointestinal tumours (benign, pre-malignant, or malignant).

A further additional aspect of the present invention relates to a new use for the sequences within the Formula I for determining whether pre-malignant tumours as defined by histology criteria contain aforementioned polynucleotides sequences that originate from members of the *E. coli* and *Shigella* species or related microorganisms. This aspect relates to the new use of the marker for detecting the presence of any one or more member of the Escherichieae family in the pre-malignant tumours such as colonic and gastric tissues as defined by histology criteria. The presence of the polynucleotide sequence in the cells of pre-malignant colonic and gastric tumours is evidence that the cells are infected by or contain DNA sequences of members of the *E. coli* and *Shigella* species. These pre-malignant tumours contain the presence of at least bacteria DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population. The polynucleotide sequence of the present invention are only found in Escherichieae family and, therefore, their presence in other species such as in eukaryotic cells, for an example, is a sign of an abnormal event. Accordingly, the present invention provides one or more molecular markers for screening patients to identify those at risk of having gastrointestinal tumours (benign, pre-malignant, malignant).

The present invention also relates to a new use for the aforementioned polynucleotide sequence(s) as a marker for determining whether malignant changes in the colonic and gastric mucosa contain marker sequences that originate from members of the *E. coli* and *Shigella* species. Malignant changes are defined by conventional histological criteria. This aspect of the invention relates to the new use of the marker for detecting the DNA presence of any one or more member of the Escherichieae family in the malignant tumours. The presence of the polynucleotide sequence in the cells of these malignant tumours is evidence that the cells are infected by or contain DNA sequences of members of the *E. coli* and/or *Shigella* species. These malignant tumours contain the presence of at least bacteria DNA sequences that are present in high, low copy numbers or present as single copies per haploid genome in a normal population. The polynucleotide sequence of the present

invention are only found in Escherichieae family and, therefore, their presence in other species such as in eukaryotic cells, for an example, is a sign of an abnormal event. Accordingly, the present invention provides one or more molecular markers for screening patients having gastrointestinal tumours (benign, pre-malignant, malignant).

The instant invention provides in a related embodiment a new use for the aforementioned polynucleotide DNA sequence(s) as markers for determining the presence of any member of the *E. coli* and *Shigella* species in metastatic cells of colonic or of gastric tumour origin residing in other tissues. This aspect of the invention relates to the new use of the marker sequence for detecting the DNA presence of any one or more member of the *E. coli* and/or *Shigella* species in the metastatic cells. The presence of the polynucleotide sequence in the cells of these malignant tumours is evidence that the cells are infected by or contain DNA sequences of members of the *E. coli* and/or *Shigella* species. These metastatic cells contain the presence of at least bacteria DNA sequences that are present in high or low copy numbers or present as single copies per haploid genome in a normal population. The polynucleotide sequence of the present invention are only found in Escherichieae family and, therefore, their presence in other species such as in eukaryotic cells, for an example, is a sign of an abnormal event. Accordingly, the present invention provides one or more molecular markers for screening patients having gastrointestinal metastatic cells.

The present invention furthermore relates in a different aspect to a new use for the formula I sequence that is found in *E. coli* and *Shigella* species or in related microorganisms, as a marker for determining cells that possess it. The invention relates to the new use of the marker sequence for detecting the DNA presence of any one or more member of the *E. coli* and/or *Shigella* species in the cells. This gene sequence is only found in Escherichieae family and therefore its presence in high, low copy numbers or as single copies per haploid genome in a normal population in eukaryotic cells, for an example, is a sign of an abnormal event that may lead to genetic instability of cell that possess it. It provides as a molecular marker for risk of and genetic instability and therefore tumourigenesis.

Even more particularly the present invention, in one aspect, provides a method for detecting the presence of *E. coli* or *Shigella* species or related microorganisms in a sample, said method comprising subjecting a nucleic acid molecule preparation from said sample to

genetic analysis using one or more *E. coli*- or *Shigella* species'-specific nucleotide sequences obtainable from one or more nucleotide sequences of Formula 1 and/or Table 1 wherein the ability for said *E. coli*- or *Shigella* species'-specific nucleotide sequences to hybridize to complementary nucleotide sequences in the nucleic acid preparation is indicative of the presence of *E. coli, Shigella* species or related microorganisms.

In a further aspect of the present invention there is provided a method for detecting the presence of *E. coli* and/or *Shigella* species or related microorganisms in a sample as hereinbefore described wherein the nucleotide sequences of Formula I comprises from nucleotide position 246 of GenBank Accession No. AE000201 to nucleotide position 6693 of GenBank Accession No. AE000203 including the nucleotide sequence of GenBank Accession No. AE000202.

Still a further aspect of the present invention provides a method for detecting the presence of *E. coli* or *Shigella* species or related microorganisms in a sample as hereinbefore described wherein the *E. coli*- and/or *Shigella* species'-specific nucleotide sequences comprises at least 8 nucleotides in length.

A related aspect of the present invention discloses a method for detecting the presence of *E. coli* or *Shigella* species or related microorganisms in a sample as hereinbefore described wherein hybridization of *E. coli*- and/or *Shigella* species'-specific nucleotide sequences to the nucleic acid preparation is detected by the presence of amplified nucleic acid products.

A further related aspect of the present invention provides a method for detecting the presence of *E. coli, Shigella* species or related microorganisms in a sample wherein hybridization of *E. coli*- and/or *Shigella* species'-specific nucleotide sequences to the nucleic acid preparation or the presence of amplified nucleic acid products is detected by a reporter molecule giving an identifiable signal.

Still yet a further related aspect of the present invention provides a method for detecting the presence of *E. coli, Shigella* species or related microorganisms in a the sample wherein the sample comprises food, water, semi-solids or semi-liquid material, mammalian tissue, tissue extract or cells of tissue or normal tissue or tissue predisposed to cancer growth or

malignancy or cellular instability.

In a particularly preferred aspect of the present the mammalian tissue is associated with colon, stomach or colorectal tissue.

A related aspect of the present invention provides a method for identifying nucleotide sequences, or their expressed products, capable of inducing or otherwise facilitating abnormal cell growth or abnormal physiology, said method comprising introducing a nucleotide sequence comprising *E. coli*- and/or *Shigella* species'-specific nucleotide sequences from the nucleotide sequences in Formula I into cells and observing morphological and/or physiological changes to said cells compared to control cells without said introduced nucleotide sequences wherein the presence of abnormal morphology and/or physiology in a cell is indicative of a nucleotide sequence from Formula I, or a polypeptide expressed therefrom, which is capable of inducing or facilitating abnormal cell growth or physiology.

In a further preferred aspect of the instant invention the abnormal cell growth or physiology is associated with cancer or a predisposition to the development of cancer or cellular instability.

A further aspect of the present invention provides a molecular probe comprising at least 8 nucleotides obtainable from the nucleotide sequences of Formula I wherein said molecular probe is capable of specifically hybridizing to *E. coli*- and/or *Shigella* species'-derived nucleic acid molecules.

Yet more particularly the present invention encompasses a use of a nucleotide sequence obtainable from the nucleotide sequence of Formula I in the manufacture of a molecular probe for the identification of *E. coli* and/or *Shigella* species and/or for the identification of a cellular instability or a cancer or tumor or a predisposition to development of same.

Still even yet more especially the present invention provides a method for testing and selecting other sequences in *E. coli*, *Shigella* species or related microorganisms in a sample, said method comprising subjecting a nucleic acid molecule preparation from said

sample to genetic analysis using one or more *E.coli* or *Shigella* species'-specific nucleotide sequences obtainable from one or more nucleotide sequences of Formula 1 and/or Table 1 wherein the ability for said *E. coli*- or *Shigella* species'-specific nucleotide sequences to hybridize to complementary nucleotide sequences in the nucleic acid preparation is indicative of an *E. coli* or *Shigella* species'-specific nucleotide sequence.

Another aspect of the instant invention provides a molecular probe of at least 8 nucleotides, identified by the methods as herein described wherein said probe comprises a sequence of nucleotides from Formula I and wherein said molecular probe is capable of specifically hybridizing to *E. coli* and/or *Shigella* species'-derived nucleic acids.

The invention still yet provides the use of a nucleotide sequence identified by the methods herein disclosed in the manufacture of a molecular probe for the identification of *E. coli*, *Shigella* species and/or for the identification of a cellular instability or a cancer or tumour or a predisposition to development of same.

A related aspect of the instant invention discloses a use of a nucleotide sequence specific to *E. coli* and/or *Shigella* species and/or related microorganism in the manufacture of a molecular probe for the identification of one or more gastrointestinal cancers or tumours or a predisposition to the development of same.

A final preferred aspect of the instant invention provides a molecular probe comprising a nucleotide sequence specific to *E. coli* and/or *Shigella* species and/or related micro-organism for the identification of one or more gastrointestinal cancers or tumours or a predisposition to same.

Other aspects, features and advantages of the present invention will become apparent from the detailed description that follows, or may be learned by practice of the invention.

For the sake of brevity, reference to specific microorganisms such as *Escherichia coli* (*E. coli*) or *Shigella* species includes reference to related microorganisms.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying figures serve to further explain the principles of the instant invention. It is to be understood, however, that the figures are designed for purposes of illustration only, and not as a definition of the limits of the invention for which reference should be made to the claims appearing at the end of the description.

Figure 1, a-c. Schematic diagram of the various locations of the genes tested that is within the polynucleotide sequence of formula I. The formula I sequence extends from nucleotide position (nt) 246 of GenBank accession #AE000201, including sequence of GenBank accession #AE000202 to nucleotide position 6693 of GenBank accession #AE000203.

Figure 2. Autoradiograph result of probe A hybridizing to a panel of bacteria DNA as listed in Table 2a grid C. Probe A consists of fragments 1,2,3 and 4 as depicted in Figure 1, a-c. Each fragment is generated by primer directed PCR carried out on K12 *E. coli* DNA and subsequently combined for ^{32}P labeling and hybridization. The gene sequence spans between nucleotide position 1163 of AE000201 through AE000202 to 503 of AE000203. The primer pairs used are: ECM-1163, torT-5750 (fragment 1, AE000201); torT-5129 AE000201, CD-1351 AE000202 (fragment 2); CD-415, ycdG 7359 (fragment 3, AE000202); ycdG-6073 AE000202, New2-503 AE000203 (fragment 4). Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 3. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 246 to 850 of AE000201. ECM-246 and ECM-850 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM NA pyrophosphate at 65°C.

Figure 4. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 1163 to 1958 of AE000201. ECM-1163 and ECM-1958 are the primers used to generate the gene

probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 5. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 7218 to 7761 of AE000201. Primers tor C-7218 and tor C-7761 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 6. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 8332 to 8891 of AE000201. Primers tor A-8332 and tor A-8891 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 7. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 10574 to 11160 of AE000201. Primers tor D-10574 and tor D-11160 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 0.1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 8. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 415 to 1351 of AE000202. CD-415 and CD-1351 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 0.1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 9. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as

listed in Table 2a grid B. The gene sequence spans between nucleotide position 3151 to 4359 of AE000202. Primers agp-3151 and agp-4359 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 10. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid C. The gene sequence spans between nucleotide position 4807 to 5235 of AE000202. Wrb-4807 and Wrb-5235 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 11. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 6073 to 7359 of AE000202. Primers ycdG-6073 and ycdG-7359 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 0.1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 12. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid A. The gene sequence spans between nucleotide position 7223 to 7794 of AE000202. 81B-7223 and 81B-7794 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 13. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 7278 to 7773 of AE000202. 81B-7278 and 81B-7754 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 14. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid C. The gene sequence spans between nucleotide position 7419 to 7985 of AE000202. OH-7419 and OH-7985 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 15. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 7562 to 7794 of AE000202. OH-7562 and 81B-7794 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 16. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid C. The gene sequence spans between nucleotide position 8160 to 9704 of AE000202. New1-8160 and New1-9704 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 17. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 9731 to 11375 of AE000202. New2-9731 and B-11375 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 5 X SSC, 0.05% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 18. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid C. The gene sequence spans between nucleotide position 9731 of AE000202 to 503 of AE000203. New 2-9731 and New2-503 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred

nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 19. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid B. The gene sequence spans between nucleotide position 5944 to 6693 of AE000203. Primers putP-5944 and putP-6693 are used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 20. Autoradiograph result of radiolabeled probe A hybridized to Enterobacter cloacae and K12 *E. coli* genomic DNA as depicted in Table 2a grid E. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 21. Autoradiograph result of radiolabeled gene probe hybridized to Enterobacter cloacae and K12 *E. coli* genomic DNA as depicted in Table 2a grid E. The gene sequence spans between nucleotide position 7562 to 7794 of AE000202. OH-7562 and 81B-7794 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 22. Autoradiograph result of radiolabeled gene probe hybridized to Enterobacter cloacae and K12 *E. coli* genomic DNA as depicted in Table 2a grid E. The gene sequence spans between nucleotide position 7223 to 7794 of AE000202. 81B-7223 and 81B-7794 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 23. Autoradiograph result of radiolabeled gene probe hybridized to bacteria DNA as listed in Table 2a grid D. The gene sequence spans between nucleotide position 7278 to 7773 of AE000202. 81B-7278 and 81B-7754 are the primers used to generate the gene probe by PCR amplification of K12 *E. coli* genomic DNA. Five hundred nanogram DNA is loaded per dot. Post hybridization wash condition is 1 x SSC, 0.1% w/v SDS, 20 mM Na

pyrophosphate at 65°C.

Figure 24. Autoradiograph result of ^{32}P radiolabeled *H. pylori* ribosomal gene probe hybridized to bacteria DNA as listed in Table 2a grid A. Primer pairs indicated are used for PCR amplification of *H. pylori* genomic DNA to generate the required gene segment. Five hundred nanogram of genomic DNA is loaded per dot. Post hybridization wash condition is 0.1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 25. Autoradiograph result of ^{32}P radiolabeled *H. pylori* ribosomal gene probe hybridized to bacteria DNA as listed in Table 2a grid D. Primer pairs indicated are used for PCR amplification of *H. pylori* genomic DNA to generate the required gene segment. Five hundred nanogram of genomic DNA is loaded per dot. Post hybridization wash condition is 0.1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C.

Figure 26. Autoradiograph result of *in vitro* simulated PCRISH. *H. pylori* ^{32}P radiolabeled ribosomal gene probe is hybridized to products generated from its primer directed PCR amplification of *H. pylori* and *E. coli* genomic DNA and total DNA of *H. pylori* and *E. coli* isolates obtained from patients' fecal specimens. The post hybridization wash condition is 5 x SSC, 0.05% w/v SDS, 20 mM Na pyrophosphate at 65°C. See Table 2a grid F.

Table 1. Oligonucleotide Primers.

Table 2a. Grids A,B,C show the different types of bacteria genomic DNA loaded onto corresponding nylon plus membrane and hybridized to random primed ^{32}P-radiolabeled gene probes. Columns W to Y for all three grids have the same panel of bacteria DNA. Column Z as indicated, has only a few bacteria DNA that is common among them. DNA of *E. coli* isolates obtained from patient fecal specimens are: 219/1, 196/1, 196/28, 197/5, 218/40, 142/31,179/36, and 117/3B. Patient's *Shigella* sonnei isolate is 219/1. 078:H11 and 0157:H7 are commercial *E. coli* strains. TG2 is a gift from Gibson TJ. Placental DNA is from commercial source (Sigma, UK). SssDNA is sonicated denatured salmon sperm DNA (Sigma, UK). These membranes exist in replicates and are hybridized to different radio-labeled gene probes. 500 ng DNA is loaded per dot.

Table 2a. Grid D correspond to nylon plus membrane that contain DNA from *E. coli* and gram positive bacteria isolated from patients' fecal specimen. Unless otherwise stated, all are *E. coli* DNA. Placental DNA, K12 and 0157:H7 are from commercial source. 114/3g is Streptococcus group D DNA, 115/TA is *Streptococcus* group G DNA, 116/TC is *Aeromonass sobria* DNA, 116/TD is *Streptococcus viridans* DNA, 117/2D is *Streptococcus* group D DNA, 154/9 is unidentified gram positive bacteria DNA and HP is *Helicobacter pylori* DNA. 500 ng DNA is loaded per dot.

Table 2a. Grid E correspond to nylon plus membranes that contain in duplicate a range of different amount of Enterobacter cloacae and K12 *E. coli* DNA. The membranes are then hybridized to different ^{32}P-radiolabeled gene probes to determine the level of crossover between the 2 different species of bacteria.

Table 2a. Grid F correspond to nylon plus membrane that contains different amount of *E. coli* DNA, *H. pylori* DNA and *H. pylori* primer directed PCR product of *E. coli* and *H. pylori* DNA. This is hybridized to ^{32}P radiolabeled *H. pylori* ribosomal gene probe (HP).

Table 2b. Bacteria source.

Table 3. Result of *in-vitro* simulation of PCRISH. The primer directed PCR amplification of bacteria DNA has resulted in amplification of product(s). This is indicated by the appearance of band(s) upon gel resolution of a given aliquot of a post PCR mixture (not shown). Post PCR mixture having multiple bands is denoted as positive if any one DNA band migrates at the expected molecular weight of the intended target product. Hybridization with corresponding radiolabeled gene probe has not picked up any non-specific bands other than the single band of the intended target.

Figures 27-34: Detection of bacteria DNA in biopsies and surgical specimens obtained from the colon by polymerase-chain-reaction-*in-situ*-hybridization technique (PCRISH). For the detection of 81B gene sequence (marker for *E. coli/Shigella* species except *Shigella boydii*) PCR *in-situ* amplification is carried out with outer primers 81B-7223 and 81B-7794. The PCR digoxigenin labeled 81B gene probe is made with inner primers 81B-7278 and 81B-7754. For detection of *H. pylori* ribosomal gene sequence, PCR in situ

amplification is carried out with the outer primers HP-178 and HP-775. The PCR digoxigenin labeled *H. pylori* probe is made with inner primers HP-228 and HP-513. Positive signals are denoted by the dark spots and an example is highlighted by the arrow. Tissue condition is defined by histologic criteria.

Figure 27: Hyperplastic polyp with no evidence of malignancy tested with 81B probe.

Figure 28: Adenomatous polyp, tubulovillous type tested with 81B probe.

Figure 29: Well differentiated adenocarcinoma tested with 81B probe.

Figure 30: Liver tissue: metastatic, poorly differentiated adenocarcinoma with primary in the gastrointestinal/pancreatico-biliary tract tested with 81B probe.

Figure 31: Normal mucosa next to tumour tested with 81B probe.

Figure 32: Normal mucosa tested with 81B probe.

Figure 33: Colonic mucosa diagnosed with proctitis tested with 81B probe.

Figure 34: Liver specimen: metastatic, poorly differentiated adenocarcinoma with primary in the gastrointestinal/pancreatico-biliary tract. Specimen tested negative with *H. pylori* ribosomal gene probe.

Figures 35-44: PCRISH detection of 81B DNA (marker for *E. coli/Shigella* except *Shigella boydii*) and *H. pylori* DNA in biopsies and surgical specimens obtained from the stomach. *H. pylori* probe has not been tested against closely related Helicobacter species, and therefore does not claim to detect only *H. pylori*. Positive signals are denoted by the dark spots and an example is highlighted by the arrow. PCR *in-situ* amplification is carried out with outer primers 81B-7223 and 81B-7794 for the detection of the 81B gene and HP-178 and HP-775 for the detection of *H. pylori* gene. The PCR digoxigenin labeled 81B gene probe is made with inner primers 81B-7278 and 81B-7754, while HP-228 and HP-513 primers are used for the *H. pylori* gene probe. Tissue condition is defined by histologic

criteria.

Figure 35: Adenocarcinoma of stomach tested with 81B probe.

Figure 36: Normal gastric mucosa adjacent to gastric tumour tested with 81B probe.

Figure 37: Liver specimen: metastatic, poorly differentiated adenocarcinoma from gastric cancer, tested with 81B probe.

Figure 38: *H. pylori* negative normal gastric mucosa tested with 81B probe.

Figure 39: *H. pylori* negative normal gastric mucosa tested with 81B probe. Specimen is the same as the one shown in Figure 38 but showing a different area.

Figure 40: Normal gastric mucosa adjacent to gastric tumour tested with *H. pylori* probe.

Figure 41: Adenocarcinoma of stomach tested with *H. pylori* probe.

Figure 42: Liver specimen: metastatic, poorly differentiated adenocarcinoma from gastric cancer tested with *H. pylori* probe.

Figure 43: Active chronic gastritis in the presence of *H. pylori* and tested with *H. pylori* probe.

Figure 44: Active chronic gastritis in the presence of *H. pylori*, tested with 81B probe.

Figure 45: GenBank Accession Nos. AE000201, AE000202 and AE000203 comprising respectively sections 91, 92 and 93 of 400 of the complete genome sequences of *Escherichia coli* K-12.

DETAILED DESCRIPTION OF THE INVENTION

The invention herein identifies *inter alia* a new use for Formula I and nucleotide sequences that are part of Formula I (Figure 1, a-c) as molecular markers. The Formula I sequence extends from nucleotide position (nt) 246 of GenBank accession #AE000201, including sequence of GenBank accession #AE000202 to nucleotide position 6693 of GenBank accession #AE000203.

The various DNA fragments for use as hybridization probes can be generated by polymerase chain reaction (PCR) using the primer sequences tabulated in Table 1. In this invention the various DNA fragments are shown to have a new use for identifying bacteria in the Escherichieae family such as *E. coli* and *Shigella* species. Each DNA fragment generated is specific for *E. coli* species and some *Shigella* species as some of the gene sequences show different specificity towards *Shigella boydii* and *Shigella Flexneri* (Table 2a, Figure 2-19).

The DNA fragments highlighted in this invention, as examples, and including Formula 1, can be use in detecting the aforementioned bacteria presence in liquids, solids, semi-solids to identify the level of sanitation, in monitoring the level/depth of infection, in studying the association of the presence of sequences within Formula I in gastrointestinal conditions or other clinical conditions where their presence/aforementioned bacteria is sought. The invention finds particular utility in monitoring for the presence of metastatic gastric and/or colon tumour cells. In addition, it can be used to study the association of such aforementioned sequences with gastrointestinal cancer risk. In addition, it can be used for studying the stability of genome sequence in cells that contains sequences of Formula I whether in parts, complete or in association with sequence upstream or downstream of Formula I as can be obtained in the published sequence of *E. coli* strains (example of source: GenBank database). Furthermore, the invention shows that when the sequences of Formula I are found inside cells of tissues, it is useful to identify specimens where the pathogen can be isolated and identified by probing with gene sequences for pathogenic properties, especially when the pathogen does not cause overt symptoms.

Accordingly, one aspect of present invention provides a method for detecting the presence

of *E. coli* or *Shigella* species or related microorganisms in a sample, said method comprising subjecting a nucleic acid molecule preparation from said sample to genetic analysis using one or more *E. coli*- or *Shigella* species'-specific nucleotide sequences obtainable from one or more nucleotide sequences of Formula 1 and/or Table 1 wherein the ability for said *E. coli*- or *Shigella* species'-specific nucleotide sequences to hybridize to complementary nucleotide sequences in the nucleic acid preparation is indicative of the presence of *E. coli*, *Shigella* species or related microorganisms.

In a particularly preferred embodiment Formula I comprises from nucleotide position 246 of GenBank Accession No. AE000201 to nucleotide position 6693 of GenBank Accession No. AE000203 including the nucleotide sequence of GenBank Accession No. AE000202.

Specificity of DNA Probe.

The invention herein describes Formula I and the various genes it encodes, which characteristically permit identification of members of the Escherichieae family especially *E. coli* species and *Shigella* species.

The various DNA gene probes are generated by primer directed PCR amplification carried out on *E. coli* genomic DNA. The choice of primer sequence is assisted with the use of the software programme, PCR PLAN (PC/Gene system, Intelligenetics, Inc. USA) as applied on the published sequence of *E. coli* obtained from GenBank database accession #AE000201, #AE000202, and #AE000203. The suitability of a gene probe will depend on its specificity. The specificity of the primer sequence and gene probe will have to be determined empirically in a step-wise systematic fashion along the Formula I sequence or by use of software programmes. The gene fragment generated by primer directed PCR or the primer sequence should preferably be without degeneracy when tested on the panel of bacteria of related species. For the gene fragment, any variation observed is further characterized so that the variation is consistent with a particular strain or species (e.g. *E. coli* versus *Shigella*) but yet having homology to the test probe. Each gene probe fragment is preferably purified after amplification to ensure specificity.

There are many ways of obtaining such bacteria-specific fragments: this can be achieved, for example, by cloning the bacteria DNA, either in parts or complete, and screening with

oligonucleotide or genomic probes of *E. coli*. The species or strain specificity of the cloned probes is determined.

The dot blot format (96 well dot-blot manifold, Bio-Rad, USA) is used here for creating a panel of bacteria genomic DNA to test for the specificity of the DNA probes (Table 2a). Reference strains are obtained from American Type Culture Collection (Table 2b. ATCC, Rockville, MD, USA). A representative bacteria from each member of the enterobacteriaceae family is chosen for the panel on which the probes are to be tested against. Gram-negative and gram-positive microorganisms obtained from subjects/ patients are used as test samples (Table 2a, grid D). The reference strain and test strains are streaked for isolation, colony purified, and verified by analytical profile index test strips (API, bioMerieux, USA) and if required additional sugar test (acid production test from D-Mannitol, Cellobiose, Lactose and D-Xylose). The panel of bacteria chosen include organisms that are likely to be found in the natural flora of food and fecal specimens or have known or suggested physiologic or genetic relatedness to *E. coli* and *Shigella* species.

The nucleotide sequence of the present invention may be a non-ribosomal sequence of *E. coli, Shigella* species or a related species.

The DNA probes are ^{32}P-labeled by random prime method (Amersham Pharmacia Biotech Inc, USA) as per supplier's instruction but with modifications. The probe specificity is refined carefully by controlling the hybridization condition and the post hybridization wash condition. It will be noticed herein, that for optimal specificity, the post hybridization salt wash condition ranges between 5 x SSC and 0.1 x SSC for the various probes. Unless specified, the wash temperature for the membranes is at 65°C.

Probe A is made up of a combination of fragments (1,2,3 and 4 of Figure 1, a-c) and covers a large part of the Formula I sequence. It will be noticed that probe A and some of the gene probes within Formula I (Table 2a, figure 2-19) are specific for the *E. coli* species and some *Shigella* species, implying that there are highly conserved and specific regions among *E. coli* species and *Shigella* species. For example, within the Escherichieae family, gene probes defined by primers pairs ECM-1163, ECM-1958 (Figure 4) and tor D-10574, tor D-11160 (Figure 7) detect *E. coli* and *Shigella* species but not non-coli Escherichia

species such as *E. vulneris* (ATCC 33821), *E. hermannii* (ATCC 33650) and *E. blattae* (ATCC 29907). Gene probes defined by primer pairs 81B-7223, 81B-7794 (Figure 12) and 81B-7278, 81B-7754 (Figure 13) detect the presence of *E. coli, Shigella flexneri* and *sonnei*, except *Shigella boydii*. Gene probe defined by putP-5944 and putP-6693 (Figure 19) detects *E. coli, Shigella boydii* and *Shigella sonnei* and other enterobacteria (showing variable homology) but does not detect *Shigella flexneri*. Probe A detect *E. coli* and *Shigella* species only (Figure 2). This invention also indicates that when a large gene probe is reduced to many smaller gene probes, the specificity of the smaller probes can be different from the parent probe and therefore must be tested out before use.

The method of the present invention encompasses the use of probes of any length however it is envisages that the *E. coli*- and/or *Shigella* species'-specific nucleotide sequences comprise at least 8 nucleotides in length.

A further aspect of the present invention provides a molecular probe comprising at least 8 nucleotides obtainable from the nucleotide sequences of Formula I wherein said molecular probe is capable of specifically hybridizing to *E. coli*- and/or *Shigella* species'-derived nucleic acid molecules.

Still even yet more especially the present invention provides a method for testing and selecting other sequences in *E. coli*, *Shigella* species or related microorganisms in a sample, said method comprising subjecting a nucleic acid molecule preparation from said sample to genetic analysis using one or more *E.coli* or *Shigella* species'-specific nucleotide sequences obtainable from one or more nucleotide sequences of Formula 1 and/or Table 1 wherein the ability for said *E. coli*- or *Shigella* species'-specific nucleotide sequences to hybridize to complementary nucleotide sequences in the nucleic acid preparation is indicative of an *E. coli* or *Shigella* species'-specific nucleotide sequence.

Some of the advantages of a smaller size probe are the ease with which they can be amplified by PCR to generate more material to work with and to carry out PCRISH.

The specificity of a smaller gene probe may be greater, equivalent or less. *E. coli* gene sequence at 7562 to 7794, AE000202 has relatively similar specificity as probe A (Figures

21, 20) to *Enterobacter cloacae* but has the added advantage of not having homology to *Shigella boydii* (Figures 15, 2). A dilution assay indicates that *Enterobacter cloacae* DNA has 50 times less hybridization signal intensity to both probe A and *E. coli* gene probe sequence (nt 7562 to 7794 of AE000202) as compared to *E. coli* DNA. Post hybridization wash condition is at 1 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate at 65°C. Other examples such as those shown in Figures 3,4 and 7 indicate higher specificity than probe A. An example of a more specific gene probe that recognizes *E. coli* and *Shigella* is that defined by primer sequence torD-10574 and torD-11160. Although the wash condition for Figure 7 is at 0.1 x SSC, 0.1% SDS, 20 mM pyrophosphate at 65°C, the probe is equally specific at 1 x SSC wash (data not included). The wash at 0.1 x SSC indicates the hybrids are relatively stable and specific.

DNA probe test, selective agar media and biochemical test

The invention herein describes the specificity of the gene probe at nt 7223 to 7794, AE000202 (Figure 12), herein called 81B, as an example, in identifying *E. coli*. This probe give 10 times less intense signal with *Enterobacter cloacae* DNA at the selected hybridization (see assay section: hybridization and post hybridization conditions) and post hybridization wash (1 x SSC at 65°C) condition (Figure 22). Despite this cross over in hybridization signal with *Enterobacter cloacae*, the primer sequence and gene probe sequence is relatively conserved. This molecular DNA test is compared to the selective isolation media and biochemical test. This comparative study is carried out on ATCC strains of *E. coli* and fecal isolates of subjects. The fecal isolates that are either positive or negative for this gene probe are isolated and tested on eosin methylene blue (EMB) lactose-sucrose agar (bioMerieux, France), MacConkey agar (Oxoid, USA) and MUG agar (Oxoid, USA), API (bioMerieux, France) and sugar test.

All 66 (100% sensitive) randomly selected colony purified fecal bacteria isolates that are positive for the probe 81B are identified to be *E. coli* with the API test strip and additional sugar test (4/66).

All 89 (100% specificity) randomly selected colony purified fecal bacteria isolates that are negative for the probe 81B are identified to be either gram negative or gram positive microorganism but not *E. coli* or *Shigella* species.

Of the patients' *E. coli*, 27 of 38 (71%) *E. coli* tested are EMB positive, 56/66 (85%) of *E. coli* are tested positive on MacConkey and MUG agar plates alone. *E. coli* that is MacConkey positive need not be MUG positive and *vice versa*. The *E. coli* identified by the probe 81B included lactose-negative *E. coli* which cannot be identified using the MPN method alone, or the MacConkey media alone or the MUG media alone. The probe detects the enteropathogenic serotype 0157:H7 and the 029:NM *E. coli* that will not be picked up by the MUG assay. The selective agar plates and biochemical test are purchased in dried form and reconstituted according to supplier's instruction.

The 81B probe will recognize *E. coli* strains even those possessing haemolytic properties, as observed on blood agar, and possessing DNA sequences that encodes known toxigenic genes, invasive genes, adherent genes and cytonecrotizing genes.

One particularly preferred aspect of the present invention discloses a method for detecting the presence of *E. coli* or *Shigella* species or related microorganisms in a sample as hereinbefore described wherein hybridization of *E. coli*- and/or *Shigella* species'-specific nucleotide sequences to the nucleic acid preparation is detected by the presence of amplified nucleic acid products.

A further related aspect of the present invention provides a method for detecting the presence of *E. coli*, *Shigella* species or related microorganisms in a sample wherein hybridization of *E. coli*- and/or *Shigella* species'-specific nucleotide sequences to the nucleic acid preparation or the presence of amplified nucleic acid products is detected by a reporter molecule giving an identifiable signal.

Tissue Infection

Several techniques employing molecular hybridization for diagnosis of bacteria infection can be used. They include southern filter or dot-blot hybridization on DNA extracted from tissues. They may be useful but they are unable to tell the investigator where precisely the infection is and if it is specific to certain cell type within a tissue. The DNA molecular probes we have identified in this invention can be used for those assays which follows the aforementioned approach as an early investigation tool. However, the use of primer

sequences and DNA gene probes can be made more informative in the technique of polymerase-chain-reaction-*in-situ*-hybridization (PCRISH). This technique can help in locating the presence of the microorganism, bearing in mind that presence of specific antigen within a test sample, may not necessarily imply that the identified pathogen is viable or necessarily causative of the clinical infectious disease at that tissue site. The causative agent at the site of mucosal disease if isolated, identified and shown to have properties that can account for the diseased state, will provide additional evidence of its role for the disease condition.

The invention herein will describe the utility of the new probes to show the association of bacteria DNA presence with histopathology of tissue samples.

Presence of Bacteria in Tissue as Detected by PCR of Tissue Total Genomic DNA

The invention herein used PCR technique in amplifying the gene sequence 81B (primers: 81B-7223, 81B-7794) to test for its presence in total DNA extracted from colon tissue (total of 1 μg per 100 μl PCR reaction mix). Five microlitres of the PCR product mixture is resolved on ethidium bromide stained agarose that has a molecular weight marker on one of the tracks. A PCR product migrating at the expected position and is hybridization positive for the ^{32}P labeled DNA probe (probe sequence correspond to nucleotide (nt) 7278 to 7773 of AE000202) is taken as positive. The inventors found the presence of the sequence as visualized under 312nm ultra-violet (uv) illumination, in all cancer patients studied, either in tumour tissues (25/29, 86%) or in adjacent normal tissues (17/23, 74%). In contrast, in 34 control patients with no cancer, but is admitted for some related gastrointestinal complaints, this sequence is found in 7/34 (21%). The P value is <0.001.

It is emphasized that confirmation of target PCR product can be carried out by hybridization of the Southern transferred membrane bound PCR products to specific radiolabeled gene probes.

This data suggests that colon tissue from patients with colon cancer harbor more bacteria than patients with no colon cancer. Such data are informative when correlated with histology, clinical diagnosis and pathogenic bacteria isolates from tissue samples devoid of contaminants.

Accordingly, yet a further related aspect of the present invention provides a method for detecting the presence of *E. coli*, *Shigella* species or related microorganisms in a sample wherein the sample comprises food, water, semi-solids or semi-liquid material, mammalian tissue, tissue extract or cells of tissue or normal tissue or tissue predisposed to cancer growth or malignancy or cellular instability.

In a particularly preferred aspect of the present the mammalian tissue is associated with colon, stomach or colorectal tissue.

A related aspect of the present invention provides a method for identifying nucleotide sequences, or their expressed products, capable of inducing or otherwise facilitating abnormal cell growth or abnormal physiology, said method comprising introducing a nucleotide sequence comprising *E. coli*- and/or *Shigella* species'-specific nucleotide sequences from the nucleotide sequences in Formula I into cells and observing morphological and/or physiological changes to said cells compared to control cells without said introduced nucleotide sequences wherein the presence of abnormal morphology and/or physiology in a cell is indicative of a nucleotide sequence from Formula I, or a polypeptide expressed therefrom, which is capable of inducing or facilitating abnormal cell growth or physiology.

In a further preferred aspect of the instant invention the abnormal cell growth or physiology is associated with cancer or a predisposition to the development of cancer or cellular instability.

The genes encoded within Formula I can be transcribed into mRNA and translated into functional proteins. These can serve as targets for detection and intervention. The detection systems that can be used include imaging, scintigraphy, immunohistological methods, enzyme-chemical/amplification detection methods, chemical methods and microchip computer assisted methods (direct/indirect). These can be carried out on patients or animals or on samples such as serum, tissues, food and liquids.

The invention can be used in the following areas:

(1) Molecular diagnosis – for example, testing patients, e.g. predisposed to cancer, cancer, having infection, for the presence of Formula I and /or its transcripts;

(2) Serological diagnosis – for example, testing patients predisposed to cancer, cancer having infection, for the presence of the protein encoded by the Formula I sequence, or that specifically binds to such a polynucleotide, or testing such patients for the presence of antibodies to such protein. A sample from the patient is preferably blood, urine or other body fluids, tissue or excretion products;

(3) Immunohistochemistry/histochemistry applications – for the diagnosis of predisposition to cancer, cancer, cellular instability, infection in tissue samples;

(4) Diagnostic imaging – in which case the antibody or probe will have an appropriate label or marker, for example, a radioactive label or marker;

(5) Therapy

(a) for example, antibodies of the present invention may form part of an immunotoxin, in order to deliver toxic agents, such as plant toxins, e.g. ricin, to the site of a malignant or even a benign tumour (see, for example, European Patent Application No. 84304801.8 – Publication No. 0145111);

(b) for example, polynucleotides of the present invention may be useful alone in therapy as anti-sense DNA or RNA. Thus, polynucleotides of the present invention optionally in a vector or in a polynucleotide analogue, which contains sequences complementary to DNA or RNA defining a protein which is differentially expressed during cancer initiation, progression and metastasis or a portion thereof can be employed to prevent expression of the said protein;

(c) for example, the metabolic pathway that the Formula I sequence or its

nearby sequence encode can be utilized to activate pro-drugs;

(6) Histological analysis – DNA or RNA or protein having an appropriate label or marker may be useful for *in-situ* detection for histological analysis; and

(7) Food and water sanitation.

Polymerase-Chain-Reaction-In-Situ-Hybridization Technology (PCRISH).

In this invention herein, we will describe the new use for DNA probes that permit the detection and localization of members of Escherichieae family (and *Shigella* species) in paraffin embedded tissue samples by polymerase-chain-reaction-*in-situ*-hybridization (PCRISH). This is useful for the study of (1) clinical conditions with bacteria infection (2) treatment and infection (3) association of bacteria in the development of cancer and (4) identifying cells at risk of or having genome that is genetically unstable.

These investigations can be carried out on routinely processed, formalin-fixed and paraffin-embedded tissues obtained from pathology files at any hospital. Such tissues have excellent preservation of tissue architecture and cellular detail and allow retrospective analysis of stored tissue blocks. To illustrate the utility of the above-mentioned probe, we will provide data on its use in the study of colon and gastric related conditions.

Sensitivity and specificity

This invention herein will provide data on how sensitivity and specificity of an assay can be dramatically improved by carrying out a hybridization step after PCR. The inventors will describe an assay that will simulate PCRISH *in-vitro*. First, PCR is carried out on randomly selected *E. coli* (number, N=3), *Shigella sonnei* (N=1), Enterobacter agglomerans (N=1), *Psedomonas aeruginosa* (N=1), *H. pylori* (N=1) from patients, TG2, K12 and the panel of ATCC bacteria total genomic DNA (Table 3). A given volume of the PCR product mixture is then resolved on an agarose gel, and the PCR products visualized by ethidium bromide staining under UV illumination. The primer targeted bacteria sequence will give a single band. Many additional bands indicates the lack of specificity of the primer sequences. The products in the gel are subsequently transferred onto nylon-plus membrane (Amersham Pharmacia Biotech Inc, USA) by the method of Southern, and

hybridized to ^{32}P random-prime labeled sequence specific probes. This hybridization step increases the detection level and also improves the specificity of the assay by hybridizing specifically to homologous sequences. Specificity is denoted by a single band when exposed to an autoradiograph. Such assays indicate whether the primers and probes are specific and can help to detect the presence of bacteria when used in PCRISH. The selection of primer and probe sequence within the Formula 1 can be empirically carried out in a step-wise systematic fashion until all the appropriate ones are selected.

Gene sequences between nt 7419 and 7985 (AE000202), herein called the OH, the 81B (nt 7223 to 7794, AE000202) and the *H. pylori* ribosomal gene (Table 3) have been tested in accordance with the present invention. By varying the PCR conditions, OH-7419 and OH-7985 primers directed PCR amplification of bacteria DNA could give rise to specific and non-specific products. Any DNA sample giving rise to PCR amplified band, whether or not in the presence of non-specific bands, and migrating in agarose gel at the expected molecular weight is denoted as positive (Table 3). Hybridization of labeled gene probe defined by the sequence between primers OH-7562 and 81B-7794 specifically picks up PCR products from *E. coli* and *Shigella* species except *Shigella boydii*. Primers 81B-7223 and 81B-7794 are more specific under the PCR conditions used and give specific bands of expected molecular weight. The labeled gene probe having the sequence between nt 7278 and 7773 specifically hybridizes to the products from *E. coli* and *Shigella* species except *Shigella boydii*. The specificity of this latter gene probe for genomic DNA of *E. coli* species is shown in Figure 23. Although there is low homology between *Enterobacter cloacae* DNA and OH gene probes (low, Figures 14,15,21) and 81B gene probes (relatively higher than OH, Figures 12,13,22), the aforementioned OH and 81B PCR primer pairs do not amplify specific products from *Enterobacter cloacae* DNA.

The dramatic increase in sensitivity and specificity is illustrated by the *H. pylori* ribosomal gene primers and gene probe. Figures 24 and 25 shows that the *H. pylori* probe has a low to moderate level of homology to enterobacteria and gram positive bacteria DNA even with stringent post hybridization wash condition of 0.1 x SSC, 0.1% SDS, 20 mM pyrophosphate at 65°C. However, upon PCR and hybridization to DNA of a selected number of enterobacteria and gram positive bacteria (Table 2a, grid D) likely to be found in the gastrointestinal tract, only *H. pylori* presence is detected (Table 3). Figure 26 and

Table 2a, grid F will show the dot-blot hybridization result where the detection limits for the presence of *H. pylori* DNA is dramatically improved after a simulated PCRISH technique. Two hundred nanogram each of total genomic HP DNA and different purified fecal isolates of *E. coli* genomic DNA from different patients, are each PCR amplified in 100 µl of reaction mixture. After PCR, 10 µl is taken out from the *H. pylori* reaction tube and diluted with 90 µl of water before a known volume is being dot-blotted onto Amersham nylon plus membranes and hybridized to random primed ^{32}P labeled *H. pylori* (HP) ribosomal probe.

Figure 26, column A, rows 1 to 4 each contained 1, 5, 10 and 50 µl of the10 fold diluted *H. pylori* post PCR reaction mixture. The hybridization signals intensified with increase in amount of PCR products. Rows 5 to 8 contained 1, 5, 10 and 50 µl of 10 fold diluted PCR product comprising of equal portions of 4 *E. coli* DNA PCR products (K12, 142/31, 179/36, 197/5 and 117/3B). No amplified PCR products are obtained with *H. pylori* primer directed amplification of *E. coli* DNA. This column shows that the *H. pylori* primers are specific for the *H. pylori* genome.

Figure 26, column B, rows 1 to 5 contained mixture of 5 µl diluted post *H. pylori* PCR reaction mixture and 5 µl (undiluted) of post PCR reaction mixture of different *E. coli* DNA isolates respectively. The result indicates that the hybridization of HP probe to the *H. pylori* PCR products is not affected by the presence of *E. coli* DNA. This is indicated by the equal intensity of the dot.

Figure 26, column C, rows 1 to 5, each contained 5 µl (undiluted) post PCR reaction mixture of different *E. coli* isolates of different patients. The HP primers did not PCR amplify *E. coli* DNA.

Figure 26, rows 6 to 8 of column B, each contained 10 ng, 50 ng and 200 ng respectively of *H. pylori* genomic DNA and in column C, each contained instead a respective equivalent of *E. coli* genomic DNA. The HP probe showed higher homology to its own gene.

Hybridization of a panel of bacterial DNA to the HP gene probe has shown that it has some

homology to the other bacteria species (Figures 24, 25). However by combining sequence specific primer directed PCR and hybridization with a probe that has low homology to other species, it is still possible to increase the sensitivity and specificity of HP detection (Figure 26, Table 2a- grid F).

Exemplification of various aspects of the invention using PCRISH in no way limits the methods of the present invention and any appropriate technique may be used to monitor or detect the presence of *E. coli* DNA in cells. This invention herein relates the use of PCRISH as a method for illustrating how pathogens can be detected *in-situ*. It does not exclude the use of *in-situ* hybridization nor *in-situ*-PCR technology, using sequences within Formula I, as an alternative. The amount of target, reporter system, PCR cycles and hybridization conditions are important determinants in deciding which method is appropriate.

Bacteria Detection In Tissue Specimens Using PCRISH Technology

A preferred method in amplifying vast amount of DNA sequences in cells/tissue and detecting the presence of the sequences *in-situ* use the PCR-*in-situ* hybridization (PCRISH) technology. PCRISH is practiced routinely by those having ordinary skill in the art and its uses in detecting infection *in-situ* is widely used and accepted.

As a start, cells or tissues are fixed with a suitable fixative (example 10% buffered formalin) for 2-72 hrs and the tissues subsequently paraffin-embedded. The paraffin-embedded tissues are then sectioned at 3–4 microns and mounted on aminoalkysilane treated glass slides. Before amplification is undertaken, the sections are deparaffinized first and then carefully permeabilized using proteolytic enzymes such as proteinase K, so that the cellular morphology is maintained and the DNA is accessible. The cell/tissue is now ready to permit access of PCR reagents into the cellular compartment where the nucleic acids are found. Amplification then occurs at the defined target intracellular sequence. Detection of amplified product is then carried out by hybridizing with a non-radioactive labeled DNA probe. We used DNA probes labeled with Digoxigenin-labeled nucleotides. This DIG-labeled probe is detected with anti-digoxigenin (anti-DIG) antibodies that are conjugated to alkaline phosphatase. This alkaline phosphatase is visualized with colorimetric (NBT and BCIP) alkaline phosphatase substrates (Boehringer Mannheim). It

is imperative to always to perform a negative control with each specimen and to understand that a negative result only indicates an inability to detect the substance analyzed. Negative tissue control does not contain the specific antigen in question. Positive control must accompany each run of test tissue sections within which the specific DNA sequence is sought.

Interpretation of Hybridized Tissue Sections.

It is known that some members of the Escherichieae family (*E. coli* and *Shigella* species) are able to adhere to, invade cells and move across cells. Therefore, there is the need to correlate histopathology with quantity and depth of bacteria presence in order to evaluate their role in gastrointestinal infection and cancer. This is done by studying the quantity and spatial distribution within the tissues of the hybridization signals generated with the DIG-labeled bacteria DNA probes.

We have categorically defined the spatial distribution of the bacterial presence into positive and negative depending on whether the signals are found (i) outside mucosa cells/lumina propria (negative), (ii) in mucosa cells (positive) and (iii) in nuclei of mucosa cells (positive). All tissues are obtained upon discovery (pre-treatment) unless stated otherwise.

Colon

The use of primer and probe DNA gene sequence of 81B for carrying out PCRISH in colonic mucosa tissue in this invention is characteristically found to identify all hyperplastic tissues (number [N] = 7; Figure 27), adenomas (N = 12; Figure 28), adenocarcinomas (N = 11; Figure 29) and metastatic colonic cells in liver specimens (N = 5; Figure 30) to be positive for the 81B marker probe. The signals are characteristically predominant in cells and in nuclei of the cells. The hyperplastic tissues, adenomas and adenocarcinomas can also have 81B signals located outside cells. Metastatic cells, with primary in the colon, found in liver specimens show distinct 81B signals in the nucleus and without significant exogenous 81B signals outside cells. Mucosa tissues of normal histology obtained adjacent to tumours (N = 8; Figure 31) are positive in certain areas and have intense signals in those areas. The majority of tissues without cancer (N = 16 ; Figure 32) and obtained from normal patients indicated for colonoscopy are negative (69%). One example of an infected tissue is that of a patient's colon tissue diagnosed to have proctitis

on histology (Figure 33). Here, the distribution of 81B signal is diffusely distributed but without the intense cellular/nuclear signal that is observed in normal tissue adjacent to tumours (Figure 31) and in tumours (benign, pre-malignant and malignant).

H. pylori probe is insignificant in metastatic cells in liver sections (N = 5; Figure 34) and serves as a negative control.

The transition from normal through pre-malignant to localized malignant and finally a metastatic stage has been histologically defined for colorectal cancer. Therefore the appearance of 81B signals at early stages and before histological changes, through to all histological changes associated with cancer and leading to the metastatic stage indicates a possible aetiological role for *E. coli/Shigella.* This invention herein identifies several uses for the *E. coli/Shigella* probe whereby markers specific for *E. coli/Shigella* can be used for screening patients for risk of colorectal cancer, colorectal cancer and infection with *E. coli/Shigella.*

Gastric

Helicobacter pylori is established as the major aetiological agent in gastritis and peptic ulcer and it is also known to infect half of the world's population. This high infection/disease ratio is however explained by host factors, socio-economic conditions of the various countries and infection by a sub-population of virulent *Helicobacter pylori* (Chu, K.M. and Branicki, F.J. *JAMA SEA* **13**:5-7, 1997). Therefore, it is clear that to understand the pathogenesis of *Helicobacter pylori* and general infection, one approach is to precisely locate the bacteria to the site of mucosal disease and to correlate it to histology and clinical presentation. Past data using immunocytochemical detection system (Cartun *et al. J Clin Pathol.* **43**:518. 1990) and *in-situ* hybridization technique have both localized the bacteria to the luminal, foveolar epithelium or periepithelial mucous layers of the gastric mucosa of helicobacter pylori associated gastritis (Van Den Berg *et al. J Clin Pathol.* **42**:995-1000, 1989; Bashir *et al. J Clin Pathol* **47**:862-866, 1994). Unlike these observations, we are able to show that *Helicobacter pylori* can be found beyond the luminal or foveolar epithelium in patients with gastritis and may assist in understanding the significance of high infection/disease ratio. This invention will highlight the resolution power of a successful PCRISH technique for studying the association of endoscopic

diagnosis of gastric conditions with histological findings, bacteria pathogenesis and clinical presentations.

The invention herein has found *E. coli*, other than *H. pylori*, in stomach related conditions such as gastritis of varying severity, ulcer and cancer.

The use of primer and probe DNA gene sequence of *H. pylori* and 81B for carrying out PCRISH in stomach tissue in this invention is found to identify all *H. pylori* positive active chronic gastritis (N = 9), *H. pylori* positive gastritis (acute on chronic) (N = 12), active chronic gastritis with no histologically detectable *H. pylori* (N = 10), chronic gastritis with no histologically detectable *H. pylori* (N = 9), mild gastritis with no histologically detectable *H. pylori* (N = 4), normal with no histologically detectable *H. pylori* (N = 4), normal tissue distant to tumour (N = 1), adenocarcinoma (N = 5) and liver specimen with poorly differentiated metastatic stomach adenocarcinoma (N = 1), positive for the 81B sequence. The distribution of the 81B sequence is such that they can be in the luminal areas with variable degree of penetration into intercellular, cellular and nuclear areas of specimens. The area of infection is patchy. For the cancer specimens, the signals are found mainly in the nuclei of tumours (Figure 35) and in normal tissue adjacent to tumours (Figure 36) they are found both in the nuclear and luminal area. For the metastatic cells in liver section, that originated from the cancer at the gastro-oesophageal junction, the signals are also mainly in the nuclei but without the exogenous signals that can sometimes be observed with the primary tumour (Figure 37). Gastric mucosa with normal histology and no histologically detectable H. pylori but positive for the 81B probe is interesting as the patient presented with non-ulcer dyspepsia, epigastric discomfort and vomiting. The 81B signals show patchy aggregates. Figures 38 and 39 are of the same specimen but showing different area.

The samples that are histologically diagnosed to be positive for *H. pylori* are confirmed by PCRISH. Although the *H. pylori* probe is found present in normal tissue distant to the adenocarcinoma of stomach (Figure 40), it showed insignificant nuclear signals in the tumour itself (Figure 41). *H. pylori* probe is not detected in metastatic tumour cells in liver (Figure 42).

This invention herein presents a use for the *E. coli/Shigella* probes in gastric related studies whereby it can be used to screen for patients with risk of gastric cancer, gastric cancer and gastric infection. When found in the presence of *H. pylori*, it supports the possibility that synergistic interaction between two types of bacteria can be a cause for gastric related conditions (active chronic gastritis, Figures 43, 44). How important the presence of *H. pylori* is to *E. coli* is yet to be determined as it is known that *E. coli* has a mechanism for acid resistance. (Lin J, *et al. Applied and Environmental Microbiology* **62**:3094-3100, 1996). In addition, *E. coli* are known to have invasive, cytonecrotizing, adherent, and toxigenic genes that can account for gastric related conditions.

Cellular Instability

Studies have suggested that apparently uninvolved mucosa adjacent to, and even remotely from colorectal cancer is abnormal morphologically and histochemically (Ngoi *et al. Cancer* **66**:953-958, 1996). In addition, the invention herein showed *E. coli/Shigella* DNA marker in the nucleus of many normal cells next to tumours, tumour cells and metastatic tumour. This suggests that *E. coli/Shigella* probe can be a marker associated with risk of cellular instability or cellular instability and tumourigenesis as it is known that bacteria DNA has tumourigenic potential.

Isolation of microorganism that is associated with tissue infection

The invention herein indicates a method whereby in addition to localization of bacteria in tissue sample, the marker sequence can further assist in identifying and isolating members/strains of the family of bacteria it represents, in tissue/fecal specimens.

To start off at the tissue level, using PCRISH, the spatial distribution of the marker (for example, 81B or OH or any other that is within the Formula 1 sequence) would indicate whether there is a pathogen present. This observation may be supported by histological diagnosis for infection, benign, pre-malignant or malignant cancer. This is then followed by detecting for positive colonies isolated from the positive specimens (for example tissue, fecal), and hybridizing to the positive colonies, separately, a panel of DNA that is known to encode either invasive, adherent, necrotizing, toxigenic or other pathogenic properties. The bacteria that is found to possess pathogenic properties as defined by a DNA sequence, can have their presence within the tissue checked by PCRISH using primers and

hybridizing probes for that property. From hereon, association studies can be carried out empirically with the various pathogenic probes to understand bacteria presence and infection, cancer risk and etiology of cancers.

With this approach, the inventors have found pathogenic *E. coli* to be the main pathogen associated with cancer risk and cancer. Herein the inventors define pathogenic *E. coli* as those strains of *E. coli* able to invade into cells and those that can cause cellular changes in cells. Since pathogenic *E. coli* is found in the instant invention as a marker of cellular instability and tissue at risk of cancer (benign tumours, pre-malignant tumours) and cancers (malignant tumour/cells), it therefore implies that any sequence of *E. coli* (pathogenic or otherwise) can be used for the above purpose.

A related aspect of the instant invention provides a molecular probe of at least 8 nucleotides, identified by the methods as herein described wherein said probe comprises a sequence of nucleotides from Formula I and wherein said molecular probe is capable of specifically hybridizing to *E. coli* and/or *Shigella* species'-derived nucleic acids.

Assays

Microbiological techniques are practiced routinely by those having ordinary skill in the art and its use is wide and accepted. Information for practicing is disclosed herein by reference: "Nga, B.H. and Lee, Y.K.: Microbiology Applications in Food Biotechnology. Elsevier Applied Science, 1990. Koneman *et al.* Introduction to Diagnostic Microbiology. JB Lippincott Company 1994. AOAC Official Methods of Analysis 1995. Pepper IL, Gerba CP and Brendecke JW: Environmental Microbiology: A laboratory Manual. Academic Press 1995. PR Hunter: Waterborne Disease: Epidemiology and Ecology. John Wiley and Sons 1997.

DNA and related technology is widely used by those having ordinary skill in the art and information for practicing DNA and related technology is disclosed herein by reference: Moseley *et al. The Journal of Infectious Diseases* **142**:892-898, 1980. Berger SL and Kimmel AR: Guide to Molecular Cloning Techniques. Methods in Enzymology. Volume 152. Academic Press Inc 1987. Tenover FC: DNA Probes for Infectious Diseases. CRC Press 1989. Sambrook J, Fritsch EF and Maniatis T: Molecular Cloning: A laboratory

manual. Second Edition. Cold Spring Harbor Laboratory Press 1989 (3 volumes); Echeverria *et al. J Clinical Microbiol* **27**: 31-334, 1989; Virginia LC and Bavoil PM: Bacteria Pathogenesis. Academic Press 1997.

Aspects of the present invention include various methods such as PCR and dot-blot, for example, for determining whether a sample contains bacteria sequence of the Escherichieae family by molecular analysis. Nucleic acid-based assays can be made exquisitely sensitive, and their specificity can be adjusted from broad to a very narrow range by careful design of the probe and precise control of hybridization conditions.

The invention relates to Formula I, DNA gene probes and primers used in the method of identifying DNA that is encoded in the DNA of members of the Escherichieae family such as the *E. coli* and *Shigella* species.

The DNA sequence based methods for determining whether a sample DNA encoding the sequence found in *E. coli* and *Shigella* species include but are not limited to polymerase chain reaction technology, Northern and Southern blot technology, dot-blot, colony-blot, PCR-*in-situ* hybridization, *in-situ*-hybridization technology and oligonucleotide hybridization technology.

The invention primarily involves methods in DNA detection and these methods are commonly used alone or in combination so as to enhance the sensitivity and specificity of the DNA sequence for detecting the presence of species of *E. coli* and *Shigella*.

It is contemplated that other sequence-based methodology for detecting the presence of bacteria DNA in samples may be employed according to the invention.

PCR Technology

A preferred method in amplifying vast amount of DNA sequences uses the polymerase chain reaction (PCR) technology. PCR technology is practiced routinely by those having ordinary skill in the art and its uses is wide and accepted. Methods for practicing PCR technology are disclosed in "McPherson MJ, Quirke P and Taylor GR: PCR. A Practical Approach. Volume 1. Oxford University Press 1994", which is incorporated herein by

reference.

The nucleotide sequence present in species of *E .coli* and *Shigella* is well known such as in GenBank database accession number AE000201, AE000202 and AE000203. To perform this method, DNA is extracted/released from cells in a sample and tested using well-known methods and readily available starting material.

The DNA is combined with the primers, free nucleotides and enzyme following published PCR protocols. The mixture undergoes a series of temperature changes. If the DNA encoding the marker sequence for *E. coli* and *Shigella* species are present, that is, if both primers hybridize to sequences on the same molecule, the molecule comprising the primers and the intervening complementary sequences will be exponentially amplified. The amplified DNA can be easily detected by a variety of well known means. If the sequence is not present, no DNA molecule will be exponentially amplified. The PCR technology therefore provides an extremely easy, straightforward and reliable method of detecting *E. coli* and *Shigella* species DNA in a sample. Publication is herein provided by reference for indicating how PCR technology is utilized for detecting pathogens in food, Gannon *et al. Applied and Environmental Microbiology* **58**:3809-3815, 1992; AOAC Official Methods of Analysis 1995.

PCR Primers

PCR primers can be designed routinely by those having ordinary skill in the art using well known DNA sequence information that can be retrieved from GenBank database. Software programs are also available to assist in designing optimal primer sequences as long as sequences of gene of interest are known. (PCR PLAN (PC/Gene system, Intelligenetics, inc. USA). Primers are generally 8-50 nucleotides, preferably 18-28 nucleotides. A pair of primers is routinely used for PCR. When performing PCR on extracted DNA samples containing targeted specific bacteria DNA, multiple copies of the DNA will be made. If the targeted DNA is not present, PCR will not generate a discrete detectable product.

PCR Product Detection

PCR amplified DNA may be detected by several well-known methods. The preferred method for detecting the presence of amplified DNA is to resolve the PCR reaction material by gel electrophoresis and staining the gel with ethidium bromide in order to see

the amplified DNA if present. A standard molecular weight marker containing a standard of equivalent size to the expected size of the amplified DNA is preferably run on the gel as a control.

In some instances, such as when small amounts of amplified DNA is generated at the first attempt, therefrom, it is desirable or necessary to perform a PCR reaction on the first PCR reaction product. A nested set of primers is used in the second PCR reaction. The nested set of primers can either hybridize to sequences downstream of the 5'primer and upstream of the 3'primer used in the first reaction or hybridize further into the first PCR product and allow a second PCR reaction.

The present invention includes oligonucleotide sequences, which are useful as primers for performing PCR methods to amplify DNA fragments that are useful for identifying the Escherichieae family in particular *E. coli* and *Shigella* species. For the detection of the sequence, labeling of DNA can either be radioactive or non-radioactive nucleotides.

Random prime labeling of DNA fragment (Megaprime DNA, Amersham Pharmacia Biotech Inc, USA)

Labeling can be carried out according to manufacturer's recommendation. Between 100 - 500 ng of DNA fragment is heated for 5 min with optimized amount of hexamer primer-solution (7 µl) before cooling rapidly to room temperature. This is then mixed with reaction mixture containing 12 µl of labeling buffer, 5 µl enzyme (1U/µl), 5µl of ^{32}P (6,000 Ci/mmol) and water to 50 µl final reaction volume. Labeling is carried out at 37°C for 3 hrs. Separation of unincorporated ^{32}P from labeled sequence is carried out by spin column method (Biospin columns: Bio-Rad Laboratories, USA).

The amount of sequence to be labeled can vary over a wide range as other parameters can influence the outcome of the invention. Examples to consider: the efficiency of the labeling system employed as other systems of labeling are available (polymerase chain reaction [PCR] labeling and use of radioactive or non-radioactive nucleotides), the nature of the sequence, either as probe A or in combination with other sequences or sub-fragments of the Formula I sequence.

PCR on total genomic DNA.

One hundred microlitres of PCR reaction mixture containing 10 mM KCl, 10 mM$(NH_4)_2SO_4$, 20 mM Tris HCl (pH 8.0), 2 mM $MgSO_4$, 0.1% Triton X-100, 200 μM (each) of dNTPs, 0.4 μM (each) of two primers, 1.5 unit of Taq polymerase and 1 μg of total genomic DNA (e.g. extracted from colon tissue) is amplified in a thermal cycler (Perkin Elmer Cetus, model TC1). Amplification is carried out in 3 stages: (i) 94°C for 3 min; (ii) 10 cycles of 30 s at 94°C, 30 s at 62°C, 1 min 30 s at 72°C, and then (iii) 35 cycles of 1 min at 92°C, 40 s at 58°C, 1 min 30 s at 72°C with an auto extension of 5 s at 72°C at the end of each cycle. Lastly, an extension time of 7 min at 72 °C.

Hybridization and post hybridization conditions

The complete/partial sequence of formula I either alone or as a mixture are radioactively labeled by random-prime method using ^{32}P-dCTP (alpha-^{32}P-dCTP, 6,000 Ci/mmol, Amersham Pharmacia Biotech Inc, USA). Nylon membranes (Hybond-N+, Amersham Pharmacia Biotech) are pre-hybridized overnight at 65°C in a buffer containing 5 x SSC (1 x SSC is 0.15 M sodium chloride, 0.015 M sodium citrate) 0.1 % SDS, 5 x Denhardt's solution (1 x Denhardt's solution is 0.02% bovine serum albumin, 0.02% polyvinyl pyrolidine 360K MW and 0.02% ficoll 400K MW in water), 200 μg/ml sonicated denatured salmon sperm DNA (sssDNA). Hybridization is carried out at 40°C for overnight in a buffer containing 5 x SSC, 20 mM sodium phosphate, pH 7.0, 1 x Denhardt's solution, 0.1% SDS, 200 μg/ml sssDNA, 25% deionized formamide and labelled probe. The probe used can be from 100 to 500 ng of labelled sequence. The membranes are washed at 65°C in 5 x SSC, 0.1% w/v SDS, 20 mM Na pyrophosphate for several washes and followed by another few washes at 65°C with 1 x SSC, 0.1% w/v SDS and 20 mM Na pyrophosphate. All washes ranged from 30 to 60 min or until the background signal is acceptably low. Autoradiographs are exposed overnight at -70°C.

Several parameters can be empirically adjusted to improve on the specificity of the hybridization assay. An increased in stringency of the assay can be achieved by increasing the percentage formamide used during hybridization and/or by decreasing the salt concentration of the post hybridization wash buffer. Use of alpha-^{32}P-dCTP, 3,000 Ci/mmol is possible as this will only affect the level of sensitivity of detection.

Polymerase-chain-reaction-in-situ-hybridization technology

Incorporated herein are references that are useful for sample preparation and understanding the principals and application of PCRISH: "Rentrop *et al. Histochemical J* **18**:271-276, 1986; Carson, F.L. Histotechnology. A Self-Instructional Text. American Society of Clinical Pathologists Press 1990; Mikel UV: Advanced Laboratory Methods in Histology and Pathology. Armed Forces Institute of Pathology 1994; Nuovo, G.J.: PCR *In Situ* Hybridization. Protocols and Applications. Raven Press, N.Y. 1994; Gu, J: *In Situ* Polymerase Chain Reaction and Related Technology. Eaton Publishing Co. USA 1995; Boehringer Mannheim Non-radioactive *In Situ* Hybridization Application Manual (2nd edition) Washington, DC., 1996; O'Leary *et al. J Pathol.* **178**:11-20, 1996.

Fixation and tissue preparation

Pathology departments routinely fix tissue that are targeted for embedding in paraffin wax and which are subsequently used for diagnosis of tissue pathology. For fixation, we used for our biopsy specimens, 10% buffered formalin for 48-72 hrs (100% formalin diluted in phosphate buffered saline [PBS] to a 10% v/v concentration). After paraffin embedding, specimens of 3-4 micron thickness are cut onto 3-aminopropyltriethoxysilane (silane) coated microscope slides. Requirement of tissue thickness is dependent on the overall size of the tissue to be tested. The inventors recommend 3 micron for biopsy specimen and 4 micron for larger specimens. Before the samples can be used for PCRISH, the sections are fixed on slides at 58°C for 1 hr 20 min, followed by deparaffinization in xylene (3 x 5 min) and final rinse in 100% (3 x 1 min) ethanol. At this stage, the sections can be left for use the following day. Just prior to use, the sections are rehydrated 2 min each through graded (70% and 50%) ethanol and water with final equilibration in PBS for 5 min.

PCR Labeling

PCR labeling of probes for PCRISH is carried out using digoxigenin dUTP (Boehringer Mannheim). This is carried out in a 100 µl reaction mixture comprising of 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris HCl (pH 8.0). 2 mM MgSO4, 0.1% Trition X-100, 50 µM DIG-11-dUTP, 150 µM dTTP, 200 µM each of dATP, dCTP, dGTP, 0.4 µM primers each, 2 units of Taq DNA polymerase (Perkin Elmer, UK) and 2 µl of first PCR product. The first PCR product is made with the outer primers for the same gene using K12 bacteria genomic DNA.

Amplification is carried out in 4 stages in a thermal cycle (GeneAmp PCR System 2400 or 9600, Perkin Elmer, UK):

(i) 94°C for 2 min, (ii) 5 cycles of 30 s at 94°C, 30s at 58°C, 90s at 72°C, (iii) 10 cycles of 30 s at 94°C, 30s at 55°C, 90 s at 72°C and (iv) 30 cycles of 60 s at 92°C, 30 s at 52°C, 90 s at 72°C, with an autoextension of 5 sec at 72°C at the end of each cycle. Final extension time is 7 min at 72°C.

Proteinase K digestion

For optimal cell permeabilization, each tissue type is empirically titrated against time at 40°C for any given concentration of proteinase K. The inventors routinely use a fixed concentration of 10 µg/ml of proteinase K in PBS and adjust the time of digestion. The enzyme is diluted from a 10 mg/ml stock into the 40°C pre-warmed PBS solution for 15 min before starting the permeabilization step. The size, type of tissue (normal or fibrous and necrotic) and how long it has been archived affect the duration of proteinase K digestion. After digestion, the sections are rinsed several times in PBS for a total of 5 min to stop the proteinase K digestion. This is followed by incubating in 10% buffered formalin for 1 min at room temperature and a further rinse in several changes of PBS to remove excess formalin. The sections are then dehydrated by incubating them, for 3 mins each, in a graded ethanol series of 70%, 85% and 100%. The slides are air dried at room temperature and free of ethanol before PCR.

In-Situ PCR

Twenty five microlitres of PCR reaction mixture containing 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris HCl (pH 8.0), 2 mM $MgSO_4$, 0.1% Triton X-100, 200 µM (each) of dNTPs, 0.6 µM (each) of two primers, 1 unit of Taq polymerase is overlayed onto each tissue section. Glass coverslip (24 x 40 mm) are then carefully put on top, taking care not to create air bubbles. The cover slips are sealed on all sides with nail vanish, then air dried before placing the slides into the thermal cycler (GeneAmp *In situ* PCR System 1000, Perkin Elmer Cetus). Amplification is carried out in 3 stages: (i) 94°C for 2 min; (ii) 10 cycles of 30 sec at 94°C, 30 sec at 58°C, 60 sec at 72°C, and then (iii) 15 cycles of 30 sec at 92°C, 30 sec at 55°C, 60 sec at 72°C with an auto extension of 5 sec at 72°C at the end

of each cycle. Lastly, an extension time of 7 mins at 72°C.

Post PCR wash

After PCR, a check for air bubbles inside the cover-slip is always made, and its location and size is recorded before incubating them 5 min each at 92% and 100% ethanol respectively. The validity/interpretation of the final data will take into account the presence or absence, size and location of the bubble. Bubbles create artifacts. The layer of varnish is carefully peeled off the coverslip and the sections rinsed several times with 2 x SSC before soaking twice in 0.5 x SSC at room temperature for 5 min each. It is important that the varnish does not touch the specimen. The section is then treated with 10 µg/ml proteinase K solution for approximately 15 s before rinsing several times with PBS and dehydrated through a graded ethanol series (1 minutes each in 70%, 85% and 100% ethanol). The slides are air dried at room temperature before hybridization.

Hybridization

Twenty five microlitre of hybridization buffer (comprising of 25% formamide, 1 x dendhardt solution, 5% (w/v) dextran sulphate, 200 µg/ml SSS DNA, 4 x SSC and 5% (v/v) DNA probe) is overlaid onto the tissue section. Coverslip is placed over each tissue section, taking care not to create air bubbles. Coverslip must be larger that the specimen, with a minimum distance of 3 mm away from all sides of the tissue. The coverslip is sealed on all sides with rubber cement and the cement is thoroughly air dried before the denaturation step. The sections are placed over a 95°C heating block for 6 min to denature the DNA, rapidly cooled on ice for 1 min and finally incubated overnight at 40°C in a humid chamber.

Post hybridization wash

After the overnight incubation, the rubber cement is carefully peeled off and the section rinsed several times in 0.5 x SSC, twice for 5 min in 0.5 x SSC at room temperature, and in 0.1 x SSC at 42°C for 20 min (2 x 5 min, followed by 1 x 10 min).

Immunological Detection Kit (Boehringer Mannheim)

This protocol follows closely the one provided by the manufacturer, but with minor

modification. Each section is dipped briefly into buffer 1 (100 mM maleic acid, 150 mM NaCl, adjusted to pH 7.5 [20°C] with NaOH , autoclaved) for equilibration and the excess liquid drained off before addition of 100 µl of buffer 2 (1% blocking reagent) is added to cover each tissue section. This is incubated in a humid chamber at room temperature for 20 min. This is followed by incubation with 100 µl of anti-DIG antibodies conjugated to alkaline phosphatase (1:500 v/v in 1% blocking reagent. Blocking reagent: a 100 mg/ml stock is prepared by dissolving the supplied powder form in buffer 1. This is then autoclaved and aliquoted into eppendorf tubes and frozen at -20°C. This stock is diluted 1:10 in buffer 1 to make the 1% blocking reagent.) A coverslip is put over the section to prevent evaporation and to ensure that the tissue section is completely covered with the antibody conjugate. This is placed in a humid chamber for 1 hr at room temperature. Sections are then incubated in buffer 1 for 20 min (2 x 5 min and 1 x 10 min) to wash off the unbound antibodies and equilibrated for another 5 min in 37°C buffer 3 (100 mM Tris HCl, 100 mM NaCl, 50 mM $MgCl_2$, adjusted to pH 9.5 [20°C] with NaOH). The colour-substrate stock solution (nitroblue tetrazolium salt/ 5-bromo-4-chloro-3-indolyl phosphate [NBT/BCIP]) is diluted 1:50 in 37°C buffer 3 just prior to use. Detection buffer 3 is drained off sections and 120 µl of the freshly prepared colour-substrate solution is distributed over the tissue to detect the hybridized probe. The sections are covered with coverslip each and left in a 37°C dark, humid chamber for at least 1 hour. Color development is carefully monitored until the reaction on the positive control slide is appropriate. Color development can range between 1 to 4 hours. After rinsing sections several times in 37°C distilled water, they are stained with nuclear fast red for 1 to 5 s depending on the tissue size. Excess stain is rinsed off by several dips into distilled water and sections are mounted in a water-based mounting medium (DPX Mountant, BDH).

The system has been tested out on cytospin cells, colon, stomach sections that undergo PCRISH to (1) a reaction mixture without a probe and (2) a reaction mixture with a probe that has no complementary sequences in the tissue. There is no endogenous activity to give artifacts.

Table 1

Primer Code	Map	Map Position	Sequence 5'- 3'	Gene
ECM-246	AE000201	246-266	atgactggtttagtaaaatgg	cspG
ECM-850	AE000201	850-830	tcaatattcactgttaacctc	sfa
ECM-1163	AE000201	1143-1163	cattgcgtaaccaatcaccgc	yccM
ECM-1958	AE000201	1958-1938	gcaagtagcacgacatttgtc	yccM
torT-5129	AE000201	5129-5148	ggtgcaagcctctacgccgc	torT
torT-5750	AE000201	5750-5731	tgccgcctctgccgcaatgg	torT
torC-7218	AE000201	7218-7242	aacttgccgagcgtgaatgggcgcg	torC
torC-7761	AE000201	7761-7737	gtggcctgcaacttgctccactcgg	torC
torA-8332	AE000201	8332-8356	tatccgatggtacgcgtggactggc	torA
torA-8891	AE000201	8891-8867	gcaatgtgcttcacatgctcgcgcc	torA
torD-10574	AE000201	10574-10593	gaccacgctgacagcacaac	torD
torD-11160	AE000201	11160-11141	ggtggtcgcactccactaac	torD
CD-415	AE000202	415-436	gctttcccccaatctttacgtg	cbpA
CD-1351	AE000202	1351-1329	gatttacgcgagataacgctatg	cbpA
agp-3151	AE000202	3151-3172	cgctaatcgccgcagctgtggc	agp
agp-4359	AE000202	4359-4336	cgctatcaaacttatccatcgggc	agp
Wrb-4807	AE000202	4807-4836	tgtgaaacgtcaaataattcctgcgctgcg	wrbA
Wrb-5235	AE000202	5235-5206	catgtacggacatattgaaacgatggcacg	wrbA
ycdG-6073	AE000202	6073-6094	ctcctgatgaacaacttctggc	ycdG
81B-7223	AE000202	7223-7240	ggatccagccccatcaga	ycdG

Primer Code	Map	Map Position	Sequence 5'- 3'	Gene
81B-7278	AE000202	7278-7297	cgtgttgaacgcccattact	ycdG
ycdG-7359	AE000202	7359-7340	tcgacctctacagagagcgg	ycdG
OH-7419	AE000202	7419-7436	acaagcagggcgcatcag	b1007
OH-7562	AE000202	7562-7590	acgaaaccagagcctcttccagttgcggg	b1007
81B-7754	AE000202	7773-7754	gcccacattactggtgtgcc	b1007
81B-7794	AE000202	7794-7774	ctgcagtgtgaccgatacgcc	b1007
OH-7985	AE000202	7985-7966	atagcagcaagctttatgcg	b1008
New1-8160	AE000202	8160-8179	cggcaagttgtgggctggag	b1008
New1-9704	AE000202	9704-9684	cgtaattattcccgctggcag	b1010
New2-9731	AE000202	9731-9749	gcgatatgagcaaaggacg	b1011
B-11375	AE000202	11375-11356	ctgtcgatgatcaaactgcg	b1012
New2-503	AE000203	503-484	gcatctccatacagaacagg	ycdc
putP-5944	AE000203	5944 - 5963	ctgggttacttcgggcagcc	putP
putP-6693	AE000203	6693-6674	cggagccgaatgatagtgcg	putP
HP-178	E25742	178-199	gctaagagatcagcctatgtcc	HP16S rRNA
HP-228	E25742	228-252	accaaggctatgacgggtatccggc	HP16S rRNA
HP-513	E25742	513-489	gcgctctttacgcccagtgattccg	HP16S rRNA
HP-775	E25742	775-751	gccctccaacaactagcatccatcg	HP16S rRNA

Table 2a. Grid A lists the different types of bacteria genomic DNA loaded onto nylon plus membrane and hybridized to random primed ^{32}P-radiolabeled gene probes.

	W	X	Y	Z
1	Placenta	Escherichia vulneris	Salmonella choleraesuis	
2	Helicobacter pylori	Hafnia alvei	Shigella flexneri	TG2
3	Aeromonas jandanei	Klebsiella pneumoniae	Serratia marcescens	
4	Citrobacter Freundii	Morganella morganii	Shigella sonnei	K12
5	Cedecea Lapagei	Proteus vulgaris	Yersinia enterocolitica	
6	Enterobacter cloacae	Providencia alcalifaciens	Escherichia blattae	078:H11
7	Escherichia hermannii	Pseudomonas aeruginosa	Enterobacter agglomerans	
8	Edwardsiella tarda	Shigella boydii	219/1 Shigella sonnei	sssDNA

Table 2a. Grid B lists the different types of bacteria genomic DNA loaded onto nylon plus membrane and hybridized to random primed ^{32}P-radiolabeled gene probes.

	W	X	Y	Z
1	Placental	Escherichia vulneris	Salmonella choleraesuis	078:H11
2	Helicobacter pylori	Hafnia alvei	Shigella flexneri	142 / 31
3	Aeromonas Jandanei	Klebsiella pneumoniae	Serratia marcescens	K12
4	Citrobacter freundii	Morganella morganii	Shigella sonnei	TG2
5	Cedecea lapagei	Proteus vulgaris	Yersinia enterocolitica	179 / 36
6	Enterobacter cloacae	Providencia alcalifaciens	Escherichia blattae	197 / 5
7	Escherichia hermannii	Pseudomonas aeruginosa	Enterobacter agglomerans	117 /3B
8	Edwardsiella tarda	Shigella boydii	219/1 Shigella sonnei	sssDNA

Table 2a. Grid C lists the different types of bacteria genomic DNA loaded onto nylon plus membrane and hybridized to random primed ^{32}P-radiolabeled gene probes.

	W	X	Y	Z
1	Placenta	Escherichia vulneris	Salmonella choleraesuis	078:H11
2	Helicobacter pylori	Hafnia alvei	Shigella flexneri	196/1
3	Aeromonas jandanei	Klebsiella pneumoniae	Serratia marcescens	K12
4	Citrobacter Freundii	Morganella morganii	Shigella sonnei	TG2
5	Cedecea Lapagei	Proteus vulgaris	Yersinia enterocolitica	196/28
6	Enterobacter cloacae	Providencia alcalifaciens	Escherichia blattae	197/5
7	Escherichia hermannii	Pseudomonas aeruginosa	Enterobacter agglomerans	218/40
8	Edwardsiella tarda	Shigella boydii	219/1 Shigella sonnei	0157:H7

Table 2a. Grid D lists the DNA of E.coli and gram positive bacteria isolated from patients' fecal specimen and loaded onto nylon plus membrane and hybridized to different random primed ^{32}P labeled gene probes.

	A	B	C	D	E	F
1	Placental	139/6	158/20	174/TM5	206/13	HP
2	114/TB	142/31	158/37	179/36	218/40	114/3g
3	117/3B	145/34	159/TM17	193/6	224/1	115/TA
4	119/1TM1	145/40	162/20	196/1	231/1	116/TC
5	129/20	152/W2-35	164/1	196/28	236/4	116/TD
6	135/35	154/1	168/4	197/5	240/28	117/2D
7	130/22	154/9	168/38	205/T18	252/22	O157:H7
8	136/36	156/TM22	172/33	205/T34	K12	Staphylococcus

Table 2a. Grid E lists the range of different amount of Enterobacter cloacae and K12 E.coli DNA loaded onto nylon plus membrane and hybridized to different types of random primed ^{32}P gene probe.

	Enterobacter Cloacae (ng)	E.coli K12 (ng)	E.coli K12 (ng)
1		1	1
2	100	2	2
3		5	5
4	100	10	10
5		20	20
6		50	50

Table 2a. Grid F lists the H. pylori primer directed PCR amplified DNA product of H. pylori and E.coli DNA and total DNA of H. pylori and E.coli isolates obtained from patients' fecal specimens. These are loaded onto nylon plus membrane and hybridized to random primed ^{32}P radiolabeled H. pylori ribosomal gene probe (HP).

	A Volume of Diluted (1:10) PCR reaction mixture after amplification	B Rows 1-5: PCR reaction mixture Rows 6-8: HP Genomic DNA	C Rows 1-5: PCR reaction mixture Rows 6-8: E.coli Genomic DNA
1	1µl (HP)	5µl of HP diluted reaction mix 5µl of K12 reaction mixture	5µl of K12 reaction mixture
2	5µl (HP)	5µl of HP diluted reaction mix 5µl of 142/31 reaction mixture	5µl of 142/31 reaction mixture
3	10µl (HP)	5µl of HP diluted reaction mix 5µl of 179/36 reaction mixture	5µl of 179/36 reaction mixture
4	50µl (HP)	5µl of HP diluted reaction mix 5µl of 197/5 reaction mixture	5µl of 197/5 reaction mixture
5	1ul (mixed E.coli isolates)	5µl of HP diluted reaction mix 5µl of 117/3B reaction mixture	5µl of 117/3B reaction mixture
6	5ul (mixed E.coli isolates)	10ng of HP genomic DNA	10ng of 117/3B genomic DNA
7	10ul (mixed E.coli isolates)	50ng of HP genomic DNA	50ng of 117/3B genomic DNA
8	50ul (mixed E.coli isolates)	200ng of HP genomic DNA	200ng of 117/3B geonmic DNA

Table 2b. Bacteria source.

Bacterial genus	Bacteria species	Source
Escherichia (E.coli)	078:H11	ATCC 35401
	078:K80:H12	ATCC 43896
	0157:H7	ATCC 43895
	029:NM	ATCC 43892
	0111	ATCC 33780
	0142:K86(B)	ATCC 23985
	TG2	gift
	K12	ATCC 29947
Escherichia (non-coli)	E. hermannii	ATCC 33650
	E. vulneris	ATCC 33821
	E. blattae	ATCC 29907
Shigella	S. flexneri (serotype 2A)	ATCC 29903
	S. sonnei	ATCC 29930
	S. boydii (serotype 2)	ATCC 8700
Edwardsiella	E. tarda (01433:H1)	ATCC 15947
Salmonella	S. choleraesuis	ATCC 43971
Citrobacter	C. freundii	ATCC 8090
Klebsiella	K. pneumoniae (Ozaenae type4)	ATCC 11296
Enterobacter	E. cloacae	ATCC 13047
Hafnia	H. alvei	ATCC 13337
Serratia	S. marcescens	ATCC 13880
Proteus	P. vulgaris	ATCC 13315
Morganella	M. morganii	ATCC 49948

Table 2b. Continue from Table 2b. Bacteria source

Bacterial genus	Bacteria species	Source
Providencia	P. alcalifaciens	ATCC 9886
Yersinia	Y.enterocolitica	ATCC 29913
Cedecea	C. lapagei	ATCC 33432
Aeromonas	A. jandanei	ATCC 49568
Enterobacter	E. agglomerans	field isolates
Pseudomonas	P. aeruginosa	field isolates
Helicobacter	H. pylori	field isolates
Human	Placental DNA	Sigma (UK)
Salmon	Salmon sperm DNA	Sigma (UK)

TG2: Gibson TJ: Studies on the Epstein-Bar-Virus Genome. Ph.D Thesis 1984. Cambridge University, UK.

Table 3. Result of in-vitro simulation of PCRISH.

Bacterial genus	Bacteria species	b1007 / b1008 genes		ycdG / b1007 genes		HP gene	
		PCR primers	Probe primers	PCR primers	Probe primers	PCR primers	Probe primers
		OH-7419 OH-7985	OH-7562 81B-7794	81B-7223 81B-7794	81B-7278 81B-7754	HP-178 HP-775	HP-228 HP-513
Escherichia (E.coli)	078:H11	+	+	+	+	-	-
	TG2	+	+	+	+	-	-
	K12	+	+	+	+	-	-
	142/31	+	+	+	+	-	-
	179/36	+	+	+	+	-	-
	197/5	+	+	+	+	-	-
Escherichia (non-coli)	E. hermannii	-	-	-	-	-	-
	E. vulneris	+	-	-	-	-	-
	E. blattae	-	-	-	-	-	-
Shigella	S. flexneri (serotype 2A)	+	+	+	+	-	-
	S. sonnei	+	+	+	+	-	-
	S.sonnei: 219/1	+	+	+	+	-	-
	S. boydii (serotype 2)	-	-	-	-	-	-
Edwardsiella	E. tarda (01433:H1)	-	-	-	-	-	-
Salmonella	S. choleraesuis	+	-	-	-	-	-
Citrobacter	C. freundii	-	-	-	-	-	-
Klebsiella	K. pneumoniae (Ozaenae type4)	-	-	-	-	-	-
Enterobacter	E. cloacae	-	-	-	-	-	-
Hafnia	H. alvei	-	-	-	-	-	-

continue from Table 3. Result of in-vitro simulation of PCRISH.

Bacterial genus	Bacteria species	b1007 / b1008 genes		ycdG / b1007 genes		HP gene	
		PCR primers	Probe primers	PCR primers	Probe primers	PCR primers	Probe primers
		OH-7419 OH-7985	OH-7562 81B-7794	81B-7223 81B-7794	81B-7278 81B-7754	HP-178 HP-775	HP-228 HP-513
Proteus	P. vulgaris	-	-	-	-	-	-
Morganella	M. morganii	-	-	-	-	-	-
Providencia	P. alcalifaciens	+	-	-	-	-	-
Yersinia	Y.enterocolitica	-	-	-	-	-	-
Cedecea	C. lapagei	+	-	-	-	-	-
Aeromonas	A. jandanei	-	-	-	-	-	-
Enterobacter	E. agglomerans	-	-	-	-	-	-
Pseudomonas	P. aeruginosa	-	-	-	-	-	-
Helicobacter	H. pylori	-	-	-	-	+	+
Human	Placental DNA	-	-	-	-	-	-
Salmon	Salmon sperm DNA	-	-	-	-	-	-

CLAIMS

1. A method for detecting cancer in a sample or predisposition to development of same, said method comprising subjecting said sample to genetic analysis using one or more *E. coli*- or *Shigella* species-specific nucleotide sequences obtainable from one or more nucleotide sequences comprising a sequence of Formula 1 or a part thereof and/or Table 1 wherein the sample is mammalian tissue or extract or cells thereof and wherein the ability of said *E. coli*- or *Shigella* species'-specific nucleotide sequences to hybridize to complementary nucleotide sequences in the nucleus of cells of said tissue, extract or cells is indicative of cancer or a predisposition to development of same.

2. The method according to claim 1 wherein the *E. coli*- and/or *Shigella* species-specific nucleotide sequences comprise at least 8 nucleotides in length.

3. The method according to claim 1 or 2 wherein hybridization of *E. coli*- and/or *Shigella* species-specific nucleotide sequences to the complementary nucleotide sequences is detected by the presence of amplified nucleic acid products.

4. The method according to any one of claims 1 to 3 wherein hybridization of *E. coli*- and/or *Shigella* species-specific nucleotide sequences to the complementary nucleotide sequences or the presence of amplified nucleic acid products is detected by a reporter molecule giving an identifiable signal.

5. The method according to any one of claims 1 to 4 wherein the tissue, extract or cells are from a patient suffering from cancer or cellular instability or a patient at risk of cancer or cellular instability.

6. The method according to any one of claims 1 to 5 wherein the cancer is gastrointestinal cancer.

7. The method according to any one of claims 1 to 5 wherein the cancer is colon cancer.

8. The method according to any one of claims 1 to 5 wherein the cancer is stomach cancer.

9. The method according to any one of claims 1 to 5 wherein the cancer is colorectal cancer.

10. The method according to claim 5 wherein the tissue or extract or cells thereof is from the gastrointestinal tract of a human patient.

11. The method according to claim 5 wherein the tissue or extract or cells thereof is from the colon of a human patient.

12. The method according to claim 5 wherein the tissue or extract or cells thereof is from the stomach of a human patient.

13. The method according to claim 5 wherein the tissue or extract or cells thereof is from the colorectum of a human patient.

14. A method for identifying nucleotide sequences, or their expressed polypeptides, capable of inducing or otherwise facilitating abnormal cell growth or abnormal physiology, said method comprising introducing a nucleotide sequence comprising *E. coli*- and/or *Shigella* species'-specific nucleotide sequence obtainable from one or more nucleotide sequences comprising a sequence of Formula I or a part thereof and/or Table 1 into cells and observing morphological and/or physiological changes to said cells compared to control cells without said introduced nucleotide sequence wherein the presence of abnormal morphology and/or physiology in a cell is indicative of a nucleotide sequence, or a polypeptide expressed therefrom, which is capable of inducing or facilitating abnormal cell growth or physiology.

15. The method according to claim 14 wherein abnormal cell growth or physiology is associated with cancer or cellular instability or a predisposition to the development of cancer or cellular instability.

16. The method according to claim 15 wherein the cancer is gastrointestinal cancer.

17. The method according to claim 15 wherein the cancer is colon cancer.

18. The method according to claim 15 wherein the cancer is stomach cancer.

19. The method according to claim 15 wherein the cancer is colorectal cancer.

20. A method for detecting cancer, or risk of cancer or cellular instability in a patient comprising subjecting a sample from the patient to genetic analysis using one or more nucleotide sequences specific to *E. coli*- or *Shigella* species and/or related microorganism and present mainly in the nucleus of cancer cells and in the normal cells adjacent to cancer cells, wherein the ability of said nucleotide sequence to hybridize to complementary nucleotide sequences in the sample is indicative of cancer, or risk of cancer or cellular instability in the patient.

21. The method according to claim 20 wherein the nucleotide sequence specific to *E. coli*- and/or *Shigella* species and/or related microorganism comprises at least 8 nucleotides in length.

22. The method according to claim 20 or 21 wherein hybridization of the nucleotide sequence specific to *E. coli*- and/or *Shigella* species and/or related microorganism to the complementary nucleotide sequences is detected by the presence of amplified nucleic acid products.

23. The method according to any one of claims 20 to 22 wherein hybridization of the nucleotide sequence specific to *E. coli*- and/or *Shigella* species and/or related microorganism to the complementary nucleotide sequences or the presence of amplified nucleic acid products is detected by a reporter molecule giving an identifiable signal.

24. The method according to any one of claims 20 to 23 wherein the sample is tissue, or extract or cells thereof from the patient or a nucleic acid preparation from said tissue, extract or cells thereof.

25. The method according to claim 24 wherein the tissue or extract or cells thereof is from the gastrointestinal tract of a human patient.

26. The method according to claim 24 wherein the tissue or extract or cells thereof is from the colon of a human patient.

27. The method according to claim 24 wherein the tissue or extract or cells thereof is from the stomach of a human patient.

28. The method according to claim 24 wherein the tissue or extract or cells thereof is from the colorectum of a human patient.

29. Use of a nucleotide sequence specific to *E. coli* and/or *Shigella* species and/or related microorganism and present mainly in the nucleus of cancer cells and in the normal cells adjacent to cancer cells for the identification of cancer or tumour or a predisposition to development of same.

30. The use according to claim 29 wherein the cancer is gastrointestinal cancer.

31. The use according to claim 29 wherein the cancer is colon cancer.

32. The use according to claim 29 wherein the cancer is stomach cancer.

33. The use according to claim 29 wherein the cancer is colorectal cancer.

34. A molecular probe comprising a nucleotide sequence specific to *E. coli* and/or *Shigella* species /or related microorganism and present mainly in the nucleus of cancer cells and in the normal cells adjacent to cancer cells for the identification of a gastrointestinal cancer or tumour or a predisposition to same.

35. The method according to any one of claims 20 to 24 wherein the cancer is gastrointestinal cancer.

36. The method according to any one of claims 20 to 24 wherein the cancer is colon cancer.

37. The method according to any one of claims 20 to 24 wherein the cancer is stomach cancer.

38. The method according to any one of claims 20 to 24 wherein the cancer is colorectal cancer.