METHODS FOR SIMPLIFYING MICROBIAL NUCLEIC ACIDS BY CHEMICAL MODIFICATION OF CYTOSINES

Technical Field

The invention relates to nucleic acid detection assays for the detection of microorganisms. The invention also relates to methods for chemical treatment of nucleic acids to reduce the complexity of microbial genomes combined with the use of specific ligands for microbial detection.

Background Art

A number of procedures are presently available for the detection of specific nucleic acid molecules. These procedures typically depend on sequence-dependent hybridisation between the target nucleic acid and nucleic acid probes which may range in length from short oligonucleotides (20 bases or less) to sequences of many kilobases (kb).

The most widely used method for amplification of specific sequences from within a population of nucleic acid sequences is that of polymerase chain reaction (PCR) (Dieffenbach, C and Dveksler, G. eds. PCR Primer: A Laboratory Manual. Cold Spring Harbor Press, Plainview NY). In this amplification method, oligonucleotides, generally 20 to 30 nucleotides in length on complementary DNA strands and at either end of the region to be amplified, are used to prime DNA synthesis on denatured single-stranded DNA. Successive cycles of denaturation, primer hybridisation and DNA strand synthesis using thermostable DNA polymerases allows exponential amplification of the sequences between the primers. RNA sequences can be amplified by first copying using reverse transcriptase to produce a complementary DNA (cDNA) copy. Amplified DNA fragments can be detected by a variety of means including gel electrophoresis, hybridisation with labelled probes, use of tagged primers that allow subsequent identification (eg by an enzyme linked assay), and use of fluorescently-tagged primers that give rise to a signal upon hybridisation with the target DNA (eg Beacon and TaqMan systems).

As well as PCR, a variety of other techniques have been developed for detection and amplification of specific nucleotide sequences. One example is the ligase chain reaction (1991, Barany, F. et al., Proc. Natl. Acad. Sci. USA 88, 189-193).

Another example is isothermal amplification which was first described in 1992 (Walker GT, Little MC, Nadeau JG and Shank D. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. PNAS 89: 392-396 (1992) and termed

Strand Displacement Amplification (SDA). Since then, a number of other isothermal amplification technologies have been described including Transcription Mediated Amplification (TMA) and Nucleic Acid Sequence Based Amplification (NASBA) that use an RNA polymerase to copy RNA sequences but not corresponding genomic DNA (Guatelli JC, Whitfield KM, Kwoh DY, Barringer KJ, Richmann DD and Gingeras TR. Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. PNAS 87: 1874-1878 (1990): Kievits T, van Gemen B, van Strijp D, Schukkink R, Dircks M, Adriaanse H, Malek L, Sooknanan R, Lens P. NASBA isothermal enzymatic in vitro nucleic acid amplification optimized for the diagnosis of HIV-1 infection. J Virol Methods. 1991 Dec; 35(3):273-86).

Other DNA-based isothermal techniques include Rolling Circle Amplification (RCA) in which a DNA polymerase extends a primer directed to a circular template (Fire A and Xu SQ. Rolling replication of short circles. PNAS 92: 4641-4645 (1995), Ramification Amplification (RAM) that uses a circular probe for target detection (Zhang W, Cohenford M, Lentrichia B, Isenberg HD, Simson E, Li H, Yi J, Zhang DY. Detection of *Chlamydia trachomatis* by isothermal ramification amplification method: a feasibility study. J Clin Microbiol. 2002 Jan; 40(1):128-32.) and more recently, Helicase-Dependent isothermal DNA amplification (HDA), that uses a helicase enzyme to unwind the DNA strands instead of heat (Vincent M, Xu Y, Kong H. Helicase-dependent isothermal DNA amplification. EMBO Rep. 2004 Aug; 5(8):795-800.)

Recently, isothermal methods of DNA amplification have been described (Walker GT, Little MC, Nadeau JG and Shank D. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. PNAS 89: 392-396 (1992). Traditional amplification techniques rely on continuing cycles of denaturation and renaturation of the target molecules at each cycle of the amplification reaction. Heat treatment of DNA results in a certain degree of shearing of DNA molecules, thus when DNA is limiting such as in the isolation of DNA from a small number of cells from a developing blastocyst, or particularly in cases when the DNA is already in a fragmented form, such as in tissue sections, paraffin blocks and ancient DNA samples, this heating-cooling cycle could further damage the DNA and result in loss of amplification signals. Isothermal methods do not rely on the continuing denaturation of the template DNA to produce single stranded molecules to serve as templates from further amplification, but on enzymatic nicking of DNA molecules by specific restriction endonucleases at a constant temperature.

The technique termed Strand Displacement Amplification (SDA) relies on the ability of certain restriction enzymes to nick the unmodified strand of hemi-modified DNA and the ability of a 5'-3' exonuclease-deficient polymerase to extend and displace the downstream strand. Exponential amplification is then achieved by coupling sense and antisense reactions in which strand displacement from the sense reaction serves as a template for the antisense reaction (Walker GT, Little MC, Nadeau JG and Shank D. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. PNAS 89: 392-396 (1992). Such techniques have been used for the successful amplification of *Mycobacterium tuberculosis* (Walker GT, Little MC, Nadeau JG and Shank D. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. PNAS 89: 392-396 (1992), HIV-1, *Hepatitis C* and HPV-16 Nuovo G. J., 2000), *Chlamydia trachomatis* (Spears PA, Linn P, Woodard DL and Walker GT. Simultaneous Strand Displacement Amplification and Fluorescence Polarization Detection of Chlamydia trachomatis. Anal. Biochem. 247: 130-137 (1997).

The use of SDA to date has depended on modified phosphorthioate nucleotides in order to produce a hemi-phosphorthioate DNA duplex that on the modified strand would be resistant to enzyme cleavage, resulting in enzymic nicking instead of digestion to drive the displacement reaction. Recently, however, several "nickase" enzyme have been engineered. These enzymes do not cut DNA in the traditional manner but produce a nick on one of the DNA strands. "Nickase" enzymes include N.Alw1 (Xu Y, Lunnen KD and Kong H. Engineering a nicking endonuclease N.Alw1 by domain swapping. PNAS 98: 12990-12995 (2001), N.BstNB1 (Morgan RD, Calvet C, Demeter M, Agra R, Kong H. Characterization of the specific DNA nicking activity of restriction endonuclease . N.BstNBI. Biol Chem. 2000 Nov;381(11):1123-5.) and Mly1 (Besnier CE, Kong H. Converting MlyI endonuclease into a nicking enzyme by changing its oligomerization state. EMBO Rep. 2001 Sep;2(9):782-6. Epub 2001 Aug 23). The use of such enzymes would thus simplify the SDA procedure.

In addition, SDA has been improved by the use of a combination of a heat stable restriction enzyme (Ava1) and Heat stable Exo-polymerase (Bst polymerase). This combination has been shown to increase amplification efficiency of the reaction from a 10^8 fold amplification to 10^{10} fold amplification so that it is possible, using this technique, to the amplification of unique single copy molecules. The resultant amplification factor using the heat stable polymerase/enzyme combination is in the order of 10^9 (Milla M. A., Spears P. A., Pearson R. E. and Walker G. T. Use of the Restriction Enzyme Ava1 and

Exo-Bst Polymerase in Strand Displacement Amplification Biotechniques 1997 24:392-396).

To date, all isothermal DNA amplification techniques require the initial double stranded template DNA molecule to be denatured prior to the initiation of amplification. In addition, amplification is only initiated once from each priming event.

For direct detection, the target nucleic acid is most commonly separated on the basis of size by gel electrophoresis and transferred to a solid support prior to hybridisation with a probe complementary to the target sequence (Southern and Northern blotting). The probe may be a natural nucleic acid or analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA) or intercalating nucleic acid (INA). The probe may be directly labelled (eg with ^{32}P) or an indirect detection procedure may be used. Indirect procedures usually rely on incorporation into the probe of a "tag" such as biotin or digoxigenin and the probe is then detected by means such as enzyme-linked substrate conversion or chemiluminescence.

Another method for direct detection of nucleic acid that has been used widely is "sandwich" hybridisation. In this method, a capture probe is coupled to a solid support and the target nucleic acid, in solution, is hybridised with the bound probe. Unbound target nucleic acid is washed away and the bound nucleic acid is detected using a second probe that hybridises to the target sequences. Detection may use direct or indirect methods as outlined above. Examples of such methods include the "branched DNA" signal detection system, an example that uses the sandwich hybridization principle (1991, Urdea, M. S., et al., Nucleic Acids Symp. Ser. 24,197-200). A rapidly growing area that uses nucleic acid hybridisation for direct detection of nucleic acid sequences is that of DNA microarrays, (2002, Nature Genetics, 32, [Supplement]; 2004, Cope, L.M., et al., Bioinformatics, 20, 323-331; 2004, Kendall, S.L., et al., Trends in Microbiology, 12, 537-544). In this process, individual nucleic acid species, that may range from short oligonucleotides, (typically 25-mers in the Affymetrix system), to longer oligonucleotides, (typically 60-mers in the Applied Biosystems and Agilent platforms), to even longer sequences such as cDNA clones, are fixed to a solid support in a grid-pattern or photolithographically synthesized on a solid support. A tagged or labelled nucleic acid population is then hybridised with the array and the level of hybridisation to each spot in the array quantified. Most commonly, radioactively- or fluorescently-labelled nucleic acids (eg cRNAs or cDNAs) are used for hybridisation, though other detection systems can be employed, such as chemiluminescence.

A rapidly growing area that uses nucleic acid hybridisation for direct detection of nucleic acid sequences is that of DNA micro-arrays (Young RA Biomedical discovery with DNA arrays. Cell 102: 9-15 (2000); Watson A New tools. A new breed of high tech detectives. Science 289:850-854 (2000)). In this process, individual nucleic acid species, that may range from oligonucleotides to longer sequences such as complementary DNA (cDNA) clones, are fixed to a solid support in a grid pattern. A tagged or labelled nucleic acid population is then hybridised with the array and the level of hybridisation with each spot in the array quantified. Most commonly, radioactively- or fluorescently-labelled nucleic acids (eg cDNAs) were used for hybridisation, though other detection systems were employed.

Traditional methods for the detection of microorganisms such as bacteria, yeasts and fungi and include culture of the microorganisms on selective nutrient media then classification of the microorganism based on size, shape, spore production, characters such as biochemical or enzymatic reactions and specific staining properties (such as the Gram stain) as seen under conventional light microscopy. Viral species have to be grown in specialised tissue or cells then classified based on their structure and size determined by electron microscopy. A major drawback of such techniques is that not all microorganisms will grow under conventional culture or cell conditions limiting the usefulness of such approaches. With bacteria, for example, such as *Neisseria meningitidis*, *Streptococcus pneumoniae* and *Haemophilus influenzae* (which all cause meningitis and amongst which *N. meningitidis* causes both meningitis and fulminant meningococcaemia) all three species are difficult to culture. Blood culture bottles are routinely examined every day for up to seven days, and subculturing is required. *H. influenzae* requires special medium containing both nicotinamide adenine dinucleotide and haemin and growth on Chocolate Agar Plates. Blood cultures require trypticase soy broth or brain heart infusion and the addition of various additives such as sodium polyanetholesulphonate. For microorganisms such as *Clostridium botulinum*, which causes severe food poisoning and floppy baby syndrome, the identification of the toxin involves injection of food extracts or culture supernatants into mice and visualization of results after 2 days. In addition, culturing of the potential microorganism on special media takes a week. *Staphylococcus aureus* enterotoxin (a cause of food poisoning as well as skin infections, blood infections, pneumonia, osteomyelitis, arthritis and brain abscesses) is detected in minute amounts by selective absorption of the toxin via ion exchange resins or Reverse Passive Latex Agglutination using monoclonal antibodies. Its relative, *S. epidermis*, leads to blood infections and contaminates equipment and surfaces in hospitals and health care machines and appliances.

Non-viral microorganisms can also be classified based on their metabolic properties such as the production of specific amino acids or metabolites during fermentation reactions on substrates such as glucose, maltose or sucrose. Alternatively, microorganisms can be typed based on their sensitivity to antibiotics. Specific antibodies to cell surface antigens or excreted proteins such as toxins are also used to identify or type microorganisms. However, all the above methods rely on the culture of the microorganism prior to subsequent testing. Culture of microorganisms is expensive and time consuming and can also suffer from contamination or overgrowth by less fastidious microorganisms. The techniques are also relatively crude in that many tests must be done on the same sample in order to reach definitive diagnosis. Most microorganisms can not be readily grown in known media, and hence they fall below levels of detection when a typical mixed population of different species of microorganism is present in the wild or in association with higher organisms.

Other methods for the detection and identification of pathogenic microorganisms are based on the serological approach in which antibodies are produced in response to infection with the microorganism. Meningococci, for example, are classifiable on the basis of the structural differences in their capsular polysaccharides. These have different antigenicities, allowing five major serogroups to be determined, (A, B, C, Y and W-135). Enzyme Linked Immunosorbent Assays (ELISA) or Radio Immuno Assay (RIA) can assess the production of such antibodies. Both these methods detect the presence of specific antibodies produced by the host animal during the course of infection. These methods suffer the drawback in that it takes some time for an antibody to be produced by the host animal, thus very early infections are often missed. In addition, the use of such assays cannot reliably differentiate between past and active infection.

More recently, there has been much interest in the use of molecular methods for the diagnosis of infectious disease. These methods offer sensitive and specific detection of pathogenic microorganisms. Examples of such methods include the "branched DNA" signal detection system. This method is an example that uses the sandwich hybridization principle (Urdea MS et al. Branched DNA amplification multimers for the sensitive, direct detection of human HIV and hepatitis viruses. Nucleic Acids Symp Ser. 1991;(24):197-200).

Another method for the detection and classification of bacteria is the amplification of 16S ribosomal RNA sequences. 16S rRNA has been reported to be a suitable target for use in PCR amplification assays for the detection of bacterial species in a variety of clinical or environmental samples and has frequently been used to identify various

specific microorganisms because 16S rRNA genes show species-specific polymorphisms (Cloud, J. L., H. Neal, R. Rosenberry, C. Y. Turenne, M. Jama, D. R. Hillyard, and K. C. Carroll. 2002. J. Clin. Microbiol. 40:400-406). However, pure culture of bacteria are required and after PCR amplification the sample still has to be sequenced or hybridized to a micro-array type device to determine the species (Fukushima M, Kakinuma K, Hayashi H, Nagai H, Ito K; Kawaguchi R. J Clin Microbiol. 2003 Jun; 41(6):2605-15). Such methods are expensive, time consuming and labour intensive.

The present inventors have developed new methods for detecting microorganisms which can be adapted to general detection or initial screening assays for any microbial species.

Disclosure of Invention

In a general aspect, the present invention relates to reducing the complexity of the base make up of a microbial genome or nucleic acid by treating microbial nucleic acid with an agent that modifies cytosine and amplifying the treated nucleic acid to produce a simplified form of the genome or nucleic acid.

In a first aspect, the present invention provides a method for simplification of a microbial genome or microbial nucleic acid comprising:

treating microbial genome or nucleic acid with an agent that modifies cytosine to form derivative microbial nucleic acid; and

amplifying the derivative microbial nucleic acid to produce a simplified form of the microbial genome or nucleic acid.

In a second aspect, the present invention provides a method for producing a microbial-specific nucleic acid molecule comprising:

treating a sample containing microbial derived DNA with an agent that modifies cytosine to form derivative microbial nucleic acid; and

amplifying at least part of the derivative microbial nucleic acid to form a simplified nucleic acid molecule having a reduced total number of cytosines compared with the corresponding untreated microbial nucleic acid, wherein the simplified nucleic acid molecule includes a nucleic acid sequence specific for a microorganism or microorganism type.

In a third aspect, the present invention provides a method for producing a microbial-specific nucleic acid molecule comprising:

obtaining a DNA sequence from a microorganism;

forming a simplified form of the microbial DNA sequence by carrying out a conversion of the microbial DNA sequence by changing each cytosine to thymine such that the simplified form of the microbial DNA comprises substantially bases adenine, guanine and thymine; and

selecting a microbial-specific nucleic acid molecule from the simplified form of the microbial DNA.

In a fourth aspect, the present invention provides a microbial-specific nucleic acid molecule obtained by the method according to the third aspect of the present invention.

In a fifth aspect, the present invention provides use of the method according to the third aspect of the present invention to obtain probes or primers to bind or amplify the microbial-specific nucleic acid molecule in a test or assay.

In a sixth aspect, the present invention provides probes or primers obtained by the fifth aspect of the present invention.

In a seventh aspect, the present invention provides a method for detecting the presence of a microorganism in a sample comprising:

obtaining microbial DNA from a sample suspected of containing the microorganism;

treating the microbial nucleic acid with an agent that modifies cytosine to form derivative microbial nucleic acid;

providing primers capable of allowing amplification of a desired microbial-specific nucleic acid molecule to the derivative microbial nucleic acid;

carrying out an amplification reaction on the derivative microbial nucleic acid to form a simplified nucleic acid; and

assaying for the presence of an amplified nucleic acid product containing the desired microbial-specific nucleic acid molecule, wherein detection of the desired microbial-specific nucleic acid molecule is indicative of the presence of the microorganism in the sample.

If the genome or microbial nucleic acid is DNA it can be treated to form a derivative DNA which is then amplified to form simplified form of DNA.

If the genome or microbial nucleic acid is RNA it can be converted to DNA prior to treating the microbial genome or nucleic acid. Alternatively, microbial RNA can be

treated to yield a derivative RNA molecule which is then converted a derivative DNA molecule prior to amplification. Methods of conversion of RNA to DNA are well known and include use of reverse transcriptase to form a cDNA.

The microbial genome or nucleic acid can be obtained from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, or single cell organism.

The microbial genome or nucleic acid can be selected from protein encoding nucleic acid, non-protein encoding nucleic acid, ribosomal gene regions of prokaryotes or single celled eukaryotic microorganisms. Preferably, the ribosomal gene regions are 16S or 23S in prokaryotes and 18S and 28S in the case of single celled eukaryotic microorganisms. The agent can be selected from bisulfite, acetate or citrate. Preferably, the agent is sodium bisulfite.

Preferably, the agent modifies an cytosine to a uracil in each strand of complementary double stranded microbial genomic DNA forming two derivative but non-complementary microbial nucleic acid molecules. In a preferred form, the cytosine is unmethylated as is typically found in microbial nucleic acid.

Preferably, the derivative microbial nucleic acid has a reduced total number of cytosines compared with the corresponding untreated microbial genome or nucleic acid.

Preferably, the simplified form of the microbial genome or nucleic acid has a reduced total number of cytosines compared with the corresponding untreated microbial genome or nucleic acid.

In one preferred form, the derivative microbial nucleic acid substantially contains bases adenine (A), guanine (G), thymine (T) and uracil (U) and has substantially the same total number of bases as the corresponding untreated microbial genome or nucleic acid.

In another preferred form, the simplified form of the microbial genome or nucleic acid is comprised substantially of bases adenine (A), guanine (G) and thymine (T).

Preferably, the amplification is carried out by any suitable means such as polymerase chain reaction (PCR), isothermal amplification, or signal amplification.

The method according to the second aspect of the present invention may further comprise:

detecting the microbial-specific nucleic acid molecule.

In a preferred form, the microbial-specific nucleic acid molecule is detected by:

providing a detector ligand capable of binding to a target region of the microbial-specific nucleic acid molecule and allowing sufficient time for the detector ligand to bind to the target region; and

measuring binding of the detector ligand to the target region to detect the presence of the microbial-specific nucleic acid molecule.

In another preferred form, the microbial-specific nucleic acid molecule is detected by separating an amplification product and visualising the separated product. Preferably, the amplification product is separated by electrophoresis and detected by visualising one or more bands on a gel.

Preferably, the microbial-specific nucleic acid molecule does not occur naturally in the microorganism.

In a preferred form, the microbial-specific nucleic acid molecule has a nucleic acid sequence indicative of a taxonomic level of the microorganism. The taxonomic level of the microorganism includes, but not limited to, family, genus, species, strain, type, or different populations from the same or different geographic or benthic populations.

In a preferred form of the method according to third aspect of the present invention, simplified forms of two or more microbial DNA sequences are obtained and the two or more sequences are compared to obtain at least one microbial-specific nucleic acid molecule.

In a preferred form of the seventh aspect of the present invention, the nucleic acid molecules are detected by:

providing a detector ligand capable of binding to a region of the nucleic acid molecule and allowing sufficient time for the detector ligand to bind to the region; and

measuring binding of the detector ligand to the nucleic acid molecule to detect the presence of the nucleic acid molecule.

In another preferred form, the nucleic acid molecules are detected by separating an amplification product and visualising the separated product.

In situations where the microorganism does not have a DNA genome or the microbial genome or nucleic acid is RNA, for example a RNA virus, the RNA viral genome can be first converted to cDNA in order to treat DNA with the agent. RNA may also be treated and the derivative RNA is converted to DNA prior to amplification.

Preferably, the derivative nucleic acid substantially contains the bases adenine (A), guanine (G), thymine (T) and uracil (U) and has substantially the same total number

of bases as the corresponding unmodified microbial nucleic acid. Importantly, the derivative nucleic acid molecule substantially does not contain cytosine (C), with the proviso that the microbial DNA was not methylated at any cytosines.

Preferably the amplified derivative nucleic acid substantially contains the bases A, T and G and has substantially the same total number of bases as the corresponding derivative nucleic acid (and unmodified microbial nucleic acid). The amplified derivative nucleic acid is termed simplified nucleic acid.

In a preferred form, the microbial-specific nucleic acid molecule has a nucleic acid sequence indicative of a taxonomic level of the microorganism. The taxonomic level of the microorganism can include family, genus, species, strain, type, or different populations from the same or different geographic or benthic populations. In the case of bacteria we can adhere to the generally recognized schema, such as; *Bacteria, Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria*. Different populations may be polymorphic for single nucleotide changes or variation that exists in DNA molecules that exist in an intracellular form within a microorganism (plasmids or phagemids), or polymorphic chromosomal regions of microorganism genomes such as pathogenicity islands.

The present invention can also be used to recognize the fluidity of microbial and viral genomes, and can be used to recognize the chimeric nature of viral genomes, which can be in independent pieces, and hence newly arising strains arise from re-assortment of genomic regions from different animals e.g. new human influenza strains as chimeras of segments that are picked up from other mammalian or avian viral genomes.

It will be appreciated that the method can be carried out *in silico* from known nucleic acid sequences of microorganisms where one or more cytosines in the original sequences is converted to thymine to obtain the simplified nucleic acid. Sequence identity can be determined from the converted sequences. Such an *in silico* method mimics the treatment and amplification steps.

When a microbial-specific nucleic acid molecule has been obtained for any given microorganism by this method, probes or primers can be designed to ensure amplification of the region of interest in an amplification reaction. Thus, when the probes or primers have been designed, it will be possible to carry out clinical or scientific assays on samples to detect a given microorganisms at a given taxonomic level.

The microbial-specific nucleic acid molecule can be unique or have a high degree of similarity within a taxonomic level. One advantage of the present invention is the

ability to greatly simplify the potential base differences between, or within, taxonomic levels, for example, of a microorganism to either an unique molecule or molecules that have close sequence similarity. Specific primers or reduced number of degenerate primers can be used to amplify the microbial-specific nucleic acid molecule in a given sample.

For double stranded DNA which contains cytosines, the treating step results in two derivative nucleic acids (one for each complementary strand), each containing the bases adenine, guanine, thymine and uracil. The two derivative nucleic acids are produced from the two single strands of the double stranded DNA. The two derivative nucleic acids preferably have no cytosines but still have the same total number of bases and sequence length as the original untreated DNA molecule. Importantly, the two derivative nucleic acids are not complimentary to each other and form a top and a bottom strand template for amplification. One or more of the strands can be used as the target for amplification to produce the simplified nucleic acid molecule. During amplification of the derivative nucleic acids, uracils in the top (or bottom strand) are replaced by thymines in the corresponding amplified simplified form of the nucleic acid. As amplification continues, the top (and/ or bottom strand if amplified) will be diluted out as each new complimentary strand will have only bases adenine, guanine, thymine.

It will be appreciated that this aspect of the invention also includes nucleic acid molecules having complementary sequences to the microbial-specific nucleic acid molecule, and nucleic acid molecules that can hybridize, preferably under stringent conditions, to the microbial-specific nucleic acid molecule.

The present invention can use probes or primers that are indicative of representative types of microorganism which can be used to determine whether any microorganism is present in a given sample. Further microbial type-specific probes can be used to actually detect or identify a given, type, subtype, variant and genotype examples of microorganism.

When a microbial-specific nucleic acid molecule has been obtained or identified for any given microorganism, probes or primers can be designed to ensure amplification of the region of interest in an amplification reaction. It is important to note that both strands of a treated and thus converted genome, (hereafter termed "derivative nucleic acid') can be analyzed for primer design, since treatment or conversion leads to asymmetries of sequence, and hence different primer sequences are required for the detection of the 'top' and 'bottom' strands of the same locus, (also known as the 'Watson' and 'Crick' strands). Thus, there are two populations of molecules, the converted

genome as it exists immediately after conversion, and the population of molecules that results after the derivative nucleic acid is replicated by conventional enzymological means (PCR) or by methods such as isothermal amplification. Primers are typically designed for the converted top strand for convenience but primers can also be generated for the bottom strand. Thus, it will be possible to carry out clinical or scientific assays on samples to detect a given microorganism.

The primers or probes can be designed to allow specific regions of derivative nucleic acid to be amplified. In a preferred form, the primers cause the amplification of the microbial-specific nucleic acid molecule.

In a seventh aspect, the present invention provides a kit for detecting a microbial-specific nucleic acid molecule comprising primers or probes according to fifth aspect of the present invention together with one or more reagents or components for an amplification reaction.

Preferably, the microorganism is selected from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, single cell organism, or any other microorganism, no matter how variously classified, such as the Kingdom *Protoctista* by Margulis, L., *et al* 1990, Handbook of Protoctista, Jones and Bartlett, Publishers, Boston USA, or microorganisms that are associated with humans, as defined in Harrisons Principles of Internal Medicine, 12th Edition, edited by J D Wilson et al., McGraw Hill Inc, as well as later editions. It also includes all microorganisms described in association with human conditions defined in OMIM, Online Mendelian Inheritance in Man, www.ncbi.gov.

The microorganism can be a pathogen, naturally occurring environmental sample, water or airborne organism, (or an organism existing or being carried in a liquid or gaseous medium), in either a mature or spore form, either extracellularly or intracellularly, or associated with a chimeric life form, or existing ectocommensally between two or more life forms, such as a microbe associated with a lichen, or a microbe associated with a bacterial film.

It is possible to assay for the presence of RNA viruses or viroids by first converting their RNA genome into a cDNA form via reverse transcription and then modifying the cDNA by the reagent. This gets over the problem of any methylation existing at cytosines in RNA viruses, as the reverse transcriptase will copy these as if they were regular cytosines.

Preferably, the agent modifies unmethylated cytosine to uracil which is then replaced as a thymine during amplification of the derivative nucleic acid. Preferably, the

agent used for modifying cytosine is sodium bisulfite. Other agents that similarly modify unmethylated cytosine, but not methylated cytosine can also be used in the method of the invention. Examples include, but not limited to bisulfite, acetate or citrate. Preferably, the agent is sodium bisulfite, a reagent, which in the presence of water, modifies cytosine into uracil.

Sodium bisulfite ($NaHSO_3$) reacts readily with the 5,6-double bond of cytosine to form a sulfonated cytosine reaction intermediate which is susceptible to deamination, and in the presence of water gives rise to a uracil sulfite. If necessary, the sulfite group can be removed under mild alkaline conditions, resulting in the formation of uracil. Thus, potentially all cytosines will be converted to uracils. Any methylated cytosines, however, cannot be converted by the modifying reagent due to protection by methylation.

The present invention can be adapted to assist in circumventing some of the emerging problems revealed by the enormous unexpected genomic variation between isolates of the same bacterial species, (2005, Tettelin , H., et al., Proc. Natl. Acad. Sci. USA. 102, 13950-13955; Genome analysis of multiple pathogenic isolates of *Streptococcus agalacticiae*: implications for the microbial "pan-genome"). All isolates of this bacterial species have a "core" genome of protein coding genes which represents approximately 80% of the gene pool, plus a dispensable genome consisting of partially shared and strain-specific protein coding genes. By treating the 23S gene(s) present within a bacterial population by the methods according to the present invention, the inventors can deal with a core non-protein coding component that is present in all bacterial isolates.

The present invention is suitable for clinical, environmental, forensic, biological warfare, or scientific assays for microorganisms where the initial identity above or at the species level is useful, in order to first determine the general group to which the organism belongs. Examples include, but not limited to, diagnosis of disease in any organism, (be it vertebrate, invertebrate, prokaryotic or eukaryotic, e.g. diseases of plants and livestock, diseases of human food sources such as fish farms and oyster farms), screening or sampling of environmental sources be they natural or contaminated, determining contamination of cell cultures or *in vitro* fertilized eggs for human blastocyst production in *in vitro* fertilization clinics or for animal breeding. Detection of microorganisms in forensic settings or in biological warfare contexts, is of particular significance.

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, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia prior to development of the present invention.

In order that the present invention may be more clearly understood, preferred embodiments will be described with reference to the following drawings and examples.

Brief Description of the Drawings

Figure 1 shows alignment of part of the *Neisseria meningitidis* and *Neisseria gonorrhoeae* iga gene before and after genomic simplification. As can be seen, prior to genomic simplification, a total of 512 probe combinations would be required for the universal detection of *Neisseria* species (74% sequence similarity) compared with only 2 combinations after simplification to form derivative nucleic acid (97% sequence similarity). (SEQ ID NO is listed after each sequence).

Figure 2 shows the use of INA probes to further increase the sequence similarity of the simplified sequences, since INA probes can be of shorter length than standard oligonucleotide probes. Combining the genomic simplification procedure with INA probes allows the selection and use of probes with 100% sequence similarity to the target sequence. (SEQ ID NO is listed after each sequence).

Figure 3 shows genomic simplification to differentiate between closely related species using alignments of the iga gene from *Neisseria* and *Haemophilus*. As can be seen, the method of the present invention allows the simplification of the genomic material in order to produce species specific probes. In addition, although simplifying the genomic DNA, it still allows differentiation between *Neisseria* and the closely related *Haemophilus* species. (SEQ ID NO is listed after each sequence).

Figure 4 shows alignment of the Streptococcal tuf gene before and after genomic simplification in 10 different species of Streptococci. Before treatment, a total of 12,288

probe combinations would be required for the universal primer of the tuf gene. After genomic simplification, only 64 probe combinations would be required for universal detection. In addition, the sequence similarity before simplification is only 67.5% which increases to 85% after simplification. (SEQ ID NO is listed after each sequence).

Figure 5 shows alignment of the Staphylococcal enterotoxin genes before and after genomic simplification. Before bisulfite treatment, a total of 1,536 probe combinations would be required for the universal primer of the Staphylococcal enterotoxin gene. After genomic simplification only 64 probe combinations would be required for universal detection. (SEQ ID NO is listed after each sequence).

Figure 6 shows alignment of the Influenza group A and B neuraminidase gene of various influenza strains before and after genomic simplification. Before treatment, a total of 2,048 probe combinations would be required for the universal primer of group A and B neuraminidase gene. After genomic simplification only 48 probe combinations would be required for universal detection. In addition, the sequence similarity before simplification is only 50% which increases to 75% after simplification. (SEQ ID NO is listed after each sequence).

Figure 7 shows alignment of the Rotavirus VP4 gene before and after genomic simplification. Before treatment, a total of 512 probe combinations would be required for the universal primer of the Rotavirus VP4 gene. After genomic simplification only 32 probe combinations would be required for universal detection. (SEQ ID NO is listed after each sequence).

Figure 8 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions of Gram positive and Gram negative bacteria, with appropriate amplicons being detected as bands of specific length by agarose gel electrophoresis. The arrow indicates the expected size of the amplicons relative to standard sized markers run in the Marker lane, (M). Using primers specific for Gram negative bacteria reveals bands only in the six Gram negative lanes, (top panel). Using primers specific for Gram positive bacteria reveals only bands in the six Gram positive lanes, (lower panel).

Figure 9 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions of *E. coli* (lane 1) and *K. pneumoniae*, (lane 3). The specificity of amplification is illustrated by the absence of amplification products from the remaining 10 species of bacteria.

Figure 10 shows the amplification product obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for Neisseria.

Figure 11 shows the amplification product obtained by PCR from a protein coding gene from the genomically simplified region of the recA gene of *E. coli*. The specificity of the amplicon is illustrated by the presence of the *E. coli* recA amplicon and its absence from the other 11 species of bacteria.

Figure 12 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for Staphylococci.

Figure 13 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for Streptococci.

Figure 14 shows the amplification products obtained by PCR from a protein coding gene from the genomically simplified region of the recA gene of *Staphylococcus epidermidis*. The two bands (arrowed) represent carry over amplicons from the first round, (upper band) and second round (lower band), PCR amplifications.

Figure 15 shows detection of amplicons using specific primers targeting the genomically simplified 23S ribosomal genes of *Chlamydia trachomatis*.

Figure 16 shows sequences of normal genomic and genomically simplified 23S rDNA sequences from *Staphylococcus epidermidis*. (SEQ ID NO is listed after each sequence).

Figure 17 shows sequences of genomic and genomically simplified sequences of the *E. coli* recA gene. (SEQ ID NO is listed after each sequence).

Mode(s) for Carrying Out the Invention

Definitions

The term "genomic simplification" as used herein means the genomic (or other) nucleic acid is modified from being comprised of four bases adenine (A), guanine (G), thymine (T) and cytosine (C) to substantially containing the bases adenine (A), guanine (G), thymine (T) but still having substantially the same total number of bases.

The term "derivative nucleic acid " as used herein means a nucleic acid that substantially contains the bases A, G, T and U (or some other non-A, G or T base or base-like entity) and has substantially the same total number of bases as the corresponding unmodified microbial nucleic acid. Substantially all cytosines in the microbial DNA will have been converted to uracil during treatment with the agent. It will

be appreciated that altered cytosines, such as by methylation, may not necessarily be converted to uracil (or some other non-A, G or T base or base-like entity). As microbial nucleic acid typically does not contain methylated cytosine (or other cytosine alterations) the treated step preferably converts all cytosines. Preferably, cytosine is modified to uracil.

The term "simplified nucleic acid" as used herein means the resulting nucleic acid product obtained after amplifying derivative nucleic acid. Uracil in the derivative nucleic acid is then replaced as a thymine (T) during amplification of the derivative nucleic acid to form the simplified nucleic acid molecule. The resulting product has substantially the same number of total bases as the corresponding unmodified microbial nucleic acid but is substantially made up of a combination of three bases (A, G and T).

The term "simplified sequence" as used herein means the resulting nucleic acid sequence obtained after amplifying derivative nucleic acid to form a simplified nucleic acid. The resulting simplified sequence has substantially the same number of total bases as the corresponding unmodified microbial nucleic acid sequence but is substantially made up of a combination of three bases (A, G and T).

The term "non-converted sequence" as used herein means the nucleic acid sequence of the microbial nucleic acid prior to treatment and amplification. A non-converted sequence typically is the sequence of the naturally occurring microbial nucleic acid.

The term "modifies" as used herein means the conversion of an cytosine to another nucleotide. Preferably, the agent modifies unmethylated cytosine to uracil to form a derivative nucleic acid.

The term "agent that modifies cytosine" as used herein means an agent that is capable of converting cytosine to another chemical entity. Preferably, the agent modifies cytosine to uracil which is then replaced as a thymine during amplification of the derivative nucleic acid. Preferably, the agent used for modifying cytosine is sodium bisulfite. Other agents that similarly modify cytosine, but not methylated cytosine can also be used in the method of the invention. Examples include, but not limited to bisulfite, acetate or citrate. Preferably, the agent is sodium bisulfite, a reagent, which in the presence of acidic aqueous conditions, modifies cytosine into uracil. Sodium bisulfite ($NaHSO_3$) reacts readily with the 5,6-double bond of cytosine to form a sulfonated cytosine reaction intermediate which is susceptible to deamination, and in the presence of water gives rise to a uracil sulfite. If necessary, the sulfite group can be removed under mild alkaline conditions, resulting in the formation of uracil. Thus, potentially all

cytosines will be converted to uracils. Any methylated cytosines, however, cannot be converted by the modifying reagent due to protection by methylation. It will be appreciated that cytosine (or any other base) could be modified by enzymatic means to achieve a derivative nucleic acid as taught by the present invention.

There are two broad generic methods by which bases in nucleic acids may be modified: chemical and enzymatic. Thus, modification for the present invention can also be carried out by naturally occurring enzymes, or by yet to be reported artificially constructed or selected enzymes. Chemical treatment, such as bisulphite methodologies, can convert cytosine to uracil via appropriate chemical steps. Similarly, cytosine deaminases, for example, may carry out a conversion to form a derivative nucleic acid. The first report on cytosine deaminases to our knowledge is 1932, Schmidt, G., Z. physiol. Chem., 208, 185; (see also 1950, Wang, T.P., Sable, H.Z., Lampen, J.O., J. Biol. Chem, 184, 17-28, Enzymatic deamination of cytosines nucleosides). In this early work, cytosine deaminase was not obtained free of other nucleo-deaminases, however, Wang et al. were able to purify such an activity from yeast and *E. coli*. Thus any enzymatic conversion of cytosine to form a derivative nucleic acid which ultimately results in the insertion of a base during the next replication at that position, that is different to a cytosine, will yield a simplified genome. The chemical and enzymatic conversion to yield a derivative followed by a simplified genome are applicable to any nucleo-base, be it purines or pyrimidines in naturally occurring nucleic acids of microorganisms.

The term "simplified form of the genome or nucleic acid" as used herein means that a genome or nucleic acid, whether naturally occurring or synthetic, which usually contains the four common bases G, A, T and C, now consists largely of only three bases, G, A and T since most or all of the Cs in the genome have been converted to Ts by appropriate chemical modification and subsequent amplification procedures. The simplified form of the genome means that relative genomic complexity is reduced from a four base foundation towards a three base composition.

The term 'base-like entity' as used herein means an entity that is formed by modification of cytosine. A base-like entity can be recognised by a DNA polymerase during amplification of a derivative nucleic acid and the polymerase causes A, G or T to be placed on a newly formed complementary DNA strand at the position opposite the base-like entity in the derivate nucleic acid. Typically, the base-like entity is uracil that has been modified from cytosine in the corresponding untreated microbial nucleic acid. Examples of a base-like entity includes any nucleo-base, be it purine or pyrimidine.

The term "relative complexity reduction" as used herein relates to probe length, namely the increase in average probe length that is required to achieve the same specificity and level of hybridization of a probe to a specific locus, under a given set of molecular conditions in two genomes of the same size, where the first genome is "as is" and consists of the four bases, G, A T and C, whereas the second genome is of exactly the same length but some cytosines, (ideally all cytosines), have been converted to thymines. The locus under test is in the same location in the original unconverted as well as the converted genome. On average, an 11-mer probe will have a unique location to which it will hybridize perfectly in a regular genome of 4,194,304 bases consisting of the four bases G, A, T and C, (4^{11} equals 4,194,304). However, once such a regular genome of 4,194, 304 bases has been converted by bisulfite or other suitable means, this converted genome is now composed of only three bases and is clearly less complex. However the consequence of this *decrease* in genomic complexity is that our previously unique 11-mer probe no longer has a unique site to which it can hybridize within the simplified genome. There are now many other possible equivalent locations of 11 base sequences that have arisen *de novo* as a consequence of the bisulfite conversion. It will now require a 14-mer probe to find and hybridize to the original locus. Although it may initially appear counter intuitive, one thus requires an *increased* probe length to detect the original location in what is now a simplified three base genome, because more of the genome looks the same, (it has more similar sequences). Thus the reduced *relative genomic complexity*, (or simplicity of the three base genome), means that one has to design *longer* probes to find the original unique site.

The term "relative genomic complexity reduction" as used herein can be measured by increased probe lengths capable of being microbe-specific as compared with unmodified DNA. This term also incorporates the type of probe sequences that are used in determining the presence of a microorganism. These probes may have non-conventional backbones, such as those of PNA or LNA or modified additions to a backbone such as those described in INA. Thus, a genome is considered to have reduced *relative complexity*, irrespective of whether the probe has additional components such as Intercalating pseudonucleotides, such as in INA. Examples include, but not limited to, DNA, RNA, locked nucleic acid (LNA), peptide nucleic acid (PNA), MNA, altritol nucleic acid (ANA), hexitol nucleic acid (HNA), intercalating nucleic acid (INA), cyclohexanyl nucleic acid (CNA) and mixtures thereof and hybrids thereof, as well as phosphorous atom modifications thereof, such as but not limited to phosphorothioates, methyl phospholates, phosphoramidites, phosphorodithiates, phosphoroselenoates, phosphotriesters and phosphoboranoates. Non-naturally occurring nucleotides include,

but not limited to the nucleotides comprised within DNA, RNA, PNA, INA, HNA, MNA, ANA, LNA, CNA, CeNA, TNA, (2'-NH)-TNA, (3'-NH)-TNA, α-L-Ribo-LNA, α-L-Xylo-LNA, β-D-Xylo-LNA, α-D-Ribo-LNA, [3.2.1]-LNA, Bicyclo-DNA, 6-Amino-Bicyclo-DNA, 5-epi-Bicyclo-DNA, α-Bicyclo-DNA, Tricyclo-DNA, Bicyclo[4.3.0]-DNA, Bicyclo[3.2.1]-DNA, Bicyclo[4.3.0]amide-DNA, β-D-Ribopyranosyl-NA, α-L-Lyxopyranosyl-NA, 2'-R-RNA, α-L-RNA or α-D-RNA, β-D-RNA. In addition non-phosphorous containing compounds may be used for linking to nucleotides such as but not limited to methyliminomethyl, formacetate, thioformacetate and linking groups comprising amides. In particular nucleic acids and nucleic acid analogues may comprise one or more intercalator pseudonucleotides (IPN). The presence of IPN is not part of the complexity description for nucleic acid molecules, nor is the backbone part of that complexity, such as in PNA.

By 'INA' is meant an intercalating nucleic acid in accordance with the teaching of WO 03/051901, WO 03/052132, WO 03/052133 and WO 03/052134 (Unest A/S) incorporated herein by reference. An INA is an oligonucleotide or oligonucleotide analogue comprising one or more intercalator pseudonucleotide (IPN) molecules.

By 'HNA' is meant nucleic acids as for example described by Van Aetschot et al., 1995.

By 'MNA' is meant nucleic acids as described by Hossain et al, 1998.

'ANA' refers to nucleic acids described by Allert et al, 1999.

'LNA' may be any LNA molecule as described in WO 99/14226 (Exiqon), preferably, LNA is selected from the molecules depicted in the abstract of WO 99/14226. More preferably, LNA is a nucleic acid as described in Singh et al, 1998, Koshkin et al, 1998 or Obika et al., 1997.

'PNA' refers to peptide nucleic acids as for example described by Nielsen et al, 1991.

'Relative complexity reduction' as used herein, does not refer to the order in which bases occur, such as any mathematical complexity difference between a sequence that is ATATATATATATAT (SEQ ID NO: 1) versus one of the same length that is AAAAAAATTTTTTT (SEQ ID NO: 2), nor does it refer to the original re-association data of relative genome sizes, (and inferentially, genomic complexities), introduced into the scientific literature by Waring, M. & Britten R. J.1966, Science, 154, 791-794; and Britten, R.J and Kohne D E., 1968, Science, 161, 529-540, and earlier references therein that stem from the Carnegie Institution of Washington Yearbook reports.

'Relative genomic complexity' as used herein refers to an unchanged position of bases in two genomes that is accessed by molecular probes (both the original and unconverted genomes have bases at invariant positions 1 to n. In the case of the 3 billion base pair haploid human genome of a particular human female, the invariant positions are defined as being from 1 to n, where n is 3,000,000,000. If in the sequence 1 to n, the i^{th} base is a C in the original genome, then the i^{th} base is a T in the converted genome.

The term "genomic nucleic acid" as used herein includes microbial (prokaryote and single celled eukaryote) RNA, DNA, protein encoding nucleic acid, non-protein encoding nucleic acid, and ribosomal gene regions of prokaryotes and single celled eukaryotic microorganisms.

The term "microbial genome" as used herein covers chromosomal as well as extrachromosomal nucleic acids, as well as temporary residents of that genome, such a plasmids, bacteriphage and mobile elements in the broadest sense. The "genome" has a core component as exemplified by *S. galactiae*, as well as possibly having coding and non-coding elements that vary between different isolates.

The term "microbial derived DNA" as used herein includes DNA obtained directly from a microorganism or obtained indirectly by converting microbial RNA to DNA by any of the known or suitable method such as reverse transcriptase.

The term "microorganism" as used herein includes phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, single cell organism, or any other microorganism, no matter how variously classified, such as the Kingdom *Protoctista* by Margulis, L., *et al* 1990, Handbook of Protoctista, Jones and Bartlett, Publishers, Boston USA, or microorganisms that are associated with humans, as defined in Harrisons Principles of Internal Medicine, 12th Edition, edited by J D Wilson et al., McGraw Hill Inc, as well as later editions. It also includes all microorganisms described in association with human conditions defined in OMIM, Online Mendelian Inheritance in Man, www.ncbi.gov.

The term "microbial-specific nucleic acid molecule" as used herein means a molecule which has been determined or obtained using the method according to the present invention which has one or more sequences specific to a microorganism.

The term "taxonomic level of the microorganism" as used herein includes family, genus, species, strain, type, or different populations from the same or different geographic or benthic populations. While in the case of bacteria the generally recognized schema, such as; *Bacteria, Proteobacteria; Betaproteobacteria; Neisseriales;*

Neisseriaceae; Neisseria is used. Different populations may be polymorphic for single nucleotide changes or variation that exists in DNA molecules that exist in an intracellular form within a microorganism (plasmids or phagemids), or polymorphic chromosomal regions of microorganism genomes such as pathogenicity islands. The fluidity of microbial and viral genomes is recognized, and includes the chimeric nature of viral genomes, which can be in independent nucleic acid pieces. Hence, newly arising strains from re-assortment of genomic regions from different animals .e.g., new human influenza strains as chimeras of segments that are picked up from other mammalian or avian viral genomes.

The term "close sequence similarity" as used herein includes the above definition of relative sequence complexity and probe lengths as a measure.

MATERIALS and METHODS

Extraction of DNA

In general, microbial DNA (or viral RNA) can be obtained from any suitable source. Examples include, but not limited to, cell cultures, broth cultures, environmental samples, clinical samples, bodily fluids, liquid samples, solid samples such as tissue. Microbial DNA from samples can be obtained by standard procedures. An example of a suitable extraction is as follows. The sample of interest is placed in 400 µl of 7 M Guanidinium hydrochloride, 5 mM EDTA, 100 mMTris/HCl pH6.4, 1% Triton-X-100, 50 mM Proteinase K (Sigma), 100 µg/ml yeast tRNA. The sample is thoroughly homogenised with disposable 1.5 ml pestle and left for 48 hours at 60°C. After incubation the sample is subjected to five freeze/thaw cycles of dry ice for 5 minutes/95°C for 5 minutes. The sample is then vortexed and spun in a microfuge for 2 minutes to pellet the cell debris. The supernatant is removed into a clean tube, diluted to reduce the salt concentration then phenol:chloroform extracted, ethanol precipitated and resuspended in 50 µl of 10 mM Tris/0.1 mM EDTA.

Specifically, the DNA extractions from Gram positive and Gram negative bacteria grown on standard agar plates (with nutritional requirements specific to each species) were performed as follows.

For DNA extraction from Gram Negative bacteria the protocol was as follows:

a) Using a sterile toothpick bacterial colonies were scraped off the culture plate into a sterile 1.5 ml centrifuge tube.

b) 180 µl of Guanidinium thiocyanate extraction buffer (7M Guanidinium thiocyanate, 5 mM EDTA (pH8.0), 40 mM Tris/Hcl pH 7.6, 1% Triton-X-100) was added and the sample mixed to resuspend the bacterial colonies.

c) 20 µl (20 mg/ml) Proteinase K was added and the samples were mixed well.

d) Samples were incubated @ 55°C for 3 hours to lyse the cells.

e) 200 µl of water was added to each sample and samples mixed by gentle pipetting.

f) 400 µl of Phenol/Chloroform/iso-amyl alcohol (25:24:1) was added and the samples vortexed for 2 X 15 seconds.

g) The samples were then spun in a microfuge at 14,000 rpm for 4 minutes.

h) The aqueous phase was removed into a clean 1.5 ml centrifuge tube.

i) 400 µl of Phenol/Chloroform/iso-amyl alcohol (25:24:1) was added and the samples vortexed for 2 X 15 seconds.

j) The samples were then spun in a microfuge at 14,000rpm for 4 minutes.

k) The aqueous phase was removed into a clean 1.5 ml centrifuge tube.

l) 800 µl of 100% ethanol was added to each sample, the sample vortexed briefly then left at -20°c for 1 hour.

m) The samples were spun in a microfuge at 14,000 rpm for 4 minutes at 4°C.

n) The DNA pellets were washed with 500 µl of 70% ethanol.

o) The samples were spun in a microfuge at 14,000rpm for 5 minutes at 4°C, the ethanol was discarded and the pellets were air dried for 5 minutes.

p) Finally the DNA was resuspended in 100 µl of 10 mM Tris/HCl pH 8.0, 1 mM EDTA pH 8.0.

q) The DNA concentration and purity were calculated by measuring the absorbance of the solution at 230, 260, 280nm.

For DNA extraction from Gram Positive bacteria the protocol was as follows:.

a) Using a sterile toothpick bacterial colonies were scraped off the culture plate into a sterile 1.5 ml centrifuge tube.

b) 180 µl of 20 mg/ml Lysozyme (Sigma) and 200 µg of Lysostaphin (Sigma) was added to each sample and the samples were mixed gently to resuspend the bacterial colonies.

c) The samples were incubated at 37°C for 30 minutes to degrade the cell wall.

d) The samples were then processed and the DNA extracted according to the QIAamp DNA mini kit protocol for Gram positive bacteria.

DNA extraction from Cytology samples from patients.

a) The sample was shaken vigorously by hand to resuspend any sedimented cells and to ensure the homogeneity of the solution.
b) 4 ml of the resuspended cells were transferred to a 15 ml Costar centrifuge tube.
c) The tubes were centrifuged in a swing-out bucket rotor at 3000 x g for 15 minutes.
d) The supernatant was carefully decanted and discarded without disturbing the pelleted cellular material.
e) The pelleted cells were resuspended in 200 µl of lysis buffer (100 mM Tris/HCl pH 8.0, 2 mM EDTA pH 8.0, 0.5% SDS, 0.5% Triton-X-100) and mixed well until the solution was homogeneous.
f) 80 µl of the sample was transferred to a 96 well sample preparation plate .
g) 20 µl of Proteinase K was added and the solution incubated at 55°C for 1 hour (this procedure results in cell lysis)

DNA extraction from urine samples

DNA was extracted from a starting volume of 1 ml of urine according to the QIAamp UltraSens™ Virus Handbook.

Bisulfite treatment of DNA samples

Bisulfite treatment was carried out according the MethylEasy™ High Throughput DNA bisulfite modification kit (Human Genetic Signatures, Australia) see also below..

Surprisingly, it has been found by the present inventors that there is no need to separate the microbial DNA from other sources of nucleic acids, for example when there is microbial DNA in a sample of human cells. The treatment step can be used for an vast mixture of different DNA types and yet a microbial-specific nucleic acid can be still identified by the present invention. It is estimated that the limits of detection in a complex DNA mixtures are that of the limits of standard PCR detection which can be down to a single copy of a target nucleic acid molecule.

Samples

Any suitable sample can be used for the present invention. Examples include, but not limited to, microbial cultures, clinical samples, veterinary samples, biological fluids, tissue culture samples, environmental samples, water samples, effluent. As the

present invention is adaptable for detecting any microorganism, this list should not be considered as exhaustive.

Kits

The present invention can be implemented in the form of various kits, or combination of kits and instantiated in terms of manual, semi automated or fully robotic platforms. In a preferred form, the MethyEasy™ or HighThroughput MethylEasy™ kits (Human Genetic Signatures Pty Ltd, Australia) allow conversion of nucleic acids in 96 or 384 plates using a robotic platform such as EpMotion.

Bisulfite treatment

An exemplary protocol for effective bisulfite treatment of nucleic acid is set out below. The protocol results in retaining substantially all DNA treated. This method is also referred to herein as the Human Genetic Signatures (HGS) method. It will be appreciated that the volumes or amounts of sample or reagents can be varied.

Preferred method for bisulfite treatment can be found in US 10/428310 or PCT/AU2004/000549 incorporated herein by reference.

To 2 µg of DNA, which can be pre-digested with suitable restriction enzymes if so desired, 2 µl (1/10 volume) of 3 M NaOH (6g in 50 ml water, freshly made) was added in a final volume of 20 µl. This step denatures the double stranded DNA molecules into a single stranded form, since the bisulfite reagent preferably reacts with single stranded molecules. The mixture was incubated at 37°C for 15 minutes. Incubation at temperatures above room temperature can be used to improve the efficiency of denaturation.

After the incubation, 208 µl 2 M Sodium Metabisulfite (7.6 g in 20 ml water with 416 ml 10 N NaOH; BDH AnalaR #10356.4D; freshly made) and 12 µl of 10 mM Quinol (0.055 g in 50 ml water, BDH AnalR #103122E; freshly made) were added in succession. Quinol is a reducing agent and helps to reduce oxidation of the reagents. Other reducing agents can also be used, for example, dithiothreitol (DTT), mercaptoethanol, quinone (hydroquinone), or other suitable reducing agents. The sample was overlaid with 200 µl of mineral oil. The overlaying of mineral oil prevents evaporation and oxidation of the reagents but is not essential. The sample was then incubated overnight at 55°C. Alternatively the samples can be cycled in a thermal cycler as follows: incubate for about

4 hours or overnight as follows: Step 1, 55°C / 2 hr cycled in PCR machine; Step 2, 95°C / 2 min. Step 1 can be performed at any temperature from about 37°C to about 90°C and can vary in length from 5 minutes to 8 hours. Step 2 can be performed at any temperature from about 70°C to about 99°C and can vary in length from about 1 second to 60 minutes, or longer.

After the treatment with Sodium Metabisulfite, the oil was removed, and 1 µl tRNA (20 mg/ml) or 2 µl glycogen were added if the DNA concentration was low. These additives are optional and can be used to improve the yield of DNA obtained by co-precipitating with the target DNA especially when the DNA is present at low concentrations. The use of additives as carrier for more efficient precipitation of nucleic acids is generally desired when the amount nucleic acid is <0.5 µg.

An isopropanol cleanup treatment was performed as follows: 800 µl of water were added to the sample, mixed and then 1 ml isopropanol was added. The water or buffer reduces the concentration of the bisulfite salt in the reaction vessel to a level at which the salt will not precipitate along with the target nucleic acid of interest. The dilution is generally about 1/4 to 1/1000 so long as the salt concentration is diluted below a desired range, as disclosed herein.

The sample was mixed again and left at 4°C for a minimum of 5 minutes. The sample was spun in a microfuge for 10-15 minutes and the pellet was washed 2x with 70% ETOH, vortexing each time. This washing treatment removes any residual salts that precipitated with the nucleic acids.

The pellet was allowed to dry and then resuspended in a suitable volume of T/E (10 mM Tris/0.1 mM EDTA) pH 7.0-12.5 such as 50 µl. Buffer at pH 10.5 has been found to be particularly effective. The sample was incubated at 37°C to 95°C for 1 min to 96 hr, as needed to suspend the nucleic acids.

Another example of bisulfite treatment can be found in WO 2005021778 (incorporated herein by reference) which provides methods and materials for conversion of cytosine to uracil. In some embodiments, a nucleic acid, such as gDNA, is reacted with bisulfite and a polyamine catalyst, such as a triamine or tetra-amine. Optionally, the bisulfite comprises magnesium bisulfite. In other embodiments, a nucleic acid is reacted with magnesium bisulfite, optionally in the presence of a polyamine catalyst and/or a quaternary amine catalyst. Also provided are kits that can be used to carry out methods of the invention. It will be appreciated that these methods would also be suitable for the present invention in the treating step.

Amplification

PCR amplifications were performed in 25 µl reaction mixtures containing 2 µl of bisulfite-treated genomic DNA, using the Promega PCR master mix, 6 ng/µl of each of the primers. Strand-specific nested primers are used for amplification. 1st round PCR amplifications were carried out using PCR primers 1 and 4 (see below). Following 1st round amplification, 1µl of the amplified material was transferred to 2nd round PCR premixes containing PCR primers 2 and 3 and amplified as previously described. Samples of PCR products were amplified in a ThermoHybaid PX2 thermal cycler under the conditions: 1 cycle of 95°C for 4 minutes, followed by 30 cycles of 95°C for 1 minute, 50°C for 2 minutes and 72°C for 2 minutes; 1 cycle of 72°C for 10 minutes.



Multiplex amplification

If multiplex amplification is required for detection, the following methodology can be carried out.

One µl of bisulfite treated DNA is added to the following components in a 25 µl reaction volume, x1 Qiagen multiplex master mix, 5-100 ng of each 1st round INA or oligonucleotide primer 1.5- 4.0 mM $MgSO_4$, 400 uM of each dNTP and 0.5-2 unit of the polymerase mixture. The components are then cycled in a hot lid thermal cycler as follows. Typically there can be up to 200 individual primer sequences in each amplification reaction

Step 1	94°C	15 minute	1 cycle
Step 2	94°C	1 minute	
	50°C	3 minutes	35 cycles
	68°C	3 minutes	
Step 3	68°C	10 minutes	1 cycle

A second round amplification is then performed on a 1 µl aliquot of the first round amplification that is transferred to a second round reaction tube containing the enzyme reaction mix and appropriate second round primers. Cycling is then performed as above.

Primers

Any suitable PCR primers can be used for the present invention. A primer typically has a complementary sequence to a sequence which will be amplified. Primers are typically oligonucleotides but can be oligonucleotide analogues.

Probes

The probe may be any suitable nucleic acid molecule or nucleic acid analogue. Examples include, but not limited to, DNA, RNA, locked nucleic acid (LNA), peptide nucleic acid (PNA), MNA, altritol nucleic acid (ANA), hexitol nucleic acid (HNA), intercalating nucleic acid (INA), cyclohexanyl nucleic acid (CNA) and mixtures thereof and hybrids thereof, as well as phosphorous atom modifications thereof, such as but not limited to phosphorothioates, methyl phospholates, phosphoramidites, phosphorodithiates, phosphoroselenoates, phosphotriesters and phosphoboranoates. Non-naturally occurring nucleotides include, but not limited to the nucleotides comprised within DNA, RNA, PNA, INA, HNA, MNA, ANA, LNA, CNA, CeNA, TNA, (2'-NH)-TNA, (3'-NH)-TNA, α-L-Ribo-LNA, α-L-Xylo-LNA, β-D-Xylo-LNA, α-D-Ribo-LNA, [3.2.1]-LNA, Bicyclo-DNA, 6-Amino-Bicyclo-DNA, 5-epi-Bicyclo-DNA, α-Bicyclo-DNA, Tricyclo-DNA, Bicyclo[4.3.0]-DNA, Bicyclo[3.2.1]-DNA, Bicyclo[4.3.0]amide-DNA, β-D-Ribopyranosyl-NA, α-L-Lyxopyranosyl-NA, 2'-R-RNA, α-L-RNA or α-D-RNA, β-D-RNA. In addition non-phosphorous containing compounds may be used for linking to nucleotides such as but not limited to methyliminomethyl, formacetate, thioformacetate and linking groups comprising amides. In particular nucleic acids and nucleic acid analogues may comprise one or more intercalator pseudonucleotides.

Preferably, the probes are DNA or DNA oligonucleotides containing one or more internal IPNs forming INA.

Electrophoresis

Electrophoresis of samples was performed according to the E-gel system user guide (www.invitrogen.doc).

Detection methods

Numerous possible detection systems exist to determine the status of the desired sample. It will be appreciated that any known system or method for detecting nucleic acid molecules could be used for the present invention. Detection systems include, but not limited to:

I. Hybridization of appropriately labelled DNA to a micro-array type device which could select for 10->200,000 individual components. The arrays could be composed of either INAs, PNAs or nucleotide or modified nucleotides arrays onto any suitable solid surface such as glass, plastic, mica, nylon , bead, magnetic bead, fluorescent bead or membrane;

II. Southern blot type detection systems;

III. Standard PCR detection systems such as agarose gel, fluorescent read outs such as Genescan analysis. Sandwich hybridisation assays, DNA staining reagents such as ethidium bromide, Syber green, antibody detection, ELISA plate reader type devices, fluorimeter devices;

IV. Real-Time PCR quantitation of specific or multiple genomic amplified fragments or any variation on that.

V. Any of the detection systems outlined in the WO 2004/065625 such as fluorescent beads, enzyme conjugates, radioactive beads and the like;

VI. Any other detection system utilizing an amplification step such as ligase chain reaction or Isothermal DNA amplification technologies such as Strand Displacement Amplification (SDA).

VII. Multi-photon detection systems.

VIII. Electrophoresis and visualisation in gels.

IX. Any detection platform used or could be used to detect nucleic acid.

Intercalating nucleic acids

Intercalating nucleic acids (INA) are non-naturally occurring polynucleotides which can hybridize to nucleic acids (DNA and RNA) with sequence specificity. INA are candidates as alternatives/substitutes to nucleic acid probes in probe-based hybridization assays because they exhibit several desirable properties. INA are polymers which hybridize to nucleic acids to form hybrids which are more thermodynamically stable than a corresponding naturally occurring nucleic acid/nucleic acid complex. They are not substrates for the enzymes which are known to degrade peptides or nucleic acids. Therefore, INA should be more stable in biological samples, as well as, have a longer

shelf-life than naturally occurring nucleic acid fragments. Unlike nucleic acid hybridization which is very dependent on ionic strength, the hybridization of an INA with a nucleic acid is fairly independent of ionic strength and is favoured at low ionic strength under conditions which strongly disfavour the hybridization of naturally occurring nucleic acid to nucleic acid. The binding strength of INA is dependent on the number of intercalating groups engineered into the molecule as well as the usual interactions from hydrogen bonding between bases stacked in a specific fashion in a double stranded structure. Sequence discrimination is more efficient for INA recognizing DNA than for DNA recognizing DNA.

Preferably, the INA is the phosphoramidite of (S)-1-O-(4,4'-dimethoxytriphenylmethyl)-3-O-(1-pyrenylmethyl)-glycerol.

INA are synthesized by adaptation of standard oligonucleotide synthesis procedures in a format which is commercially available. Full definition of INA and their synthesis can be found in WO 03/051901, WO 03/052132, WO 03/052133 and WO 03/052134 (Unest A/S) incorporated herein by reference.

There are indeed many differences between INA probes and standard nucleic acid probes. These differences can be conveniently broken down into biological, structural, and physico-chemical differences. As discussed above and below, these biological, structural, and physico-chemical differences may lead to unpredictable results when attempting to use INA probes in applications were nucleic acids have typically been employed. This non-equivalency of differing compositions is often observed in the chemical arts.

With regard to biological differences, nucleic acids are biological materials that play a central role in the life of living species as agents of genetic transmission and expression. Their *in vivo* properties are fairly well understood. INA, however, is a recently developed totally artificial molecule, conceived in the minds of chemists and made using synthetic organic chemistry. It has no known biological function.

Structurally, INA also differs dramatically from nucleic acids. Although both can employ common nucleobases (A, C, G, T, and U), the composition of these molecules is structurally diverse. The backbones of RNA, DNA and INA are composed of repeating phosphodiester ribose and 2-deoxyribose units. INA differ from DNA or RNA in having one or more large flat molecules attached via a linker molecule(s) to the polymer. The flat molecules intercalate between bases in the complementary DNA stand opposite the INA in a double stranded structure.

The physico/chemical differences between INA and DNA or RNA are also substantial. INA binds to complementary DNA more rapidly than nucleic acid probes bind to the same target sequence. Unlike DNA or RNA fragments, INA bind poorly to RNA unless the intercalating groups are located in terminal positions. Because of the strong interactions between the intercalating groups and bases on the complementary DNA strand, the stability of the INA/DNA complex is higher than that of an analogous DNA/DNA or RNA/DNA complex.

Unlike other nucleic acids such as DNA or RNA fragments or PNA, INA do not exhibit self aggregation or binding properties.

As INA hybridize to nucleic acids with sequence specificity, INA are useful candidates for developing probe-based assays and are particularly adapted for kits and screening assays. INA probes, however, are not the equivalent of nucleic acid probes. Consequently, any method, kits or compositions which could improve the specificity, sensitivity and reliability of probe-based assays would be useful in the detection, analysis and quantitation of DNA containing samples. INA have the necessary properties for this purpose.

RESULTS

The detection of microorganisms (such as bacterial, viral or fungal strains) is often hampered by the large number of individual strains of microorganism within that species.

The general in silico principles of the invention are taught using the bacteria *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Streptococcus sp* and *Staphylococcus* (Figures 1 to 5). The general principles of the invention have been taught using the Influenza virus and Rotavirus (Figures 6 and 7).

The general biochemical data for teaching and supporting the invention is described in Figures 8 to 18 using clinically relevant Gram positive as well as Gram negative bacteria.

Bacteria

Figure 1 shows a 34 nucleotide region of the *iga* protease gene in *N. meningitides* and the corresponding locus in *N. gonorrhoeae* (as these regions exist in their natural bacterial genomes) (full classification; *Bacteria; Proteobacteria; Betaproteobacteria;*

Neisseriales; Neisseriaceae; Neisseria meningitides, Z2491 Serogroup A and full locus characteristics; *iga*, IgA1 protease; GeneID 906889. Locus Tag NMA0905; RefSeq accession # NC_003116.1; PMID 10761919; Parkhill J et al., 2000, *Nature*, 404, 502-506). There is 74% sequence similarity between these two *Neisseria* 34 nucleotide sequences. PCR-based primers made to amplify these regions in both bacterial species would require degenerate primers with 512 possible combinations. The common sequence used for part of the PCR amplification would be the 34 nucleotide sequence GYAATYW AGGYCGYCTY GAAGAYTAYA AYATGGC (SEQ ID NO: 3) where the standard code for designating different positions is given below; N = A, G, T or C; D = A, G or T; H = A, T or C; B = G, T or C; V = G, A or C; K = G or T; S = C or G; Y = T or C; R = A or G; M = A or C and W = A or T.

However, when the bacterial DNA from these two species is treated with the bisulfite reagent, (resulting in the conversion of cytosines to thymines), the naturally occurring sequences are converted to derivative sequences that have no coding potential and do not exist in nature. The derivative sequences are now 97% sequence similar. PCR-based primers designed to allow PCR amplification of both these bacterial loci in a single test now only require only 2 primer combinations. The combination would be based on the sequence GTAATTW AGGTTGTTTT GAAGATTATA ATATGGT (SEQ ID NO: 4), where only the base at position 7 is either an adenine or a thymine (denoted **W**). Thus, the bisulfite conversion reduces the relative genomic complexity from 512 to 2 primer types. This massive reduction simplifies the amplification of the same locus from related bacterial species.

Further advantages accrue from optionally using INA probes to amplify regions from these two bacterial species, again using the same locus. Figure 2 illustrates the same 34 nucleotide region of the *iga* genes of *N. meningitides* and *N. gonorrhoeae* as depicted in Figure 1, with the added demonstration of the extent to which probe length and complexity can be reduced even further using INA probes. A short INA 16 mer sequence AGGYCGYCTY GAAGAY (SEQ ID NO: 5) would require 16 possible primer combinations to detect this region, but after conversion with bisulfite, a unique primer sequence, AGGTTGTTTT GAAGAT (SEQ ID NO: 6) would be sufficient. The advantage of the INA molecule is that; owing to the intercalating pseudonucleotides that are incorporated into its backbone, hybridization to the correct locus is much more easily distinguished from non specific binding, owing to the increased Tm of the INA relative to a standard oligonucleotide. It will be appreciated, however, that standard oligonucleotides will still perform adequately.

When closely related bacterial species cause similar clinical symptoms, bisulfite converted DNA can again be used to design simpler probes to assay for presence of specific bacterial types. Figure 3 shows the DNA alignments of the *iga* gene in three bacterial species, one of which, *Haemophilus influenzae* is from a different taxonomic group. Bisulfite treatment of the bacterial DNA resulted in a much smaller number of probe combinations. This comparison illustrates the importance of being able to assay for unrelated species in one test. Both *N. meningitides* and *H. influenzae* cause meningitis, so it is advantageous to be able to assay in the one test for all microbes that cause the same clinical symptoms.

The analysis of a large number of different bacterial species from the same taxonomic group is again facilitated by the present invention. Figure 4 shows a 40 nucleotide segment of the *tuf* gene in 10 bacterial species of the *Streptococcus* group namely *S. oralis, S. mitis, S. dysgalactiae, S. cristatus, S. gordonii, S. parauberis, S. pneumoniae, S. bovis, S. vestivularis and S. uberis.* This region has approximately 68% sequence similarity between the 10 species and requires 12,288 primer combinations in order to simultaneously assay for the 10 species in the one test. The bisulfite converted sequence between these species has 85% sequence similarity and now only requires 64 possible primer combinations.

The analysis of different strains belonging to the same bacterial species is also simplified by the invention. Figure 5 illustrates a 23 nucleotide segment of the *Staphylococcal aureus* enterotoxin gene *se*. The natural sequence of this gene region has only 56% sequence similarity between all 7 strains and requires 1536 primer combinations, whereas the bisulfite converted sequence has 74% sequence similarity and requires only 64 primer combinations.

Viral nucleic acid analyses and relative genomic complexity reduction

The principle of relative genomic complexity reduction can also be applied to viral groups, such as Influenza virus which has a DNA genome, as well as to viral groups which have RNA genomes, (as the RNA can be converted to DNA by reverse transcriptase and then bisulfite treated accordingly). To illustrate application for viral detection, the neuraminidase gene of strains of influenza virus, (Family Orthomyxoviridae), and the surface protein encoding *VP4* gene of rotavirus strains, (Family Reoviridae), both viruses having a segmented RNA genome, have been used. The taxonomy of influenza viruses is complex, with types A, B and C for example being based on antigenic characteristics, and with further subtypes being based on site of

origin, year of isolation, isolate number and subtype. This reinforces the need in the first instance to be able to identify influenza viruses as a group, and only then to drill down to analyse sub-sub-classification levels.

The taxonomy of rotaviruses is also complex. The number of rotavirus serotypes is large with two main serotypes being recognized, the P and G serotypes. There are minimally 14 different G serotypes and their unambiguous detection is of importance in paediatric medicine. It is estimated that by the age of three, nearly every child worldwide has already been infected at least once by Rotavirus, even though these infections may be subclinical and have only mild effects on the gastrointestinal tract.

The consequences of infection by influenza at the clinical level are well known, with significant morbidity and mortality nearly every winter. However there can be massive secondary complications following infection, especially by *Streptococcus pneumoniae, Hemophilus influenzae* and *Staphylococcus aureus*. It is very clearly advantageous to be able to simultaneously analyse for both viral infections and bacterial infections since pneumonial complications can arise from mixed features of bacterial and viral infections, and prompt antibiotic treatment can be an effective therapy.

The relative genomic complexity reduction in 9 different influenza strains is shown in Figure 6. A 20 nucleotide region of the neuraminidase gene of influenza virus is shown in its DNA form. There is 50% sequence similarity between these 9 isolates. After bisulfite conversion, the sequence similarity has increased to 75%. In its original form it would require 2048 possible primer combinations to analyse these 9 strains, whereas after bisulfite conversion only 48 primer combinations are needed.

The relative genomic complexity reduction in the VP4 gene of 3 different rotavirus strains is shown in Figure 7. A 20 nucleotide region of the VP4 gene has 52% sequence similarity before conversion and 74% after conversion. The number of primer combinations reduces from 512 to 32.

The molecular data supporting the in silico approach of simplifying microbial genomes as a means of detecting microorganisms is illustrated in Figures 8 through 15 using clinically relevant microbial species that are commonly encountered in hospital and pathology testing units.

It is a distinct advantage, and a clinical imperative for the rapid detection of contaminating microorganisms, if the initial decision could be made between the presence of Gram positive or Gram negative bacteria in a sample. The method described herein provides such a test using the 23S ribosomal genes of different

bacterial species to generate a set of primers that allow either Gram positive or Gram negative bacteria to be detected by utilising such primers on simplified genomes via an amplification reaction. The 23S sequences are ideal for such high level distinctions, since they occur in all bacterial species, unlike some protein coding sequences which are optional additions to some bacterial genomes, such as seen in the previous *S. galactiae* example. Many protein coding microbial sequences are akin to genomic "flotsam and jetsam", and their usefulness lies in differentiating between lower level taxonomic categories such as different microbial strains, types or isolates, or in the case of viruses, between different types or newly arisen mutations. The normal and simplified genomic sequences of both of these components, the non protein coding ribosomal RNA genes, and the protein coding recA gene of bacteria are given in Figures 15 and 16 respectively. The primer sequences used to perform the amplification reactions for the 23S bacterial amplicons are given in Table 1. The primer sequences used to perfom the amplification reactions for the recA amplicons are given in Table 2. All primers are made to bisulfite treated DNA and are shown in the 5' to 3' orientation.

Table 1 sets out suitable bacterial primers sequences used in amplifying bisulfite simplified DNA from the 23S ribosomal RNA gene(s) using alignments to generate primers for the detection of Gram positive (Pos), Gram negative (Neg). In addition primers were designed for specific detection of Mycoplasma spp (Myc), Staphylococcus spp (Staph), Streptococcus spp (Strep), Neisseria spp (NG), Chlamydia (CT), and Escherichia coli and Klebsiella pneumoniae (EC).

The following symbols designate the following base additions; N = A, G, T or C; D = A, G or T; H = A, T or C; B = G, T or C; V = G, A or C; K = G or T; S = C or G; Y = T or C; R = A or G; M = A or C and W = A or T.

All primers used were based on bisulfite simplified DNA sequences.

Table 1 Bacterial primers

23S Primers	Sequence 5'-3'	SEQ ID NO
Pos-R1F1	GGTTTTTTTTGAAATAGTTTTAGGGTTA	7
Neg-R1F1	GGTTTTTTTTGAAARTTATTTAGGTAGT	8
Pos-R1F2	TGGKAGTTAGAWTGTGRRWGATAAG	9
Neg-R1F2	TGGGAGATAKATRGTGGGTGTTAAT	10

23S Primers	Sequence 5'-3'	SEQ ID NO
Pos-R1F3	GGATGTGGDRTTKTKWAGATAA	11
Neg-R1F3	TGAWGTGGGAAGGTWTAGATAG	12
Pos-R1R1	HCAATMHHACTTCAMMMCMMYT	13
Neg-R1R1	WCAAHHCACCTTCAHMAACYTAC	14
Pos-R1R2	ACCAACATTCTCACTYMTAAWMAMTCCAC	15
Neg-R1R2	ATCAACATTCACACTTCTAATACCTCCAA	16
W-Pos-R1F1	GGTTTTTTTYGAAATAGTTTTAGGGTTA	17
W-Neg-R1F1	GGTTTTTTTYGAAARTTATTTAGGTAGT	18
W-Pos-R1F2	YGGKAGTTAGAWYGYGRRWGATAAG	19
W-Neg-R1F2	YGGGAGATAKAYRGYGGGTGTTAAT	20
W-Pos-R1F3	GGATGTGGDRTTKYKWAGATAA	21
W-Neg-R1F3	YGAWGTGGGAAGGTWTAGATAG	22
W-Pos-R1R1	HCRATMHHRCTTCRMMMCMMYT	23
W-Neg-R1R1	WCRAHHCACCTTCAHMRACYTAC	24
W-Pos-R1R2	ACCRACATTCTCACTYMTAAWMAMTCCAC	25
W-Neg-R1R2	ATCAACATTCRCACTTCTAATACCTCCAA	26
Pos-R2F1	KTTRAGAAAAGTWTTTAGDDAGRK	27
Neg-R2F1	TTTARGAAAAGTTWTTAAGTWTTA	28
Pos-R2F2	AGDTRAGRWGAGDATTTTWAGGTKR	29
Neg-R2F2	GGKTRGGWWGAGAATWTTAAGGTGT	30
Pos-R2R1	AATYTMYMATTAAAACAATACMCAA	31
Neg-R2R1	AATCTCAAAWAAAAACAAYMYMACC	32
Pos-R2R2	ACMHACATCTTCACWMAYAYTAYAAYTTCACC	33
Neg-R2R2	MAYTACATCTTCACAACMAHWTCAAYTTCACT	34
Pos-R2R3	CMATAYYAAAYTACAATAAAACTC	35
Neg-R2R3	CAATAYMAAACTAYAATAAAAATT	36

23S Primers	Sequence 5'-3'	SEQ ID NO
Pos-R3F1	GGTGAARTTRTARTRTKWGTGAAGATGTDKG	37
Neg-R3F1	AGTGAARTTGAWDTKGTTGTGAAGATGTART	38
Pos-R3F2	GATWGGATGGAAAGATTTTRTRGAG	39
Neg-R3F2	KGTWAGATGGAAAGATTTTGTGAAT	40
Pos-R3R1	HYMAYMMWAYHAAAATAATATCC	41
Neg-R3R1	TCAAMMMYWMMAAAATAATATTT	42
Pos-R3R2	AWCCATTCTAAAAAAACCTTTAAACA	43
Neg-R3R2	AACCAWWMYWAAMHMACCTTCAWACT	44
EC-F1	GTTGGTAAGGTGATATGAATTGTTATAA	45
EC-F2	TTATTATTAATTGAATTTATAGGTTA	46
EC-F3	GAGGAGTTTAGAGTTTGAATTAGTRTG	47
EC_R1	TATATACAAAACTATCACCCTATATC	48
EC-R2	TCATCAAACTCACAACAYATAC	49
NG-F1	TTGAGTAAGATATTGATGGGGGTAA	50
NG-F2	TATGGTTAGGGGGTTATTGTA	51
NG-R1	AATCTATCATTTAAAACCTTAACC	52
NG-R2	CCTAACTATCTATACCTTCCCACT	53
NG-R3	CACTCCCCTACCATACCAATAAACC	54
CT-R1F1	GTATGATGAGTTAGGGAGTTAAGTTAAA	55
CT-R1F2	GGTGAGGTTAAGGGATATATA	56
CT-R1F3	AAAAGAGTGAAGAGTTGTTTGGTTTAGATA	57
CT-R1R1	TCCAAACCTTTTTCAACATTAACT	58
CT-R1R2	CCCTAAAATTATTTCAAAAAAAACAAAA	59
CT-R2F1	TTAGTGGGGGTTTATTGGTTTATTAATGGA	60
CT-R2F2	TAAGGAAGTGATGATTTGAAGATAGTTGGA	61
CT-R2R1	ACACCTTCTCTACTAAATACT	62

23S Primers	Sequence 5'-3'	SEQ ID NO
CT-R2R2	TATACCATAAATCTTCACTAATATC	63
CT-R3F1	TTGTGTAGATGATGGAGTAGTAGGTTA	64
CT-R3F2	GAATGATGGAGTAAGTTAAGTATGTGGA	65
CT-R3R1	TAAAAATTATTTCTTAAAAACCTCACT	66
CT-R3R2	AAATTATCTCACACACCTTAAAATAT	67
CT-R4F1	AATGTTAAAAGGTTAAAGGGATAT	68
CT-R4F2	TATTGAATTTAAGTTTTGGTGAATGGTT	69
CT-R4R1	CCAATATTTCAACATTAACTCCCACTCTC	70
CT-R4R2	ATATCCATCTTCCAAATTCATAAAATAAT	71
CT-R4R3	TAAACAACAACAATTCCACTTTCC	72
Myc-R1F1	ATAGGAAAAGAAAWTGAAWGWGATTTTG	73
Myc-R1F2	GTGTAGTGGTGAGTGAAAGTGGAATAGG	74
Myc-R1R1	TAAACAAMTTCMMTCAAAATAACATTTYYCAA	75
Myc-R1R2	CTAATTAATATTTAAACTTACCC	76
Myc-R2F1	TTTTGAAATTATATGTTTATAATGT	77
Myc-R2F2	AAGTATGAGTTGGTGAGTTATGATAGT	78
Myc-R2R1	CCTCCAMTTAWTYATAATCTYAC	79
Myc-R2R2	CACCWAAAYAACACCATCATACATT	80
Myc-R3F1	TGTAGTTAGATAGTGGGGTATAAGTTTTA	81
Myc-R3F2	AGGGGAAGAGTTTAGATTATTAAA	82
Myc-R3R1	ATAACTTCAWCYCMWATACAACACTCAT	83
Myc-R3R2	ATCAATTTAAAAAATTCTCACTCYCAAA	84
Myc-R4F1	TTTTTATWATTGGATTTGGGGWTAAA	85
Myc-R4F2	TKKTWWTTAGTATTGAGAATGA	86
Myc-R4F3	TGTAAATTWATTTTGTAAGTTWGT	87
Myc-R4F4	GAATGAGGGGGGATTGTTTAATT	88

23S Primers	Sequence 5'-3'	SEQ ID NO
Myc-R4R1	TCTATAACCAAAACAATCAAAAAATA	89
Myc-R4R2	CATTACACCTAACAAATATCTTCACC	90
Myc-R5F1	ATWWATAGGTTGAATAGGTRAGAAAT	91
Myc-R5F2	ATAGTGATTTGGTGGTTTAGTATGGAAT	92
Myc-R5R1	CAAACCTACTTCAACTCAAAAATAAAATAAAT	93
Myc-R5R2	ACAACAATTTAAACCCAACTCACATATCT	94
Myc-R5R3	AAAAYAAMWCTYTTCAATCTTCCTAYAAA	95
Strep-R1F1	ATWWTTGTTAAGGDWRTGARRAGGAAG	96
Strep-R1F2	TAGRAGGGTAAATTGARGWGTTTA	97
Strep-R1F3	TKATTTGGGAARRTWRGTTAAAGAGA	98
Strep-R1R1	TCTCTTCAACTTAACCTCACATCAT	99
Strep-R1R2	ATAATTTCAAATCTACAWCMWAAT	100
Strep-R2F1	RATKTATTGGAGGATTGAATTAGGG	101
Strep-R2F2	ATGTTGAAAAGTGTTTGGATGAT	102
Strep-R2R1	TCTAAAATYAATAAWCCAAAATAAMCCCCTC	103
Strep-R2R2	ACTACCAAYHATAWHTCATTAAC	104
Strep-R3F1	AGGTTGAKATTTTTGTATTAGAGTA	105
Strep-R3F2	RWAGTGATGGAGGGATGTAGTAGGTTAAT	106
Strep-R3R1	CTTTTCTYAACAATATAACATCACT	107
Strep-R3R2	CTCTCAMTCACCTAAAACTACTCA	108
Staph-R1F1	AGAAGTTGATGAAGGATGTTATTAATGA	109
Staph-R1F2	GTTATTGATATGTGAATWTATAGTATRTT	110
Staph-R1R1	CAAAAYTHTTACCTTCTYTAATYC	111
Staph-R1R2	CAACAAAATTYCACATACTCCAT	112
Staph-R2F1	GATTTGATGTAAGGTTAAGTAGT	113
Staph-R2F2	TTGGTTAGGTTGAAGTTTAGGTAATATTGAA	114

23S Primers	Sequence 5'-3'	SEQ ID NO
Staph-R2F3	GATTTATGTTGAAAAGTGAGTGGATGAATTGA	115
Staph-R2R1	CCTYTTTCTAACTCCCAAATTAAATTAAT	116
Staph-R3F1	GAAGTTGTGGATTGTTTTTTGGATA	117
Staph-R3F2	AAGGGTGTTGAAGTATGATTGTAAGGATAT	118
Staph-R3R1	TACAMTCCAAYMACACACTTCACCTATCCTA	119
Staph-R3R2	CAACAATATAAAATCAACAACTCAAA	120
Staph-R4F1	AGGAGTGGTTAGTTTTTGTGAAGTTA	121
Staph-R4F1	ACAAATTAAAAAWCCAACACAACT	122
Staph-R4F2	TAACACTATCTCCCACCAYAATMAAT	123

Table 2 sets out bacterial primer sequences used in amplifying simplified DNA from the recA protein coding gene using alignments from *Staphylococcus aureus* (SA), *Staphylococcus epidermidis* (SE), *Serratia marscesens* (SM), *Escherichia coli* (EC) and Yersinia enterocolitica (YE) for unique bacterial typing.

Table 2 Bacterial primer sequences used in amplifying simplified DNA from the recA protein coding gene

RecA Specific	Sequence	SEQ ID NO
A-SA-F1	TAGGTTGTTGAGTTTTAATTATA	124
A-SA-F2	GAAGTATAAAGTAATGGTGGGGTG	125
A-SA-R1	TACAATATCAACTACACCACTTCTAACAAAT	126
A-SA-R2	TAATAAAAATAACAATTATATTT	127
A-SE-F1	AAGGTTGTAGAGTATTAAGTATTTTAAG	128
A-SE-F2	GTTGATAATGTATTAGGGGTTGGA	129
A-SE-F3	ATATGGATTTGAAAGTTTAGGTAAGATG	130
A-SE-R1	TACTACTAAATCAACAACAACAATATCCACA	131
A-SE-R2	CTTAATACTTAAAACATTAATCT	132

RecA Specific	Sequence	SEQ ID NO
A-SM-F1	GAGAATAAGTAAAAGGTGTTAGTTGTG	133
A-SM-F2	GATTTTTATTGGTTTATTGTTATTTGATATTGTT	134
A-SM-R1	CAAATAATCAATATCAACACCCAACTTTTTC	135
A-SM-R2	TACACACCACCAAACCCATATAC	136
A-EC-F1	GAAAATAAATAGAAAGTGTTGGTG	137
A-EC-F2	TGTTTTTATTGGATATTGTGTTT	138
A-EC-R1	CAATAACATCTACTACACCAAAACAC	139
A-EC-R2	CATATTAAACTACTTCAAATTACCC	140
A-YE-F1	TATGTGTTTTGGTGAAGATTGTTTA	141
A-YE-F2	TTTTGATATTGTATTGGGGGTG	142
A-YE-F3	GGTTTGTTAATGGGGTGTATTGTTGAG	143
A-YE-R1	CATACTCTACATCAATAAAA	144

Table 1 shows the bacterial primer sequences used in amplifying bisulfite simplified DNA from the 23S ribosomal RNA gene(s) using multiple alignments to generate optimal primers for the detection of Gram positive (denoted Pos), and Gram negative (denoted Neg), bacteria. In addition primers were also designed for specific detection of groups of species as well as for individual species. The designations for these bacterial primer groups are as follows; *Escherichia coli* and *Klebsiella pneumoniae* (EC), Neisseria spp (NG), Chlamydia (CT), Mycoplasma spp (Myc), Streptococcus spp (Strep) and Staphylococcus spp (Staph). The F and R sub designations refer to forward and reverse primers respectively. In addition, where more than one possible base is necessary at a given nucleotide position, the base degeneracy is given by the following code; N = A, G, T or C; D = A, G or T; H = A, T or C; B = G, T or C; V = G, A or C; K = G or T; S = C or G; Y = T or C; R = A or G; M = A or C; and W = A or T. To reiterate, all primers used in this invention are based on bisulfite simplified DNA sequences.

Table 2 shows bacterial primers sequences used in amplifying bisulfite simplified DNA from the recA protein coding gene using alignments from *Staphylococcus aureus* (SA), *Staphylococcus epidermidis* (SE), *Serratia marscesens* (SM). *Escherichia coli* (EC) and *Yersinia enterocolitica* (YE) for unique bacterial typing.

Figure 8 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions of Gram positive and Gram negative bacteria, with appropriately sized amplicons being detected as bands of specific length by agarose gel electrophoresis. The arrow indicates the expected size of the amplicons relative to standard sized markers run in the Marker lane, (M). Using primers specific for Gram negative bacteria reveals bands only in the six Gram negative lanes 1 through 6, (top panel), for *Escherichia coli*, *Neisseria gonorrheae*, *Klebsiella pneumoniae*, *Moraxella catarrhalis*, *Pseudomonas aeruginosa* and *Proteus vulgaris*. Using primers specific for Gram positive bacteria reveals only bands in the six Gram positive lanes, 7 through 12 (lower panel) for *Enterococcus faecalis*, *Staphylococcus epidermidis*, *Staphylococcus aureus*, *Staphylococcus xylosis*, *Streptococcus pneumoniae* and *Streptococcus haemolyticus*

Figure 9 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions designed to detect amplicons from only two Gram negative bacterial species, (in this example) *E. coli* and *K. pneumoniae*. The specificity of the amplification methodology is illustrated by the presence of amplicons in lanes 1 and 3, representing *E. coli* and *K. pneumoniae*, and the absence of amplification products in lane 2, as well as from lanes 4 through 12, these 10 empty lanes representing the remaining 10 species of bacteria used in the test.

Figure 10 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for only one bacterial group, Neisseria. The specificity of the genomic simplification methodology is illustrated by the presence of an amplicon only in lane 2, representing *Neisseria gonorrheae*, and the absence of an amplification product in lane 1, as well as from lanes 3 through 12, these 11 empty lanes representing the remaining 11 species of bacteria used in the test.

For analysis of individual microbial species, protein coding genes can also be used where appropriate, with the proviso that different strains of microorganism are not polymorphic for their presence/absence of the gene sequence in question.

Figure 11 illustrates the use of primers to the bacterial recA gene of *E. coli*. The specificity of the amplicon is illustrated by the presence of the correctly sized amplicon in lane 1 and its absence from the remaining lanes 2 through 12, representing other 11 species of bacteria.

The data of Figure 12 further illustrate the specificity of primers that reveal the membership of a larger bacterial group, such as Staphylococci. The amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions

using primers specific for Staphylococci reveal amplicons only in lanes 8, 9, and 10, representing *Staphylococcus epidermidis*, *Staphylococcus aureus* and *Staphylococcus xylosis*. The absence of an amplification product in lanes 1 through 7, as well as from lanes 11 and 12, attest to the specificity of the reaction. The 9 empty lanes representing the 9 species of non Staphylococcal bacteria used in the test.

Figure 13 shows the amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for Streptococcal bacteria. The amplification products obtained by PCR from the genomically simplified 23S ribosomal gene regions using primers specific for Streptoococci reveal amplicons only in lanes 11 and 12, representing *Streptococcus pneumoniae* and *Streptococcus haemolyticus*. The absence of an amplification product in lanes 1 through 10, reveal the specificity of the reaction. These 10 empty lanes representing the 10 species of non Streptococcal bacteria used in the test.

Figure 14 shows the amplification products obtained by PCR from a protein coding gene from the genomically simplified region of the recA gene of *Staphylococcus epidermidis*, (lane 8). The two bands (arrowed) represent the carry over amplicons from the first round, (upper band) and second round (lower band), PCR amplifications. The absence of amplicons in lanes 1 through 7, and 9 through 12 show the specificity of the method and emphasizes the point that protein coding genes can be utilized in particular circumstances instead of the non coding components of the genome, to achieve detection of only one bacterial species.

Figure 15 shows detection of amplicons using specific primers targeting the genomically simplified 23S ribosomal genes of Chlamydia PCR reactions were carried out in duplicate due to the low amounts of starting DNA. Lane number 5 was DNA extracted from the urine of a known negative individual. The presence of a band in any of the duplicates was considered a positive reaction for the presence of Chlamydia DNA.

Figure 16 shows the normal nucleotide sequence of the 23S ribosomal RNA gene from *E. coli* and the same sequence after genomic simplification, where for illustrative purposes all cytosines have been replaced with thymines.

Figure 17 shows the normal nucleotide sequence of the recA gene from *E. coli* and the same sequence after genomic simplification, where for illustrative purposes all cytosines have been replaced with thymines.

In summary, the bisulfite-treated DNA from microbial sources, when amplified using genomically simplified primers, be they oligonucleotides or modified nucleic acids

such as INAs provide an unsurpassed detection system for finding microorganisms of any type within a sample, be that sample from human clinical material or at another extreme from an environmental source such as contaminated water. The present invention has been demonstrated for a wide range of different bacterial species, and for a clinically relevant virus. The detection of single celled eukaryotic microorganisms such as the yeast *Saccharomyces cerevisiae* or its relatives is a simple extension of the method. It requires similar genomic sequence sources, such as the 18 or 28S ribosomal sequences, or as shown, protein coding sequences that are specific for a given species, type, strain or mutant or polymorphism.

The practical implications of the detection system according to the present invention are also important. While the principles described in detail herein have been demonstrated using PCR for amplification, readouts can be engaged via any methodology known in the art. With the current emphasis on microarray detection systems, one would be able to detect a far greater range of microorganisms using genomically simplified DNA since the bisulfite treatment reduces the genomic complexity and hence allows for more classes of micro organisms to be tested on microarrays with a smaller number of detectors (features).

If for example a microarray was to be constructed to detect 250,000 or so different microorganisms in one test, current methodology could not provide an adequate pragmatic detection platform, as it would be swamped by physical limitations of the detector platform. However, with genomic simplification, a small microarray could detect 1000 or so different high level bacterial categories. The positives from such a test could then be evaluated using another array, simply containing representatives of those groups that were positive in the initial test.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Claims:

1. A method for obtaining a microbial specific nucleic acid of a microbial genome or microbial nucleic acid comprising:

reducing the complexity of the microbial genome or microbial nucleic acid by generating a simplified form of the microbial genome or microbial nucleic acid in which substantially all of the positions naturally occupied by cytosines are occupied by uracil or thymine; and

obtaining a microbial specific nucleic acid from the simplified form of the nucleic acid.

2. The method according to claim 1 comprising converting microbial RNA to DNA prior to carrying out the method.

3. The method according to claim 1 comprising carrying out the method on microbial RNA to yield a simplified RNA molecule then converting the simplified RNA to form a simplified DNA molecule.

4. The method according to any one of claims 1 to 3 wherein cytosine is occupied by uracil.

5. The method according to any one of claims 1 to 3 wherein cytosine is occupied by thymine.

6. The method according to any one of claims 1 to 3 wherein the microbial genome or nucleic acid is treated with an agent selected from bisulphite, acetate or citrate that modifies cytosine to uracil.

7. The method according to any one of claims 1 to 3 wherein the microbial genome or nucleic acid is treated with an agent selected from bisulphite, acetate or citrate that modifies cytosine to uracil and amplifying the derivative microbial nucleic acid to produce a microbial genome or microbial nucleic acid having a reduced complexity.

8. The method according to claim 7 wherein amplification is carried out by polymerase chain reaction (PCR), isothermal amplification, or signal amplification.

9. The method according to any one of claims 6 to 8 wherein the agent is sodium bisulphite.

10. The method according to any one of claims 1 to 9 wherein the microbial specific nucleic acid comprises one or more nucleotide sequences unique to a microorganism.

11. The method according to any one of claims 1 to 10 wherein the microbial genome or nucleic acid is obtained from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, or single cell organism.

12. The method according to any one of claims 1 to 11 wherein the microbial genome or nucleic acid is selected from protein encoding nucleic acid, non-protein encoding nucleic acid, ribosomal gene regions of prokaryotes or single celled eukaryotic microorganisms.

13. The method according to claim 12 wherein the ribosomal gene regions are 16S or 23S in prokaryotes and 18S or 28S in single celled eukaryotic microorganisms.

14. A method for obtaining or identifying a microbial specific nucleic acid sequence of a microbial genome or microbial nucleic acid comprising:

obtaining the nucleic acid sequence of the microbial genome or microbial nucleic acid;

reducing the base complexity of the nucleic acid sequence by changing substantially all cytosines in the sequence to thymine to form a simplified nucleic acid sequence containing substantially only bases selected from adenine (A), guanine (G) and thymine (T); and

obtaining or identifying a microbial specific nucleic acid sequence from the simplified nucleic acid sequence.

15. The method according to claim 14 wherein a microbial specific nucleic acid sequence contains one or more sequences unique to a microorganism.

16. The method according to claim 14 or 15 wherein the microbial genome or nucleic acid is obtained from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, or single cell organism.

17. The method according to any one of claims 14 to 16 wherein the microbial genome or nucleic acid is selected from protein encoding nucleic acid, non-protein encoding nucleic acid, ribosomal gene regions of prokaryotes or single celled eukaryotic microorganisms.

18. The method according to claim 17 wherein the ribosomal gene regions are 16S or 23S in prokaryotes and 18S or 28S in single celled eukaryotic microorganisms.

19. A method for simplification of a microbial nucleic acid comprising:

treating microbial nucleic acid from a ribosomal gene region of a prokaryote or a single celled eukaryotic microorganism with an agent that modifies cytosine to uracil

to form derivative microbial nucleic acid; and

producing a simplified form of the microbial genome or nucleic acid.

20. The method according to claim 19 wherein the ribosomal gene region is 16S or 23S in a prokaryote or 18S or 28S in single celled eukaryotic microorganism.

21. The method according to claim 19 or 20 wherein the agent is selected from bisulphite, acetate or citrate.

22. The method according to claim 21 wherein the agent is sodium bisulphite.

23. The method according to any one of claims 19 to 22 wherein the agent modifies a cytosine to a uracil in each strand of complementary double stranded microbial DNA forming two derivative but non-complementary microbial nucleic acid molecules.

24. The method according to any one of claims 19 to 23 wherein the derivative microbial nucleic acid has a reduced total number of cytosines compared with the corresponding untreated microbial nucleic acid.

25. The method according to any one of claims 19 to 23 wherein the simplified form of the microbial nucleic acid has a reduced total number of cytosines compared with the corresponding untreated microbial genome or nucleic acid.

26. The method according to any one of claims 19 to 25 wherein the derivative microbial nucleic acid substantially contains bases adenine (A), guanine (G), thymine (T) and uracil (U) and has substantially the same total number of bases as the corresponding untreated microbial nucleic acid.

27. The method according to any one of claims 19 to 26 wherein the simplified form of the microbial nucleic acid is comprised substantially of bases adenine (A), guanine (G) and thymine (T).

28. The method according to any one of claims 19 to 27 wherein the simplified form of the microbial nucleic acid is produced by amplifying the derivative microbial nucleic acid.

29. The method according to claims 28 wherein amplification is carried out by polymerase chain reaction (PCR), isothermal amplification, or signal amplification.

30. A method for producing a microbial-specific nucleic acid molecule comprising:

treating a sample containing microbial derived DNA with an agent that modifies cytosine to uracil to form derivative microbial nucleic acid; and

producing a simplified nucleic acid molecule from at least part of the derivative microbial nucleic acid, the simplified nucleic acid molecule having a reduced total

number of cytosines compared with the corresponding untreated microbial nucleic acid, wherein the simplified nucleic acid molecule includes a nucleic acid sequence specific for a microorganism or microorganism type.

31. The method according to claim 30 wherein the microorganism is selected from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, or single cell organism.

32. The method according to claim 30 or 31 wherein the microbial nucleic acid is selected from protein encoding nucleic acid, non-protein encoding nucleic acid, ribosomal gene regions of prokaryotes or single celled eukaryotic microorganisms.

33. The method according to claim 32 wherein the ribosomal gene regions are 16S or 23S in prokaryotes and 18S or 28S in single celled eukaryotic microorganisms.

34. The method according to any one of claims 30 to 33 wherein the agent modifies unmethylated cytosine.

35. The method according to any one of claims 30 to 34 wherein the agent is selected from bisulphite, acetate or citrate.

36. The method according to claim 35 wherein the agent is sodium bisulphite.

37. The method according to any one of claims 30 to 36 wherein the simplified nucleic acid molecule is produced by amplifying the derivative microbial nucleic acid.

38. The method according to claim 37 wherein amplification is carried out by polymerase chain reaction (PCR), isothermal amplification, or signal amplification.

39. The method according to any one of claims 30 to 38 further comprising:

detecting the microbial-specific nucleic acid molecule.

40. The method according to claim 39 wherein the microbial-specific nucleic acid molecule is detected by Real-Time PCR.

41. The method according to claim 39 wherein the microbial-specific nucleic acid molecule is detected by a microarray detection system.

42. The method according to claim 39 wherein the microbial-specific nucleic acid molecule is detected by:

providing a detector ligand capable of binding to a target region of the microbial-specific nucleic acid molecule and allowing sufficient time for the detector ligand to bind to the target region; and

measuring binding of the detector ligand to the target region to detect the presence of the microbial-specific nucleic acid molecule.

43. The method according to claim 42 wherein the microbial-specific nucleic acid molecule is detected by separating an amplification product and visualising the separated product.

44. The method according to claim 43 wherein the amplification product is separated by electrophoresis and detected by visualising one or more bands on a gel.

45. The method according to any one of claims 30 to 44 wherein the simplified nucleic acid molecule has substantially no cytosines.

46. The method according to claim 45 wherein the microbial-specific nucleic acid molecule does not occur naturally in the microorganism.

47. The method according to any one of claims 30 to 46 wherein the microbial-specific nucleic acid molecule has a nucleic acid sequence indicative of a taxonomic level of the microorganism.

48. The method according to claim 47 wherein the taxonomic level of the microorganism includes family, genus, species, strain, type, or different populations from the same or different geographic or benthic populations.

49. A method for selecting a sequence of a microbial-specific nucleic acid molecule comprising:

obtaining a DNA sequence from a microorganism;

forming a simplified form of the microbial DNA sequence by carrying out a conversion of the microbial DNA sequence by changing each cytosine to thymine such that the sequence of the simplified form of the microbial DNA comprises substantially bases adenine, guanine and thymine, wherein the conversion mimics the action of an agent that converts cytosine to uracil; and

selecting a microbial-specific nucleic acid sequence from the simplified form of the microbial DNA sequence.

50. The method according to claim 49 wherein the conversion is carried out in silico.

51. The method according to claim 49 or 50 wherein simplified forms of two or more microbial DNA sequences are obtained and the two or more sequences are compared to obtain at least one microbial-specific nucleic acid molecule.

52. A microbial-specific nucleic acid molecule obtained by the method according to any one of claims 49 to 51.

53. Use of the method according to any one of claims 49 to 51 to obtain probes or primers to bind or amplify the microbial-specific nucleic acid molecule in a test or assay.

54. A method for detecting the presence of a microorganism in a sample comprising:

treating the microbial nucleic acid with an agent that modifies cytosine to uracil to form derivative microbial nucleic acid;

providing primers capable of allowing amplification of a desired microbial-specific nucleic acid molecule to the derivative microbial nucleic acid;

carrying out an amplification reaction on the derivative microbial nucleic acid to form a simplified nucleic acid; and

assaying for the presence of an amplified nucleic acid product containing the desired microbial-specific nucleic acid molecule, wherein detection of the desired microbial-specific nucleic acid molecule is indicative of the presence of the microorganism.

55. The method according to claim 54 wherein the microorganism is selected from phage, virus, viroid, bacterium, fungus, alga, protozoan, spirochaete, or single cell organism.

56. The method according to claim 54 or 55 wherein the agent is selected from bisulphite, acetate or citrate.

57. The method according to claim 56 wherein the agent is sodium bisulphite.

58. The method according to any one of claims 54 to 57 wherein amplification is carried out by polymerase chain reaction (PCR), isothermal amplification, or signal amplification.

59. The method according to any one of claims 54 to 58 wherein the microbial-specific nucleic acid molecule is detected by Real-Time PCR.

60. The method according to any one of claims 54 to 58 wherein the microbial-specific nucleic acid molecule is detected by a microarray detection system.

61. The method according to any one of claims 54 to 58 wherein the nucleic acid molecule is detected by:

providing a detector ligand capable of binding to a region of the nucleic acid molecule and allowing sufficient time for the detector ligand to bind to the region; and

measuring binding of the detector ligand to the nucleic acid molecule to detect the presence of the nucleic acid molecule.

62. The method according to any one of claims 54 to 58 wherein the nucleic acid molecule is detected by separating an amplification product and visualising the separated product.

SEQUENCE LISTING

<110> Human Genetic Signatures
<120> Detection of Microorganisms
<130> 205591675
<160> 226
<170> PatentIn version 3.1

<210> 1
<211> 14
<212> DNA
<213> artificial
<400> 1
atatatatat atat 14

<210> 2
<211> 14
<212> DNA
<213> artificial
<400> 2
aaaaaaattt tttt 14

<210> 3
<211> 34
<212> DNA
<213> Neisseria
<400> 3
gyaatywagg ycgyctygaa gaytayaaya tggc 34

<210> 4
<211> 34
<212> DNA
<213> artificial
<400> 4
gtaattwagg ttgttttgaa gattataata tggt 34

<210> 5
<211> 16
<212> DNA
<213> artificial
<400> 5
aggycgycty gaagay 16

<210> 6
<211> 16
<212> DNA
<213> artificial
<400> 6
aggttgtttt gaagat 16

<210> 7
<211> 28

<212> DNA
<213> bacterial
<400> 7
ggtttttttt gaaatagttt tagggtta 28

<210> 8
<211> 28
<212> DNA
<213> bacterial
<400> 8
ggtttttttt gaaarttatt taggtagt 28

<210> 9
<211> 25
<212> DNA
<213> bacterial
<400> 9
tggkagttag awtgtgrrwg ataag 25

<210> 10
<211> 25
<212> DNA
<213> bacterial
<400> 10
tgggagatak atrgtgggtg ttaat 25

<210> 11
<211> 22
<212> DNA
<213> bacterial
<400> 11
ggatgtggdr ttktkwagat aa 22

<210> 12
<211> 22
<212> DNA
<213> bacterial
<400> 12
tgawgtggga aggtwtagat ag 22

<210> 13
<211> 22
<212> DNA
<213> bacterial
<400> 13
hcaatmhhac ttcammmcmm yt 22

<210> 14
<211> 23
<212> DNA

<213> bacterial
<400> 14

wcaahhcacc ttcahmaacy tac 23

<210> 15
<211> 29
<212> DNA
<213> bacterial
<400> 15

accaacattc tcactymtaa wmamtccac 29

<210> 16
<211> 29
<212> DNA
<213> bacterial
<400> 16

atcaacattc acacttctaa tacctccaa 29

<210> 17
<211> 28
<212> DNA
<213> bacterial
<400> 17

ggtttttty gaaatagttt tagggtta 28

<210> 18
<211> 28
<212> DNA
<213> bacterial
<400> 18

ggtttttty gaaarttatt taggtagt 28

<210> 19
<211> 25
<212> DNA
<213> bacterial
<400> 19

yggkagttag awygygrrwg ataag 25

<210> 20
<211> 25
<212> DNA
<213> bacterial
<400> 20

ygggagatak ayrgygggtg ttaat 25

<210> 21
<211> 22
<212> DNA
<213> bacterial
<400> 21

ggatgtggdr ttkykwagat aa 22

<210> 22
<211> 22
<212> DNA
<213> bacterial
<400> 22

ygawgtggga aggtwtagat ag 22

<210> 23
<211> 22
<212> DNA
<213> bacterial
<400> 23

hcratmhhrc ttcrmmmmcmm yt 22

<210> 24
<211> 23
<212> DNA
<213> bacterial
<400> 24

wcrahhcacc ttcahmracy tac 23

<210> 25
<211> 29
<212> DNA
<213> bacterial
<400> 25

accracattc tcactymtaa wmamtccac 29

<210> 26
<211> 29
<212> DNA
<213> bacterial
<400> 26

atcaacattc rcacttctaa tacctccaa 29

<210> 27
<211> 24
<212> DNA
<213> bacterial
<400> 27

kttragaaaa gtwtttagdd agrk 24

<210> 28
<211> 24
<212> DNA
<213> bacterial
<400> 28

tttargaaaa gttwttaagt wtta 24

<210> 29
<211> 25
<212> DNA
<213> bacterial
<400> 29

agdtragrwg agdattttwa ggtkr 25

<210> 30
<211> 25
<212> DNA
<213> bacterial
<400> 30

ggktrggwwg agaatwttaa ggtgt 25

<210> 31
<211> 25
<212> DNA
<213> bacterial
<400> 31

aatytmymat taaaacaata cmcaa 25

<210> 32
<211> 25
<212> DNA
<213> bacterial
<400> 32

aatctcaaaw aaaaacaaym ymacc 25

<210> 33
<211> 32
<212> DNA
<213> bacterial
<400> 33

acmhacatct tcacwmayay tayaayttca cc 32

<210> 34
<211> 32
<212> DNA
<213> bacterial
<400> 34

maytacatct tcacaacmah wtcaayttca ct 32

<210> 35
<211> 24
<212> DNA
<213> bacterial
<400> 35

cmatayyaaa ytacaataaa actc 24

<210> 36
<211> 24
<212> DNA
<213> bacterial
<400> 36
caataymaaa ctayaataaa aatt 24

<210> 37
<211> 31
<212> DNA
<213> bacterial
<400> 37
ggtgaarttr tartrtkwgt gaagatgtdk g 31

<210> 38
<211> 31
<212> DNA
<213> bacterial
<400> 38
agtgaarttg awdtkgttgt gaagatgtar t 31

<210> 39
<211> 25
<212> DNA
<213> bacterial
<400> 39
gatwggatgg aaagattttr trgag 25

<210> 40
<211> 25
<212> DNA
<213> bacterial
<400> 40
kgtwagatgg aaagattttg tgaat 25

<210> 41
<211> 23
<212> DNA
<213> bacterial
<400> 41
hymaymmway haaaataata tcc 23

<210> 42
<211> 23
<212> DNA
<213> bacterial
<400> 42
tcaammmywm maaaataata ttt 23

<210> 43
<211> 26

<212> DNA
<213> bacterial
<400> 43

awccattcta aaaaaacctt taaaca 26

<210> 44
<211> 26
<212> DNA
<213> bacterial
<400> 44

aaccawwmyw aamhmacctt cawact 26

<210> 45
<211> 28
<212> DNA
<213> bacterial
<400> 45

gttggtaagg tgatatgaat tgttataa 28

<210> 46
<211> 26
<212> DNA
<213> bacterial
<400> 46

ttattattaa ttgaatttat aggtta 26

<210> 47
<211> 27
<212> DNA
<213> bacterial
<400> 47

gaggagttta gagtttgaat tagtrtg 27

<210> 48
<211> 26
<212> DNA
<213> bacterial
<400> 48

tatatacaaa actatcaccc tatatc 26

<210> 49
<211> 22
<212> DNA
<213> bacterial
<400> 49

tcatcaaact cacaacayat ac 22

<210> 50
<211> 25
<212> DNA

<213> bacterial
<400> 50

ttgagtaaga tattgatggg ggtaa 25

<210> 51
<211> 21
<212> DNA
<213> bacterial
<400> 51

tatggttagg gggttattgt a 21

<210> 52
<211> 24
<212> DNA
<213> bacterial
<400> 52

aatctatcat ttaaaacctt aacc 24

<210> 53
<211> 24
<212> DNA
<213> bacterial
<400> 53

cctaactatc tataccttcc cact 24

<210> 54
<211> 25
<212> DNA
<213> bacterial
<400> 54

cactccccta ccataccaat aaacc 25

<210> 55
<211> 28
<212> DNA
<213> bacterial
<400> 55

gtatgatgag ttagggagtt aagttaaa 28

<210> 56
<211> 21
<212> DNA
<213> bacterial
<400> 56

ggtgaggtta agggatatat a 21

<210> 57
<211> 30
<212> DNA
<213> bacterial
<400> 57

aaaagagtga agagttgttt ggtttagata 30

<210> 58
<211> 24
<212> DNA
<213> bacterial
<400> 58

tccaaacctt tttcaacatt aact 24

<210> 59
<211> 28
<212> DNA
<213> bacterial
<400> 59

ccctaaaatt atttcaaaaa aaacaaaa 28

<210> 60
<211> 30
<212> DNA
<213> bacterial
<400> 60

ttagtggggg tttattggtt tattaatgga 30

<210> 61
<211> 30
<212> DNA
<213> bacterial
<400> 61

taaggaagtg atgatttgaa gatagttgga 30

<210> 62
<211> 21
<212> DNA
<213> bacterial
<400> 62

acaccttctc tactaaatac t 21

<210> 63
<211> 25
<212> DNA
<213> bacterial
<400> 63

tataccataa atcttcacta atatc 25

<210> 64
<211> 27
<212> DNA
<213> bacterial
<400> 64

ttgtgtagat gatggagtag taggtta 27

<210> 65
<211> 28
<212> DNA
<213> bacterial
<400> 65

gaatgatgga gtaagttaag tatgtgga 28

<210> 66
<211> 27
<212> DNA
<213> bacterial
<400> 66

taaaaattat ttcttaaaaa cctcact 27

<210> 67
<211> 26
<212> DNA
<213> bacterial
<400> 67

aaattatctc acacacctta aaatat 26

<210> 68
<211> 24
<212> DNA
<213> bacterial
<400> 68

aatgttaaaa ggttaaaggg atat 24

<210> 69
<211> 28
<212> DNA
<213> bacterial
<400> 69

tattgaattt aagttttggt gaatggtt 28

<210> 70
<211> 29
<212> DNA
<213> bacterial
<400> 70

ccaatatttc aacattaact cccactctc 29

<210> 71
<211> 29
<212> DNA
<213> bacterial
<400> 71

atatccatct tccaaattca taaaataat 29

<210> 72
<211> 24
<212> DNA
<213> bacterial
<400> 72

taaacaacaa caattccact ttcc 24

<210> 73
<211> 28
<212> DNA
<213> bacterial
<400> 73

ataggaaaag aaawtgaawg wgattttg 28

<210> 74
<211> 28
<212> DNA
<213> bacterial
<400> 74

gtgtagtggt gagtgaaagt ggaatagg 28

<210> 75
<211> 32
<212> DNA
<213> bacterial
<400> 75

taaacaamtt cmmtcaaaat aacatttyyc aa 32

<210> 76
<211> 23
<212> DNA
<213> bacterial
<400> 76

ctaattaata tttaaactta ccc 23

<210> 77
<211> 25
<212> DNA
<213> bacterial
<400> 77

ttttgaaatt atatgtttat aatgt 25

<210> 78
<211> 27
<212> DNA
<213> bacterial
<400> 78

aagtatgagt tggtgagtta tgatagt 27

<210> 79
<211> 23

<212> DNA
<213> bacterial
<400> 79

cctccamtta wtyataatct yac 23

<210> 80
<211> 25
<212> DNA
<213> bacterial
<400> 80

caccwaaaya acaccatcat acatt 25

<210> 81
<211> 29
<212> DNA
<213> bacterial
<400> 81

tgtagttaga tagtggggta taagtttta 29

<210> 82
<211> 24
<212> DNA
<213> bacterial
<400> 82

aggggaagag tttagattat taaa 24

<210> 83
<211> 28
<212> DNA
<213> bacterial
<400> 83

ataacttcaw cycmwataca acactcat 28

<210> 84
<211> 28
<212> DNA
<213> bacterial
<400> 84

atcaatttaa aaaattctca ctcycaaa 28

<210> 85
<211> 26
<212> DNA
<213> bacterial
<400> 85

tttttatwat tggatttggg gwtaaa 26

<210> 86
<211> 22
<212> DNA

<213> bacterial
<400> 86

tkktwwttag tattgagaat ga 22

<210> 87
<211> 24
<212> DNA
<213> bacterial
<400> 87

tgtaaattwa ttttgtaagt twgt 24

<210> 88
<211> 23
<212> DNA
<213> bacterial
<400> 88

gaatgagggg ggattgttta att 23

<210> 89
<211> 26
<212> DNA
<213> bacterial
<400> 89

tctataacca aaacaatcaa aaaata 26

<210> 90
<211> 26
<212> DNA
<213> bacterial
<400> 90

cattacacct aacaaatatc ttcacc 26

<210> 91
<211> 26
<212> DNA
<213> bacterial
<400> 91

atwwataggt tgaataggtr agaaat 26

<210> 92
<211> 28
<212> DNA
<213> bacterial
<400> 92

atagtgattt ggtggtttag tatggaat 28

<210> 93
<211> 32
<212> DNA
<213> bacterial
<400> 93

caaacctact tcaactcaaa aataaaataa at 32

<210> 94
<211> 29
<212> DNA
<213> bacterial
<400> 94
acaacaattt aaacccaact cacatatct 29

<210> 95
<211> 29
<212> DNA
<213> bacterial
<400> 95
aaaayaamwc tyttcaatct tcctayaaa 29

<210> 96
<211> 27
<212> DNA
<213> bacterial
<400> 96
atwwttgtta aggdwrtgar raggaag 27

<210> 97
<211> 24
<212> DNA
<213> bacterial
<400> 97
tagragggta aattgargwg ttta 24

<210> 98
<211> 26
<212> DNA
<213> bacterial
<400> 98
tkatttggga arrtwrgtta aagaga 26

<210> 99
<211> 25
<212> DNA
<213> bacterial
<400> 99
tctcttcaac ttaacctcac atcat 25

<210> 100
<211> 24
<212> DNA
<213> bacterial
<400> 100
ataatttcaa atctacawcm waat 24

<210> 101
<211> 25
<212> DNA
<213> bacterial
<400> 101

ratktattgg aggattgaat taggg 25

<210> 102
<211> 23
<212> DNA
<213> bacterial
<400> 102

atgttgaaaa gtgtttggat gat 23

<210> 103
<211> 31
<212> DNA
<213> bacterial
<400> 103

tctaaaatya ataawccaaa ataamcccct c 31

<210> 104
<211> 23
<212> DNA
<213> bacterial
<400> 104

actaccaayh atawhtcatt aac 23

<210> 105
<211> 25
<212> DNA
<213> bacterial
<400> 105

aggttgakat ttttgtatta gagta 25

<210> 106
<211> 29
<212> DNA
<213> bacterial
<400> 106

rwagtgatgg agggatgtag taggttaat 29

<210> 107
<211> 25
<212> DNA
<213> bacterial
<400> 107

cttttctyaa caatataaca tcact 25

<210> 108
<211> 24
<212> DNA
<213> bacterial
<400> 108
ctctcamtca cctaaaacta ctca 24

<210> 109
<211> 28
<212> DNA
<213> bacterial
<400> 109
agaagttgat gaaggatgtt attaatga 28

<210> 110
<211> 29
<212> DNA
<213> bacterial
<400> 110
gttattgata tgtgaatwta tagtatrtt 29

<210> 111
<211> 24
<212> DNA
<213> bacterial
<400> 111
caaaaythtt accttctyta atyc 24

<210> 112
<211> 23
<212> DNA
<213> bacterial
<400> 112
caacaaaatt ycacatactc cat 23

<210> 113
<211> 23
<212> DNA
<213> bacterial
<400> 113
gatttgatgt aaggttaagt agt 23

<210> 114
<211> 31
<212> DNA
<213> bacterial
<400> 114
ttggttaggt tgaagtttag gtaatattga a 31

<210> 115
<211> 32

<212> DNA
<213> bacterial
<400> 115

gatttatgtt gaaaagtgag tggatgaatt ga 32

<210> 116
<211> 29
<212> DNA
<213> bacterial
<400> 116

cctytttcta actcccaaat taaattaat 29

<210> 117
<211> 25
<212> DNA
<213> bacterial
<400> 117

gaagttgtgg attgtttttt ggata 25

<210> 118
<211> 30
<212> DNA
<213> bacterial
<400> 118

aagggtgttg aagtatgatt gtaaggatat 30

<210> 119
<211> 31
<212> DNA
<213> bacterial
<400> 119

tacamtccaa ymacacactt cacctatcct a 31

<210> 120
<211> 26
<212> DNA
<213> bacterial
<400> 120

caacaatata aaatcaacaa ctcaaa 26

<210> 121
<211> 26
<212> DNA
<213> bacterial
<400> 121

aggagtggtt agtttttgtg aagtta 26

<210> 122
<211> 24
<212> DNA

<213> bacterial
<400> 122
acaaattaaa aawccaacac aact 24

<210> 123
<211> 26
<212> DNA
<213> bacterial
<400> 123
taacactatc tcccaccaya atmaat 26

<210> 124
<211> 23
<212> DNA
<213> bacterial
<400> 124
taggttgttg agttttaatt ata 23

<210> 125
<211> 24
<212> DNA
<213> bacterial
<400> 125
gaagtataaa gtaatggtgg ggtg 24

<210> 126
<211> 31
<212> DNA
<213> bacterial
<400> 126
tacaatatca actacaccac ttctaacaaa t 31

<210> 127
<211> 23
<212> DNA
<213> bacterial
<400> 127
taataaaaat aacaattata ttt 23

<210> 128
<211> 28
<212> DNA
<213> bacterial
<400> 128
aaggttgtag agtattaagt attttaag 28

<210> 129
<211> 24
<212> DNA
<213> bacterial
<400> 129

gttgataatg tattaggggt tgga 24

<210> 130
<211> 28
<212> DNA
<213> bacterial
<400> 130
atatggattt gaaagtttag gtaagatg 28

<210> 131
<211> 31
<212> DNA
<213> bacterial
<400> 131
tactactaaa tcaacaacaa caatatccac a 31

<210> 132
<211> 23
<212> DNA
<213> bacterial
<400> 132
cttaatactt aaaacattaa tct 23

<210> 133
<211> 27
<212> DNA
<213> bacterial
<400> 133
gagaataagt aaaaggtgtt agttgtg 27

<210> 134
<211> 34
<212> DNA
<213> bacterial
<400> 134
gatttttatt ggtttattgt tatttgatat tgtt 34

<210> 135
<211> 31
<212> DNA
<213> bacterial
<400> 135
caaataatca atatcaacac ccaacttttt c 31

<210> 136
<211> 23
<212> DNA
<213> bacterial
<400> 136
tacacaccac caaacccata tac 23

<210> 137
<211> 24
<212> DNA
<213> bacterial
<400> 137
gaaaataaat agaaagtgtt ggtg 24

<210> 138
<211> 23
<212> DNA
<213> bacterial
<400> 138
tgtttttatt ggatattgtg ttt 23

<210> 139
<211> 26
<212> DNA
<213> bacterial
<400> 139
caataacatc tactacacca aaacac 26

<210> 140
<211> 25
<212> DNA
<213> bacterial
<400> 140
catattaaac tacttcaaat taccc 25

<210> 141
<211> 25
<212> DNA
<213> bacterial
<400> 141
tatgtgtttt ggtgaagatt gttta 25

<210> 142
<211> 22
<212> DNA
<213> bacterial
<400> 142
ttttgatatt gtattggggg tg 22

<210> 143
<211> 27
<212> DNA
<213> bacterial
<400> 143
ggtttgttaa tggggtgtat tgttgag 27

<210> 144
<211> 20
<212> DNA
<213> bacterial
<400> 144
catactctac atcaataaaa 20

<210> 145
<211> 34
<212> DNA
<213> bacterial
<400> 145
gtaatcaagg tcgtcttgaa gactacaaca tggc 34

<210> 146
<211> 34
<212> DNA
<213> bacterial
<400> 146
gcaatttagg ccgcctcgaa gattataata tggc 34

<210> 147
<211> 34
<212> DNA
<213> bacterial
<400> 147
gtaattaagg ttgttttgaa gattataata tggt 34

<210> 148
<211> 34
<212> DNA
<213> bacterial
<400> 148
gtaatttagg ttgttttgaa gattataata tggt 34

<210> 149
<211> 16
<212> DNA
<213> bacterial
<400> 149
aggycgycty gaagay 16

<210> 150
<211> 16
<212> DNA
<213> bacterial
<400> 150
aggttgtttt gaagat 16

<210> 151
<211> 16
<212> DNA
<213> bacterial
<400> 151
taactacgga agatca 16

<210> 152
<211> 16
<212> DNA
<213> bacterial
<400> 152
taattatgga agatta 16

<210> 153
<211> 16
<212> DNA
<213> bacterial
<400> 153
gtaatcaagg tcgtct 16

<210> 154
<211> 16
<212> DNA
<213> bacterial
<400> 154
gtaattaagg ttgttt 16

<210> 155
<211> 16
<212> DNA
<213> bacterial
<400> 155
gcaatttagg ccgcct 16

<210> 156
<211> 16
<212> DNA
<213> bacterial
<400> 156
gtaatttagg ttgttt 16

<210> 157
<211> 40
<212> DNA
<213> bacterial
<400> 157
aagctcttga aggtgactct aaatacgaag acatcatcat 40

<210> 158
<211> 40

<212> DNA
<213> bacterial
<400> 158
aagcccttga aggtgacact aaatacgaag acatcgttat 40

<210> 159
<211> 40
<212> DNA
<213> bacterial
<400> 159
aagctcttga aggtgactca aaatacgaag atatcatcat 40

<210> 160
<211> 40
<212> DNA
<213> bacterial
<400> 160
aagctcttga aggtgatact aagtacgaag acatcatcat 40

<210> 161
<211> 40
<212> DNA
<213> bacterial
<400> 161
aagctcttga aggtgactct aaatacgaag atatcatcat 40

<210> 162
<211> 40
<212> DNA
<213> bacterial
<400> 162
aagctcttga aggcgataca gcacatgaag atatcatcat 40

<210> 163
<211> 40
<212> DNA
<213> bacterial
<400> 163
aagctcttga aggtgactct aaatacgaag acatcgttat 40

<210> 164
<211> 40
<212> DNA
<213> bacterial
<400> 164
aagctcttga aggtgacact cagtacgaag atatcatcat 40

<210> 165
<211> 40
<212> DNA

<213> bacterial
<400> 165
aagctcttga aggtgattct aaatacgaag acatcatcat 40

<210> 166
<211> 40
<212> DNA
<213> bacterial
<400> 166
aagctcttga aggtgattct aaatacgaag acatcatcat 40

<210> 167
<211> 40
<212> DNA
<213> bacterial
<400> 167
aagcycttga aggygaywcw vmryaygaag ayatcrtyat 40

<210> 168
<211> 40
<212> DNA
<213> bacterial
<400> 168
aagtttttga aggtgatttt aaatatgaag atattattat 40

<210> 169
<211> 40
<212> DNA
<213> bacterial
<400> 169
aagtttttga aggtgatatt aaatatgaag atattgttat 40

<210> 170
<211> 40
<212> DNA
<213> bacterial
<400> 170
aagtttttga aggtgattta aaatatgaag atattattat 40

<210> 171
<211> 40
<212> DNA
<213> bacterial
<400> 171
aagtttttga aggtgatatt aagtatgaag atattattat 40

<210> 172
<211> 40
<212> DNA
<213> bacterial
<400> 172

aagtttttga aggtgatttt aaatatgaag atattattat 40

<210> 173
<211> 40
<212> DNA
<213> bacterial
<400> 173

aagtttttga aggtgatata gtatatgaag atattattat 40

<210> 174
<211> 40
<212> DNA
<213> bacterial
<400> 174

aagtttttga aggtgatttt aaatatgaag atattgttat 40

<210> 175
<211> 40
<212> DNA
<213> bacterial
<400> 175

aagtttttga aggtgatatt tagtatgaag atattattat 40

<210> 176
<211> 40
<212> DNA
<213> bacterial
<400> 176

aagtttttga aggtgatttt aaatatgaag atattattat 40

<210> 177
<211> 40
<212> DNA
<213> bacterial
<400> 177

aagtttttga aggtgatttt aaatatgaag atattattat 40

<210> 178
<211> 40
<212> DNA
<213> bacterial
<400> 178

aagtttttga aggtgatwtw rwrtatgaag atattrttat 40

<210> 179
<211> 23
<212> DNA
<213> bacterial
<400> 179

tacaacgaca ataaaacggt tga 23

<210> 180
<211> 23
<212> DNA
<213> bacterial
<400> 180

tataatgata ataaaatggt tga 23

<210> 181
<211> 23
<212> DNA
<213> bacterial
<400> 181

tacggagata ataaagttgt tga 23

<210> 182
<211> 23
<212> DNA
<213> bacterial
<400> 182

tatggagata ataaagttgt tga 23

<210> 183
<211> 23
<212> DNA
<213> bacterial
<400> 183

tacaacgaca ataaaacggt tga 23

<210> 184
<211> 23
<212> DNA
<213> bacterial
<400> 184

tataatgata ataaaatggt tga 23

<210> 185
<211> 23
<212> DNA
<213> bacterial
<400> 185

tacaacgaca ataaaacggt tga 23

<210> 186
<211> 23
<212> DNA
<213> bacterial
<400> 186

tataatgata ataaaatggt tga 23

<210> 187
<211> 23
<212> DNA
<213> bacterial
<400> 187
tatagagata ataaaacgat taa 23

<210> 188
<211> 23
<212> DNA
<213> bacterial
<400> 188
tatagagata ataaaatgat taa 23

<210> 189
<211> 23
<212> DNA
<213> bacterial
<400> 189
tacagagata ataaaactat taa 23

<210> 190
<211> 23
<212> DNA
<213> bacterial
<400> 190
tatagagata ataaaattat taa 23

<210> 191
<211> 23
<212> DNA
<213> bacterial
<400> 191
tacaatgaca ataaaatggt tga 23

<210> 192
<211> 23
<212> DNA
<213> bacterial
<400> 192
tataatgata ataaaatggt tga 23

<210> 193
<211> 23
<212> DNA
<213> bacterial
<400> 193
tayrrhgaya ataaarykrt tra 23

<210> 194
<211> 23

<212> DNA
<213> bacterial
<400> 194
tatrrwgata ataaartkrt tra 23

<210> 195
<211> 20
<212> DNA
<213> viral
<400> 195
tgtgtgtgca gggataattg 20

<210> 196
<211> 20
<212> DNA
<213> viral
<400> 196
tgtatatgta gggacaattg 20

<210> 197
<211> 20
<212> DNA
<213> viral
<400> 197
tgtgtttgta gagacaactg 20

<210> 198
<211> 20
<212> DNA
<213> viral
<400> 198
tgtatatgta gggacaattg 20

<210> 199
<211> 20
<212> DNA
<213> viral
<400> 199
tgtgtttgca gagataattg 20

<210> 200
<211> 20
<212> DNA
<213> viral
<400> 200
tgcatttgca gggacaattg 20

<210> 201
<211> 20
<212> DNA

<213> viral
<400> 201

tgcacttgca gggataattg 20

<210> 202
<211> 20
<212> DNA
<213> viral
<400> 202

tgcgtttgcc gagataattg 20

<210> 203
<211> 20
<212> DNA
<213> viral
<400> 203

tgtgcctgta gagataacag 20

<210> 204
<211> 20
<212> DNA
<213> viral
<220>
<221> misc_feature
<222> (1)..(10)
<223> n

<400> 204

tgyryntgym grgayaaywg 20

<210> 205
<211> 20
<212> DNA
<213> viral
<400> 205

tgtgtgtgta gggataattg 20

<210> 206
<211> 20
<212> DNA
<213> viral
<400> 206

tgtatatgta gggataattg 20

<210> 207
<211> 20
<212> DNA
<213> viral
<400> 207

tgtgtttgta gagataattg 20

<210> 208
<211> 20
<212> DNA
<213> viral
<400> 208
tgtatatgta gggataattg 20

<210> 209
<211> 20
<212> DNA
<213> viral
<400> 209
tgtgtttgta gagataattg 20

<210> 210
<211> 20
<212> DNA
<213> viral
<400> 210
tgtatttgta gggataattg 20

<210> 211
<211> 20
<212> DNA
<213> viral
<400> 211
tgtatttgta gggataattg 20

<210> 212
<211> 20
<212> DNA
<213> viral
<400> 212
tgtgtttgtt gagataattg 20

<210> 213
<211> 20
<212> DNA
<213> viral
<400> 213
tgtgtttgta gagataatag 20

<210> 214
<211> 20
<212> DNA
<213> viral
<400> 214
tgtrtdtgtw grgataatwg 20

<210> 215
<211> 19

<212> DNA
<213> viral
<400> 215

ctaaattcgc tccgattta 19

<210> 216
<211> 19
<212> DNA
<213> viral
<400> 216

ttaaatttgt tttgattta 19

<210> 217
<211> 19
<212> DNA
<213> viral
<400> 217

caaaattgac ccagactta 19

<210> 218
<211> 19
<212> DNA
<213> viral
<400> 218

taaaattgat ttagattta 19

<210> 219
<211> 19
<212> DNA
<213> viral
<400> 219

ttaaattcgt taagattca 19

<210> 220
<211> 19
<212> DNA
<213> viral
<400> 220

ttaaatttgt taagattta 19

<210> 221
<211> 19
<212> DNA
<213> viral
<400> 221

ywaaattsry ymmgaytya 19

<210> 222
<211> 19
<212> DNA

<213> viral
<400> 222
twaaattkrt twwgattta 19

<210> 223
<211> 2922
<212> DNA
<213> bacterial
<400> 223
gattaagtta ttaagggcgc acggtggatg ccttggcact agaagccgat gaaggacgtt 60
actaacgacg atatgctttg ggtagctgta agtaagcgtt gatccagaga tttccgaatg 120
ggggaaccca gcatgagtta tgtcatgtta tcgatatgtg aatttatagc atgtcagaag 180
gcagacccgg agaactgaaa catcttagta cccggaggaa gagaaagaaa aatcgattcc 240
ctgagtagcg gcgagcgaaa cgggaagagc ccaaaccaac aagcttgctt gttggggttg 300
taggacactc tatacggagt tacaaaagaa catgttagac gaatcatctg gaaagatgaa 360
tcaaagaagg taataatcct gtagtcgaaa acatattctc tcttgagtgg atcctgagta 420
cgacggagca cgtgaaattc cgtcggaatc tgggaggacc atctcctaag gctaaatact 480
ctctagtgac cgatagtgaa ccagtaccgt gagggaaagg tgaaaagtac cccggaaggg 540
gagtgaaaga gaacttgaaa ccgtgtgctt acaagtagtc agagcccgtt aatgggtgat 600
ggcgtgcctt ttgtagaatg aaccggcgag ttacgatctg atgcaaggtt aagcagcaaa 660
tgcggagccg cagcgaaagc gagtctgaat agggcgttga gtatttggtc gtagacccga 720
aaccaggtga tctacccttg gtcaggttga agttcaggta acactgaatg gaggaccgaa 780
ccgacttacg ttgaaaagtg agcggatgaa ctgagggtag cggagaaatt ccaatcgaac 840
ttggagatag ctggttctct ccgaaatagc tttagggcta gcctcaagtg atgattattg 900
gaggtagagc actgtttgga cgaggggccc ctctcgggtt accgaattca gacaaactcc 960
gaatgccaat taatttaact tgggagtcag aacatgggtg ataaggtccg tgttcgaaag 1020
ggaaacagcc cagaccacca gctaaggtcc caaaatatat gttaagtgga aaaggatgtg 1080
gcgttgccca gacaactagg atgttggctt agaagcagcc atcatttaaa gagtgcgtaa 1140
tagctcacta gtcgagtgac actgcgccga aaatgtaccg gggctaaaca tattaccgaa 1200
gctgtggatt gtcctttgga caatggtagg agagcgttct aagggcgtcg aagcatgatc 1260
gcaaggacat gtggagcgct tagaagtgag aatgccggtg tgagtagcga aagacgggtg 1320
agaatcccgt ccaccgattg actaaggttt ccagaggaag gctcgtccgc tctgggttag 1380
tcgggtccta agctgaggcc gacaggcgta ggcgatggat aacaggttga tattcctgta 1440
ccacctagta tcgttttaat cgatgggggg acgcagtagg ataggcgaag cgtgctgttg 1500
gagtgcacgt ccaagcagta aggctgagtg ttaggcaaat ccggcactca taaggctgag 1560
ctgtgatggg gagaggaaat tgtttcctcg agtcgttgat ttcacactgc cgagaaaagc 1620
ctctagatag ataacaggtg cccgtaccgc aaaccgacac aggtagtcaa gatgagaatt 1680
ctaaggtgag cgagcgaact ctcgttaagg aactcggcaa aatgaccccg taacttcggg 1740
agaaggggtg ctctttaggg ttcacgccca gaagagccgc agtgaatagg cccaagcgac 1800
tgtttatcaa aaacacaggt ctctgctaaa ccgtaaggtg atgtataggg gctgacgcct 1860
gcccggtgct ggaaggttaa gaggagtggt tagcttctgc gaagctacga atcgaagccc 1920
cagtaaacgg cggccgtaac tataacggtc ctaaggtagc gaaattcctt gtcgggtaag 1980
ttccgacccg cacgaaaggc gtaacgattt gggcactgtc tcaacgagag actcggtgaa 2040

atcatagtac ctgtgaagat gcaggttacc cgcgacagga cggaaagacc ccgtggagct 2100
ttactgtagc ctgatattga aattcggcac agcttgtaca ggataggtag gagcctttga 2160
aacgtgagcg ctagcttacg tggaggcgtt ggtgggatac taccctagct gtgttggctt 2220
tctaacccgc accacttatc gtggtgggag acagtgtcag gcgggcagtt tgactggggc 2280
ggtcgcctcc taaaaggtaa cggaggcgct caaaggttcc ctcagaatgg ttggaaatca 2340
ttcatagagt gtaaaggcat aagggagctt gactgcgaga cctacaagtc gagcagggtc 2400
gaaagacgga cttagtgatc cggtggttcc gcatggaagg gccatcgctc aacggataaa 2460
agctaccccg gggataacag gcttatctcc cccaagagtt cacatcgacg gggaggtttg 2520
gcacctcgat gtcggctcat cgcatcctgg ggctgtagtc ggtcccaagg gttgggctgt 2580
tcgcccatta aagcggtacg cgagctgggt tcagaacgtc gtgagacagt tcggtcccta 2640
tccgtcgtgg gcgtaggaaa tttgagagga gctgtcctta gtacgagagg accgggatgg 2700
acatacctct ggtgtaccag ttgtcgtgcc aacggcatag ctgggtagct atgtatggac 2760
gggataagtg ctgaaagcat ctaagcatga agcccccctc aagatgagat ttcccaactt 2820
cggttataag atccctcgaa gatgacgagg ttaataggtt cgaggtggaa gcgtggtgac 2880
acgtggagct gacgaatact aatcgatcga agacttaatc aa 2922

<210> 224
<211> 2922
<212> DNA
<213> bacterial
<400> 224

gattaagtta ttaagggtgt atggtggatg ttttggtatt agaagttgat gaaggatgtt 60
attaatgatg atatgttttg ggtagttgta agtaagtgtt gatttagaga tttttgaatg 120
ggggaattta gtatgagtta tgttatgtta ttgatatgtg aatttatagt atgttagaag 180
gtagatttgg agaattgaaa tattttagta tttggaggaa gagaaagaaa aattgatttt 240
ttgagtagtg gtgagtgaaa tgggaagagt ttaaattaat aagtttgttt gttggggttg 300
taggatattt tatatggagt tataaaagaa tatgttagat gaattatttg gaaagatgaa 360
ttaaagaagg taataatttt gtagttgaaa atatattttt ttttgagtgg attttgagta 420
tgatggagta tgtgaaattt tgttggaatt tgggaggatt atttttaag gttaaatatt 480
ttttagtgat tgatagtgaa ttagtattgt gagggaaagg tgaaaagtat tttggaaggg 540
gagtgaaaga gaatttgaaa ttgtgtgttt ataagtagtt agagtttgtt aatgggtgat 600
ggtgtgtttt ttgtagaatg aattggtgag ttatgatttg atgtaaggtt aagtagtaaa 660
tgtggagttg tagtgaaagt gagtttgaat agggtgttga gtatttggtt gtagatttga 720
aattaggtga tttatttttg gttaggttga agtttaggta atattgaatg gaggattgaa 780
ttgatttatg ttgaaaagtg agtggatgaa ttgagggtag tggagaaatt ttaattgaat 840
ttggagatag ttggtttttt ttgaaatagt tttagggtta gttttaagtg atgattattg 900
gaggtagagt attgtttgga tgaggggttt tttttgggtt attgaattta gataaatttt 960
gaatgttaat taatttaatt tgggagttag aatatgggtg ataaggtttg tgtttgaaag 1020
ggaaatagtt tagattatta gttaaggttt taaaatatat gttaagtgga aaaggatgtg 1080
gtgttgttta gataattagg atgttggttt agaagtagtt attatttaaa gagtgtgtaa 1140
tagtttatta gttgagtgat attgtgttga aaatgtattg gggttaaata tattattgaa 1200
gttgtggatt gttttttgga taatggtagg agagtgtttt aagggtgttg aagtatgatt 1260
gtaaggatat gtggagtgtt tagaagtgag aatgttggtg tgagtagtga aagatgggtg 1320

agaattttgt ttattgattg attaaggttt ttagaggaag gtttgtttgt tttgggttag 1380
ttgggtttta agttgaggtt gataggtgta ggtgatggat aataggttga tatttttgta 1440
ttatttagta ttgttttaat tgatgggggg atgtagtagg ataggtgaag tgtgttgttg 1500
gagtgtatgt ttaagtagta aggttgagtg ttaggtaaat ttggtattta taaggttgag 1560
ttgtgatggg gagaggaaat tgtttttttg agttgttgat tttatattgt tgagaaaagt 1620
ttttagatag ataataggtg tttgtattgt aaattgatat aggtagttaa gatgagaatt 1680
ttaaggtgag tgagtgaatt tttgttaagg aatttggtaa aatgattttg taatttttggg 1740
agaaggggtg ttttttaggg tttatgttta gaagagttgt agtgaatagg tttaagtgat 1800
tgtttattaa aaatataggt ttttgttaaa ttgtaaggtg atgtataggg gttgatgttt 1860
gtttggtgtt ggaaggttaa gaggagtggt tagtttttgt gaagttatga attgaagttt 1920
tagtaaatgg tggttgtaat tataatggtt ttaaggtagt gaaatttttt gttgggtaag 1980
ttttgatttg tatgaaaggt gtaatgattt gggtattgtt ttaatgagag atttggtgaa 2040
attatagtat ttgtgaagat gtaggttatt tgtgatagga tggaaagatt ttgtggagtt 2100
ttattgtagt ttgatattga aatttggtat agtttgtata ggataggtag gagtttttga 2160
aatgtgagtg ttagtttatg tggaggtgtt ggtgggatat tattttagtt gtgttggttt 2220
tttaatttgt attatttatt gtggtgggag atagtgttag gtgggtagtt tgattggggt 2280
ggttgttttt taaaaggtaa tggaggtgtt taaaggtttt tttagaatgg ttggaaatta 2340
tttatagagt gtaaaggtat aagggagttt gattgtgaga tttataagtt gagtagggtt 2400
gaaagatgga tttagtgatt tggtggtttt gtatggaagg gttattgttt aatggataaa 2460
agttatttttg gggataatag gtttattttt tttaagagtt tatattgatg gggaggtttg 2520
gtattttgat gttggtttat tgtattttgg ggttgtagtt ggttttaagg gttgggttgt 2580
ttgtttatta aagtggtatg tgagttgggt ttagaatgtt gtgagatagt ttggttttta 2640
tttgttgtgg gtgtaggaaa tttgagagga gttgttttta gtatgagagg attgggatgg 2700
atatattttt ggtgtattag ttgttgtgtt aatggtatag ttgggtagtt atgtatggat 2760
gggataagtg ttgaaagtat ttaagtatga agtttttttt aagatgagat tttttaattt 2820
tggttataag attttttgaa gatgatgagg ttaataggtt tgaggtggaa gtgtggtgat 2880
atgtggagtt gatgaatatt aattgattga agatttaatt aa 2922

<210> 225
<211> 1062
<212> DNA
<213> bacterial
<400> 225

atggctatcg acgaaaacaa acagaaagcg ttggcggcag cactgggcca gattgagaaa 60
caatttggta aaggctccat catgcgcctg ggtgaagacc gttccatgga tgtggaaacc 120
atctctaccg gttcgctttc actggatatc gcgcttgggg caggtggtct gccgatgggc 180
cgtatcgtcg aaatctacgg accggaatct tccggtaaaa ccacgctgac gctgcaggtg 240
atcgccgcag cgcagcgtga aggtaaaacc tgtgcgttta tcgatgctga acacgcgctg 300
gacccaatct acgcacgtaa actgggcgtc gatatcgata acctgctgtg ctcccagccg 360
gacaccggcg agcaggcact ggaaatctgt gacgccctgg cgcgttctgg cgcagtagac 420
gttatcgtcg ttgactccgt ggcggcactg acgccgaaag cggaaatcga aggcgaaatc 480
ggcgactctc acatgggcct tgcggcacgt atgatgagcc aggcgatgcg taagctggcg 540
ggtaacctga agcagtccaa cacgctgctg atcttcatca accagatccg tatgaaaatt 600

ggtgtgatgt tcggtaaccc ggaaaccact accggtggta acgcgctgaa attctacgcc 660
tctgttcgtc tcgacatccg tcgtatcggc gcggtgaaag agggcgaaaa cgtggtgggt 720
agcgaaaccc gcgtgaaagt ggtgaagaac aaaatcgctg cgccgtttaa acaggctgaa 780
ttccagatcc tctacggcga aggtatcaac ttctacggcg aactggttga cctgggcgta 840
aaagagaagc tgatcgagaa agcaggcgcg tggtacagct acaaaggtga gaagatcggt 900
cagggtaaag cgaatgcgac tgcctggctg aaagataacc cggaaaccgc gaaagagatc 960
gagaagaaag tacgtgagtt gctgctgagc aacccgaact caacgccgga tttctctgta 1020
gatgatagcg aaggcgtagc agaaactaac gaagattttt aa 1062

<210> 226
<211> 1062
<212> DNA
<213> bacterial
<400> 226

atggttattg atgaaaataa atagaaagtg ttggtggtag tattgggtta gattgagaaa 60
taatttggta aaggttttat tatgtgtttg ggtgaagatt gttttatgga tgtggaaatt 120
atttttattg gtttgttttt attggatatt gtgtttgggg taggtggttt gttgatgggt 180
tgtattgttg aaatttatgg attggaattt tttggtaaaa ttatgttgat gttgtaggtg 240
attgttgtag tgtagtgtga aggtaaaatt tgtgtgttta ttgatgttga atatgtgttg 300
gatttaattt atgtatgtaa attgggtgtt gatattgata atttgttgtg tttttagttg 360
gatattggtg agtaggtatt ggaaatttgt gatgttttgg tgtgttttgg tgtagtagat 420
gttattgttg ttgattttgt ggtggtattg atgttgaaag tggaaattga aggtgaaatt 480
ggtgattttt atatgggttt tgtggtatgt atgatgagtt aggtgatgtg taagttggtg 540
ggtaatttga agtagtttaa tatgttgttg atttttatta attagatttg tatgaaaatt 600
ggtgtgatgt ttggtaattt ggaaattatt attggtggta atgtgttgaa attttatgtt 660
tttgtttgtt ttgatatttg ttgtattggt gtggtgaaag agggtgaaaa tgtggtgggt 720
agtgaaattt gtgtgaaagt ggtgaagaat aaaattgttg tgttgtttaa ataggttgaa 780
ttttagattt tttatggtga aggtattaat ttttatggtg aattggttga tttgggtgta 840
aaagagaagt tgattgagaa agtaggtgtg tggtatagtt ataaaggtga gaagattggt 900
tagggtaaag tgaatgtgat tgtttggttg aaagataatt tggaaattgt gaaagagatt 960
gagaagaaag tatgtgagtt gttgttgagt aatttgaatt taatgttgga ttttttttgta 1020
gatgatagtg aaggtgtagt agaaattaat gaagattttt aa 1062